Mainstreaming Data-Driven Approaches to Inclusive Service Delivery:

An Operational Toolkit for World Bank Task Teams

Working Draft
Version 1.0
May 2023

Contents

Acknowledgments

This Operational Toolkit was financed by the Human Rights, Inclusion and Empowerment Umbrella Trust Fund as part of the "Mainstreaming a Rights-Based Approach to Information and Data in Bank operations" program. The Toolkit aims to support World Bank task teams in incorporating rights-based approaches to the use of data for more inclusive, effective service delivery by governments. It is a "living document" that will be updated as needed based on user feedback and country experiences.

The toolkit was developed by a team comprising Emily Kallaur, Luis Esquivel, Ravi Kumar, Craig Hammer and Stephen Davenport. The team would also like to thank Yoshihiro Saito, Marcela Rozo, Manav Bhattarai, Mamata Ghimire, and Moatez Chaouachi for their helpful advice and support in the development of the toolkit.

Comments are welcome and should be directed to the World Bank Data Desk or rkumar17@worldbank.org.

Abbreviations

ADB	Asian Development Bank
AfDB	African Development Bank
CPF	Country Partnership Framework
CSO	Civil Society Organization
D4P	Data for Policy
DE4A	Digital Economy for Africa Initiative
DECA	Digital Economy Country Assessment
DGRA	Digital Government Readiness Assessment
DLI	Disbursement-Linked Indicator
FCDO	Foreign, Commonwealth & Development Office (UK)
GIZ	German Agency for International Cooperation
GRID	Green, Resilient and Inclusive Development
GTMI	GovTech Maturity Index
HRIETF	Human Rights, Inclusion, and Empowerment Trust Fund
ICT	Information and Communications Technologies
IDA	International Development Association
INDS	Integrated National Data System
NGO	Non-Governmental Organization
NSO	National Statistical Office
NSS	National Statistics Systems
ODIN	Open Data Inventory
ODRA	Open Data Readiness Assessment
OECD	Organization for Economic Cooperation and Development
PDO	Project Development Objective
RTI	Right-to-Information
SCD	Systematic Country Diagnostic
SCI	Statistical Capacity Indicators
SPI	Statistical Performance Indicators
TTL	Task Team Leader
USAID	US Agency for International Development
VPU	Vice Presidential Unit
WDR	World Development Report

1.Purpose of this toolkit

Data is essential to the World Bank's twin goals of reducing poverty and promoting shared prosperity, especially by enabling more inclusive, efficient, and effective service delivery, and thus strengthening economic and social human rights. The links between human rights and the data agenda are manifold. Data and information can play a key role in enabling governments to take a rights-based approach to service delivery, by giving decision makers and administrators better insight into service delivery access and outcomes among different groups of constituents. From a citizen's perspective, data can help people to understand their rights and how to exercise them. Data can facilitate increased government accountability and inclusion by providing a basis for discussion of government programs and policies, making it possible for civil society actors to participate actively in public affairs. Moreover, access to information as a human right in itself is enshrined in Article 19 of the Universal Declaration of Human Rights, which provides that "everyone has the right to … seek, receive and impart information and ideas through any media and regardless of frontiers." The increasing access to and use of data also raises new rights-related issues around privacy and security that need to be addressed.

The World Bank has recognized the importance of human rights principles to achieving its twin goals.[1] This is demonstrated by the work conducted under the Human Rights, Inclusion and Empowerment Trust Fund (HRIETF), an umbrella multi-donor trust fund administered by the World Bank that aims to increase and strengthen the understanding and application of human rights principles across the World Bank Group's work.[2] The development of this toolkit was possible thanks to the Umbrella's guidance and financial support. Furthermore, the toolkit is aligned with the Umbrella's strategic priorities, particularly around "Governance, Inclusive Institutions and Empowerment" and "Social Inclusion".

The 2021 World Development Report (WDR) *Data for Better Lives* puts forward a vision for achieving development objectives by harnessing the power of the data revolution. The amount of data continues to increase exponentially around the world, and if used creatively, data offers previously unimaginable ways to increase productivity, improve socioeconomic outcomes and strengthen human rights. The report argues that leveraging data effectively depends on establishing a new "social contract for data" that builds trust in the data ecosystem and ensures that benefits are shared equitably. All of this should take place within a data governance framework that supports an integrated national data system (INDS) capable of ensuring data quality and accessibility for sharing, reuse, and repurposing by all stakeholders.[3]

In this context, the toolkit aims to assist World Bank staff in advancing the vision of *Data for Better Lives* by actively and purposefully incorporating data-related components into World Bank

[1] See for example the 2006 legal opinion drafted by Robert Dañino in his role as World Bank Senior Vice-President and General Counsel, in which he wrote that "The [WB] Articles of Agreement permit, and in some cases require, the World Bank to recognize the human rights dimensions of its development policies and activities since it is now evident that human rights are an intrinsic part of the Bank's mission." Available at https://opil.ouplaw.com/view/10.1093/law-oxio/e215.013.1/law-oxio-e215-regGroup-1-law-oxio-e215-source.pdf.

[2] See https://www.worldbank.org/en/programs/humanrights/overview.

[3] World Bank 2021a.

operations, making the link between data and a rights-based development agenda more concrete. Overall, "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"[4]. This requires that countries take a holistic approach that prioritizes strengthening institutions, data analytics capacity, stakeholder engagement, and the broader enabling environment as part of a coherent INDS transformation, rather than pursuing one-off or short term investments in single datasets or stand-alone data systems.

This document provides World Bank task teams with concrete ideas and a methodology for mainstreaming data-related solutions in World Bank operations. The toolkit considers specific development challenges or technical problems, potential entry points for data-related solutions, political economy and technical analysis, and binding constraints to reform. By using this methodology, task team leaders (TTLs) and Bank data experts will be able to design and implement relevant data-related solutions to address specific challenges. These solutions could involve the collection or production of new data, the enhancement of existing data, or could instead address any aspect(s) of the broader data ecosystem, as discussed in section 3.2.

Data and information have become critical tools to generate value and to address development challenges, with the world at the threshold of what has been called the Fourth Industrial Revolution. The global increase in data production and digital connectivity, and exponential advances in data analytics capabilities, have coalesced to become a powerful engine for change in both positive and negative ways. To evolve accordingly, governments need to fill development data gaps and turn data into actionable information. This means establishing or strengthening key institutions to scale up data-driven decision-making and support improved outcomes for the poor and marginalized, taking a rights-based approach to the provision of information. Moreover, "for data to maximize value, the data should have adequate coverage (be complete, frequent, and timely), be of high quality (be accurate, comparable, and granular), be easy to use (be accessible, understandable, and interoperable), and be safe to use (be impartial, confidential, and appropriate)."[5]

Stronger and more transparent data ecosystems can facilitate citizen engagement in governance processes, service delivery and the exercise of human rights, as well as potentially generating economic benefits. The World Bank has promoted global efforts to open data and has taken a range of actions to make its own investments and knowledge products more transparent.[6] These include the development and launch of data.worldbank.org in 2010, with thousands of datasets readily available to anyone with access to the Internet. Open data also has economic value - in 2013, McKinsey's research on open data suggested that there was as much as three trillion dollars to be unlocked in the global economy through open data initiatives.[7]

This toolkit contributes to the World Bank's Green, Resilient, and Inclusive Development (GRID) approach by helping countries design and implement data-driven policies and projects, and take advantage of the ongoing data revolution to pursue the goals of poverty reduction and shared prosperity through the linked priorities of inclusion and sustainability. The toolkit is intended primarily for World Bank staff and TTLs who are interested in leveraging data to improve service

[4] Ibid.
[5] Joliffe et al. 2021.
[6] World Bank 2017a.
[7] Manyika et al. 2013.

delivery with a rights-based approach and improve development outcomes via World Bank operations. The toolkit will also be of interest to donors, development partners, policymakers, civil society, and academia.

The toolkit is organized as follows. Section 2 summarizes the definition of data, and the theory of change around the potential for data to support development, as presented in the 2021 WDR. Section 3 defines the scope of data-related solutions envisioned by this toolkit, and discusses data in World Bank operations. Section 4 lays out the methodology for incorporating data solutions into operations. Finally, Annex 1 provides examples of results indicators for data-related activities from actual World Bank projects to help TTLs as they develop new operations in coordination with clients. Annexes 2 and 3 provide overviews of the World Bank's Statistical Performance Indicators, and guidance for data diagnostics in Systematic Country Diagnostics (SCDs), respectively.

2.Data for better service delivery

2.1 Definition of data

There are various definitions of data, as this term has evolved in the past few decades. The 2021 WDR acknowledges this lack of uniformity regarding the concept of data, and aligns with Carriere-Swallow and Haksar´s definition: "data can be quantitative or qualitative in nature, and may be stored on analog (that is, paper, stone tablets) or digital media". Furthermore, the 2021 WDR establishes the difference between data and information, as "data must be processed, structured, and analyzed to be converted into information".

Data can be divided, depending on its main purpose, into public and private intent data. Public intent data refers to data that is primarily generated by the public sector, although CSOs, academia and international organizations can also contribute to this process. The WDR classifies public intent data in six categories: (a) administrative data; (b) census; (c) sample surveys; (d) citizen-generated data; (e) machine-generated data; and (f) geospatial data. This data aims to improve decision-making by the state and improve the relationship between the state and society. These kinds of data can improve lives if they lead to the creation of better-targeted service delivery and policies, and prioritization of scarce resources, particularly to help marginalized populations.[8] In situations where official data does not exist on neglected or underrepresented populations, citizens or civil society groups may be able to shine light by generating data and demanding action or accountability.[9]

In contrast, private intent data refers to "data collected and curated by the private sector for commercial purposes".[10] 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. While World Bank operations might benefit from implementing solutions around both kinds of data, public intent data and the role of government in the data ecosystem are the focus of this toolkit.

[8] World Bank 2021a.

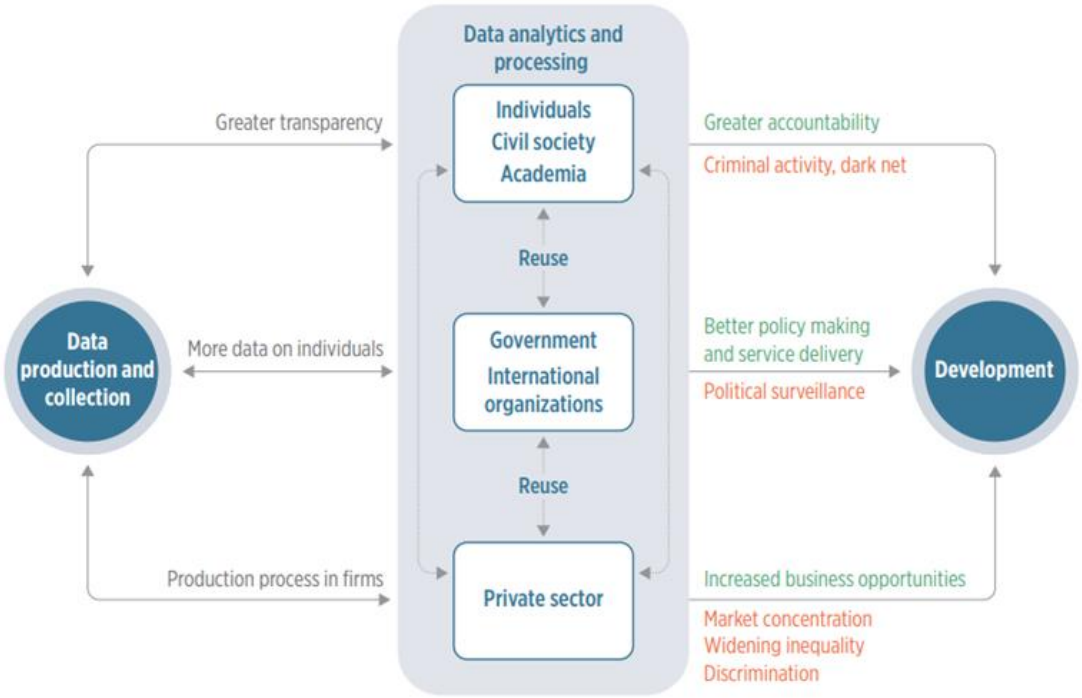
[9] An example of creatively generating data where official data do not exist: Milusheva et al. (2021) did an experiment using machine learning techniques on tweets from Twitter to compile data on car crashes, to be used for traffic planning purposes.

[10] World Bank 2021a.

2.2 Impact of data and its use

Data has significant potential to foster development, if used responsibly. Governments, civil society and the private sector increasingly initiate and manage data initiatives to accomplish their objectives. Such initiatives can concern collection, production, use, and/or dissemination of data, and can support the government at the sub-national and national levels in designing, implementing and monitoring policies and programs. According to the WDR's theory of change, shown in Figure 1, leveraging data successfully can enable 1) greater accountability; 2) better policy making; and 3) greater business opportunities in the private sector. Each of these channels is discussed below.

Figure 1. How data can support development: A theory of change (WDR 2021)



Source: Reproduced from World Bank 2021a.

2.2.1 Greater accountability

Data can help to foster accountability and trust in government by making citizens more aware of government actions, results and performance (Box 1).[11] In a successful scenario, increased provision of open government data helps strengthen government accountability and thus improves service delivery.[12] Data can empower citizens and civil society to actively and effectively participate in public decision making and social accountability exercises, and to demand better services. When data is publicly available and can be used and reused, a process that can be helpfully facilitated by intermediaries, it may lead to increased public oversight of government policies and programs and potentially help reduce corruption. For example, in Nigeria, the BudgIT budget transparency initiative

[11] See https://odimpact.org/ for a set of case studies that showcase ways open data is changing the world. This includes examples such as Brazil's open budget transparency portal, Burundi's work to make health spending and performance transparent, and Slovakia's efforts to strengthen trust in government through open data. Another example is the e-procurement system in Ukraine, ProZorro, which helped the government save nearly $6 billion from 2017 to 2021; see Yukins and Kelman 2022.

[12] Jelenic 2019.

reached 2.5 million Nigerians and "exposed a 41 million naira (US$113,575) investment that claimed to be funding a non-existent youth center in Kebbi State".[13] Data can also be produced by non-government stakeholders, including civil society, academia, media outlets, and the private sector (e.g., CSOs can conduct surveys and crowdsource data). Civil society and citizens may be able to use data to advocate for action or monitor government projects to improve efficiency and implementation of programs and projects, creating feedback loops.

Box 1. Budgets that everybody can understand and track: Budget Stories in Moldova

In Moldova, the NGO-led initiative Budget Stories uses information generated and published by the government to provide citizens and other civil society organizations with information and data related to the national budget. Budget Stories recognizes that accessing budget-related data can be daunting to everyday citizens, even if the information is publicly available, given its technical nature. Budget Stories therefore uses infographics and visualizations to organize information in a way that is easy for everyone to understand. They provide data covering the whole budget process, from design and approval to execution, among other information. This has enabled the public to track government expenditures, potentially contributing to greater accountability in the management of public finances.

Sources: Budget Stories website (https://www.budgetstories.md/) and Neagu 2013. Screenshot from budgetstories.md, April 12, 2023.

