# Data for Development

## An Evaluation of World Bank Support for Data and Statistical Capacity

IEG
Independent Evaluation Group
WORLD BANK GROUP
World Bank • IFC • MIGA

WHAT WORKS

1818 H Street NW
Washington, DC 20433
Telephone: 202-473-1000
Internet: www.worldbank.org

Attribution—Please cite the report as: World Bank. 2018. *Title.* Independent Evaluation Group,
Washington, DC: World Bank.

Cover Photo:

# Data for Development

## AN EVALUATION OF WORLD BANK SUPPORT FOR DATA AND STATISTICAL CAPACITY

**AN INDEPENDENT EVALUATION**



Careful observation and analysis of program data and the many issues impacting program efficacy reveal what works as well as what could work better. The knowledge gleaned is valuable to all who strive to ensure that World Bank goals are met and surpassed.

# contents

## Boxes



## Figures



## Table



## Appendixes

# abbreviations

| | |
|---|---|
| ADP | Accelerated Data Program |
| CPF | country partnership framework |
| CRVS | civil registration and vital statistics |
| DFID | U.K. Department for International Development |
| DPF | development policy financing |
| FY | fiscal year |
| GPS | Global Positioning System |
| IBRD | International Bank for Reconstruction and Development |
| ICP | International Comparison Program |
| ICR | Implementation Completion and Results Report |
| ICRR | Implementation Completion and Results Report Review |
| IDA | International Development Association |
| IEG | Independent Evaluation Group |
| IFC | International Finance Corporation |
| IMF | International Monetary Fund |
| IT | information technology |
| MAPS | Marrakech Action Plan for Statistics |
| NSDS | national strategy for the development of statistics |
| NSO | national statistical office |
| OECD | Organisation for Economic Co-operation and Development |
| PARIS21 | Partnership in Statistics for Development in the 21st Century |
| PPAR | project performance assessment report |
| SCD | systematic country diagnostic |
| SCI | Statistical Capacity Indicator |
| SDG | Sustainable Development Goal |
| STATCAP | Statistical Capacity Building Program |
| TFSCB | Trust Fund for Statistical Capacity Building |
| UN | United Nations |
| UNFPA | United Nations Population Fund |

All dollar amounts are U.S. dollars unless otherwise indicated.

# acknowledgments

This evaluation was prepared by an Independent Evaluation Group (IEG) team led by Soniya Carvalho (co-team leader) and Rasmus Heltberg (co-team leader) under the overall direction of Caroline Heider, director-general, Evaluation, and with the guidance and supervision of Marie Gaarder, manager, Corporate and Human Development, and Auguste Tano Kouame, director, Human Development and Economic Management. The team comprised Jose Ramon Albert, Andrew Bent, Sankalpa Dashrath, Ann Flanagan, Andrew Flatt, John Heath, Javier Horovitz, Basil Kavalsky, Nidhi Khattri, Chad Leechor, Eduardo Maldonado, Joan Nelson, Brian Ngo, Estelle Raimondo, Swizen Rubbani, and Bahar Salimova.

Project reviews and background papers were authored by Morten Jerven. Administrative support was provided by Marie Charles, Dung Thi Kim Chu, and Rich Kraus. The editor was Amanda O'Brien.

The report was peer reviewed by Martin Ravallion (professor, Georgetown University).

The team is grateful to the many staff who engaged with us and gave generously of their time, especially the counterparts in the Development Data Group.

Overview

# highlights

1 Data and evidence are the foundation of development policy and effective program implementation, and countries need data to formulate policy and evaluate progress. This evaluation's objective was to assess how effectively the World Bank has supported development data production, sharing, and use, and to suggest ways to improve its approach.

2 This evaluation defines development data as data produced by country systems, the World Bank, or third parties on countries' social, economic, and environmental issues.

3 At the global level, the World Bank has a strong reputation in development data and has been highly effective in data production. It produces influential, widely used data and cross-country indicators that fill important niches, benchmark countries, and stimulate research and policy action.

4 The World Bank has also taken a prominent leadership role in global data partnerships so far. However, the World Bank needs to determine its future role carefully because the global partnership landscape is becoming more uncertain—as old partnerships phase

out, the complementarity of new partnerships is unclear. This makes the World Bank's future role especially pivotal because the sustainability of funding from global data partnerships at both the national level and for some global data efforts is at risk. Without sustained funding, past progress will be in jeopardy, as observed in some countries where data quality worsened when trust fund support ended.

**5** At the national level, the World Bank has been mostly effective at fostering its client countries' data production through its own financing and through financing from small trust fund grants. It has been less effective in promoting data sharing; while the World Bank has used its leverage in some of its client countries, it needs to do a better job at encouraging other countries to share data. The World Bank has been even less effective in promoting data use by governments and citizens.

**6** The World Bank's systemwide approach to building the capacity of national statistical organizations yielded significant successes in countries where it was deployed, and it should now add a focus on building subnational capacity and strengthening client countries' administrative data systems.

**7** Big data offers big opportunities, but it also has risks. The World Bank needs to make sure it clearly understands when and how big data can complement traditional data in answering key development questions related to its mission, and use big data analytics appropriately to underpin its own decisions and to ensure that it supports its country clients effectively in big data use. The World Bank still needs to address the implications for organizing big data work internally, entering into corporate agreements with private providers (typically the producers of big data), and seriously considering and addressing privacy and ethical concerns related to big data use.

## Promoting Data for Development at the Global Level

**THE WORLD BANK** has a strong, global reputation in development data. It produces influential, widely used data and cross-country indicators that fill important niches, benchmark countries, and stimulate research and policy action.

The World Bank has taken leading roles in global partnership programs that filled gaps in the global statistical system. Since 1999, it has helped establish, run, and fund ($50.9 million) global data partnership programs that have made important contributions and mostly balanced global and national data needs. The World Bank's success is attributable to technical expertise, the ability to link global needs to national needs, an ability to sustain initiatives for the long term, and its well-aligned partnership engagements.

A coherent architecture existed for the older generation of partnerships for statistical capacity building, but coherency is missing for the new partnerships involving data innovation. Some of the newer global initiatives appear duplicative. Opportunity exists for consolidating data innovation partnerships, setting clearer goals, and identifying future funding for major data partnership engagements.

## Country-Level Data Support

The World Bank supported data production, sharing, and use through lending and small trust fund grants in a large number of countries. This support was mostly effective for data production, but was less effective for data sharing and even less for data use. Commitments for data activities averaged about $90 million per year, increasing in the second half of the fiscal year (FY) 06–15 reference period.

The World Bank had in-depth engagement in statistical reforms in fewer countries. In-depth statistical capacity–building efforts addressed data supply constraints and helped move countries away from scenarios where data scarcity and low data quality are associated with low use, low data literacy, and little demand and funding for data.

In countries where the World Bank and its partners used a systemwide approach to statistical capacity building, there were significant successes. Reforms paired improvements in the institutional and legal environment for data production with investments in the physical and human capital required to produce quality, timely, and reliable data. However, many countries are still data deprived, and experience major gaps in the quality and availability of data, especially regarding routine administrative data. Client countries now expect support that is more coordinated and long-term from the World Bank and its partners in building and strengthening data systems, particularly support beyond national statistical offices (for example, for administrative data systems and subnational statistical systems).

The World Bank has an important role to play in continuing to do what has worked well in the past: collecting select global data on prices, poverty, and other specific areas; supporting household survey collection and methodology development; and coordinating and funding support for national statistical organizations.

## Growing a User-Centered Data Culture

Support for national statistical systems enhanced data production more than it promoted in-country data sharing and use. The World Bank influenced several countries to share data and microdata publicly and worked with partners to improve microdata cataloging and metadata development. However, several countries refuse to share data for political reasons, quality concerns, or a reluctance to lose a revenue source. The World Bank has occasionally raised data sharing issues at high levels of policy dialogue, but it needs to use its leverage fully in client countries that are reluctant to share data openly.

The World Bank could do much better to encourage governments to use data, even though their ultimate use is not necessarily within the World Bank's control. Weaknesses in promoting data use have been a major issue for the past 10–15 years, but efforts in this area are scattered. Only 27 of the 201 projects reviewed for this evaluation supported activities to build data use capacity. The World Bank has a well-established approach to building the capacity of data producers, but it has not yet formulated a conceptual model for assessing user capacity. It could promote enhanced data use, for example, by understanding the different kinds of data users and their needs and motivations, and by including both government and nongovernment data users in the design of its projects.

The next step is to work toward a user-centered data culture, understood as reciprocity between the agencies that produce, share, and use data. Low data literacy and weak research communities are constraints in poorer countries, yet people interviewed in many countries told the Independent Evaluation Group (IEG) that they want to know how their region, city, or community is doing relative to others in their country. Decision makers in central and local governments need this information to set priorities and compel action. The user-centered data culture will take years to develop, but by working with a broad menu of clients, the World Bank can nurture an ecosystem of data use.

## Exploring Big Data's Potential

Big data are extremely large data sets resulting from the growing digitization of our lives. Big data activities at the World Bank so far have been ad hoc and the result of individual initiative instead of a coordinated institutional approach. The ad hoc approach has been helpful in facilitating small-scale exploration and experimentation, but is unlikely to work well if the World Bank decides to scale up its big data work. Scaling up would require a more coordinated approach, clearly defined responsibilities for big data within the organization, sufficient data science expertise, systematic cataloging, a centralized repository, and the removal of barriers to combining all forms of relevant data (from

geospatial to social media to traditional) in answering key development questions. The World Bank should also consider when and where it would make sense to grow the big data capacity of national statistical organizations.