2.2.2 Better policy-making and service delivery

Government decision-making requires accurate, detailed, representative, and timely data and information. Greater availability of data, and the ability to process it effectively, can be critical to improved policy-making and service delivery (Box 2). Again, data can be generated by the government itself (i.e., census data), academia and NGOs (data stemming from their research or contact with the citizenry), and even the private sector (data on traffic accidents by insurance companies). Ideally, to be useful for service delivery management, it should be possible to disaggregate data by different user groups based on geography, gender, income status, vulnerability, etc. This enables the government to understand who is able to access services (or not) and how their experiences differ. The data should also ideally be linked to actual service delivery events so that government officials can follow up with

[13] SDSN TreNDS 2018.

specific service providers to take corrective action. Data must also be representative of users if the intent is for the government to have a reliable picture of the average service delivery experience. For instance, citizen feedback that requires the user to proactively submit a comment or complaint can be useful in identifying issues, but will not be representative of the average user experience.[14] Data proactively collected by the government and data proactively submitted by citizens can complement each other to improve the government's understanding of the user experience.

Box 2. Phone-based service delivery monitoring in India

An example from India demonstrates the power of proactive data collection by government as a means to improve service delivery. In 2018, the government of Telengana State began making lump-sum cash transfers to farmers to enable them to purchase agricultural inputs such as seeds and fertilizer at the beginning of the growing season rather than having to incur debt. The timely delivery of the payments was key to achieving the program's goals. In a large-scale experiment, some of the government officials responsible for making the payments were told that a call center would contact farmers to ask about their experience receiving the transfers, while a control group of officials did not receive this treatment. The results showed that the desired outcome (on-time delivery of payments) was 2.4 percent more likely for the farmers receiving payments from officials in the phone monitoring group (a statistically significant difference). The results suggested that "the cost per dollar of [incremental] benefits delivered on time was less than one cent" as a result of the phone monitoring, representing a highly cost-effective investment in data to improve service delivery.

Source: Muralidharan 2021.

2.2.3 Improved business opportunities

Public data can help drive private sector growth and innovation. The use of data can improve decision making by companies, and spur business opportunities, hence promoting economic development and creating jobs (Box 3). Data that is useful for businesses can be generated within the private sector or by the government. Some of the world's most valuable companies are data-driven. Opportunities for data-driven enterprises are growing in areas as diverse as healthcare and climate change response. To take advantage of these trends, a wave of new start-ups along the data value chain—from data collection to intelligence—is providing innovative, low-cost goods and services to individuals, businesses, and governments around the world.[15]

Globally, the data economy has led to the development of new business models and enabled the private sector to innovate on product development, customer relations, and other key business functions by using data.[16] For many businesses, data has become an asset as valuable as its people,

[14] See Case Study 12 in Beschel et al. 2018.
[15] Filippov 2014.
[16] World Bank 2019a.

technology, and capital.[17] Indeed, according to one study, firms that adopt data-driven decision making increase their output and productivity by 5–6 percent.[18]

Box 3. Business opportunities through data

In Malaysia, iMoney aggregates financial sector data to help users make data-driven decisions, from finding the right credit card to purchasing the best broadband service. ByteDance, a Beijing start-up, provides personalized news aggregation by analyzing content, user data, and users' interaction with the content. Sixtuple, a Bangalore-based start-up, is digitizing images of pathology slides and running cloud-based image processing to analyze visual medical data. The Global Entrepreneurship Network predicts that artificial intelligence, blockchain, and data analytics will be among the technologies leading the entrepreneurial revolution in the near future.

Source: Startup Genome and Global Entrepreneurship Network 2018.

2.2.4. Impact of data in key development sectors

Data plays a key role in all sectors, with different types of data contributing in different ways. For example, administrative or program-generated data in the health or education sectors can help authorities design better policies, but can also help service providers (such as healthcare workers and teachers) do their daily jobs better by understanding how health outcomes are changing, or how students are making progress. When performance indicators are released to the public, people can make informed decisions, e.g., which hospital to use, or which school to select for their children. Financial data on allocations and spending supports better resource monitoring, which can help reduce corruption and spending leakages, and improve value for money. Financial data can be important for holding policymakers accountable for sticking to their promises, for example related to financial commitments on climate change mitigation. At the local level, it can enable citizens to monitor whether infrastructure improvements or textbook purchases are going as planned. On the climate front, coordinated data collection and sharing contributes to enhancing collaboration among different stakeholders and to better risk modeling, adaptation planning, and crisis response. And data can directly save lives when properly used during times of crisis, as we have seen with the COVID-19 pandemic.

Specific sectoral examples are given below as illustrations of ways data can contribute to policy making, service delivery, and crisis response, and thus ultimately to the exercise of basic human rights.

Education: In Brazil, the platform QEdu takes data from official sources and processes it in order for other key stakeholders (policy makers, journalists, parents) to analyze and use the data. Through visualization tools, QEdu makes this data more accessible.[19] The impact has been felt in various arenas. For example, it has become a reliable source of information for parents about the schools where their children are enrolled or will be enrolled. Furthermore, in terms of policy making, QEdu data is regularly used by state-level ministries of education to inform their decisions. In 2014, after data gathered and

[17] Porter and Heppelmann 2015.
[18] Brynjolfsson, Hitt, and Kim 2011.
[19] Open Data Impact Map. QEdu. Available at https://opendataimpactmap.org/lac.

analyzed by QEdu found that a significant number of schools in São Paulo faced problems with violence and crime, the State responded by increasing the presence of police in specific schools.[20]

Health: The 2014 Ebola outbreak in Sierra Leone was a major public health crisis. In order to face the challenge and minimize its impact, the government in collaboration with aid agencies and CSOs realized the importance of data sharing in a context where this kind of flow of information was not the norm. Shared data included information about cases, availability of medical facilities and healthcare workers, resources to fight the epidemic, relevant medical information for citizens and so on. Furthermore, much of this information was geo-referenced and useful maps were developed. As a result, healthcare workers, citizens, policy makers and aid agencies were able to make informed decisions that contributed to more effective control of the epidemic.[21]

Climate change: In Costa Rica, the government created the National Climate Change Metrics System.[22] This platform incorporates data from a wide range of relevant agencies. Through this collaboration, data generation, processing and analysis will contribute to more accurate information about climate change drivers, risks and responses.[23] Ultimately, this will also enhance government reporting mechanisms under the Paris Agreement. Similarly, in Chile, government data on climate change has enabled CR2, a research institution, to build models that are now used by a wide range of stakeholders.[24]

Social development: As a result of standardized data collection methodologies applied across dozens of countries, a more accurate picture of the extent of violence against women and girls (VAWG) around the world has emerged.[25] Having representative, internationally comparable data has supported research and policy responses at the country level. For example, World Bank teams have used data collected by the United Nations Population Fund (UNFPA) to inform operations, and in 2019 the government of Peru leveraged the analysis of VAWG data and spending to create a results-oriented budget plan. The Gender-Based Violence Information Management System (GBVIMS) is a global effort to standardize service-based data to shed light on when and how victims seek services, in order to find ways to reduce barriers. This issue area in particular also highlights the importance of following ethical and safety guidelines to ensure that the act of data collection does not itself pose a danger to survivors.

Economic development: The Aclímate Colombia platform, created through a multi-stakeholder effort led by a CSO and including government and private sector partners, brings various data sources together. The platform helps farmers adapt to changing weather patterns by enabling them to make informed decisions, such as on planting choices.[26] It has received praise for its concrete impact on farmers' livelihoods. For example, it enabled a farmers' growing association to broadcast a specific

[20] Portal QEdu, Ernesto Martins Faria, 2014. Available at https://www.slideshare.net/iGovExplica/ernesto-martins-faria-portal-q-edu.
[21] Verhulst and Young 2016.
[22] National Climate Change Information System. SINAMMEC Costa Rica. Available at http://www.sinamecc.go.cr/.
[23] Grinspan and Worker 2021.
[24] Ibid.
[25] This paragraph draws on "Spotlight 2.1" in World Bank (2021), a case study on deploying data to curtail violence against women and girls.
[26] The example in this paragraph comes from a case study written by Young and Verhulst (2017).

message to 170 farmers in Cordoba with "detailed, granular information" on when and what to plant given a projected drought, thus helping them avoid major losses.

2.2.5 Potential for negative impacts

However, for data to bring about all of these benefits, it is critical to mitigate risks that arise from the increased production and accessibility of sensitive information. As shown in Figure 1, alongside its positive effects, the data revolution has the potential to cause serious negative impacts. World Bank teams and clients need to be mindful of the potential for unintended consequences as they design data-related interventions. Increased production and availability of individually-identifiable information, for example, creates opportunities for politically-motivated use of personal data, or surveillance by public or non-public entities. According to the WDR, as long as "public accountability is strong and state actors can be presumed to act in the broader public interest," these risks can be managed, but such conditions cannot be assumed. Cybercrime that exploits sensitive data, sometimes facilitated by the "dark net", has a staggering annual cost. The private sector can also potentially misuse and exploit personal data, and data-driven businesses often experience increasing returns to scale that increase market concentration.

Risks that accompany the increased production and use of data underscore the need for a social contract for data, as outlined in the WDR. This needs to be enforced via a data governance framework that "can strengthen trust in the data system, thereby incentivizing the use of data-driven products and services, increasing their value, and ensuring a more equitable distribution of benefits". At the same time, the WDR emphasizes the necessary balancing act between enablers (which facilitate data access and reuse) and safeguards (that prevent data misuse), such that data can be used effectively *and* responsibly. Furthermore, "safeguards must differentiate between personal data, requiring a rights-based approach with individual protection, and nonpersonal data, allowing a balancing of interests in data reuse".[27] Other aspects of the enabling environment, such as freedom of the press, an independent national statistical office, and the presence of data intermediaries are also important to creating a climate of transparency and accountability.

3. Data in World Bank operations

3.1 Strategic entry points for data in World Bank operations

A systematic review of the World Bank's efforts to mainstream data in operations has not been undertaken, but common operational entry points based on the World Bank's strategic priorities can be identified. Recent corporate and sectoral strategies have emphasized the need for being data-driven, leveraging technology, and building capacity to use data to achieve development goals and/or to invest in human capital. While data-related investments and solutions can potentially be relevant in any operation, a brief (non-comprehensive) overview of common entry points for mainstreaming data use into World Bank operations is presented below. They are discussed under the following focus areas:

- Governance and accountability, IDA replenishments, and the Data for Policy (D4P) package

[27] See Chapter 6 of the WDR 2021.

- Climate change
- Operationalization of the Green, Resilient and Inclusive Development (GRID) Strategy and the 2021 WDR
- Mobilizing technology for development
- GovTech
- Gender
- Sectoral strategies
- Citizen engagement
- New procurement framework and open contracting
- Fragility and conflict prevention
- World Bank Evolution Roadmap
- World Bank Group Data Roadmap

3.1.1 Governance and accountability, IDA replenishments, and the Data for Policy (D4P) package

The World Bank Group has long focused on issues related to social accountability, governance and corruption.[28] Institutional weaknesses in the world's poorest countries pose significant challenges to ending extreme poverty, boosting shared prosperity, and achieving the Sustainable Development Goals. In 2019, as part of IDA19's special theme on Governance and Institutions, the World Bank recognized the importance of "accurate, timely, granular and accessible data…for policy making, efficient resource allocation and effective service delivery", especially in low-income contexts.[29] Furthermore, the World Bank has highlighted that national statistical systems, one of the key producers of data in the context of policy-making, often face challenges "such as inadequate and unreliable financing, limited use of data by policy makers, insufficient institutional capacity, [and] limited access to public data", among others.[30]

"Building better data and analytics" is one of the four pillars of IDA19's special theme on Governance and Institutions, which incorporates a policy commitment to improve data for evidence-based policymaking. Under this pillar, the World Bank committed to supporting institutions and building statistical capacity to reduce gaps in the availability of core data for evidence-based policy making.[31]

The Data for Policy (D4P) initiative, announced as part of IDA19 and launched that year, serves as a vehicle to fulfill this commitment, and is being led by the Poverty Global Practice. D4P was designed to close core data gaps in 30 IDA countries, with a focus on 5 areas: (i) household surveys; (ii) enterprise surveys; (iii) agricultural data; (iv) price data; and (v) administrative data. As of 2023, this target has been met (and exceeded), as 31 IDA countries have benefited from D4P operations. The commitment was renewed and updated in IDA20, where investment priorities are being guided by

[28] See for example the World Bank's Governance and Anticorruption (GAC) strategy from 2007, available at http://www1.worldbank.org/publicsector/anticorrupt/corecourse2007/GACMaster.pdf.
[29] The special theme on Governance and Institutions in IDA 19 is described at http://documents1.worldbank.org/curated/en/696731563778743629/pdf/IDA19-Second-Replenishment-Meeting-Special-Theme-Governance-and-Institutions.pdf.
[30] Dabalen, Himelein, and Castelan 2020.
[31] See pillar 4 in https://documents.worldbank.org/en/publication/documents-reports/documentdetail/696731563778743629/ida19-second-replenishment-meeting-special-theme-governance-and-institutions.

countries' Statistical Performance Indicators (SPI) profiles (see Annex 2) and support recommendations from the 2021 WDR, *Data for Better Lives*.

The FY23-24 pipeline portfolio includes substantial attention to data and statistics. To date, investments in data and statistics include $1.1 billion in IDA funding and $0.2 billion in IBRD funding, which comprises regional projects, P4Rs, and regional Series of Projects (IPF financing).[32] Catalytic and complementary financing will flow through the Global Data Facility trust fund. Technical support for operational scale is being delivered by the Living Standards Measurement Study (LSMS) program.

Planned financing includes regional investment approaches to support harmonizing data/statistics standards and INDS support. There are presently five regional D4P projects in the portfolio, and two more in the pipeline. The most ambitious is a $820 million Central/West Africa regional project supporting 11 countries. Similarly ambitious regional data/statistics projects are under development for East Africa and Southern Africa. Regional support for sectoral data and statistics is also being scaled up, including for agricultural data via the 50x2030 Data Smart Agriculture initiative.

D4P focuses on strengthening the capacities of National Statistical Systems (NSSs) around five key statistical operations: household surveys and census; enterprise surveys and census; agricultural surveys; administrative data; and price data; plus, two complementary data systems, national accounts and Big Data. D4P aims not only to help governments design and monitor public policies in general, but specifically to increase the availability of data that is sufficiently granular to be used to identify vulnerable groups and to be disaggregated by gender and disability status. World Bank support through D4P is geared toward enhanced data collection, increased utilization of data by governments for policymaking as well as increased data access and transparency for civil society, and improvements in the wider NSS. D4P also supports harmonization of statistical efforts via international standards.

IDA20 builds on IDA19 by identifying Governance and Institutions as a cross-cutting issue.[33] It promises to "deepen commitments that reinforce fiscal sustainability and accelerate digital governance to improve service delivery, statistical capacity, and institutional strengthening."[34] The World Bank's Statistical Performance Indicators (see Annex 2) will be used to measure high-level outcomes, and target intermediate outcomes include:

- IDA countries provided with statistical capacity building support by the World Bank for the implementation of household surveys
- Countries collect disability data with IDA support
- IDA countries publishing annual and timely public debt reports

Transparency- and governance-related projects more broadly can serve as entry points to mainstream attention to data, depending on the need and context. The World Bank supports its

[32] Includes AFR: Southern Africa ($259M); Eastern Africa ($381M); West & Central Africa - two regional projects w/RECs and 14 countries ($600m); Angola ($58M); Republic of Congo ($10M); EAP: Lao PDR ($25M); LAC: OECS Countries ($27M); Colombia ($75M); Ecuador ($85M); SAR: India ($110M); ECA: Uzbekistan ($50M).
[33] See https://ida.worldbank.org/en/topics/cross-cutting/governance-and-institutions#2.
[34] See the following for an overview of IDA20 special themes and cross-cutting issues, and how results will be measured: https://thedocs.worldbank.org/en/doc/12b9564a2b4343813b91eb2a8a2d38a6-0410012022/original/IDA20-Cross-Cutting-Issues-06-22-2022.pdf.

client countries in implementing Right to Information Acts[35] to increase public access to information and encourage its use. This is in addition to releasing the World Bank's own data through initiatives such as Open Data, the Open Knowledge Repository, and Open Finances, as part of its corporate commitment to the Access to Information Policy. Investment in building capacity for non-state actors to use data, and in frameworks for government transparency, can be instrumental in the World Bank's efforts to support the access to information agenda.

3.1.2 Climate change

The World Bank's "Outlook 2050: Strategic Directions Note" argues for a "whole of economy" approach to help its client countries decarbonize and develop sustainably. The World Bank emphasizes building the capacity of its stakeholders to collect and use data to monitor the impact of climate change and to develop data-driven solutions across sectors, including in food systems, energy, transport, water systems and low-carbon cities.[36] The Strategic Directions Note suggests numerous entry points to use data for identifying a problem or building related skills, e.g.:

- While the data and monitoring tools needed to monitor illegal logging and combat corruption have significantly improved, there is a need to strengthen the capacity to use them, and to support improved data analytics.
- The World Bank has a Global Program on Sustainability[37] that promotes the "generation and use of reliable data on natural capital and ecosystem services, including global data and tools, country-level support for natural capital accounting, and sustainable finance."[38]
- The World Bank commits to scaling up the use of data and information technologies to manage water demand and help create climate-smart cities.
- The World Bank commits to working with countries to support the strengthening of capacity and tools for systems-level planning to further help in the development of national plans and long term strategies for decarbonization and resilience.

3.1.3 Operationalization of the Green, Resilient and Inclusive Development (GRID) Strategy, and the 2021 WDR

In light of COVID-19, climate change, and other conflicts converging to exacerbate global challenges and reverse impressive development gains made over the last few decades, the World Bank's GRID strategy outlines a vision for "promoting economic growth that goes hand in hand with environmental goals and social inclusion". This represents a departure from the World Bank's previous approaches to these overarching issues by emphasizing that they are inextricably linked, with environmental outcomes determining socioeconomic outcomes and vice versa.

As the Bank implements its GRID strategy[39] and operationalizes the WDR 2021's vision of a new social contract for data, there are opportunities to identify gaps in service delivery to the poor and marginalized. The GRID strategy recognizes institutional strengthening and technological innovation

[35] Lemieux and Trapnell 2016.
[36] Mukhi et al. 2020.
[37] See https://www.worldbank.org/en/programs/global-program-on-sustainability.
[38] See Mukhi et al. (2020), page 56.
[39] A document titled "From COVID-19 Crisis Response to Resilient Recovery - Saving Lives and Livelihoods while Supporting Green, Resilient and Inclusive Development (GRID)" was prepared by the World Bank Group for the virtual April 9, 2021 Development Committee Meeting, and is available at https://thedocs.worldbank.org/en/doc/9385bfef1c330ed6ed972dd9e70d0fb7-0200022021/green-resilient-and-inclusive-development-grid.

as cross-cutting enablers of its vision. It points to the need for gender-disaggregated data, better risk information systems for climate and disaster resilience, data-driven analytical tools to prioritize investments in a context of resource scarcity, and improved statistical and data systems for monitoring impacts on poverty and inequality. Operationalization of the GRID strategy and the WDR vision presents enormous opportunities to mainstream data in operations to respond to client needs and increase equitable service delivery, while strengthening human rights.

3.1.4 Mobilizing technology for development

The World Bank's Digital Development Global Practice (GP) takes an ecosystem approach to digital transformation, in close collaboration with colleagues across the World Bank Group. It strives to help client countries build and grow their digital economies and is responding to growing demand: in FY2021 the Bank had almost USD 5 billion in commitments to operations with significant digital components, comprising projects in a wide range of sectors.[40] Digital Development work focuses on the following five key elements:

- Digital infrastructure
- Digital financial services and digital identification
- Digital innovation and entrepreneurship
- Digital platforms
- Digital literacy and skills

The 2016 World Development Report, *Digital Dividends*, provides analytical underpinnings for work in the digital sector. This report emphasizes the importance of "analog complements", including the regulatory environment, workers' skills, and institutional accountability, to enable investments in digital technology to generate development gains more broadly. Moreover, the 2019 World Development Report, *The Changing Nature of Work*, also makes a strong and urgent case for investing in human capital to help communities and countries make the most of economic opportunities through technology.[41] The World Bank's Digital Economy for Africa Initiative (DE4A)[42] has identified digital skills as one of its five foundational pillars. For example, the DE4A initiative aims to ensure that all 15-year-old students in Africa have basic digital skills and that there are 100,000 graduates from advanced digital programs every year.[43] The International Finance Corporation estimates that 230 million jobs in sub-Saharan Africa will require digital skills by 2030.[44] All these initiatives and activities require data, representing entry points for using data in operations.

Various programs, analytical work, and multistakeholder partnerships complement the World Bank's Digital Development operational portfolio and create opportunities for data initiatives. There is the Digital Development Partnership[45] (DDP), which convenes public and private sector partners to advance safe and inclusive digital transformation. The Identification for Development Initiative (ID4D)[46] works across sectors to support implementation of digital identification, enabling

[40] For an overview of the World Bank's work on Digital Development, see https://www.worldbank.org/en/topic/digitaldevelopment/overview#2.
[41] World Bank 2019b.
[42] See https://www.worldbank.org/en/programs/all-africa-digital-transformation.
[43] From a Digital Economy for Africa Initiative presentation, June 2019, available at http://pubdocs.worldbank.org/en/312571561424182864/062519-digital-economy-from-africa-initiative-Tim-Kelly.pdf.
[44] IFC 2019.
[45] https://www.worldbank.org/en/programs/digital-development-partnership
[46] https://id4d.worldbank.org/

governments to gather accurate and timely data on service delivery access and outcomes. Digital IDs provide a foundation for more inclusive service provision and data-driven policymaking. The report "Unraveling Data's Gordian Knot," a companion report to the 2021 WDR, focuses on striking the balance between data enablers and data safeguards to unlock the potential of data sharing while protecting people from data abuse and misuse.[47]

The Digital Development GP has developed a number of tools and platforms to support operationalization of the data agenda. The Digital Development Toolkits Series[48] targets policy-makers, practitioners, regulators, researchers and others and aims to help them identify digital development challenges, analyze solutions, and learn from practical country examples. The series includes the Digital Government Readiness Assessment as well as a number of others focusing on aspects ranging from infrastructure to regulation and promoting gender equity. There is also work underway to develop a new online access infrastructure for World Bank data and diagnostic tools on digital topics with online benchmarking and visualization functionality, which will be accompanied by guidance notes.

The Digital Development work program is marked by an increasing focus on the links between digital development, data, and climate change. Climate is one of the focus areas of the GP's strategy, with an emphasis on greening digital infrastructure, and "harnessing the power of digital technologies and data for climate action across sectors." Digital Economy assessment tools (DE4x)[49] are available to help country teams evaluate the key levers of the development of the digital economy, with the latest version of the methodology to include looking at data on the relationship between digital technologies and climate change adaptation/mitigation. Digital Development's analytical workstream includes programmatic ASAs on Green Digital Development, and Digital Development Opportunities for Climate Change, which will look at topics such as data and analytics related to weather and disaster information, and to coordination in response, relief and recovery efforts.

3.1.5 GovTech

GovTech is an approach to public sector modernization that emphasizes citizen-centric public services that are universally accessible; a whole-of-government approach to digital government transformation; and simple, efficient and transparent government systems.[50] GovTech relies on digitization and data to achieve better results for citizens. The World Bank's GovTech Strategy focuses on four pillars of support to governments: 1) core government systems; 2) public service delivery; 3) citizen engagement; and 4) GovTech enablers (including digital identification, leadership and skills, strategy and regulations, institutions, and innovation).

All four of the GovTech pillars provide operational entry points for data-oriented work. For instance, under core government systems, the World Bank supports financial management information systems, e-procurement systems, tax and customs administration systems, and other systems that generate data critical to resource management and public sector accountability. Digitizing public service delivery creates opportunities for improvements in collection and management of administrative data to improve service quality and inclusiveness. Digital citizen engagement initiatives can lead to new types of information and feedback loops enabled by technology. Support for digital

[47] World Bank 2020.
[48] https://www.worldbank.org/en/topic/digitaldevelopment/brief/digital-development-toolkits
[49] See for example https://www.worldbank.org/en/programs/de4lac/digital-economy-framework and https://www.worldbank.org/en/programs/all-africa-digital-transformation/country-diagnostics.
[50] See https://www.worldbank.org/en/programs/govtech.

skills, digitization strategies, and institutional capacity building overlap with efforts to strengthen data availability and use.

3.1.6 Gender

Achieving gender equality in all spheres of life - in terms of economic opportunity, political voice, personal autonomy, asset ownership, education and healthcare access and quality, and beyond - requires policymakers and non-government stakeholders to closely monitor trends and outcomes by gender. Increasing the collection and analysis of gender-disaggregated data across all sectors and programs is critical, and implies many operational entry points for data-related initiatives. As noted in the World Bank's Gender Strategy (2016-23), the current availability of country-level gender-disaggregated data is very uneven, and SCDs should assess gender data gaps as part of the diagnostic process. The Strategy mentions four priority areas in particular: physical and financial asset ownership and control; time use; employment; and welfare.[51]

Progress is being made, but not quickly enough. According to UN Women, as of June 2022, 42 percent of the gender data that is needed to monitor the gender-specific dimensions of the SDGs is available, compared to 26 percent in 2016. However, at this rate it will take 22 more years to close the gap.[52] The COVID-19 pandemic has exacerbated the challenge by siphoning funding away from gender data initiatives.

Given the World Bank's track record in designing data collection methods and investing in country systems for gender data, the Bank is positioned to support the gender data agenda in several key areas in particular. These include:

- Gender data collection through existing and new methods across thematic areas; improving data quality
- Data harmonization, international standards, open data, and data dissemination
- Capacity building on data collection, data literacy, and data use for government officials, the private sector, and journalists

The IDA20 special theme on Gender and Development underscores the World Bank's commitment to strengthening institutional capacity around data for decision making. It will be important to use the D4P umbrella as a way to harmonize the World Bank's various initiatives on gender data. Teams can use genderdata.worldbank.org as a starting point to identify what data is available, and should contact the World Bank Gender Group for additional guidance on aligning with ongoing programs.

3.1.7 Sectoral strategies for inclusive service delivery

Projects to strengthen capacity for service delivery in education, healthcare, water and sanitation, and other areas include many operational entry points for leveraging data more effectively. Sectoral strategies may include plans for management information systems, citizen report cards or other forms of citizen feedback, service registries, digital service delivery, and other priorities involving data production and use. Better information systems support policy planning, real-time responsiveness in service delivery, holding specific service providers accountable for performance, and broader

[51] World Bank 2015.

[52] Encarnacion, Emandi, and Seck 2022.

monitoring of trends. Disaggregated data can help governments and civil society stakeholders track service delivery access, quality, and outcomes for different groups, including women, low-income households, members of minority ethnic or religious communities, LGBTI individuals, persons living with disabilities, and other groups that may have traditionally faced marginalization.

3.1.8 Citizen engagement

Ninety-nine percent of the World Bank's Investment Project Financing approved in FY2020 had a citizen-oriented design, reflecting the Bank's commitment to mainstreaming citizen engagement.[53] The World Bank defines citizen engagement as the two-way interaction between citizens and governments or the private sector within the scope of World Bank Group interventions, with the intention that citizen engagement can improve development outcomes.[54] This commitment is reflected in the Corporate Scorecard, which includes two Tier 3 indicators that rely on beneficiary feedback ("stakeholder feedback on Bank Group effectiveness and impact on development results" and "stakeholder feedback on Bank Group knowledge").[55]

However, evaluations of the World Bank's corporate commitment to engaging citizens for improved development outcomes have highlighted the need to strengthen the capacity of governments and citizens to engage effectively.[56] Depending on the context, data components can be integrated into World Bank operations to improve the quality of citizen engagement by supporting stakeholders' ability to collect, find, access, analyze and use information to improve the development outcomes of World Bank projects. Such approaches need to address both the supply and demand sides for data, and ensure that the data being provided are valuable to citizens and will prompt a government response. Too often, data initiatives have been supply driven, without enough attention to what happens after the data are published. When citizens engage and then fail to see the government take action, it can do more harm than good by undermining trust and citizen willingness to participate in future initiatives.

Moreover, citizen engagement efforts, if not undertaken with a deliberate focus on inclusion, can reinforce existing power inequities and dynamics rather than enabling voice for all. Enhanced use of data can help stakeholders ensure that citizen engagement initiatives reach a broader swath of stakeholders.

3.1.9 New procurement framework and open contracting

In 2016, the World Bank launched a new procurement framework to better respond to the needs of client countries. As part of the framework, the World Bank uses a new electronic procurement planning and tracking platform. This new platform, in addition to facilitating more efficient procurement-related decision-making processes, enables the World Bank to be more open with its procurement data.[57] Open contracting is the process through which data related to public contracting is published in an open, accessible, and timely fashion, so that citizens and businesses can engage to

[53] For more on the World Bank's work on citizen engagement, see https://www.worldbank.org/en/topic/citizen-engagement#3.
[54] The World Bank's commitment to citizen engagement is outlined in the "Strategic Framework for Mainstreaming Citizen Engagement in World Bank Operations," available at http://documents1.worldbank.org/curated/en/266371468124780089/pdf/929570WP0Box380ategicFrameworkforCE.pdf.
[55] See https://scorecard.worldbank.org/en/scorecard/tier3.
[56] World Bank 2018.
[57] Hunja 2015.

curb corruption and deliver better results.[58] Through World Bank projects and its new e-procurement framework, the World Bank supports the publication of public contracting data. Projects related to procurement and contracting can also serve as entry points to use data to address a challenge in a given sector or area, or for related initiatives, such as regarding beneficial ownership transparency. As another example, projects related to COVID-19 response might entail procurement of vaccines and other healthcare products, which would then generate expenditure data that could be made available for public scrutiny.[59] This requires building the capacity of civil servants to collect, process, and release data, as well as building the capacity of citizens to analyze data and communicate their findings.

3.1.10 Fragility and conflict prevention

Weak institutional capacity can be a chronic driver of fragility, conflict and violence,[60] and a more forward-looking approach is needed. The World Bank Group's strategy on conflict prevention recommends that the Bank build the capacity of client countries to enhance systems of accountability and service delivery, and foster citizen engagement. Using data to find gaps between aspirations and opportunities - and gaps in service delivery access and quality - especially among marginalized populations can potentially help reduce risks associated with fragility and conflict. The G7 Leaders' Summit 2022 featured discussions around strengthening anticipatory action in humanitarian assistance, with a Foreign Ministers' statement[61] emphasizing the need for more forward-looking humanitarian assistance to address record-high levels of need. Anticipatory mechanisms include data-driven efforts such as early warning systems, risk management initiatives, and pre-agreed financing and plans.

3.1.11 World Bank Evolution Roadmap

In 2022, the World Bank Group released a draft "Evolution Roadmap" that includes strengthening the focus on outcomes by increasing investment in the data, impact evaluation and results architecture. The Evolution Roadmap aims to help the institution clarify and update its mission and vision, and reform its operating model, to address the world's increasingly complex development challenges. The Roadmap acknowledges that the goal of ending extreme poverty and boosting shared prosperity around the world is increasingly out of reach due to a range of structural trends (such as climate change and demographic shifts) as well as shocks (such as COVID-19 and the war in Ukraine). Development challenges increasingly cross-national borders and require global public goods and a coordinated international approach.