A major challenge has been the lack of a widely-shared understanding and appreciation among World Bank staff of when and how big data can complement traditional data in answering questions related to its mission. Furthermore, the lack of corporate agreements with government and private big data producers has complicated the World Bank's access to big data. Finally, the World Bank needs to deal with the complex issues of ensuring privacy and ethical use of big data for itself and its country clients.

## Conclusions and Recommendations

This evaluation finds the World Bank has been highly effective in producing influential data globally and until recently in promoting global data partnerships. It was mostly effective at the country level in supporting data production, promoting open data, encouraging some country clients to share data, and building the capacity of national statistical organizations in countries where it adopted a systemwide approach. It was less effective in adapting to the changed global partnership landscape where the complementarity of new partnerships is less clear. It was also less effective in fully using its leverage to encourage data sharing by client countries which have been reluctant to do so, and even less effective in promoting data use in government decision making, building subnational data capacity, strengthening country clients' administrative data systems, and staying at the forefront in analyzing the potential and pitfalls of big data for development.

IEG recommends the World Bank now take the following actions:

Recommendation 1: Implement goals and priorities reflecting the findings of this evaluation with regard to the World Bank's support to global data and global partnerships, country data capacity, and a user-centered data culture.

Steps to be considered by World Bank Management could include:

- articulating goals and priorities;
- specifying accountabilities for the implementation of new and existing goals and priorities; and
- ensuring sufficient management oversight so that the new and existing goals and priorities are implemented.

Recommendation 2: Mobilize and deliver additional support to countries' statistical systems, using a more comprehensive model of statistical capacity building that also factors in needs and opportunities to strengthen administrative data systems.

Recommendation 3: Step up engagements with global partners and client governments on long-term funding for development data.

Steps to be considered by World Bank Management could include:

- requiring country partnership frameworks (CPFs) to explicitly indicate how the systematic country diagnostic (SCD)–identified knowledge and data gaps; which are most relevant to CPF objectives, will be addressed;
- elevating attention to funding for data in the policy dialogue with client governments; and
- initiating high-level discussions on establishing a global umbrella mechanism for long-term financing of data.

Recommendation 4: Scale up promotion of data sharing and data use.

Steps to be considered by World Bank Management could include:

- ensuring that all data financed by the World Bank are shared with the World Bank;
- developing and using a list of essential data items that countries are expected to share with the World Bank;
- incentivizing governments to more openly share data with the public, for example, by more prominently using a ranking of countries on open data performance; and
- scaling-up promotion of government and citizen demand for data and the voice of data users in the kinds of data that are produced.

Recommendation 5: Implement coordinated actions so that World Bank operations benefit from big data's insights and clients receive appropriate support for big data use.

Steps to be considered by World Bank Management could include:

- reviewing opportunities to scale up the use of big data for development;
- specifying accountabilities for implementation of the coordinated actions; and
- ensuring sufficient management oversight so that the coordinated actions are implemented.

# management response

**WORLD BANK MANAGEMENT** welcomes the Independent Evaluation Group (IEG) report *Data for Development: An Evaluation of World Bank Support for Data and Statistical Capacity,* and appreciates the opportunity to provide comments on the approach paper and early draft of the report. The report is timely, comprehensive, constructive, and well written. It provides a useful review of the World Bank's work in supporting countries to produce, share, and use data. It is well balanced in its analysis of successes and weaknesses and offers ideas on how to address the remaining and emerging challenges.

Management agrees that international demand for data is increasing while new technological developments are revolutionizing data production methods and use patterns and offering expanding opportunities. At the same time, the evidence-based approach is under some threat from policies restricting access to data. At such a juncture, it is important to strengthen the World Bank's data-related work. Management believes that statistical development is a critical area of policy reform. If the World Bank wants to deliver on its twin goals, maximize its impact on policy advice, and promote greater transparency and accountability, it needs to support its client countries to produce, disseminate, and use more and better-quality data.

Management broadly concurs with the conclusions and recommendations of the report. Management responses to specific recommendations in the report are presented in the attached Management Action Record matrix.

## World Bank Management Comments

World Bank's role in development data. Management appreciates the recognition of the World Bank's global reputation in development data activities and high effectiveness in producing influential, widely used data that fill important niches, benchmark countries, and stimulate research and policy action.

Leading role in data partnerships. The report acknowledges the World Bank's leadership in global data partnerships. As noted in the report, continued efforts in this area are required to raise additional funding to close data gaps and ensure sustained progress.

Support to countries' statistical capacity. Management concurs with the report's finding that the World Bank's systemwide approach to building support to countries' statistical capacity has been largely successful within the scope of the relatively low-level financial resources allocated to this task.

Support for national statistical systems. Management agrees that the World Bank should support national statistical *systems* and not just national statistical *organizations.* The report recommends statistical support to be extended to sectoral ministries and subnational governments requiring a more comprehensive model of statistical capacity building and support for expanded data

dissemination and use. The key question is, *What is a reasonable expectation of the World Bank in effectively delivering on the data agenda, given its capacity and resource constraints and its comparative advantage?* There are trade-offs, which implies the need to prioritize and be selective both in the World Bank's country engagements and its partnerships. In this context, management has adopted the "rolling approach" to prioritization in the *World Bank Group Strategic Actions Program for Addressing Development Data Gaps* endorsed by senior management through the World Bank Group Development Data Council, on September 29, 2015. Management also believes that the World Bank should embed support for governments' data management capabilities and systems in sector-specific projects or through cross-sectoral engagements such as e-government or government modernization-type projects.

With regard to the report's references to the role and work of the World Bank Group Data Council, management would like to refer to progress achieved since the Data Council's creation in 2014.

- Foremost, development data issues have been elevated to the attention of the Development Committee, which declared that development data should be a core component of World Bank Group operations. This enabled significant progress in defining the World Bank's priorities for development data and how to address them through the approval of the Strategic Actions Program for addressing development data gaps and its four key action plans for (i) household surveys, (ii) price statistics, (iii) civil registration and vital statistics, and (iv) geospatial data, with additional areas currently in pipeline, including population census, jobs, gender, firm-level data, and big data.
- The World Bank Group Data Council facilitated the coordination among staff of different parts of World Bank Group to address methodological issues and offer solutions (including through Doing Development Differently and technical working groups). In particular, it helped to establish the World Bank Group Household Survey Working Group as a global leader on household survey research and technical assistance.
- The World Bank Group Data Council allowed an increase in technical assistance and lending on some data issues across Regions. It also enabled the development and launch of three indicators related to the Strategic Actions Program in the IDA18 Results Measurement System.
- The World Bank Group Data Council also made development data one of the World Bank Group's five strategic priorities for fundraising with external donors (the "A list"). Being in the A list implies that the Strategic Actions Program is excluded from the moratorium for donor fundraising. This helped World Bank Group gain respect and trust from external partners with a clear data governance structure, which has become a model for development organizations and donors around the world.
- The World Bank Group Data Council endorsed new World Bank Group protocols for producing poverty estimates and for household survey data collection, quality assurance, and standard setting at the country level.
- The World Bank Group Data Council also endorsed a new methodology for diagnosing development data gaps in each client country, which is included in the guidelines for World Bank Group Systematic Country Diagnostics.

- The World Bank Group Data Council endorsed the creation of the Development Data Hub, a World Bank Group–wide data set catalog and repository that provides a means for effective curating, searching, accessing, sharing, and using of World Bank-collected development data. An initial budget allocation was secured and contributed to the development of the Hub. (The beta version of the dataset catalog is available at https://datacatalogbetastg.worldbank.org.)
- Finally, the Data Council mandated creation of the Analytics and Geospatial Working Group (AGWG), tasked with identifying how the World Bank could better make use of geospatial data. The AGWG is both the governing body and coordinating body for geospatial operations at the World Bank. It comprises representatives of every Global Practice, ensuring that its recommendations are representative and that decisions taken have broad-based support. The Geospatial Operations Support Team (GOST) was then formed to work toward the priorities and strategic aims identified by AGWG. The first year of GOST yielded concrete results against the shortcomings identified in the IEG report.

Lending support for data activities. The evaluation report finds that World Bank lending support for data activities has been low (on average $90 million per year) and that reliance on trust funds is not sustainable. Management is aware of this important issue. Although funding for statistical capacity building through both lending and trust funds has been steadily growing in recent years, sustainability over time remains a concern, particularly in the Africa Region, where data deprivation is highest as well as in other Regions that have demonstrated progress.

Systematic Country Diagnostics Data. Management highlights the importance of data diagnostics in Systematic Country Diagnostics (SCDs). Although most SCDs to some extent discuss those data issues most critical for identifying a country's development priorities and progress toward the World Bank Group twin goals, this was not done in a systematic and standardized format until recently. Starting in calendar year 2017, SCD teams have been encouraged to use the data diagnostic template endorsed by the World Bank Group Data Council. The template was referenced in the revised SCD guidance note (issued in December 2016) as a means to record data gaps systematically using a standardized format. Management concurs that Country Partnership Frameworks would benefit from a more systematic presentation of data gaps from drawing on SCDs, with the understanding that the World Bank Group program can only address the SCD-identified gaps aligned with client countries' strategic objectives and the World Bank's comparative advantages. More generally, management will encourage teams to recognize the potential role of the data diagnostic template as a platform to organize the data conversation at the country level and promote coordination among teams working on data issues.