Evolving the operational model will require significant data-driven work, including to:

- Analyze the overlap and trade-offs between global challenges and country development outcomes
- Assess countries' exposure to, as well as contribution to, global challenges
- Increase capacity for impact evaluation for both public sector and private sector interventions
- Invest in country-level data generation and capacity, including to help expand existing and build new datasets and dashboards to monitor global issues and increase transparency

[58] For more information, see https://www.open-contracting.org/what-is-open-contracting/.
[59] Ul-Aflaha, McNeil, and Kumagai (undated).
[60] For the World Bank Group Strategy for Fragility, Conflict, and Violence (2020-25), see http://documents1.worldbank.org/curated/en/844591582815510521/pdf/World-Bank-Group-Strategy-for-Fragility-Conflict-and-Violence-2020-2025.pdf.
[61] G7 2022.

- Develop a new Corporate Scorecard with revised results indicators to measure progress towards an expanded mission, updated definitions of poverty reduction and shared prosperity, and an increased focus on climate adaptation and mitigation
- Strengthen the focus on crisis preparedness, disaster risk reduction, and crisis response, which will require a robust, dedicated program of data collection and analytics on new/emerging and existing crises - particularly on crisis anticipation, monitoring, and response analytics

A "Playbook" could be developed with accompanying data/analytics to guide support to the Evolution Roadmap from inception, which would help the Bank identify various response options and take a more agile approach.

3.1.12 World Bank Group Data Roadmap

The World Bank first inaugurated a Bank-wide data governance architecture back in 2014, and the Development Economics Group (DEC) and Information and Technology Solutions (ITS) spearheaded a refresh in 2020. The result is a new World Bank-wide data governance architecture led by all four World Bank Managing Directors; operational and corporate data priorities are identified and implementation is driven by the Vice President-level World Bank Data Council. This new data governance architecture has launched a three-pronged World Bank Data Framework (data for countries and global priorities; data for organizational impact; and the enabling environment for data), and has identified and is now implementing a multi-year action plan covering operational and corporate data priorities, summarized in the inaugural World Bank Data Roadmap.

3.2 Data as part of a larger ecosystem

Realizing the benefits of a data-driven public sector requires not only data itself, but an "integrated national data system" (INDS), defined by the WDR as an "intentional, whole-of-government, multistakeholder approach to data governance" (WDR 2021; see also Box 4). This system is country-specific and depends on the country's level of data maturity as well as other contextual factors. For example, countries at a low level of data maturity should prioritize foundational elements such as "ensuring that data producers have adequate resources, capacities, and infrastructure; putting in place data protection regulation; and recognizing the importance of data." Countries seeking to build on a solid foundation should begin emphasizing data flows, e.g. by "incentivizing data sharing, instituting common standards, and ensuring that data users have the data literacy to effectively work with data." Later, the goal becomes optimization of the system.

Box 4. Implementing an Integrated National Data System (INDS)

The INDS is built around data, which must be:

Produced in a relevant, timely, accurate manner, at a sufficient level of granularity to inform policy decisions.

Protected from misuse by regulations that prevent harm. Protection is a catalyst for trust and participation.

Open and able to flow between stakeholders. Common standards are key to the frictionless flow of data, while also facilitating international data transfer.

Quality controlled to safeguard the integrity of the data themselves. This requires sound methodological foundations in data production.

Used and reused beyond their original purpose by different stakeholders. This includes the routine use of data in planning and decision making across government entities.

***

As it develops, an INDS will integrate multiple participants. They are:

Government entities, which produce public intent data, and also use other data sources. They act as data stewards in setting the rules for the whole INDS.

Civil society organizations (CSOs), nongovernmental organizations (NGOs), and individuals, which all produce and use data to empower themselves and to hold the public and private sector to account.

Academia, including academic institutions, think tanks, and research organizations that produce and use data, generate public knowledge, and educate people on data use.

Private sector companies, which often produce data as part of their business operations. Some of this can be valuable to public policy and public interest.

International and regional organizations, which sometimes require members to report data, for example SDG data for the United Nations. They can help by setting standards to make data more comparable, and often act as donors to support data production.

***

The INDS is supported by four pillars:

Infrastructure policies, such as equal access to the internet, a vibrant competitive internet provider market, and internet exchange points.

Laws and regulations that protect individuals, ensure cybersecurity, and manage institutions. Regulation should be independent, but stewarded by the government.

Economic policies, such as government strategy for data governance. Policies are crucial to establishing the value of data, helping them flow across borders and between companies.

Institutions that are set up to govern and safeguard data and monitor compliance. Watchdogs monitor public and private sector compliance.

***

The INDS is sustained by five foundations:

Human capital, meaning talented people with the right skills to use data, safeguard them, design policies, and hold power to account.

Trust in each other and the system to uphold the social contract for data to maximize value and prevent misuse.

Funding for data production, exchange, and use. This includes competitive salaries for people working in data roles, and funding for technological infrastructure.

Incentives for institutions and individuals to produce, protect, and share data. Sometimes mandates are required for transparency.

Data demand and a culture of data use. Valuing data is crucial for the right data to be produced.

Source: Excerpted from the 2021 World Development Report online materials, available at https://wdr2021.worldbank.org/stories/improving-data-systems/.

Data-related activities in World Bank operations thus represent a wide range of interventions that cut across sectors and themes to address data supply, demand, and the enabling context. Different aspects supported by World Bank operations, any of which may comprise a data-related solution for the purposes of this toolkit, include:

- *DATA + DATA SYSTEMS AND INITIATIVES*
    - *Data-driven systems, products, or methods that respond to users inside and/or outside the government*
    - *Developing or strengthening data collection or analytical methods*
    - *Data collection and acquisition from traditional or alternative/non-traditional sources*
    - *Geospatial data creation, analysis and visualization*
    - *Data use by government officials or by non-government actors*
    - *Data curation and data management: data management, documentation, and preservation; data cleaning, consistency, and data quality assurance*
    - *Data dissemination, accessibility, and discoverability: data platforms, websites, dashboards, or information systems (MISs); data-driven tools or applications (mobile/web-based apps)*
- *DATA LEADERSHIP and STRATEGIES*
    - *Political and institutional leadership*
    - *Strategies on strengthening data quality and availability, and using it effectively*
    - *Data-driven strategies to improve service delivery outcomes*
    - *Data management plans*
    - *Change management*
- *DATA GOVERNANCE AND INSTITUTIONS*
    - *Policy, legal, and regulatory framework around the collection, processing, use, portability, and disclosure of data or statistics*
    - *Data diagnostic or data assessments*
    - *Data rights*
    - *Institutional development or strengthening*
    - *Capacity building on data or statistics*
    - *Data ethics*
- *DATA ENABLERS AND STANDARDS*
    - *Data protection and data security*
    - *Digital identification*
    - *Data standards and interoperability frameworks*
    - *Digital service delivery standards*

    - *Geospatial information framework and data/service standards*
- *DATA SKILLS and DEMAND*
    - *Data skills and literacy of public servants, civil society, private businesses / entrepreneurs, media, and citizens*
    - *Building demand and capacity for data use*
- *DATA INFRASTRUCTURE AND ECONOMY*
    - *Physical infrastructure underlying data systems*
    - *Mobile and broadband connectivity*
    - *Data economy*

4. Methodology for mainstreaming data-driven approaches in World Bank operations

This section outlines a methodology for designing data-focused operations, or finding data-related entry points within larger projects at any point in the World Bank project cycle. By aligning with the project cycle (Figure 2) the methodology aims to help operational teams look for opportunities to address development challenges by bridging data and information gaps at any stage. The approach is holistic, considering all elements of an INDS and the enabling environment for a data-driven public sector, as discussed in Section 3. The resulting proposal for a data-related intervention may address one or more of these elements.

Teams may need to adjust the order of the steps in the methodology if the proposed intervention will be part of a larger or existing operation. For example, stakeholder consultations may come at a different point in the process if they have to follow the schedule of a larger project that may already be underway. The methodology is therefore best considered as a loose framework for identifying opportunities, and to help ensure that no critical aspects of planning, implementation, and monitoring are overlooked. Moreover, it is an iterative process - various assessments help to inform the scope and nature of the intervention being developed, and in turn, as the intervention begins to take shape, this informs the kinds of questions that need to be asked at the next stage.

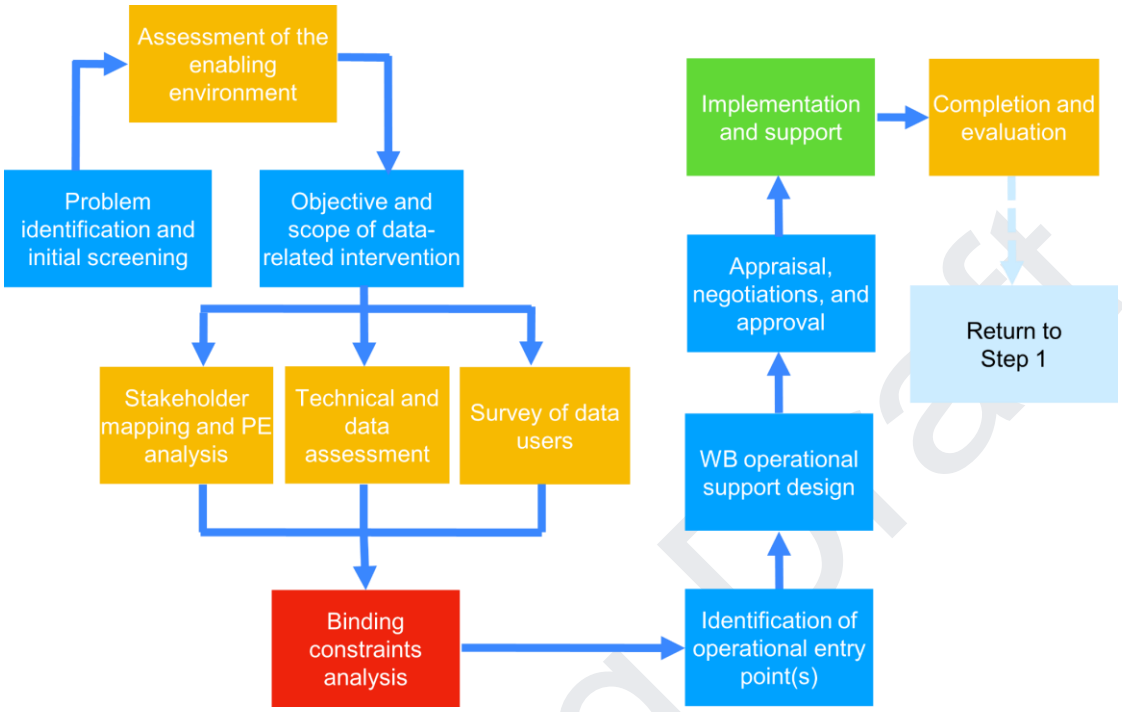
Figure 2. World Bank project cycle

Source: Reproduced from World Bank 2021b.

The methodology employs a problem-driven approach based on the idea that data solutions should respond to clearly defined development challenges (or technical problems), rather than being supply-driven. The methodology begins with defining a problem, and asking to what extent it could be addressed by identifying and bridging data gaps. It considers the enabling environment and political economy context and looks for binding constraints that could limit the impact of data interventions. After a thorough diagnosis of the issues, the methodology proposes the design of a concrete intervention, including the choice of World Bank instruments and modalities, to strengthen service

delivery through more effective incorporation of data. Figure 3 provides an overview of the methodology, which is explained in detail below within the framework of the project cycle.

Figure 3. Incorporating data in WB operations: flowchart of toolkit methodology



With a problem-driven approach, the idea is to initially keep options open as far as the nature and scope of a data-related intervention. It is not necessary to have a detailed idea in mind at the outset of the data-related solution that will ultimately be proposed. Once a problem is defined and various assessments have been conducted, it will be time to conceptualize an intervention that may or may not include any new data per se. Perhaps a new data system or data collection effort is the next obvious step. Or, the most impactful intervention may be to support a change in the governance framework (e.g., enacting a personal data protection law), improvement of data skills among citizens or public sector workers, greater transparency of existing data, or something else related to the enabling environment for effective data use. For this reason, the methodology talks about "data-related interventions" rather than focusing only on "data solutions" per se.

4.1 Identification stage

4.1.1 Problem identification

The first and most important step in this methodology is for the TTL to be able to answer these two questions clearly:

- What is the service delivery problem or development challenge that the World Bank aims to address in coordination with the client?
- Can this problem be addressed through a data-related intervention?

Once there are clear responses to these questions, a series of diagnostics will be launched. While the answer to the first question relies on the TTL's experience, knowledge and relationship with the client,

this toolkit provides guidance on the second question, including potential engagement with in-house or external data experts.

The TTL will determine the technical problem together with the client. This will stem from ongoing policy dialogue and diagnostics that the World Bank conducts in coordination with the client, as well as other relevant information sources and analysis. This may include SCDs and CPFs, preparatory work undertaken in the development of new operations, and formal and informal conversations with the client as well as other development partners.

It is important that the TTL and the client are able to break down the main technical issue into different components, as this will help the team to identify specific aspects of the problem that may improve with data-related interventions. For instance, a service delivery problem may reflect a combination of issues related to capacity (an insufficient number of appropriately skilled personnel), political will (public servants, particularly in leadership positions, do not have the incentives to tackle the identified problem), infrastructure or materials, weaknesses in the governance framework, information flows (public servants do not have timely access to information that will enable them to tackle the problem), impediments to access (service users do not have the information or the resources to access services), or other dimensions. The underlying factors will obviously vary depending on the sector, governance context, World Bank priorities, and other factors. For example, problems with public funds not reaching rural schools for infrastructure improvements will differ from problems in the health sector related to limited capacity to provide basic healthcare to specific population groups. For a non-comprehensive list of potential data-related entry points, see section 3.2.

Since each problem will be different in nature, clearly defining and describing its different components will be critical to assessing the relevance of data solutions. While data solutions can be helpful to address a wide range of technical problems, there are challenges for which they are not appropriate nor effective. Furthermore, it should be noted that data-related solutions are often only one element of a broader strategy that includes different kinds of support.

4.1.2 Initial screening to consider whether a data-related intervention is appropriate

Data solutions are most likely to be appropriate when the problem to be tackled relates to information flows. Problems related to information flows mainly reflect instances where greater availability and dissemination of government information is needed (e.g., citizens need to know about the availability of medication in specific health centers); citizen feedback needs to be incorporated into public decision-making (e.g., the government needs real-time information on how well a social program is reaching citizens); improved financial or programmatic information would improve public administration (e.g., an adequate financial or sectoral management information system is lacking); or increased exchange of data and information among government agencies is required (e.g., different databases need to be integrated to enable a unified digital citizen ID).

There may or may not be obvious obstacles to implementing a data-driven solution, however, oftentimes, there are clear barriers related to political will or constraining regulatory frameworks that could represent insurmountable obstacles for the effectiveness of a data-related solution. In the absence of evident insurmountable obstacles, various constraints and challenges will be considered later. Once a problem has been clearly defined, the TTL can reach out to a World Bank data

expert for help in determining, in coordination with the client, whether a data solution is appropriate (see Box 5).

Box 5. The World Bank Data Lab

The World Bank's Data Lab (https://wbdatalab.org/data-resources-directory/) provides a starting point for staff looking for colleagues and internal resources to help get a project going. It organizes information according to an "operations-focused data services map" that divides inquiries into four categories: 1) I need to collect new data; 2) I'm looking for existing data; 3) I need to manage data; and 4) I need insights from data. Through the Lab, it is possible to connect with Data Lab Leads who are director-nominated focal points for teams across all VPUs. It also includes a directory of team leads for specific data-related initiatives throughout the World Bank.