Access to country data. Management fully agrees that access to country data is an important issue while also recognizing that access to data is worsening in some countries. To overcome constraints in access, the report recommends that the World Bank assume a more forceful stance with client countries such as by making funding arrangements conditional on data sharing. Some questions arise about evidence on the virtues of data sharing conditionality as opposed to other alternatives such as sustained collaboration, building trust, and setting up positive incentives for statistical agencies to be more forthcoming with access to data. However, management agrees this may be

an issue where *context* should dictate the best solution; for example, where development project objectives have been successfully used to promote more data sharing or any results from World Bank experiences in exercising leverage.

Local demand for data. Management does not agree with the report's finding that local demand for data is generally low. Although the rationale for instances of low demand is correct, management believes that demand for good quality data is high. Management agrees that surfacing local demand for data could be strengthened if the World Bank focused more deliberately and systematically on supporting the demand side of data as well as supply-side actors.

Complementary role for big data. The World Bank recognizes the importance of big data and its promise to accelerate development outcomes as well as to potentially close data gaps in fragile environments. However, it remains unclear why big data is highlighted so extensively in the report. Much more work must be done on closing data gaps with "traditional" data than with that of big data. Traditional data are also often needed to draw inferences from big data, posing a need for the World Bank to strike the right balance in a resource-constrained environment. The generic consensus of management is that big data could be treated as a new source of data, complementing rather than substituting for traditional forms of data where the World Bank has developed a comparative advantage.

Conflation of big data and geospatial data. The report's use of these terms suggests that they are interchangeable or that geospatial data is a subset of big data.[1] Although some data sets are both big and geospatial (for example, call detail records, GPS traces), many are either just big (for example, web logs, government expenditure data) or just geospatial (for example, administrative boundaries, forest cover, zonal statistics). Conflating geospatial data and big data is not just a technical detail; the two terms require different staff skill sets to be harnessed effectively. They are relevant in different scenarios, solve different problems, and have different levels of applicability to World Bank operations. The World Bank is taking a nuanced, tailored approach to each. This is partly the reason for separate working groups looking at and managing the topics.

[1] This inaccuracy is present throughout chapter 5 in the section titled World Bank Support for Geospatial and Other Forms of Big Data, where the list of sectors is identical to those being supported in an operational context by the Geospatial Operations Support Team; and perhaps most importantly in bullet list of specific innovations using big data, where every example is an application of geospatial technology rather than traditional big data.

# management action record

## Implement Strategies to Support Data

IEG FINDINGS AND CONCLUSIONS The World Bank has been an effective leader on development data for global audiences. It produces influential, widely used data and cross-country indicators that fill important niches, benchmark countries, and stimulate research and policy action. The World Bank's solid reputation is attributable to technical expertise, its ability to link global and country needs, initiatives that it sustained for the long term, and successful, well-funded partnerships in which the World Bank took a prominent leadership role.

The World Bank's data efforts were more coherent in the era of the Millennium Development Goals. The number of other actors on data has been growing over time along with ambitions, which raises questions about the clarity of the World Bank's role and mission on data. The *World Bank Group Strategic Actions Program for Addressing Development Data Gaps* and its associated action plans articulate clear goals for data production and innovation. Goals and priorities also need to be spelled out for other major elements of the World Bank's work on data, especially for its engagements in partnerships; data access, sharing, and use; and the main types of administrative data systems. Issues of costs, financing, and lines of accountability for these elements of the World Bank's data work also need to be clarified.

IEG RECOMMENDATIONS **Recommendation 1: Implement goals and priorities reflecting the findings of this evaluation with regard to the World Bank's support to global data and global partnerships, country data capacity, and a user-centered data culture.**

Steps to be considered by World Bank management could include

- articulating goals and priorities;
- specifying accountabilities for the implementation of new and existing goals and priorities; and
- ensuring sufficient management oversight so that the new and existing goals and priorities are implemented.

ACCEPTANCE BY MANAGEMENT Agreed.

MANAGEMENT RESPONSE At the global level, management will explore opportunities to (i) influence selected political summits and global forums that focus on issues experiencing significant data gaps and (ii) advance World Bank Group development data priorities.

At the institutional level, goals and priorities have been articulated in the Strategic Actions Program, as recognized in the Independent Evaluation Group (IEG) report findings. Implementation of the goals and priorities has been outlined in four specific action plans to date: (i) Household Surveys, (ii) Prices, (iii) Civil Registration and Vital Statistics, and (iv) Geospatial Data. The action plans lay out technical accountabilities, costs, and financing sources. They are living documents that may be adjusted during implementation to accommodate course corrections or adapted to new priorities or areas of strategic focus, such as fragile and conflict-affected states.

Additionally, management has improved its governance arrangements for development data to strengthen the links to senior-level operational decision making and commits to reviewing its effectiveness periodically. New governance arrangements for the World Bank Group Data Council were announced on March 17, 2017, with the aim of operationalizing the Strategic Actions Program, action plans, and other Data Council decisions.[1] The newly created Development Data Council will make decisions related to World Bank Group development data agenda, with the guidance and support of the Matrix Vice Presidents.

[1] https://hubs.worldbank.org/news/Announcement/Pages/Putting-Data-Priorities-to-Work-17032017-172205.aspx

## Strengthen Statistical Systems

IEG FINDINGS AND CONCLUSIONS At the national level, the World Bank has been mostly effective at fostering data production by client countries through lending, trust funds, and technical assistance. However, progress is slow and uneven and many countries are still data deprived, especially regarding administrative data systems.

The World Bank's systemwide approach to building the capacity of national statistical organizations yielded significant successes in countries where it was deployed. However, the approach does not give sufficient attention to building subnational capacity and strengthening country clients' administrative data systems.

IEG RECOMMENDATIONS **Recommendation 2: Mobilize and deliver additional support to countries' statistical systems, using a more comprehensive model of statistical capacity building that also factors in needs and opportunities to strengthen administrative data systems.**

ACCEPTANCE BY MANAGEMENT Agreed.

MANAGEMENT RESPONSE The Systematic Country Diagnostic (SCD) guidance note now incorporates specific guidance on data, including a data diagnostic template that systematically records data gaps in key areas necessary for the country to adopt evidence-based development policies and monitor its development goals. The diagnostic pays particular attention to data relevant for monitoring development goals related to the World Bank Group twin goals and the Sustainable Development Goals most relevant for the country, including administrative and other nonsurvey data. Management will continue its efforts to encourage inclusion of this data diagnostic in SCDs and to inform the World Bank's engagement under Country Partnership Frameworks (CPFs) with SCD findings on data gaps. Management will review the data diagnostic template to more explicitly cover gaps in administrative and geospatial data. In addition, management will explore ways in which to further leverage the SCD Data Diagnostic and other tools to prioritize, promote coordination, and enhance complementarities among different in-country data initiatives.

Management will also continue its efforts to encourage corporate initiatives to recognize that development data capacity building must reflect the multifaceted sources of data (for example, administrative data, big data) becoming available to support development.

More broadly, management will continue to encourage improvements to World Bank Group systems and to better monitor and document progress of the Bank Group development data agenda, for example, through the recently created thematic code for data projects and Analytical and Advisory Services.

## Engage for the Long Term

IEG FINDINGS AND CONCLUSIONS No mechanism exists for medium-to-long-term financing for data even though the funding needs for data are significant. Producing data is a core government function, but several countries do not appreciate the value of data, fund it poorly, and are reluctant to borrow for it. Trust-funded programs were central to past successes, positioning the World Bank as a premier global funder and coordinator of data and allowing it to also engage in countries without a lending program for data. However, trust funding for core data work is dependent on only a few donors and faces uncertain prospects. The sustainability of past gains in statistical capacity is at risk in some countries. Therefore, mobilization of domestic and donor funding for data should be a top priority. World Bank senior management should seek to raise global awareness to data financing. World Bank Country Directors should ensure more consistent treatment of data issues and data funding in country programs.

IEG RECOMMENDATIONS **Recommendation 3: Step up engagements with global partners and client governments on long-term funding for development data.**

Steps to be considered by World Bank management could include

- requiring CPFs to explicitly indicate how the SCD-identified knowledge and data gaps, which are most relevant to CPF objectives, will be addressed;
- elevating attention to funding for data in the policy dialogue with client governments; and
- initiating high-level discussions on establishing a global umbrella mechanism for long-term financing of data.

ACCEPTANCE BY MANAGEMENT Agreed.

MANAGEMENT RESPONSE At the global level, as management explores opportunities to strategically advance the World Bank Group development data priorities in selected global forums, it will proactively coordinate with partners to seek additional financing for development data activities.

At the institutional level, management will assess gaps in development data priorities and related financial needs and develop a financing framework that identifies potential sources of financing to help close these gaps. Management will continue to emphasize the importance of closing critical data gaps to invest in better-quality and timely data as the foundation to evidence-based policy making.

At the country level, management will continue to explore partnership opportunities with donors and encourage public/private sector partnerships to coordinate and increase country-specific sources of funds for development data activities.

## Promote Data Sharing and Use

IEG FINDINGS AND CONCLUSIONS The World Bank had a positive role in promoting data sharing by some of its client countries, but it is unreasonable for countries to receive World Bank support for collecting data without a requirement for sharing that data with the World Bank and with the public (subject to privacy restrictions). The World Bank now needs to ensure that it uses its leverage fully to encourage universal data sharing. The World Bank has paid far less attention to promoting government and citizen data use so far, and therefore success is limited.

IEG RECOMMENDATIONS Recommendation 4: Scale up promotion of data sharing and data use.

Steps to be considered by World Bank management could include

- ensuring that all data financed by the World Bank are shared with the World Bank;
- developing and using a list of essential data items that countries are expected to share with the World Bank;
- incentivizing governments to more openly share data with the public, for example, by more prominently using a ranking of countries on open data performance; and
- scaling-up promotion of government and citizen demand for data and the voice of data users in the kinds of data that are produced.

ACCEPTANCE BY MANAGEMENT Agreed.

MANAGEMENT RESPONSE At the global level, management will seek to leverage global partnerships to support data use, including through selected forums.

At the institutional level, under the broader framework of the Bank Group Access to Information Policy and Information Security Policy, management is developing procedural guidance for World Bank Group staff involved in data activities. This includes guidance on development data acquisition, storage, dissemination, and open data. Additional guidance will be provided as new priorities emerge. Management is also seeking to complement this guidance with useful templates for Bank Group staff such as a template for memoranda of understanding and model legal agreements to enable Bank Group access to one or more data sets. Management will highlight the user focus in all data interventions.

At the country level, management will coordinate with partners to capture and disseminate information on countries' open data performance (produced by authoritative sources).

Additionally, management will explore opportunities to leverage its convening power at all levels to strengthen operational partnerships with stakeholder groups working to improve development data such as bilateral donors, civil society, and the private sector.

## Coordinate for Big Data

IEG FINDINGS AND CONCLUSIONS Big data offers big opportunities, but it also has risks. The World Bank needs to make sure it clearly understands when and how big data can complement traditional data when answering key development questions related to its mission and use big data analytics appropriately to underpin its own decisions and to ensure that it supports its country clients effectively in big data use. The World Bank still needs to address the implications for organizing big data work internally, entering into corporate agreements with producers of big data, supporting clients in big data use, and addressing privacy and ethical concerns related to big data use.

IEG RECOMMENDATIONS **Recommendation 5: Implement coordinated actions so that World Bank operations benefit from big data's insights and clients receive appropriate support for big data use.**

Steps to be considered by World Bank Management could include:

- reviewing opportunities to scale up the use of big data for development;
- specifying accountabilities for implementation of the coordinated actions; and
- ensuring sufficient management oversight so that the coordinated actions are implemented.

ACCEPTANCE BY MANAGEMENT Agreed.

MANAGEMENT RESPONSE Management recognizes the spirit of this recommendation and the importance of integrating different data such as joining conventional data with administrative data, with geospatial data, with big data, and with other frontier data. Management also recognizes the importance of strengthening macro/micro data linkages.

To support big data use specifically, management will encourage collaboration among World Bank Group teams and disciplines, seek to leverage global partnerships, and explore new technology platforms. To help facilitate these efforts, a World Bank Group Big Data Working Group has been created. Management agrees to use a widely accepted taxonomy of big data, making it clear that there are multiple types of big data. Management will prioritize actions across these different types of big data, being explicit about what concrete activities it proposes to do for each type of data.

In addition, management recognizes the importance of geospatial data as a World Bank Group priority area. Consequently, management has actively supported the Geospatial Operations Support Team (GOST), which was formed to help bring to scale promising trials of geospatial insight. In its first year, GOST has coordinated staff activity on geospatial data, taken the lead on geospatial data curation, partnered with key industry players, and helped mainstream the use of geospatial analytics.

Finally, management is also working to develop job streams for data scientists and statisticians to support more systematized recruitment and career development of technical specialists with an inclusive range of skills and experience, including with big data.

# report to the board from the committee on development effectiveness subcommittee

The subcommittee of the Committee on Development Effectiveness (CODE) met to consider report by the Independent Evaluation Group (IEG) entitled *Data for Development: An Evaluation of World Bank Support for Data and Statistical Capacity* and World Bank management's draft response.

The CODE subcommittee welcomed the report and was pleased that management broadly concurred with IEG's findings and recommendations and that the World Bank Group Development Data Council endorsed the report. The subcommittee highlighted the importance of data in meeting the Sustainable Development Goals and the need for a vision that sets out how to get there by 2030. Members were encouraged to learn that the Bank Group has a comparative advantage on global development data and has mostly been effective in supporting countries in data production. They acknowledged constraints on internal resources and that the new data template approved by the Development Data Council was being rolled out and would help assess data gaps.

Members noted the importance of supporting client countries to develop capacity to generate, use, and share data and asked how this could be implemented most effectively. In light of limited lending support and reliance on trust funds, they discussed how resources could be best deployed to ensure long-term sustainability of data activities and how to improve client interest and commitment to the data agenda. Some members stressed the importance of the Country Partnership Framework process as a policy dialogue that could promote the allocation of domestic resources to statistical capacity building, of focusing on both national and subnational statistics offices, of knowledge transfer and technology, and of the need to use International Development Association resources in low-income and fragile countries.

# 1

# Evaluating Data for Development

**DATA AND EVIDENCE** are the foundation of development policy and the effective implementation of programs. To varying degrees, countries use data for economic and sectoral policy making and for planning, implementation, monitoring, targeting, and administration of policies and programs. The global community also uses data to varying degrees for programming assistance and tracking progress. Much research on development issues relies on data. The agenda first set by the Millennium Development Goals (2000–15) and now by the Sustainable Development Goals (2016–30) has ramped up the demand for data to monitor progress toward targets.

The supply of data often has not kept up with demand. Half of the World Bank's member countries lack the data necessary to measure progress toward the twin goals of ending extreme poverty by 2030 and promoting shared prosperity (Serajuddin and others 2015). Data users have serious concerns about data quality and timeliness, especially in low-income countries, and demand is unmet for disaggregated data for local planning.

The World Bank has a long history of promoting data. The World Development Indicators database began as a statistical appendix to the 1978 *World Development Report* (added at President Robert S. McNamara's urging, almost as an afterthought); it is now "the most widely used knowledge product of the World Bank" and, thanks to the World Bank's Open Data initiative, its data are freely accessible to all (Besley and others 2015). Building on the start made in the 1970s, the World Bank's role in promoting development data became more prominent through the years, driven, for example, by President James D. Wolfensohn's 1996 vision of the World Bank as a Knowledge Bank, efforts to monitor the Millennium Development Goals in the 2000s, and a movement toward managing for results and evidence-based policy making. The country rankings in *Doing Business,* an annual report launched in 2004, are a benchmark that receives close attention from governments around the world.

## Evaluating Data Production, Sharing, and Use

There is no policy or corporate procedure on development data in the World Bank except a longstanding Operational Policy on debt data and an ongoing process to produce: (i) a procedure governing the new Development Data Hub, (ii) a Procedure on Data Acquisition (from vendors, other international organizations, and countries) and (iii) an Open Data strategy.

The World Bank has an Actions Program to address data gaps: *Strategic Actions Program for Addressing Development Data Gaps* (World Bank 2015). As part of the Actions Program, four Action Plans have been completed (civil registration and vital statics (CRVS), price data, household surveys, and geospatial data), three Action Plans are in preparation (economic statistics, gender, and population census), and two Action Plans are planned (jobs data and firm-level data).

This evaluation asks, "How effectively has the World Bank supported the production, sharing, and use of development data?" It reviews World Bank support for developing countries' capacity and data systems, data for the national and global public good, engagements in international partnerships, and technological innovations, particularly relating to big data. The World Bank supports data production, sharing, and use through lending, technical assistance, and trust fund grants. This evaluation covers all three forms of support, though not in equal depth. (Appendix A describes IEG's methodology for this evaluation.) The reference period is from 2004 (since the launch of the Marrakech Action Plan for Statistics) through the end of 2016, a period in which the World Bank's approach to data support underwent change (World Bank 2011). The World Bank's own use of data for decision-making purposes is not the primary focus of this evaluation, mainly because other recent audits and evaluations are summarized in *Results and Performance of the World Bank Group 2016* (Managing for Development Results). Compared with the World Bank, the International Finance Corporation (IFC) and the Multilateral Investment Guarantee Agency provide relatively little data support and are outside the scope of this evaluation.

### Country Data Systems

This evaluation defines *development data* as data produced by country systems, the World Bank, or third parties on countries' social, economic, and environmental issues. Development data come in several forms. *Administrative data* are the by-product of routine public services delivered by either local or central government (registration of births, marriages, and deaths; issuing drivers' licenses; registration of land titles; and recording vaccinations). *Census and survey data* are data collected periodically for the whole population and purposively for a sample. *Economic data* on prices and interest rates, employment, trade, and national income are in a category of their own. *Big data* derive from data sets distinguished by size and the speed of their generation. The private sector often generates big data. *Open data* refers to features including open and free availability, access, and reuse.

To guide its inquiry, the Independent Evaluation Group (IEG) developed a list of ingredients for successful national data systems of the future (table 1.1).