TTLs and Bank data experts can use the following questions to determine the relevance of a data-related intervention for a given technical problem. Again, any intervention should be part of a coherent INDS transformation, and should contribute to enhanced domestic capacity for data production, analysis, and use - standalone datasets or systems are not to be encouraged. Considerations may include:

- Is the necessary data/information already being produced?
- Are relevant agencies receiving the data/information in a timely fashion?
- Are relevant data users and/or decision makers able to make effective use of the data?
- If the existing data are insufficient, what type of data would be useful, and who or what would be the source?
- Would the problem be improved if the quality of available data were improved - if data were more timely, representative, accurate, or comprehensive?
- Would the problem be improved if the available data were more accessible or transparent to relevant stakeholders?
- Is existing data/information processed in order to generate relevant inputs for decision-making?
- Is the problem in a sector or area that entails substantial interactions between government and citizens?
- Do these interactions require greater exchange of data/information between citizens and government, whether from government to citizens or from citizens to government?

If a data-related intervention seems appropriate, it is useful to ask a few screening questions before proceeding:

- Are there specific stakeholders in the government with strong incentives to prevent the kinds of improvements in data or information flows under consideration?
- Do these stakeholders possess enough political clout to limit implementation of a data solution?
- Are there obvious deficiencies that impede data/information production, flows, or use in any of the general areas of:
    - Leadership and Strategy
    - Governance and Institutions
    - Data Skills and Demand for Data

- Data Enablers and Standards
- Data Infrastructure

For instance, barriers might include state secrets legislation that might hinder agencies from disseminating data/information. Note that deficiencies and challenges (for example, a regulatory framework that lacks specific policies and legislations conducive to improving information flows, such as RTI laws) are not necessarily insurmountable obstacles and may prompt ideas for DLIs to include in operations (more on constraints in section 4.2.6).

4.2 Preparation stage

Once it has been determined that a data-related intervention is appropriate and seemingly feasible, a series of diagnostics are launched and the process to design an effective solution begins.

4.2.1 Assessment of the enabling environment

Task team members can begin by reviewing various diagnostic reports, including those that are public as well as internal World Bank or government reports if available, to start building a picture of the state of the data ecosystem as it applies to the defined problem. The assessment should consider relevant overarching technical enablers (infrastructure, connectivity, availability of data skills), operational (business processes), institutional, and strategic elements. In the interest of conserving time and resources, the assessment relies mainly on a desk review of secondary sources and is intended to be rapid. A more detailed technical assessment that gets into the specifics of existing data and systems will come later in the process, once it has been determined that prospects for a data-related solution are strong enough to justify allocating more resources for project preparation.

To compile a picture of the local and sectoral context that is as comprehensive as possible, it is recommended that teams study reports from diverse sources and from different points of view. The review would ideally include reports and literature drawn from technical assistance work, civil society organizations, academia, the private sector, the media, and relevant industry associations. Ideally, consultations within the World Bank with operational staff who have experience in the country should complement the information gleaned from documentation.

Here are indicative examples of sources to be considered. These examples are not exhaustive and teams are highly encouraged to look widely depending on the context and sector.

- Research, diagnostic, and operational reports from multinational development partners and initiatives such as the World Bank, IMF, UN, OECD, ADB, AfDB, Open Government Partnership, and PARIS21; bilateral partners such as USAID, FCDO, GIZ; or foundations such as the Rockefeller Foundation, Gates Foundation, Mo Ibrahim Foundation, and others.
- Relevant country strategies such as National Statistical Development Strategies, National Development Strategies, digital transformation strategies, data strategies, and privacy policies.
- Papers from academia on the availability and use of data, data skills, ICT infrastructure, data governance, privacy policies, political economy analysis, and related topics.

- Civil society reports such as those from Transparency International, RTI International, Open Knowledge Foundation, Open Data Watch, Reporters without Borders, the World Wide Web Foundation, and others.

Some specific resources of note include:

- The World Bank's Open Government Data Toolkit, which includes a number of resources to help task teams, governments, and other stakeholders plan and implement open data initiatives.[62]
- The World Bank's Statistical Performance Indicators (SPIs), which measure the capacity and maturity of national statistical systems by assessing the use of data, the quality of services, the coverage of topics, the sources of information, and the infrastructure and availability of resources.[63]
- The World Bank's Statistical Capacity Indicator (SCI) scores, which is a composite score assessing the capacity of a country's statistical system.[64]
- The Open Data Inventory (ODIN) produced by Open Data Watch, which "assesses the coverage and openness of official statistics to identify gaps, promote open data policies, improve access, and encourage dialogue between national statistical offices (NSOs) and data users."[65]
- The Data Maturity Framework from Carnegie Mellon University, a questionnaire to help in planning a data-driven social impact project.[66]
- The Global Data Regulation Diagnostic (2021), a detailed assessment of laws and regulations on data governance covering both safeguards and enablers for data governance across 80 countries.[67]
- The World Bank's Data Governance Indicators Dataset, developed to support preparation of the 2021 WDR.
- The World Bank's GovTech Maturity Index (GTMI), which measures the key aspects of four GovTech focus areas—supporting core government systems, enhancing service delivery, mainstreaming citizen engagement, and fostering GovTech enablers.[68]
- The World Press Freedom Index produced by Reporters Without Borders.[69]

If the available reports and literature do not provide sufficiently comprehensive, timely diagnostics, teams will need to undertake their own assessment of opportunities and constraints related to using data effectively for better service delivery. Teams should consider whether it would be useful to use an existing diagnostic tool in collaboration with the client, depending on the scope and nature of the defined problem. The context might warrant a government-wide diagnostic tool such as a Digital Government Readiness Assessment (DGRA)[70] if there is a need for a comprehensive survey on digitization of government operations, an Open Data Readiness Assessment (ODRA)[71], a Global Data

[62] http://opendatatoolkit.worldbank.org/en/
[63] https://www.worldbank.org/en/programs/statistical-performance-indicators
[64] https://datatopics.worldbank.org/statisticalcapacity/
[65] https://odin.opendatawatch.com/
[66] http://www.datasciencepublicpolicy.org/our-work/tools-guides/datamaturity/
[67] Chen 2021.
[68] https://www.worldbank.org/en/programs/govtech/gtmi
[69] https://rsf.org/en/index
[70] https://openknowledge.worldbank.org/handle/10986/33674
[71] http://opendatatoolkit.worldbank.org/en/

Regulation diagnostic[72], or a Digital Economy Country Assessment (DECA). If the defined problem is based in a specific sector or theme, rather than affecting the entire government, then the assessment would have a narrower focus.

The following sample questionnaire highlights key aspects that should be part of an assessment, but can be adapted or expanded as needed:

1. **Leadership and Strategy**
    a. Does the government have a data strategy? Are there data management and/or change management plans?
    b. Do relevant frontline/sectoral/ministry agencies have a data strategy?
    c. Do these strategies have measurable goals linked to improving service delivery quality and inclusiveness?
    d. Are relevant stakeholders, secretaries, under-secretaries and others aware of and supportive of the strategy?
    e. Is there political will to implement the relevant strategies and to use data to improve service delivery? (If no, explore how the World Bank can use its technical expertise and convening power to explore alternative approaches to improving service delivery)
    f. What are the barriers to improving service delivery in this sector and making it inclusive?
    g. Is there potential for data-driven discourse within and outside of government to help overcome these barriers?
    h. If so, what kind of reforms would be politically feasible?
    i. What are the barriers to using data related to this service or sector?

2. **Governance and Institutions**
    a. Are there adequate resources, processes, technology and mandates in place to successfully implement relevant strategies and programs?
    b. Is the relevant ministry/agency or specific department within the agency adequately staffed, structured and funded to support the implementation of the strategy?
    c. Are privacy and cybersecurity policies and regulations in place and adequately enforced?
    d. Are policies and regulations around transparency and access to information in place and adequately enforced? Does the government or relevant ministry have an open data policy, open data portal or plans to launch any?
    e. Are other aspects of the data governance framework (policy, legal, regulatory framework) adequate to ensure safe and ethical use of data?
    f. To what extent is there coordination and information sharing horizontally (across government entities) and vertically (between national and various subnational levels of government)? (This can relate to coordination of data governance frameworks, e.g., are there conflicting data security regulations across different jurisdictions? What about data production/collection, e.g., are different government entities requesting that citizens provide duplicate information? Is data shared among relevant entities?)

[72] Chen 2021.

3. Data Enablers and Standards

a. What data related to the defined problem currently exists and who produces it (ministries, National Statistical Office, state or local government entity, service provider, development partners, private sector, other non-state actors)? Are there any relevant surveys available to help TTLs?
b. Are data protection and data security regulations and policies adequate and enforced?
c. Are there processes in place to monitor and evaluate relevant service delivery activities?
d. Are there any existing relevant data standards or interoperability frameworks? Is there an information framework for geospatial data? Is there a data management plan? Would a new data-related intervention need to coordinate with any other data efforts at the national, sectoral, or local level?
e. Are relevant GovTech technical enablers in place, which may include (depending on the defined problem) digital identification or digital payment capabilities?
f. Are there digital service delivery standards?

4. Data Skills and Demand for Data

a. Does the defined problem relate to a service in high demand among citizens or other key stakeholders (e.g. businesses, CSOs)? (Generally, the higher the demand among citizens, and the stronger the interaction between government agencies and this service, the higher chance of service delivery improvement.)
b. Is there high interest from media outlets in this sector/issue? (Interest from media can serve as a proxy for interest from citizens and hence increased likelihood of improvement.)
c. Is there high interest from development partners in this sector/issue?
d. What is the current perception of service delivery among citizens? (Media reports can serve as a proxy for this; if budget allows, it is recommended that task teams conduct social media sentiment analysis.)
g. What is the level of data literacy among the target population/users? (This can include specific subsets of citizens, civil society, public sector employees, private sector workers, media, etc.)
h. Are there universities or academic institutions training citizens in IT, data science, AI, or providing basic data literacy training?

5. Data Infrastructure and Economy

a. Where does the country rank in terms of telecommunications infrastructure and access[73], and how is it characterized in the GovTech Maturity Index (GTMI) and Statistical Performance Indicators (SPIs)?
b. Is there adequate digital and data infrastructure in place to help the government make the service more inclusive and efficient? (Look at the GTMI database for information on whether there is a government cloud, interoperability framework, enterprise architecture, web services/APIs, etc.)

[73] See https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx.

c. What is the level and reliability of digital connectivity and access to digital devices among the target data users (within and outside government)?

Once consultations have been held and the available documentation has been reviewed, the team can prepare an initial assessment report. The report would outline the current situation, key insights from other reports and diagnostics, and relevant initiatives that are already underway. Then the team could propose preliminary ideas about the nature of a data-related intervention that might be appropriate.

4.2.2 Objective and scope of data-related intervention

At this stage, a service delivery problem has been defined in general terms, and the team has concluded that there is potential to use a data-related intervention to address the problem. Furthermore, the assessment of the enabling environment has provided the team with an initial understanding of the context in which the intervention would be implemented. By tentatively defining the objectives, outputs, outcomes, and limitations of a potential intervention, the team will be better prepared to conduct consultations and analysis during the next stages of this process.

4.2.2.1 Strategic objectives, outcomes and outputs

The next step is for the task team, in collaboration with the client, to begin to conceptualize a data-related intervention by defining its strategic objective, in the context of the service delivery problem. The objective should be feasible and responsive to the problem and the enabling environment. Furthermore, the team should describe the concrete outputs and outcomes that are envisioned. At this point, these are all tentative ideas, and the team should remain flexible about the nature of the intervention, as subsequent steps will provide more information about what is realistic and potentially impactful. Nevertheless, it is important to formulate a starting point.

Example: **Strategic Objective:** Improve accountability around equitable delivery of healthcare, where accountability rests on a framework of timely data, feedback loops, and enhanced transparency, to inform resource allocation and administrative decisions.

Outputs:

- Review of healthcare data collection methodologies and business processes, as well as quality of existing and new datasets
- Support for development of IT tools and capacity (including skills development)
- Technical assistance supporting implementation of new methodologies, greater transparency, and user feedback mechanisms

Outcomes: Increased accessibility and usefulness of administrative data and user feedback for decision making and program management, and increased public transparency

4.2.2.2 Theory of change

As part of this process, teams should specify how the outputs and outcomes will contribute to achieving the strategic objective.

Example: **Theory of Change:** Ensuring equitable delivery of healthcare depends on high-quality data on health system usage and outcomes among different groups, and on government accountability for performance, which can be developed through enhanced capacity for data-driven management and decision making, more systematic consultation with citizens, and greater public access to information and data.

4.2.2.3 Scope and limitations of a data-related intervention

Finally, the team will outline the scope and limitations of the potential data intervention. This is important in order to manage stakeholder expectations, including those of the client, the World Bank, and others. The scope of a data-driven intervention can be stated in the following terms:

- *Geographic reach:* a data solution can be implemented across the country, or in specific regions, cities, etc.
- *Reach within the public administration:* a data solution can help address issues relevant to the whole public administration, or specific sectors or cross-cutting themes. Additionally, its coverage may impact all public officials, or only certain groups such as frontline public officials, teachers, etc.
- *Target beneficiaries:* the beneficiaries can be citizens at large, or specific groups such as vulnerable groups, refugees, etc.
- *Target users:* the individuals or organizations/entities making use of the data. The users may or may not be the same as the beneficiaries. For example, if the intervention is to make information on vaccine availability more accessible, the users and beneficiaries are citizens in need of vaccines. If the intervention aims to generate administrative data to improve schools, the users are government analysts and decision makers, while the beneficiaries are students and their families.
- *Components of processes:* a data solution can help to improve certain processes, and may address the process as a whole, or specific components.

By defining the scope and limitations of the initiative, the team will narrow the focus of the subsequent steps and make progress toward a project plan that is achievable and realistic.

4.2.3 Stakeholder mapping and political economy analysis

Next, the team should conduct a stakeholder mapping exercise to understand the various stakeholders that may be directly or indirectly impacted. The stakeholder mapping will consider what, if any, role each stakeholder might play, and how the incentives stakeholders face might shape their behavior. The stakeholder mapping process includes the following steps:

1. Make a list of key stakeholders (individuals in significant leadership positions, and institutions/agencies), grouped by category:
    a. Government, potentially including: IT ministry, digital government agency, statistical agency, ATI authorities, sectoral ministries, and others;
    b. Civil society, potentially including: local NGOs, academia, media, and citizens as a whole or by group (based on geographical location, gender, vulnerability, disability status, whether they are parents, etc.);
    c. Private sector: companies and industry groups, civil society organizations; and

   d. External partners such as donors and international NGOs.
2. Create a short description of each of the stakeholders listed, including their main interests in the defined problem.
3. Identify relevant incentives - does each stakeholder want to solve the defined problem, or do they benefit from the status quo? What incentives do they face to support it (positive and negative)?
4. If a data-related initiative were successfully implemented, how would each stakeholder be impacted (positively or negatively)?
5. What potential role might each stakeholder play in a data-related initiative related to the defined problem (types of roles include leading, planning, implementing, advising, financing, monitoring, opposing, undermining, etc.)? What would their level of engagement be (high, medium, low)?
6. How much influence does each stakeholder have in determining the results of a potential data solution?
7. How are these stakeholders connected with each other, and what potential alliances might lend support to a data solution (or might block its successful implementation)?

After answering these questions, teams can prepare a stakeholder and political economy brief capturing the responses to the above questions and highlighting strengths, weaknesses, opportunities, and threats represented by the stakeholder map. The analysis would be grounded in the specific sector and context of interest. This brief would enable the team to develop a politically and economically responsive data solution and lay the groundwork for a multistakeholder engagement strategy. At this point, teams may need to adjust the scope and objective (see section 4.2.2) to respond to the political economy analysis. For example, if the initiative concerns transparency and there is strong resistance to the open data agenda, it may be more productive to frame the intervention as part of the digital economy or public sector modernization agendas.