TABLE 1.1 | The Shape of Successful National Data Systems in the Future

| Institutions Based on | Organizations that Have | Data that Are | Users Who Are | Data Uses |
|---|---|---|---|---|
| Open data laws<br>Rights to privacy<br>Accountability to users<br>Broad outreach to society<br>Harmonized data conventions | Budgetary autonomy<br>Trained staff<br>Adequate installations<br>Connected databases<br>Early warning systems<br>International partnerships | Up to date<br>Disaggregated<br>Easy to manipulate and visualize<br>Accessible in remote areas<br>Georeferenced<br>Contestable<br>From integrated data sets | Connected<br>Data literate<br>Diverse (e.g., academics, civil society organizations, media, and local and central governments) | Planning<br>Policy making<br>Monitoring<br>Targeting<br>Research<br>Advocacy<br>Lobbying<br>Citizen empowerment |

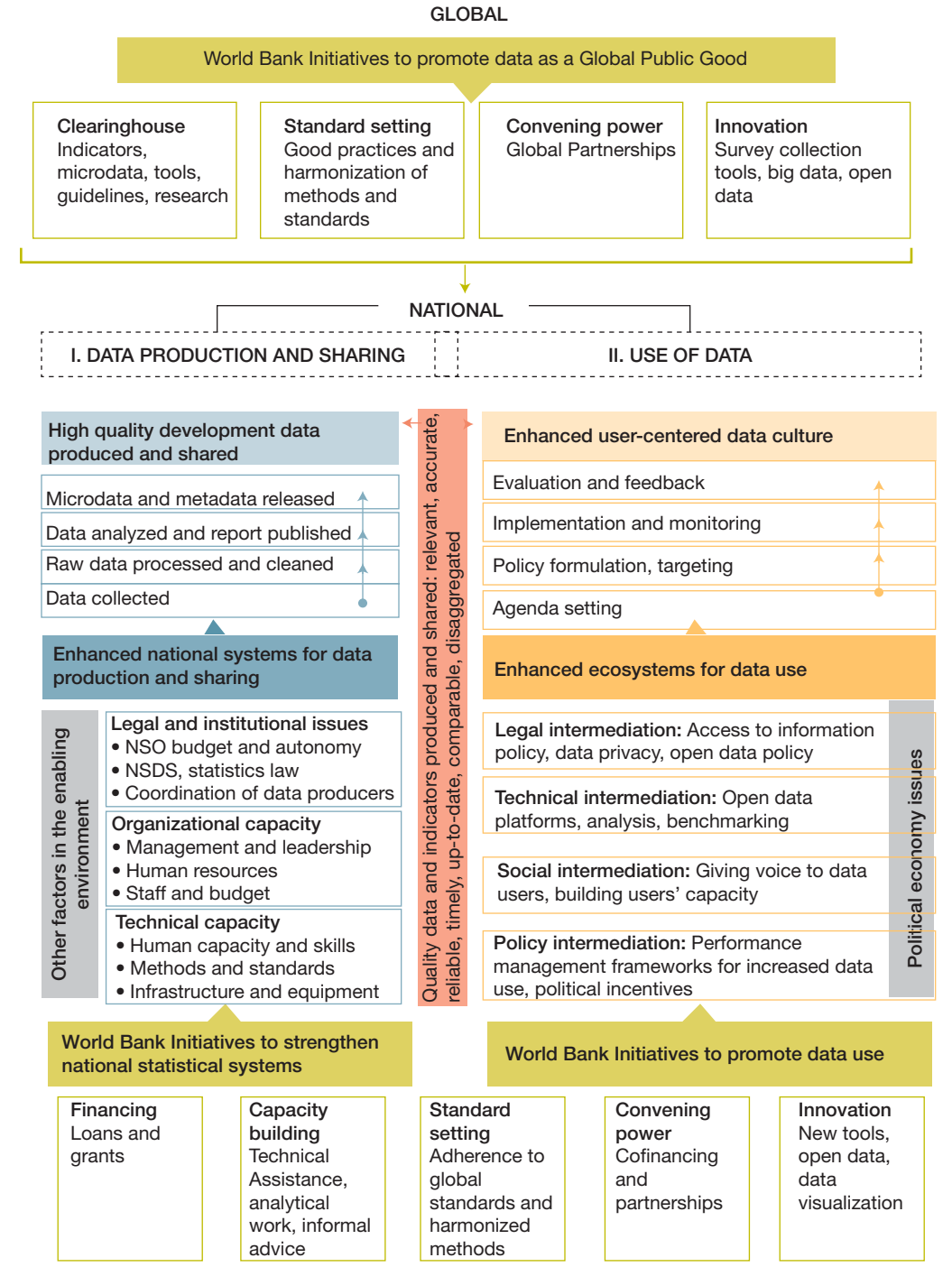
## Methodology

The evaluation is based on an intervention logic that was iteratively reconstructed in dialogue with the literature review, the portfolio analysis, and evidence from case studies (figure 1.1; appendix A). Briefly, the logic implies that to nurture data use, the type of data supplied must be relevant to user needs. Supply can potentially elicit data use and demand, though it is not a sufficient condition for it. If data are of good quality, relevant to citizen needs, and widely shared, their uses might proliferate. People will use data more and become demanding consumers. As rising demand boosts supply, a feedback loop (or virtuous circle) develops that leads to a self-sustaining, user-centered data culture. However, this will happen only if governments and their partners ensure that the data produced are in line with user priorities and if governments commit to sharing data with their people and are willing to use it for policy making.

Better data and greater data use can influence decision making and development outcomes positively. How and when that happens depends on a host of complicated factors (including political) that this evaluation did not pursue. The overarching evaluation question inspired four lines of inquiry that guided the data collection and analysis and the framing of findings and recommendations (box 1.1). The evaluation reviews the World Bank's contributions to development data in individual client countries and its support to data production and partnerships serving the global community, based on the premise that development data are an essential global public good that could be under-produced if left to individual countries.

The evaluation's initial building blocks consisted of a literature review, development of a theory of change, World Bank staff interviews with key informants, and a portfolio review. The findings from this foundational work informed the selection of evaluation instruments and country case studies and helped frame the survey questions. At the global level, the evaluation conducted a structured review of development data partnerships, and structured surveys of targeted World Bank staff (721 responded, or 30 percent) and country stakeholders (506 responded, or 26 percent). A questionnaire obtained the views of 31 development partners. At the country level, the evaluation included 11 case studies of the World Bank's role in country systems involving statistics production,

FIGURE 1.1 | Intervention Logic for Development Data



**Box 1.1 | Four Lines of Inquiry Guiding the Evaluation**

- Has the World Bank contributed effectively to data for the global public good and data partnerships? (chapter 2)
- How effectively has the World Bank helped countries strengthen data production? (chapter 3)
- How effectively has the World Bank promoted data sharing and use in countries? (chapter 4)
- Is the World Bank keeping up with technological innovations, particularly those relating to big data? (chapter 5)

sharing, and use. Various data collection and analysis modalities underlie the cases (field-based and desk-based cases, and project performance assessment reports (PPARs)). IEG purposively selected the countries where case studies took place based on criteria such as the level of World Bank funding, the diversity of support (from lending to advisory services), and the inclusion of large and small countries. The countries selected were Afghanistan, Bolivia, Ghana, Jordan, Kenya, India, Indonesia, Peru, Rwanda, Tanzania, and Ukraine. Furthermore, IEG conducted a structured survey of stakeholders in the national statistical systems of 24 countries that yielded 506 respondents, a 26 percent response rate. Appendix A provides further details on the evaluation methodology.

## References


Besley, Timothy, Peter Henry, Christina Paxson, and Christopher Udry. 2015. *Evaluation Panel Review of DEC: A Report to the Chief Economist and Senior Vice President*. World Bank, Washington, DC.

Scott, Christopher. 2005. *Measuring Up to the Measurement Problem: The Role of Statistics in Evidence-Based Policy Making*. Paris: PARIS21 (Partnership in Statistics for Development in the 21st Century).

Serajuddin, Umar, Hiroki Uematsu, Christina Wieser, Nobuo Yoshida, and Andrew Dabalen. 2015. "Data Deprivation: Another Deprivation to End." Policy Research Working Paper WPS 7252, World Bank, Washington, DC.

UN (United Nations). 2014. *A World that Counts: Mobilising the Data Revolution for Sustainable Development*. New York: UN.

World Bank. 2011. *Marrakech Action Plan for Statistics, Partnership in Statistics for Development in the 21st Century, and Trust Fund for Statistical Capacity Building.* Global Program Review, Volume 5, Issue 3. Washington, DC: World Bank.

———. 2015. *World Bank Group Strategic Actions Program for Addressing Development Data Gaps: 2016–2025.* Washington, DC: World Bank.

———. 2017. *Results and Performance of the World Bank Group 2016*. Washington, DC: World Bank.

# 2

## highlights

## Global Development Data

1 The World Bank has been an effective leader and partner in development data for global audiences, achieving synergy between producing data for the global public good and serving country clients.

2 Support for data production was more intense than for data use.

3 The World Bank should articulate clear goals for its engagements in global data partnerships and maintain a coherent, focused approach.

**IN 1990,** development professionals looking for globally comparable data would buy the *World Development Report* to access its statistical appendix, which contained the *World Development Indicators*. The 1990 *World Development Report* also presented the first estimates of global poverty, based on household surveys for only 22 countries (World Bank 1990). Today, a simple Internet search gives people access to the relevant *World Development Indicators* in milliseconds, and more than 1,000 household surveys from 159 countries—more than 2 million randomly sampled households representing 87 percent of the developing world's population—are the basis of global poverty estimates (World Bank 2017). The World Bank has been at the center of a quantum leap in the past 25 years in the quantity and availability of development data.

This chapter addresses the World Bank's role and contributions to the global data agenda. It explores the World Bank's role and accomplishments on supporting data as a global public good as well as its support to global data partnerships (chapters 3 and 4 cover support to individual countries).[1]

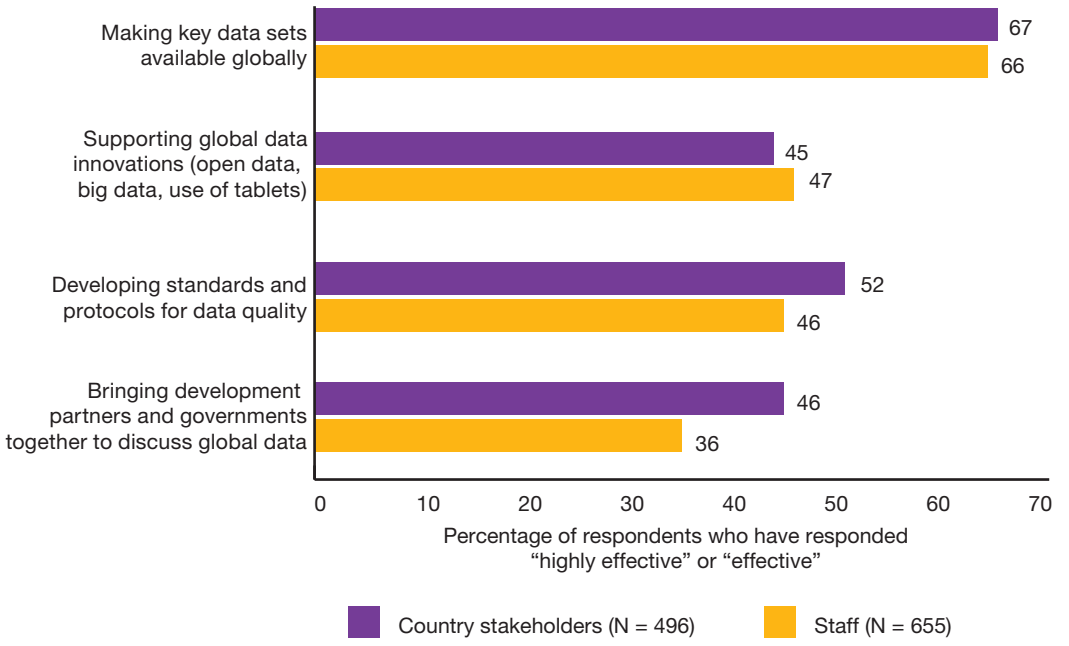
## Development Data for the Global Good

The World Bank's position as a leader and valued partner in development data is broadly recognized and appreciated. IEG's structured surveys and literature review found that the World Bank is generally expected to use its global reach and financial, analytical, and convening powers to support data, which is widely seen as an essential but underprovided public good. In interviews, surveys, and country case studies, staff and stakeholders generally expressed a high degree of satisfaction with the World Bank's global data contributions, and expectations that it should do more to ensure high-quality data for all countries.

In IEG's structured survey of World Bank staff and country stakeholders, more than 60 percent rate the World Bank as highly effective or effective in making key data sets available globally. About half of respondents in each group rate the World Bank as highly effective or effective in developing standards and protocols to ensure global data quality. Between one-third and half of the respondents gave favorable ratings to the World Bank's performance in supporting global data innovations (such as open data, big data, or the use of mobile devices for surveys) and in bringing development partners and governments together to discuss global data issues (figure 2.1).

The World Bank produces influential, widely used global data and cross-country indicators that fill important niches, benchmark countries, and stimulate research and policy action. Examples include *Doing Business*, the Global Findex database, global poverty indicators, the Statistical Capacity Indicator (SCI), and the International Comparison Program (ICP) produced by a dedicated partnership program housed at the World Bank which is possibly the largest statistical operation in the world and allows price comparisons across countries and time through purchasing power

FIGURE 2.1 | Survey Responses on the Effectiveness of World Bank Global Data Support




*Source:* IEG structured survey of World Bank staff and country stakeholders, 2016.

*Note:* In estimating the percentages, IEG excluded "Do not know and No opinion" responses from the denominator.

parities.[2] Though controversial at times and criticized by some on methodological grounds, all of these data and indicators are influential in their respective areas and may have contributed to greater data usage at the global level. *Doing Business* attracts unrivaled media and high-level attention. The Evaluation Panel Review of the Development Economics Vice Presidency (DEC) noted that the World Bank's "leadership in the ICP project shows the potential for the World Bank to create a position at the heart of the global statistics community." It recommended that the World Bank "commit to the ICP, which is a flagship example of global cooperation in data and statistics work where the World Bank Group plays a leading role" (Besley and others 2015). The global poverty measures, Global Findex, and the SCI fill data gaps and provide useful platforms for assessing poverty, financial inclusion, and statistical systems, respectively, in a manner that is comparable across countries. The World Bank's work on household surveys, including the Living Standards Measurement Study, propelled a virtual explosion of multipurpose surveys that help fill the void created by the weakness of other statistics sources.[3]

The global practices and regional vice presidencies lead or take part in many informal and organizational data partnerships that support data production, dissemination, and use sectors and topics. The IEG team identified 34 such partnerships (based on a web search, interviews, and information from DEC). Of these, the World Bank housed 12 partnerships and the other 22 reside

elsewhere or are simply informal alliances that work on data issues. Partners include the International Monetary Fund (IMF), other multilateral development banks, and United Nations (UN) entities.[4] The partnerships housed at the World Bank focus mostly on data collection, dissemination, and benchmarking in health, energy, education, and other sectors.

The United Nations Statistical Commission is at the apex of the global statistical system and has a broad mandate to promote statistics, coordinate specialized agencies, improve methods, and also the adoption of global standards for statistics.[5] The World Bank is positioned in this global landscape as a member or observer in many international statistical bodies, a major program funder and implementer, and a support provider for statistical capacity. Although it has wisely avoided formal data standard setting, the World Bank has helped foster good practices (for example, on poverty measurement and survey design, where it has helped harmonize indicators and standards).[6]

Using its convening power to support global statistical efforts, the World Bank helped establish (and is a member of) data partnership programs that made important contributions and balanced global and national data needs well (box 2.1). The World Bank had a significant role in thought leadership and coordination, and provided technical, operational, and administrative expertise and staff time. Since 1999, it has provided $50.9 million of funding through its Development Grant Facility, 54 percent of which went to partnerships housed outside of the World Bank—Partnership in Statistics for Development in the 21st Century (PARIS21) and Open Data for Development. The rest of the funding went to the Marrakech Action Plan for Statistics (MAPS) secretariat and the ICP housed at the World Bank. Interviews and external evaluations of DEC and global data partnerships hosted there show that DEC has been a strong anchor for much of this effort and a competent host for prominent global data partnerships.[7]

### Box 2.1 | Major Partnership Programs for Development Data

IEG selected nine data partnerships for review because they are formal, relatively prominent, and meet one or more of the following criteria: have a pivotal role in statistical capacity building, address important data gaps in the global statistical landscape, and promote new or innovative approaches.

**1968: International Comparison Program** is a partnership of the statistical offices of up to 199 countries, housed at the World Bank. The program produces internationally comparable price and volume measures for gross domestic product.

**1980: The Living Standards Measurement Study Program** is a household survey program focused on generating high-quality data, improving survey methods, building capacity, and facilitating the household survey data use.

**1999: Trust Fund for Statistical Capacity Building** is a multidonor trust fund that aims to improve the capacity of developing countries to produce and use statistics, with an

*(Box continues on the following page.)*

Box 2.1 | Major Partnership Programs for Development Data (continued)

overall objective of supporting effective decision making for development. The trust fund supports projects aiming to strengthen national statistical systems in priority areas and develop statistical capacity sustainably, including data openness and accessibility in line with the Open Data Initiative and innovative approaches to improve data collection.

1999: Partnership in Statistics for Development in the 21st Century (PARIS21) is a partnership to promote better use and production of statistics throughout the developing world. PARIS21, a worldwide network, is committed to evidence-based decision making through the improvement of institutional and technical capacity, thus stimulating, meeting, and improving national demand through comprehensive national plans for improvement.

2004: Marrakech Action Plan for Statistics is a global plan for improving development statistics, agreed to at the 2004 Second International Roundtable on Managing for Development Results in Morocco. Eight programs have been developed with the UN and other international agencies to put the identified actions into practice.

2009: Statistics for Results Facility is a World Bank–managed multidonor initiative to support statistical development in developing countries. The initiative and its catalytic fund promote statistical capacity building and support better policy formulation and decision making through improvements in the production, availability, and use of official statistics.

2011: The Global Financial Inclusion (Global Findex) database is a comprehensive database on financial inclusion that provides in-depth data on how individuals save, borrow, make payments, and manage risks. The first Global Findex database was launched in 2011 in partnership with Gallup and with funding from the Bill and Melinda Gates Foundation, and a second edition was launched in 2014.