4.2.4 Technical and data assessment

At this stage, the team should conduct a technical and data assessment to evaluate existing infrastructure, systems, data initiatives, data quality and availability, and capacity related to the defined problem. Any red flags about government readiness that have surfaced in the previous steps should be examined in more detail during the technical assessment. The team may need to bring in other experts for this step if team members do not have sufficient data science expertise to judge the quality of existing technical capacity and systems. The technical assessment will require close coordination with the government's relevant data and systems experts.

Designing a feasible project plan depends on a solid understanding of what functionality or performance is possible with current systems, taking current staff capacity into account. The technical assessment should consider whether a data-related solution can make use of existing technical systems and capacity, or whether a new system (such as a database, website, app, or other solution) is required. A third possibility is that an existing system can be modified or expanded, such as by adding a new module or feature, but sometimes it is ultimately more cost-effective to start from scratch. In addition to hardware, software, and infrastructure, the technical assessment should consider the skills and capacity of existing government staff. Are there enough adequately skilled staff to take on a project of the scope that is envisioned, and if not, what would be required to build a team

with the necessary capacity? In some cases recruitment may be a significant challenge, and interviews with key informants may shed light on how difficult it would be to assemble a team.

The assessment should look closely at the data that already exists related to the defined problem, if there is any, and how often it is being produced, as well as any new or planned data initiatives. The team can assess the accuracy, comprehensiveness, and timeliness of the data, as well as how representative it is of the relevant population. The purpose of the data assessment is to ensure there is clarity on what data is currently being produced (and what plans are underway) to ensure that the project plan addresses gaps and problems appropriately, and avoids duplicating existing efforts.

After conducting the technical and data assessment, the team should cross-reference the findings with those from the assessment of the enabling environment (section 4.2.1), which covered relevant government policies, regulations, and strategies. It is important to look at the technical assessment in light of the existing governance framework to consider not only what solutions are technically possible, but also what is legally possible, and where policy or regulatory changes are needed (comparing the existing governance framework to regional or global policies may be useful). Again, it may be that the proposed intervention does not involve actual data, but rather focuses on some aspect of the enabling context.

4.2.5 Survey of data users

While the stakeholder mapping and political economy analysis focused on key individuals and organizations relevant to implementation of a data-related solution, this section focuses on one specific group: the intended data users. First, task teams should identify the key data user group(s), including users both inside and outside the government. Then, the goal is to better understand what users would need for the data-related initiative to have the desired impact on the service delivery problem. This step should also take the World Bank's Strategic Framework for Mainstreaming Citizen Engagement in World Bank Operations[74] into account.

A survey of data users can help task teams to understand current data availability and usage practices, and the challenges faced by data users in their work related to the identified service delivery problem. This includes the extent to which they use government information, the main channels they use to access it, and/or why they do not use existing government information. The survey will gather information about awareness and capacity to access and use relevant government information, and explore potential ways to ease access. Furthermore, it will contribute to pinpointing data gaps.

Task teams can create a survey according to the technical problem being addressed and the type of data solution(s) being considered. It would mainly focus on the following general areas:

1. Government data and information that users already access.
2. Main purpose of using government data and information.
3. Frequency with which government data and information are accessed and used.
4. Data and information that users do not access but that would be useful to them.
5. Channels or mechanisms through which they access data and information, including existing non-government information that they find useful.

[74] See https://www.worldbank.org/en/topic/citizen-engagement#2.

6. Other potential mechanisms that could be leveraged.

In terms of conducting the survey, different approaches can be considered depending on the number of users, including both qualitative and quantitative methods. For example, if the target users are all within one government agency, interviews with a representative selection of users may be the simplest method. If citizens are the target users, it might be appropriate to consider low-resource methods such as online surveys targeted through social media, or more resource-intensive methods such as phone or in-person questionnaires through polling firms. These are not mutually exclusive, as a combination could be used. Survey results will provide key inputs into the design of a data-related intervention.

4.2.6 Binding constraints analysis

Through a binding constraints analysis, the data team will synthesize and build upon information and findings from the assessment of the enabling environment, the stakeholder mapping and political economy analysis, the survey of data users, and the technical and data assessment. The team has now looked at the problem and the initial conditions from many angles. Now it is time to take all of these findings into account to determine what is feasible. By conducting a binding constraints analysis (Box 6), the team will identify the main issues that have the potential to undermine the effectiveness of data-related solutions. Furthermore, this analysis enables the team to develop and prioritize mitigating measures. The critical question to keep in mind is, why has the proposed solution not been attempted before, or if it has, why was it not successful? If the team has not already done so, at this stage it is critical to conduct semi-structured interviews with key stakeholders identified through the stakeholder mapping to get a comprehensive understanding of relevant challenges and opportunities.

Box 6. What is binding constraints analysis?

Binding constraints analysis identifies underlying barriers to reform, whether technical, political, institutional, or otherwise. This analysis will then contribute to the development of a data-related proposal that will be tailored to local needs. Steps include:

Step 1. Formulating a clear statement of the problem. See section 4.1.

Step 2. Developing a decision tree. The decision tree will lay out all the possible explanations for the problem and identify the binding constraints to resolving the problem. Two to five high-level constraints are identified at the outset as a starting point and are then articulated into increasingly specific sub-levels or branches. Each level should offer an explanation for the problem identified above. The process is often iterative: information emerges that influences earlier assessments and adjustments are required. As the decision tree develops, the team may discover that some constraints have the same underlying cause.

Step 3. Identifying binding constraints. Binding constraints typically meet two conditions: (1) when addressed they should result in immediate gains; and (2) if they are not addressed first, other reform efforts are also likely to fail. Identification of binding constraints must rely on the evidence gathered, and supplementary consultations as needed, though the team's subjective analysis will also be necessary, and should be transparent and open to discussion. It is useful to identify any historical changes in the binding constraint, when possible. If a factor is a binding constraint, then historical changes in the constraining factor should have produced significant changes in actual outcomes.

Source: World Bank Innovations in Tax Compliance team, unpublished draft.

Binding constraints related to a data-related intervention can stem from either the supply or the demand side of data and information. From the supply side, constraints can relate to:

- ***Political economy***. This refers mainly to whether there is political will to move forward with a data solution or not. Are there high level champions of the proposed intervention? Are there stakeholders in opposition who have enough political clout to essentially veto certain actions? To determine this, the team will rely on the results of the stakeholder mapping and political economy analysis (section 4.2.3) and look at the political economy landscape and the incentives of key actors. Then, the team can develop strategies to foster "commitment, coordination, and cooperation"[75], for example by leveraging the World Bank's convening power.
- ***Lack of capacity.*** There may be a lack of appropriate skills or capacity in the civil service or among the officials ultimately responsible for implementing the proposed intervention. Even if there is political will, a lack of capacity, if not addressed, can impede successful implementation. The technical assessment (section 4.2.4) will be critical in determining if capacity shortcomings exist, and for the team to develop mitigating measures.
- ***Limited infrastructure.*** Data and information solutions require specific infrastructure in order to function properly and enable the collection, processing, use, and publication of data. The technical assessment will generate the necessary inputs for the team to understand what limitations there are in terms of infrastructure, and ensure that these are addressed.
- ***Inadequate governance framework.*** Policy and regulatory frameworks - from legislation on matters related to the right to information or privacy, to secondary regulations - provide the governance context for a data solution. The team, with inputs from the assessment of the enabling environment, will determine whether the governance framework is adequate or not, and propose appropriate modifications.

Constraints to the effectiveness of investing in data can also emerge from the demand side of data and information. In other words, whether users of a potential data solution can and will access and process information will be a key determinant for success. Constraints stemming from the demand side can be grouped as follows:

- ***Limited access to data/information.*** If citizens (or other non-state actors) are the intended data users, low levels of data literacy and access among the target group may pose binding constraints. Data solutions often entail citizens accessing information in order to take action or provide feedback related to public services. Access problems can stem from a lack of information disclosure, lack of accessible channels (e.g., digital-only options for rural communities with limited internet connectivity), or lack of an accessible format (e.g., information not available in the local language, in PDF rather than reusable format, etc.). Through the analytical survey of data users (see 4.2.5) the team will be able to understand these limitations, and by linking them to supply side constraints, a robust set of mitigating measures can be developed.

[75] As described in WDR 2017, *Governance and the Law*. See World Bank 2017b.

- ***Limited awareness or interest.*** At times, information is readily available, but target users might not know that it is there, or know how it could be helpful for them. The analytical survey of data users will shed light on this and help the team in designing awareness-raising campaigns that stimulate interest. At the same time, a user-centric approach is iterative. If users consistently fail to engage with the data, it may be that there is an underlying problem that needs to be addressed. It may be that 1) the information is not high priority to users, or they do not have time to pay attention to it; 2) they do not believe their use of it will have any impact (lack of trust in government responsiveness); 3) they doubt its quality and accuracy; 4) the information is too old to be useful; 5) they have concerns about privacy or retaliation if they engage; or another reason.
- ***Lack of capacity.*** Finally, the use of data solutions either to generate or process information requires some technical capacity. Target users, particularly if they include vulnerable groups, may not possess the technical skills to engage with a potential data solution. Relying on the analytical survey of data users, the team will be able to pinpoint these constraints and embed capacity building components into the proposed data solution, and/or incorporate analog components or outreach mechanisms (e.g., using phone as well as online surveys).

In sum, the binding constraints analysis will enable the data team to identify the main constraints that need to be taken into account. An agile approach and mentality is useful. If the constraints are particularly intractable, it may mean that the intervention as envisioned is not feasible and should be abandoned, or that major modifications to the strategic objective, outputs, and outcomes should be considered. If the decision is to move forward, the team should outline ways to mitigate or address the constraints during implementation. Any potential data-related solution should emerge from the binding constraints analysis.

4.2.7 Identification of operational entry point(s)

At this point, TTLs should review relevant operational entry points for the proposed intervention. However, as mentioned earlier, this review can be conducted at a different stage, if it is more convenient depending on the status of the project. The World Bank's Country Partnership Framework, Systematic Country Diagnostic, and regional strategic priority notes could point to options. Similarly, country strategies such as national development strategies, sectoral strategies, data strategies, and digitization strategies can provide the rationale and entry points for the Bank to engage. There may also be international initiatives that suggest entry points, for example if the country is a member of the Open Government Partnership (OGP) or the Extractive Industries Transparency Initiative (EITI).

The next step, choosing a Bank financing mechanism, requires reflection on the pros and cons of different instruments given the context:

- **Program for Results:** This instrument leverages the country's institutions and processes, and disbursement of funds is linked directly to the achievement of specific program results (Disbursement Linked Indicators, or DLIs), which could be a data-related solution(s).
- **Development Policy Operations:** This instrument emphasizes country ownership and alignment, stakeholder consultation, donor coordination, and results. It requires the creation of legally binding prior actions in the financing agreement to achieve specific program objectives.

- **Investment Project Financing:** When more World Bank involvement and oversight is preferred, this instrument supports a broad range of sectoral and government-wide activities that can provide entry points for data-related solutions.
- **Advisory Services and Analytics (ASA):** Advisory services and technical assistance, either as a stand-alone undertaking or in conjunction with one of the above operations, is a common way to support client governments in implementing data-related solutions.

4.2.8 World Bank operational support design

Once an appropriate financing instrument and entry point has been identified, the team can prepare a thoroughly informed data-related intervention that is feasible and impactful. It may be a standalone data-related project or a component of a larger existing or planned project, building on all of the consultations and assessments conducted to this point. Revisiting the strategic objective, outputs, and outcomes identified earlier and adjusting as needed forms the basis for 1) the project plan and 2) the definition of relevant results indicators, or DLIs. By clearly defining the anticipated impact of the intervention on service delivery and citizens, even if it is an upstream initiative focused on government systems and processes, the resulting intervention will come from a citizen-centric approach.

The project plan needs to identify a clear "owner" of the initiative on the client side, and to consider long-term sustainability - will the intervention last beyond the World Bank's financial and operational involvement? What are the ongoing financial and technical requirements (e.g., updating data per the planned schedule, processing and analyzing data, etc.)? Many data initiatives focus almost exclusively on the initial production or publication of data, without attention to ongoing needs, such as consistent communication with stakeholders, change management within government agencies, or technical system maintenance. How can the data-related solution be incorporated into government business processes and systems to improve the chance of sustainability? How will the initiative be monitored? After answering these questions, the team can prepare a proposed budget to include in the relevant project documents.

4.3 Appraisal stage and 4.4 Negotiations and board approval stage

At this point, project documents will go through the World Bank's review and approval process, and inputs from peer reviewers, the Country Management Unit, and World Bank management will prompt revisions to the project plan and budget. Other data-related interventions may be suggested and incorporated into the plan and results indicators. At the review stage, teams may also begin to think about adding data-related components to projects that do not have them already, perhaps prompted by questions around the availability of data for World Bank or government monitoring of the project, or due to the need for a stronger citizen engagement strategy. In this case, teams can undertake an accelerated version of the toolkit to this point to integrate data-related aspects into the project plan.

4.5 Implementation and support stage

4.5.1 Agile approach to implementation, and attention to change management

Once implementation begins, it is key to adopt an agile approach that responds to the evolving political economy context and to user needs as they become more clear. It will be critical for task teams to keep the problem definition and strategic objective in mind when adapting the project plan as conditions change. For example, the team may discover that existing data and systems are not as robust as believed, and more resources may be necessary to bolster them. When making decisions about how to adapt the workplan, teams will need to keep the problem definition and strategic objective front of mind to determine whether addressing such weaknesses is worth the investment, or if it would be more impactful to shift course and try tackling the problem in a different way instead. Teams may need to revisit assumptions from the preparation phase if warranted.

During the implementation stage teams should maintain a citizen-centric mindset. This means tracking whether relevant data are being used (e.g., website usage statistics), how often, for what, and by whom. Are data analytics underway? Are data resulting from the intervention being fed into decision-making processes, and what is the result? Are there any preliminary indications of improvement in service delivery?

Attention to non-technical issues is as important as technical ones. The team should monitor all aspects of the intervention, including technical status and progress, operational implications (impact on government business processes), strategic relevance (impact on larger government objectives), and institutional impact (state and non-stage capacity to engage, communication with stakeholders, etc.). Investing in change management is important and it may take time to overcome bureaucratic inertia and entrenched ways of doing things, as well as shifting from a government-centric to a citizen-centric mindset. Per the WDR 2021, "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."

4.5.2 Monitoring and the results matrix

Throughout implementation, task teams will collect and monitor relevant data and progress within the framework of the results matrix.[76] Examples of results indicators from actual World Bank projects are included in Annex 1. It is possible that teams may discover data gaps and potential for data-oriented interventions through the process of project monitoring. It may be possible to find new entry points at this stage, and to add data-related solutions through a project restructuring, or to begin to conceptualize a pipeline project that would build on the activities in the project at hand.

Working in fragility, conflict, and violence (FCV) settings poses a heightened challenge. For guidance on monitoring and evaluation in FCV countries, refer to the Geo-Enabling initiative for Monitoring and Supervision (GEMS)[77], which was launched by the World Bank's FCV group to help clients and Bank

[76] For general guidance, see Roberts and Khattri (2012), "Designing a Results Framework for Achieving Results: A How-To Guide". Useful resources are also available from the Development Impact Evaluation (DIME) group at https://www.worldbank.org/en/research/dime/data-and-analytics.