2014: Open Data for Development is a program designed to help developing countries use open data standards, and understand and exploit the benefits of open data. Its objectives are to support developing countries in planning, executing, and running open data initiatives; increase open data use in developing countries; and grow the evidence base on open data's effects for development.

2015: The Global Partnership for Sustainable Development Data is a network of governments, civil society, and businesses working together to strengthen the inclusivity, trust, and innovation in how data are used to promote sustainable development around the world.

The World Bank fostered successful innovations in data collection, sharing, and use, in particular, pertaining to household surveys, demonstrating the complementarities between research and support for data production, sharing, and use. PovcalNet is an extremely popular online tool that automates poverty calculations and allows users to replicate the World Bank's estimates. The World Bank helped develop and promote Survey Solutions, a free, computer-assisted personal interviewing software that eliminates the need for pen-and-paper surveys, incorporates automatic data consistency checks, and speeds up the time from fieldwork to publication of data. The World Bank conducts extensive research on survey methodology. With its partners in the International Household Survey Network, the World Bank developed tools and guidelines for data cataloging and archiving and engaged more than 60 countries in the Accelerated Data Program to document, archive, and disseminate microdata. It developed ADePT, a software platform for economic analysis that automates and standardizes the production of analytical reports from various types of surveys, thus raising efficiency and reducing human errors. These innovative tools are free to download, well disseminated, and respected by professionals in the field. In country visits, IEG saw several examples of counterpart uptake of these tools.

During 2007–16, the World Bank Group considerably stepped up its efforts to increase the availability and use of gender data by supporting the capacity of client countries to produce gender statistics, preparing tools to help produce and analyze gender data, and establishing partnerships. The World Bank's Gender Action Plan bolstered the gender data focus, along with a commitment made as part of the 17th Replenishment of the International Development Association (IDA17). That commitment, to "roll out statistical activities to increase sex-disaggregated data and improve gender statistical capacity in at least 15 IDA countries" between fiscal years 2015 and 2017, was met. The World Bank is an active member of the UN-convened Inter-Agency and Expert Group on Gender Statistics and has provided financial and technical assistance to national statistical offices (NSOs) and line ministries to collect and use gender data. The World Bank made financial contributions to the UN Statistics Division's gender statistics program through the MAPS program. More and better data disaggregated by gender is a key part of the World Bank Group's current gender strategy (many SDG indicators require gender disaggregation).

The World Bank Open Data website is a preeminent global clearinghouse for development data, containing an extensive and user-friendly compilation of indicators, microdata, tools, and guidelines that attracts high web traffic.[8] The World Bank's two most visited websites are the English and Spanish language data sites, which account for around one-third of all traffic to World Bank websites, and five of the World Bank's 12 most visited websites pertain to data. The Open Data Initiative, launched in 2010, was a milestone for free data sharing in development. The World Bank also publishes a range of globally oriented publications that monitor and analyze poverty, shared prosperity, progress toward achieving the Millennium Development Goals, and other indicators.

## Partnerships for Statistical Capacity Building

Over the period 2004–11, the World Bank played a pivotal role in defining the global support architecture for statistical capacity building in its client countries. It spearheaded MAPS in 2004

and the Busan Action Plan for Statistics in 2011 (both of which provided coherence to global statistical capacity–building efforts), contributed to PARIS21, and established the Trust Fund for Statistical Capacity Building and the Statistics for Results Facility. All of these programs emphasized the processes and systems underlying the development of statistical capacity. External evaluations and IEG's Global Program Review show that these partnership programs performed well and made progress toward their goals to improve the capacity of developing countries to compile and use statistics to support management for development results (World Bank 2011). The strongest progress was on production of statistics and on national statistical development strategies.

The major data partnerships housed at the World Bank have collectively received $250 million in donor contributions from 2000 to 2016, for which the World Bank has been the trustee and the implementing agency. The single biggest donor (65 percent of total contributions) is the U.K. Department for International Development (DFID), which focused on general statistical capacity building. The next largest donor is the Bill and Melinda Gates Foundation (21 percent), which focused on the Living Standards Measurement Study and Global Findex.

These partnership programs provided knowledge, networking, technical assistance, and advocacy at the country level; some also allocated grant funding that supported more than 80 countries and regional initiatives. The Regional Program for Improving Household Surveys and Measurement of Living Conditions in Latin America and the Caribbean is often cited as a successful partnership that made a difference in promoting household survey production (for example, Beegle and others, 2016).

External evaluations (World Bank 2011; PARIS21 2015) and interviews show that PARIS21 has been a small but key actor in statistical capacity building. It gained the trust of statistical offices through its training and diagnostic work, developed a strong network, and broadly delivered on its mandate: it was successful at raising awareness of the importance of statistics, helped countries develop national statistical data systems, and is a valuable and value-adding part of the architecture of data and statistics development and cooperation. The World Bank provided financial support to PARIS21 through the Development Grant Facility, but the facility has now ended (as part of a larger cost-cutting exercise), putting the ability to sustain the programs' achievements in jeopardy (the facility provided direct grant support for high-value, innovative global partnership programs to client countries that other funding sources could not adequately support.)

In conclusion, the World Bank channeled support for statistical capacity building through partnership programs it helped convene, support, and execute. These partnerships represented a relevant, coherent articulation of efforts in the past and aligned well with the global development agenda and the World Bank's country priorities. Support was more intense for national statistical data systems and data production than for data use and data users in developing countries. The literature review, interviews, and country cases suggest that engagement with data users was either feeble or nonexistent, and no strategies existed for stimulating demand for data from government, civil society, private sector, academia, and media.

## New Partnerships for Data: Innovation and Proliferation

The World Bank is also a member of a newer generation of partnerships focused on data innovation and housed elsewhere. Much of this effort is experimental with no clear architecture and funding mechanism.

The Open Government Partnership is a multilateral initiative that aims to secure concrete commitments from governments to promote transparency, empower citizens, fight corruption, and harness new technologies to strengthen governance. IEG found that the governments of Indonesia and Tanzania were committed to this partnership, and that the World Bank supported national open government initiatives in these countries effectively. This led, for example, to greater fiscal transparency and increased use of government administrative records.

Based on IEG's experience with evaluating global partnerships, there are grounds to expect that setting up new programs outside of established institutions would lead to lengthy delays and high costs (IEG 2015a). For example, the Global Program for Sustainable Development Data housed at the UN Foundation—established as part of the post-2015 development agenda—has made little progress to date; according to interviews, it has produced few outputs, and the governance structure is still provisional. Yet the World Bank supports this new partnership and has recently created the Trust Fund for Innovations in Development Data to promote a common funding source for scalable innovations in data production and use. Locating the funding source (Trust Fund for Innovations in Development Data) at the World Bank and the governance mechanism (Global Partnership for Sustainable Development Data) outside of the World Bank seems impractical and does not promote alignment with the World Bank's country engagements.

The new partnerships for data innovation are not framed around an articulated architecture, unlike the previous generation of partnerships for statistical capacity building that all united in support of national strategies for the development of statistics (NSDS). It is unclear why so many separate global initiatives are needed or how they relate to each other. Data innovation partnerships could be consolidated and their goals clarified, and the World Bank could engage in them more selectively. The *World Bank Group Strategic Actions Program for Addressing Development Data Gaps* (World Bank 2015a) identifies how some partnerships will contribute to the plan, but does not address many others, nor does it identify funding sources for the anticipated increase in support to data production.

Even though the size and effectiveness of the World Bank's contribution to informal partnerships and interagency working groups is hard to assess, these engagements reflect the breadth of data-related work across the World Bank and the proclivity to collaborate with other partners. Many good initiatives focus on producing and sharing globally comparable sectoral data, but several databases that a global practice collected and set up at considerable expense found little use and eventually shut down. The *World Bank Group Strategic Action Program for Addressing Development Data Gaps* also does not address the World Bank's role and contribution to informal partnerships led by the global practices and regions, and the overall picture that emerges is one of initiatives that are individually relevant, but sometimes disjointed.

## The World Bank's Role in the Global Statistical Landscape

The demand for data to monitor Sustainable Development Goal (SDG) indicators greatly exceeds the supply. The 17 SDG goals have 169 targets and 230 indicators (of which about half lack sound methodologies, adequate country coverage, or both). Several of these indicators may not be relevant for national policy making and will likely be unrealistic for countries to collect. Most of those interviewed by IEG saw SDG monitoring as more of a risk than an opportunity for statistical systems in developing countries. Many were concerned that the SDG agenda is setting up statistical systems for failure—that is, countries' NSOs could unfairly come to be seen as having failed at rising to the herculean challenge of SDG monitoring, though some also see SDG monitoring as an opportunity to give the NSOs more prominence.[9]

Asked to reflect on World Bank priorities going forward, 54 percent of staff included "making key data sets available globally" in their top five areas of strategic thrust—a higher proportion than for any other area. Fifty percent of country stakeholders chose "global availability of data sets" in their five preferred areas (appendix C). In write-in comments to the structured surveys conducted for this evaluation, staff, stakeholders, and partners noted the many and diverse data gaps that deserve more attention, with no clear pattern regarding sectors, data type, and balance between international comparability and individual countries' data needs. Likewise, staff are quick to note major data gaps, with emphasis on those gaps that most affect their own sector or line of work (for example, infrastructure data, household surveys, or enterprise data). The tension between international comparability and individual countries' data needs can be real, and given that resources are finite, the World Bank may consider developing a methodology to weigh the costs and benefits of country specificity versus cross-country comparability.