[77] See https://www.worldbank.org/en/topic/fragilityconflictviolence/brief/geo-enabling-initiative-for-monitoring-and-supervision-gems.

teams to use context-appropriate tools and methods for digital data collection and analysis. There is also the World Bank's operational guidance[78] for monitoring and evaluation (M&E) in climate and disaster resilience-building operations.

Multistakeholder and user consultations should also be an ongoing part of the monitoring process using qualitative (meetings, semi-structured interviews) and/or quantitative (survey) methods. It may be useful to form a user group or a multistakeholder steering committee, if this has not already been done earlier, to provide input and representative feedback on a repeat basis to inform implementation.

4.6 Completion/validation and evaluation stage

In this final stage, and in coordination with the client, other key stakeholders, and users, the team will reflect on lessons learned. Through the preparation of the Implementation Completion and Results (ICR) Report, the team will document these lessons. The findings may suggest ways to continue the work through other existing or pipeline operations, or to apply them to other countries with similar objectives.

[78] World Bank 2017c.

References

Beschel, Robert P., Blair James Cameron, Jana Kunicova, and C. Bernard Myers. 2018. "Improving Public Sector Performance : Through Innovation and Inter-Agency Coordination (English)." Global Report Public Sector Performance Washington, D.C.: World Bank Group. http://documents.worldbank.org/curated/en/833041539871513644/Improving-Public-Sector-Performance-Through-Innovation-and-Inter-Agency-Coordination.

Brynjolfsson, Erik, Lorin M. Hitt, and Heekyung Hellen Kim. 2011. "Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?" SSRN Electronic Journal, April 22, 2011. https://ssrn.com/abstract=1819486.

Chen, Rong. 2021. "Mapping Data Governance Legal Frameworks Around the World: Findings from the Global Data Regulation Diagnostic (English)." Policy Research Working Paper, No. WPS 9615. Washington, DC: World Bank Group.

Dabalen, Andrew, Kristen Himelein, and Carlos Rodriguez Castelan. 2020. "Data for Policy (D4P) Initiative." World Bank Poverty and Equity Notes Number 23. June 2020. https://openknowledge.worldbank.org/bitstream/handle/10986/33857/Data-for-Policy-D4P-Initiative.pdf?sequence=1&isAllowed=y.

Castelan, Carlos Rodriguez, Kristen Himelein, and Andrew Dabalen. 2020. "Evolution in the data ecosystem - an idea that's got legs." *Data Blog* (Blog), April 6, 2020. https://blogs.worldbank.org/opendata/evolution-data-ecosystem-idea-thats-got-legs.

Encarnacion, Jessamyn, Ramya Emandi, and Papa Seck. 2022. "It will take 22 years to close SDG gender gaps." Sept. 6, 2022. UN Women blog. Available at https://data.unwomen.org/features/it-will-take-22-years-close-sdg-gender-data-gaps#:~:text=As%20of%20June%202022%2C%20we,increase%20of%203%20percentage%20points.

Filippov, Sergey. 2014. "Data-Driven Business Models: Powering Startups in the Digital Age." Digital Insights.https://www.researchgate.net/publication/270508502_Data-Driven_Business_Models_Powering_Startups_in_the_Digital_Age.

G7. 2022. "G7 Foreign Ministers' statement on strengthening anticipatory action in humanitarian assistance." Press release, May 13, 2022. https://reliefweb.int/report/world/g7-foreign-ministers-statement-strengthening-anticipatory-action-humanitarian.

Grinspan, Delfina and Jesse Worker. 2021. "The Payoffs of open climate data." *Insights* (World Resources Institute blog), March 9, 2021. https://www.wri.org/insights/payoffs-open-climate-data.

Hunja, Robert. 2015. "STEPing ahead with procurement reform." *Governance for Development* (blog), October 8, 2015. https://blogs.worldbank.org/governance/steping-ahead-procurement-reform.

IFC (International Finance Corporation). 2019. "Digital Skills in Sub-Saharan Africa: Spotlight on Ghana." Washington, DC: IFC. https://www.ifc.org/wps/wcm/connect/ed6362b3-aa34-42ac-ae9f-c739904951b1/Digital+Skills_Final_WEB_5-7-19.pdf?MOD=AJPERES.

Jelenic, Michael Christopher. 2019. “From Theory to Practice: Open Government Data, Accountability, and Service Delivery.” Policy Research Working Paper No. 8873, World Bank, Washington, DC. https://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-8873.

Joliffe, Dean et al. 2021. “Under What Conditions are Data Valuable for Development?”. World Bank Policy Research Working Paper 9811. World Bank, October 2021. https://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-9811#:~:text=The%20conditions%20are%20that%20data,%2C%20confidential%2C%20and%20appropriate).

Lemieux, Victoria L., and Stephanie E. Trapnell. 2016. *Public Access to Information for Development: A Guide to the Effective Implementation of Right to Information Laws*. Directions in Development. Washington, DC: World Bank. doi: 10.1596/978-1-4648-0879-1. License: Creative Commons Attribution CC BY 3.0 IGO. http://documents1.worldbank.org/curated/en/983941467996646873/pdf/106580-PUB-ADD-SERIES-DOI-PUB-DATE-ISBN-ABSTRACT-AUTHORS-PUBLIC.pdf.

Manyika, James, et. al. 2013. “Open Data: unlocking innovation and performance with liquid information.” McKinsey. 2013. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information.

Milusheva, S., R. Marty, G. Bedoya, S. Williams, E. Resor, and A. Legovini. 2021. “Applying machine learning and geolocation techniques to social media data (Twitter) to develop a resource for urban planning.” PLoS ONE 16(2): e0244317. https://doi.org/10.1371/journal.pone.0244317.

Mukhi, Neha, Suneira Rana, Sara Mills-Knapp, and Eskedar Gessesse. 2020. “World Bank Outlook 2050: Strategic Directions Note. Supporting Countries to Meet Long-term Goals of Decarbonization.” World Bank, Washington, DC. https://openknowledge.worldbank.org/bitstream/handle/10986/33958/149871.pdf?sequence=3&isAllowed=y.

Muralidharan, Karthik, Paul Niehaus, Sandip Sukhtankar, and Jeffrey Weaver. 2021. "Improving Last-Mile Service Delivery Using Phone-Based Monitoring." *American Economic Journal: Applied Economics* 13 (2): 52-82. DOI: 10.1257/app.20190783.

Neagu, Victor. “Boosting Budget Transparency in Moldova.” *Voices* (blog), December 5, 2013. https://blogs.worldbank.org/voices/boosting-budget-transparency-moldova.

Porter, Michael E. and James E. Heppelmann. 2015. “How Smart, Connected Products are Transforming Companies.” *Harvard Business Review,* October 2015. https://hbr.org/2015/10/how-smart-connected-products-are-transforming-companies.

Roberts, Dawn and Nidhi Khattri. 2012. “Designing a Results Framework for Achieving Results : A How-to Guide.” Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/32158.

SDSN TReNDS. 2018. “BudgIT Empowers Nigerian Citizens Through Open Data.” https://www.data4sdgs.org/sites/default/files/2018-09/BudgIT%20Case%20Study_Final.pdf.

Startup Genome and Global Entrepreneurship Network. 2018. "State of Startup Ecosystems 2018." https://startupgenome.com/reports/state-of-startup-ecosystems-2018.

Ul-Aflaha, McNeil, and Kumagai (undated). "Building Blocks and New Frontiers for Open Government." Washington, DC: World Bank. https://openknowledge.worldbank.org/bitstream/handle/10986/33345/Building-Blocks-and-New-Frontiers-for-Open-Government.pdf?sequence=1&isAllowed=y#page=26&zoom=100,0,0.

Verhulst, Stefaan and Andrew Young. 2016. "Battling Ebola in Sierra Leone: Data Sharing to Improve Crisis Response." GovLab and Omidyar Network, January 2016. Available at https://odimpact.org/case-battling-ebola-in-sierra-leone.html.

World Bank Group. 2015. "World Bank Group Gender Strategy (FY16-23) : Gender Equality, Poverty Reduction and Inclusive Growth." Washington, DC: World Bank. https://openknowledge.worldbank.org/entities/publication/95293946-99d3-5ef3-92a0-1fd78a3f11e6.

World Bank. 2017a. "World Bank Support for Open Data 2012-2017." Washington, DC: World Bank. https://openknowledge.worldbank.org/entities/publication/4e9ee0fe-ea11-53ca-8646-3e21f5f0b91c.

World Bank. 2017b. *World Development Report 2017: Governance and the Law*. Washington, DC: World Bank. https://www.worldbank.org/en/publication/wdr2017.

World Bank. 2017c. "Operational guidance for monitoring and evaluation (M&E) in climate and disaster resilience-building operations (English)." Washington, D.C. : World Bank Group. http://documents.worldbank.org/curated/en/692091513937457908/Operational-guidance-for-monitoring-and-evaluation-M-E-in-climate-and-disaster-resilience-building-operations

World Bank. 2018. *Engaging Citizens for Better Development*. Independent Evaluation Group, Washington, DC: World Bank. https://ieg.worldbankgroup.org/sites/default/files/Data/Evaluation/files/Engaging_Citizens_for_Better_Development_Results_FullReport.pdf.

World Bank. 2019a. "Information and Communications for Development 2018: Data-Driven Development." Information and Communications for Development. Washington, DC: World Bank. doi:10.1596/978-1-4648-1325-2.

World Bank. 2019b. *World Development Report 2019: The Changing Nature of Work*. Washington, DC: World Bank. http://documents1.worldbank.org/curated/en/816281518818814423/pdf/2019-WDR-Report.pdf.

World Bank. 2020. "Unraveling Data's Gordian Knot: Enablers & Safeguards for Trusted Data Sharing in the New Economy." World Bank, Washington, DC. License: Creative Commons Attribution CC BY 3.0 IGO.

World Bank. 2021a. *World Development Report 2021: Data for Better Lives*. Washington, DC: World Bank. https://www.worldbank.org/en/publication/wdr2021.

World Bank. 2021b. “World Bank Sustainable Development Bond Framework.” Updated March 2021. https://thedocs.worldbank.org/en/doc/43b360bfda1e6e5b8a094ef2ce4dff2a-0340012021/original/World-Bank-IBRD-Sustainable-Development-Bond-Framework.pdf.

World Bank. 2022. Kenya, Rwanda, Tanzania, East African Community, and Africa Union - Eastern Africa Regional Statistics Program-for-Results Project (English). Washington, D.C. : World Bank Group. http://documents.worldbank.org/curated/en/966141652556689210/Kenya-Rwanda-Tanzania-East-African-Community-and-Africa-Union-Eastern-Africa-Regional-Statistics-Program-for-Results-Project.

Young, Andrew and Stefaan Verhulst. 2017. “Aclímate Colombia: Open Data to Improve Agricultural Resiliency.” USAID, mSTAR, FHI360, and GovLab, July 2017. Available at https://odimpact.org/files/case-aclimate-colombia.pdf.

Yukins, Christopher R. and Steven Kelman. 2022. “Overcoming Corruption and War - Lessons from Ukraine’s ProZorro Procurement System.” NCMA Contract Management Magazine, July 2022, p. 22. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4230322.

Annex 1. Examples of data-related project result indicators and verification criteria

Overview

This annex provides examples of specific data-related result indicators from World Bank projects. They are drawn from the results frameworks of World Bank Project Appraisal Documents (PADs) for the purpose of illustrating a broad range of targets and verification criteria related to strengthening the use of data for development gains. The sections below correspond to the categories of data-related interventions outlined in the main document.

Most of the sections provide a list of individual indicators from a range of projects. The section on Governance and Institutions takes a different approach than the others since it is a more holistic category - instead of listing individual indicators it provides the Project Development Objective (PDO) and PDO-level indicators for a number of projects that are entirely focused on data. It also includes non-World Bank examples of government targets and objectives related to the legal and policy framework for data governance, with suggested wording for prior actions/DLIs that can be adapted by task teams as needed.

The examples included here are not intended to be comprehensive and are not necessarily endorsed as best practice. Rather, they are intended to serve as a foundation for brainstorming about different ways to mainstream data in World Bank operations across different sectors and contexts.

1. Data + Data Systems and Initiatives

Examples of results indicators from World Bank projects include:

- Number of climate relevant datasets that contribute data to the climate data platform (number) [Digital Maldives for Adaptation, Decentralization, and Diversification: P177040]
- Improved hospitalization and mortality surveillance (35 hospitals piloted hospitalization and mortality surveillance) [Systems Reform Endeavours for Transformed Health Achievement in Gujarat Program for Results: P178252]
- Integrated dashboard for municipal transfers with operational information and accessible to the public (Yes/No) [Progestão Mato Grosso: P178339]
- Data collection that supports understanding on women's use and needs of shelters (Survey with questions on women's usage and needs of shelters, and their perceptions of safety and security in shelters will be conducted) [Bangladesh Resilient Infrastructure for Adaptation and Vulnerability Reduction: P173312]
- Rapid labor market survey results and reports are available and publicly disclosed (Text) [Guyana Strengthening Human Capital through Education Project: P177741]
- Tracer study results are available and published (Text) [Ibid.]
- TVET Management Information System, linked to the EMIS, is operational (Text) [Ibid.]
- Community Engagement: Survey of participant satisfaction administered, and feedback addressed (Text) [Ibid.]

- Efficiency gains in data production as measured by per-interview costs of surveys (Percentage) [Angola Strengthening Statistical Capacity: P178043]
- Conduct cartography for housing and population census (Text) [Strengthening the Statistical System of Uzbekistan: P173450]
- Number of properties updated, validated, or registered in the digital cadastre register (Number) [Kosovo: Real Estate & Geospatial Infrastructure Project: P164555]
- Prior action: The Recipient has, in accordance with paragraph 16 (x) of the Letter of Development Policy, collected data on the condition of national, district, urban and community access roads and established a baseline, and has established and operationalized a monitoring and evaluation framework to measure key road sector performance indicators [Uganda PRSC9: P097325]
- *Disbursement-Linked Indicator (DLI):* Increasing accessibility and usability of statistical and administrative data (Verification: increase in national Open Data Inventory [ODIN] score from *x* to *y*) [Jordan Inclusive, Transparent, and Climate Responsive Investments Program for Results: P175662]
- *Development Policy Loan (DPL) results indicator:* Establishment of an early warning system that allows the general public to report forest fires (Yes/No) [Paraguay Green and Resilient DPL: P178285]

2. Leadership and Strategy

Examples of results indicators from World Bank projects include:

- Publication of a Data Strategy and Action Plan (Yes/No) [Burundi Digital Foundations Project: P176396]
- Increased usage of data for monitoring as measured by the number of public programs of the National Development Plan that are monitored following established standards and protocols (Number) [Angola Strengthening Statistical Capacity: P178043]
- Instruments for climate planning and action informed by improved climate data and analytics (number) [Digital Maldives for Adaptation, Decentralization, and Diversification: P177040]
- Published barometer reports on flagship services (*specific sectors*) and ministerial decisions taken based on report recommendations (Yes/No) [GovTech: Digital Transformation for User Centric Public Services, Tunisia: P168425]
- Statistics Council meetings held (Number) [Strengthening the Statistical System of Uzbekistan: P173450]
- *DLI:* Development of coordinated, data informed, and results-oriented financing and budget plan for road safety (Text) [India State Support Program for Road Safety: P177668]
- *DLI:* Strengthened systems for data quality and performance tracking for comprehensive primary health care (Verification: Master Plan developed and adopted; data quality roadmap including verification system developed; public reporting) [Systems Reform Endeavours for Transformed Health Achievement in Gujarat Program for Results: P178252]

3. Governance and Institutions

As mentioned in the overview of this annex, this section takes a different approach than the others, providing a high-level snapshot (the Project Development Objective, or PDO, and PDO-level indicators)

from a few World Bank projects focused on strengthening data-related capacity and the use of data for decision making. Following these examples, a list of non-World Bank examples related to data governance and institutions is included, with suggestions of generic prior actions/DLIs that can be adapted as needed.

Harmonizing and Improving Statistics in West Africa (P169265)

Project Development Objective (PDO): Strengthen the statistical systems of participating countries [Ghana, Liberia, Sierra Leone, Togo, Cabo Verde, Burkina Faso, Cote d'Ivoire] and regional bodies in Africa to harmonize, produce, disseminate and enhance the use of core economic and social statistics. PDO indicators are:

- Modified Statistical Capacity Index (SCI 2.0) (disaggregated by participating countries) (percentage).
- Share of beneficiary countries (National Statistical Office) that score at least 30 out of 50 for the Institutional Capacity Index (ICI) (percentage).

The following indicators will also be used to monitor production and dissemination activities:

- Share of comparable core economic and social statistics indicators produced and available for open access (percentage).
- Share of comparable core economic and social statistics datasets produced and available for open access (disaggregated by country) (percentage).
- Share of statistical products produced and published by the NSO following international standards and a predetermined release calendar (disaggregated by country) (percentage).

Central Africa Republic Data for Decision Making (P179053)

Project Development Objective (PDO): To increase the capacity of the national statistical institute (ICASEES) to produce and publicly disseminate statistics through data recovery, institutional development, and support to data production. The main expected outcomes of the project are the following:

- A secure electronic data archive exists (yes/no)
- Rebased national accounts are being published (yes/no)
- Statistical data sets are publicly disseminated online (number)

Strengthening the National Statistical System in Ecuador Project (P178564)

PDO: To improve the national statistical capacity of Ecuador in the production and dissemination of timely and high-quality economic and sociodemographic statistics for evidence-based policymaking. The PDO indicators are:

- Improved statistical capacity as measured by a composite score derived from the World Bank Statistical Performance Indicator (SPI) methodology. This indicator is aimed to measure the attribute of improvement in the national statistical capacity of Ecuador.
- Increased use of statistics as measured by the number of downloads of datasets supported by the project.

- Increased transparency on data production methodologies and the quality of analysis measured by the publishing of 12 relevant reports based on statistics supported by the project.
- Increased availability of sector administrative data for statistical purposes.
- Number of gender indicators produced using recently adopted international standards on labor statistics and enhanced measurement.
- Number of Risk and Climate Change indicators produced using recently adopted international environmental standards.
- Satisfaction rate of statistical users of the statistics produced within the framework of the project.
- Percentage of adoption by INEC of relevant recommendations made by users of the statistics produced within the framework of the project aimed to increase their timeliness, coverage, and relevance.

OECS Data for Decision Making (P174986)

PDO: To improve the capacity of Participating Eastern Caribbean Countries to (i) produce and publicly disseminate statistical data for country and regional level analytics; and (ii) provide immediate and effective response to an Eligible Emergency. The Project will strengthen NSSs through (i) statistical modernization and capacity building and (ii) data collection, analysis, and dissemination, with a focus on the population and housing census, living conditions surveys, labor market surveys, and agricultural census. PDO level indicators are:

- Customized Statistical Performance Indicator, average value for Participating Eastern Caribbean countries (baseline 42.5; target 90.0)
- Average number of household survey datasets per Participating Eastern Caribbean country accessible via the OECS Regional Microdata Catalog (baseline 0; target 10).
- Number of OECS regional publications featuring harmonized statistics across Participating Eastern Caribbean countries (baseline 0; target 4).
- Number of regionally harmonized methods for data collection and analysis published by the Regional Data Governance Council (baseline 0; target 5)

Eastern Africa Regional Statistics Program-for-Results (P176371)

PDO: To strengthen the regional harmonization, dissemination and use of core economic and social statistics for Kenya, Rwanda and Tanzania. This is aligned with the Strategy for Harmonization of Statistics in Africa 2017 – 2026 (SHaSA2). The achievement of the PDO level results will be measured by three high-level indicators. The indicators cover harmonization, availability and use of statistics:

- Harmonization: Number of published statistical dimensions that are harmonized according to the applicable regional technical guidelines by participating countries.
- Availability: Average score for the customized Statistical Performance Index (SPI) across participating countries.
- Use: Number of strategy and policy documents utilizing or referring to data or statistics produced by the operation across participating countries.

The PforR will be structured into three results areas, all of which contribute directly to laying the foundations of a solid national and regional data system (Table 1).

Table 1. Results Areas (from the Project Appraisal Document, P176371)

		Kenya	Rwanda	Tanzania
	Results Area 1: Harmonization, Quality, Dissemination & Use			
DLI 1.1	Improved harmonization and quality of statistics	5	3.0	4
DLI 1.2	Improved dissemination and use of statistics	5	4.0	4.5
	Results Area 2: Availability & New Data Sources			
DLI 2.1	Improved availability of household-based survey data	27	12.5	15.5
DLI 2.2	Improved availability of enterprise-based survey data	24	4.0	8
DLI 2.3	Improved availability of agricultural statistics	12	8	8
DLI 2.4	Improved availability of administrative data	9.5	18.5	24.5
DLI 2.5	Improved availability of price & macroeconomic statistics	14.5	3.5	4
DLI 2.6	Gained experience with big data sources and new technologies	4	2.5	3.5
	Results Area 3: Infrastructure & Institutional Development			
DLI 3.1	Upgraded IT and statistical infrastructure	8	2.5	7.0
DLI 3.2	Enhanced institutional capacity and development	11	1.5	3.0
	Total	**120**	**60**	**82**

Source: World Bank 2022.

***Suggestions/e*xamples related to data governance and institutions from outside the World Bank**

Potential Prior Action/DLI: **"To operationalize open data: the Government published a Decree, in form and substance satisfactory to XXX, establishing an Open Data Policy pursuant to paragraph XXX of the Program."** See, e.g.:

- New Zealand: Cabinet Declaration on Open and Transparent Government (2011)
- United States: Executive Order: Making Open and Machine Readable the New Default for Government Information (2013), establishing the US Open Data policy – Managing Information as an Asset (2013)
- Mexico: Executive Decree on Open Data (2015)
- Australia: Prime Minister: Government Public Data Policy Statement (2015)
- Other country examples (executive decrees): France, Spain, Argentina, and Brazil

Potential Prior Action/DLI: **"Open data legislation is enacted, including satisfactory provisions for [insert explicit priorities here, e.g. "specific responsibilities had by NSO / other public agencies with respect to implementation of open data initiative under the auspices of this legislation."]** See, e.g.:

- South Korea: Act on Promotion of the Provision and use of Public Data (2013)
- Germany: E-Government Act (which includes explicit Open Data provisions under Section 12a) (2017)
- United States: Open, Public, Electronic, and Necessary ("OPEN") Government Data Act (2019), enacted as part of the Foundations for Evidence-Based Policymaking Act of 2018; previous legislation includes the Digital Accountability and Transparency ("DATA") Act (2014), which specifically opened more data on federal expenditures
- New Zealand: In the process of exploring new data and statistics legislation which explicitly includes open data, and which would aim to provide a consistent approach to the production of official statistics and the safe management and use of government data for research and analysis

Potential Prior Action/DLI: **"The Government created [or designated] an autonomous inter-government administrative and coordination agency [or national institute for statistics] with the responsibility of oversight and implementation of a nationwide open data initiative."** See, e.g.:

- South Korea: Act on Promotion of the Provision and use of Public Data (2013). This Law includes provisions for the creation of a high-level Open Data Strategy Council which is mandated to shape/guide the implementation of the open data initiative (and which includes representatives of central government, local government, the private sector and civil society). This Law further establishes the National Open Data Center as a coordinating body and to provide policy and technical support, including operating the national open data platform (data.go.kr), under the auspices of the National Information Society Agency. The Law also details that the Government of South Korea would develop a 3-year Master Open Data Plan to implement the national open data policy (including specific guidance for central administrative agencies and local authorities to develop annual Open Data Implementation plans).
- Germany: E-Government Act (which includes explicit Open Data provisions under Section 12a) (2017). This law provides for the creation of a central agency as both a contact point for regional authorities and to support national public authorities regarding the provision of data as open data.
- Other country examples (creation of new agency): France, Italy, Bulgaria
- Other country examples (designation of existing agency):
    - Ministry of the Interior (Netherlands, Czech republic)
    - Office of the Prime Minister (UK, South Korea)
    - National Archives (Sweden)

4. Data Skills and Demand for Data

Examples of results indicators from World Bank projects include:

- The number of staff in INE and other data producers of the NSS trained as part of the integrated training plan, disaggregated by gender (Number) [Angola Strengthening Statistical Capacity: P178043]
- Biannual publication of user satisfaction score with the statistical products produced by the INE (based on web and telephone surveys) (Number) [Ibid.]
- Increase in number of downloads of datasets and reports supported by the project in the most recent six-month period relative to the 6 months before project start date (Percentage) [Ibid.]
- Number of user-producer consultations, workshops, and feedback reports with relevant feedback integrated in project interventions (Number) [Ibid.]
- Increased user satisfaction [with statistical production] (Percentage) [Strengthening the Statistical System of Uzbekistan: P173450]
- Training needs assessment and training plan [related to the statistical system] (Text) [Ibid.]
- Increase in the share of beneficiaries who report that the project has established effective engagement processes (Text) [Ibid.]
- Number of annual hits on the NSDI Geoportal (Number (Thousand)) [Kosovo: Real Estate & Geospatial Infrastructure Project: P164555]

5. Data Enablers and Standards

Examples of results indicators from World Bank projects include:

- Data controllers and processors registered with the new Data Protection Authority (Number) [Burundi Digital Foundations Project: P176396]
- Publication of a cybersecurity and cybercrime national strategy and roadmap (Yes/No) [Ibid.]
- Cybersecurity Center established and operational with staff in place and procedures published in a format that is readily available to the public and industry (Yes/No) [Digital Nepal Acceleration (DNA) Project: P176543]
- National Cybersecurity Incidence Response Team provided with technical assistance on its core functional capabilities (Yes/No) [Mongolia: Smart Government II Project: P176631]
- (Gender informed) Policy framework and regulation to enforce data protection legislation are established (Yes/No) [Congo Digital Acceleration Project: P175592]
- Digital ID roadmap and operational plan drafted (Yes/No) [Ibid.]
- Number of people who have been issued with a new, digitally enabled ID (number, and of which % female) [Digital Maldives for Adaptation, Decentralization, and Diversification: P177040]
- Implementation of the general statistical business process model (GSBPM), the General Activity Model for Statistics (GAMSO), a data quality framework, and an Integrated NSS data hosting platform (Number) [Angola Strengthening Statistical Capacity: P178043]
- The quality of administrative data will be increased as a result of the project as measured through the Word Bank's quality administrative records assessment tool (HECRA) (Number) [Ibid.]
- Subscription to Special Data Dissemination Standard (SDDS) (Text) [Strengthening the Statistical System of Uzbekistan: P173450]
- Open Data format (Percentage) [Ibid.]
- *Prior action results indicator:* Number of social programs integrated into the Beneficiary registry and Management Information System for Social Programs (number of programs) [Dominica COVID-19 Response and Recovery DPC: P174927]

6. Data Infrastructure and Economy

Examples of results indicators from World Bank projects include:

- Mobile broadband penetration rate (Percentage) [Mozambique Digital Acceleration Project: P176459]
- To increase broadband internet access, especially to underserved communities (of which women, of which rural) [Burundi Digital Foundations Project: P176396]
- PCM indicator: Private capital attracted into rural broadband under Subcomponent 1.1 (Amount(USD)) [Digital Nepal Acceleration (DNA) Project: P176543]
- Cost-efficiency savings from use of the shared cloud computing platform by GoM agencies (Amount(USD)) [Mongolia: Smart Government II Project: P176631]
- CO2 e-emission reduction by GoM's national and disaster recovery data centers (Percentage) [Ibid.]
- Digital government functions and services using data centers supported by the Project (Number) [Digital Nepal Acceleration (DNA) Project: P176543]

- Number of Public Data Centers upgraded (Number) [Congo Digital Acceleration Project: P175592]

Annex 2. Overview of the World Bank Statistical Performance Indicators (SPIs)

The following description of the World Bank Statistical Performance Indicators (SPIs) is reproduced from the following site, where it is possible to explore the data or learn more about the framework: https://www.worldbank.org/en/programs/statistical-performance-indicators.

Reliable, usable, high-quality statistics are vital for global prosperity and progress. The Statistical Performance Indicators (SPI) provide an open-source framework for assessing the performance of statistical systems and the efforts to improve them.

The SPI framework assesses the maturity and performance of national statistical systems in five key areas, called pillars. The five pillars are:

Data Use: Statistics have value only if they are used. So the first pillar is data use. A successful statistical system produces data that are used widely and frequently.

Data Services: A range of services connects data users to producers and facilitate dialogues between them, thus building trust and a sense of value.

Data Products: The dialogues between users and producers drive the design and range of statistical products and their accuracy, timeliness, frequency, comparability, and levels of disaggregation. The products signal whether countries are able to produce indicators related to the 17 Sustainable Development Goals.

Data Sources: To create useful products, the statistical system needs to draw on sources inside and outside the government. Data collection thus goes beyond the typical censuses and surveys to include administrative and geospatial data as well as data generated by private firms and citizens.

Data Infrastructure: A mature statistical system has well-developed hard infrastructure (legislation, governance, standards) and soft infrastructure (skills, partnerships) as well as the financial resources to deliver useful—and widely used—data products and services.

Each of these pillars is supported by four or five dimensions and uses defined methods and indicators, all available as open data and open code.

Annex 3. Data Diagnostics in World Bank Systematic Country Diagnostics (SCDs)

A guidance note is available to help World Bank task teams undertaking Systematic Country Diagnostics (SCDs) to assess the availability of data and statistics in each country. Creation of the guidance note was prompted by recognition of the significant data gaps that limit SCD teams as they try to analyze development challenges and priorities. With a view to strengthening evidence-based decision making, the guidance note recommends that TTLs complete a Data Diagnostic Template to serve as a "constructive starting point for World Bank support to fill the data gaps identified".

In brief, the guidance note offers instructions for completing the eight sections of the Data Diagnostic Template:

- **Section 1, General information about the statistical system:** Information about the legal status of the National Statistical Office (NSO), whether or not there is a statistics law and a current National Statistical Development Plan (NSDS, or similar strategy document)
- **Section 2, Micro data:** Types of surveys conducted in recent years (e.g., household surveys, labor force surveys, censuses, etc.), with key metadata for each
- **Section 3, Macro data:** Availability and quality of macroeconomic indicators, e.g., national accounts, data on public spending/government finance, price indices (CPI, PPI), balance of payments and sectoral data
- **Section 4, Compliance with WBG core data standard:** Whether the country has key data available including household surveys of income or consumption meeting certain criteria, a yearly Purchasing Power Parity price survey, and civil registration and vital statistics
- **Section 5, Statistical Capacity Indicators:** The Statistical Capacity Indicator (SCI) covers three dimensions of statistical capacity: methodology, source data and periodicity
- **Section 6, Openness indicators:** Overall country score on the Open Data Barometer and the Global Open Data Index
- **Section 7, Country priorities:** Assess the availability of data specifically related to making policy decisions about the country's priority areas (as indicated for example in a national development plan, in the SCD itself, or in the country's adoption of one or more SDG indicators) and monitoring progress
- **Section 8, Data gaps identified and recommended actions/solutions:** Identification of types of micro and macro data that are either unavailable, outdated or of insufficient quality, with suggested actions and solutions to fill the gaps