The World Bank has committed itself to increasing support for poverty data and adopted a corporate target of supporting a new household survey every three years in 78 data-deprived countries, starting in 2020. This commitment is in line with IEG's recommendations in the evaluation *The Poverty Focus of Country Programs: Lessons from World Bank Experience* (World Bank 2015b), and will support measurement of progress toward the twin goals and selected SDG targets. However, funding is uncertain.

Trade-offs can exist between global and national data priorities, but are not too great to overcome. Domestic policy makers may want better economic statistics, geographically disaggregated indicators, and surveys and censuses of economic establishments for taxation purposes. Donors and international organizations can favor social statistics, global monitoring data, and household surveys (which they sometimes commission in an uncoordinated fashion). In practice, the World Bank often managed this trade-off well. It has helped build statistical capacity (chapter 3), and the household surveys it promotes are designed for multiple purposes, not just poverty monitoring.

The World Bank has an important role to play in coordinating and funding support for national statistical systems. Data from PARIS21 and the literature review and interviews conducted for this evaluation point to a fragmented, redundant, and insufficiently funded global statistical community in which agency-specific interests sometimes take precedence over country needs. The World Bank could coordinate and fund general statistical support to countries and contribute to partnerships that serve global coordination and leadership roles.

## Conclusions

The World Bank has earned a solid reputation in the field of development data based on its research and technical expertise, the ability to link global needs to national needs, initiatives that it sustained for the long term, and its well-aligned and successful partnerships. The World Bank has performed well on data for the global public good because of its strong ability to engage with countries' statistical systems through the full range of its financial and knowledge instruments, and by working closely with global partners. The best data initiatives and partnership engagements filled clear niches—adequate staff and sustained funding from internal sources and trust funds maintained them for decades. This type of long-term engagement helped build the World Bank's comparative advantage in household surveys. The implication going forward is that the World Bank should consider the long-term sustainability of its data initiatives.

The World Bank and its partners will need to protect the gains made under MAPS and the Busan Action Plan for Statistics, which provided legitimacy and funding for statistical capacity building.[10] The risk is that the existing, well-functioning partnership architecture will stop receiving adequate funding as donors' attention shifts to a newer generation of less clearly articulated data partnerships with lofty ambitions, overlapping goals, and insufficient funding. The Development Grant Facility phased out, thus ending the World Bank's financial support for PARIS21, with potentially adverse consequences for the small, but well-regarded program with a solid record of accomplishment. This could reverse past gains in statistical capacity.

The World Bank and its partners will also need to work toward maintaining coherent global efforts. The World Bank has not spelled out priorities for its engagements with the global statistical community, and for how its formal and informal data partnerships will support the proposed scaling-up of World Bank investments in data.

---

[1] This report is about data as a public good, meaning data that are non-rival and non-excludable. One person or country's enjoyment of data does not affect its enjoyment by others and no person or country can be excluded from sharing its benefits. Some data are clearly global public goods (international price comparisons, for example), other data are clearly national public goods (population numbers by district, for example), and some are both global and national public goods. Data for the private good (proprietary firm data, for example) are not covered in this report.

[2] This is not an exhaustive list. The World Bank also produces data on member countries' debt, the Worldwide Governance Indicators, the Country Policy and Institutional Assessment data, enterprise surveys, Service Delivery Indicators, the Atlas of Social Protection: Indicators of Resilience and Equity, and more.

[3] Another example of data partnership is a joint World Bank–IFC initiative that in 2008 launched GEMX, a private sector–led global bond index that tracks emerging market local currency sovereign bonds. This index is still published, but was not widely adopted (World Bank 2016).

[4] An example of a new partnership (with the Bank of Italy, the Food and Agriculture Organization of the United Nations (UN), and the International Fund for Agricultural Development) is the Center for Development Data focused on methodological innovation in household surveys and agricultural statistics located in Rome.

[5] Setting standards is an official mandate of the UN Statistical Commission. The legitimacy of formal UN representative intergovernmental processes enables it to build consensus for formal principles and technical standards. Many World Bank projects seek to help countries comply with these global standards.

[6] The Commission on Global Poverty, convened by the World Bank, provided useful recommendations on technical issues in poverty measurement (World Bank 2017).

[7] The hosting function means that the partnership programs' secretariats are located in the Development Economics Vice Presidency (DEC) and are legally part of the World Bank. Data partnerships hosted at DEC are commendable for undertaking regular external evaluations.

[8] The data portal at http://data.worldbank.org has a rich collection of data and tools.

[9] There is also an SDG target on enhancing capacity building support for data.

[10] Recent evaluations of the UN system and of the UN Population Fund made similar points (UN 2016; UNFPA 2016).

## References


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

Besley, Timothy, Peter Henry, Christina Paxson, and Christopher Udry. 2015. *Evaluation Panel Review of DEC. A Report to the Chief Economist and Senior Vice President*. World Bank, Washington, DC.

PARIS21 (Partnership in Statistics for Development in the 21st Century). 2015. *Light Evaluation of PARIS21, Final Report*. Paris: PARIS21 Secretariat.

UN (United Nations). 2016. *Evaluation of the Contribution of the United Nations Development System to Strengthening National Capacities for Statistical Analysis and Data Collection to Support the Achievement of the Millennium Development Goals and Other Internationally Agreed Development Goals.* Report no. JIU/REP/2016/5. New York: UN.

UNFPA (United Nations Population Fund). 2016. *Evaluation of UNFPA Support to Population and Housing Census Data to Inform Decision-Making and Policy Formulation (2005–2014).* New York: UN.

World Bank. 1990. *World Development Report 1990: Poverty.* New York: Oxford University Press.

———. 2011. *Marrakech Action Plan for Statistics, Partnership in Statistics for Development in the 21st Century, and Trust Fund for Statistical Capacity Building.* Global Program Review Volume 5, Issue 3. Washington, DC: World Bank.

———. 2015a. *Opportunities and Challenges from Working in Partnership: Findings from IEG's Work on Partnership Programs and Trust Funds*. Washington, DC: World Bank.

———. 2015b. *The Poverty Focus of Country Programs: Lessons from World Bank Experience.* Washington, DC: World Bank.

———. 2015c. *World Bank Group Strategic Actions Program for Addressing Development Data Gaps: 2016–2025.* Washington, DC: World Bank.

———. 2016. *The World Bank Group's Support to Capital Market Development*. Washington, DC: World Bank.

———. 2017. *Monitoring Global Poverty. Report of the Commission on Global Poverty*. Washington, DC: World Bank.

# 3

## Building the Data Capacity of Countries

## highlights

1. The World Bank used its own financing and financing from small trust fund grants to engage a very large number of countries on data activities.

2. Improvements in data availability, quality, and timeliness are observable in the few countries where the World Bank engaged in-depth on institutional reforms and capacity strengthening.

3. Progress is slow and uneven, many countries are still data deprived, and others continue to have weak data systems, especially regarding administrative data.

4. Country clients need support that is more coordinated and long-term from the World Bank in strengthening their administrative data systems and supporting statistical capacity building beyond national statistical offices.

**THIS CHAPTER** examines the World Bank's role and contributions to countries' data production and statistical capacity building. Although each type of data (for example, household surveys, census, and price data) undeniably requires its own set of skills, techniques, methods, and protocols, this chapter focuses on the building blocks that are the basis for many data-related activities. Statistical capacity as defined by Partnership in Statistics for Development in the 21st Century (PARIS21) "Is the sustainable ability of countries to meet user (government, policy makers, researchers, citizens, and business) needs for high-quality data and statistics (that is, timely, reliable, accessible, and relevant)" (PARIS21 2015). Capacity building has four aspects: institutions (including laws and enabling environment), human capital (knowledge, skills, and staff incentives), organizations (budget, infrastructure, leadership, collaboration, and coordination between statistical stakeholders), and data systems and technologies. The chapter highlights the evolving model and expanding scope of World Bank support to country data systems while focusing more extensively on the core approach to capacity building of National Statistical Offices that has been prevalent until recently.

The evidence underlying this chapter is from a review of past evaluations and project documents, surveys and interviews with a large number of partners and clients, and 11 in-depth case studies of statistical capacity–building initiatives. Only a small subset of World Bank statistical capacity–building projects were subject to a formal evaluation.[1] Therefore, this chapter focuses on the extent to which project designs are in line with well-established good practices rather than on detailed analysis of results achieved, except for the case study countries where in-depth analysis was possible.

## A Reliable Partner of National Statistical Systems

IEG consulted with 276 external stakeholders through interviews and 506 stakeholders through surveys and found that overall, the World Bank is perceived as a trusted government partner with sought-after statistical expertise, one that benefits from a far-reaching convening power, is active in a wide range of development areas, and has a distinct role as a funding organization. The World Bank forged this solid reputation through a variety of technical and financial engagements to support countries' data production, sharing (to a lesser extent), and use (to a limited extent).

The portfolio review conducted for this evaluation found that between 2005 and 2015, World Bank commitments for data activities averaged about $90 million per year and increased in the latter half of the evaluation period. The World Bank is still the largest provider of development cooperation in statistics with 37 percent of the total global commitment and 53 percent of the country-specific commitments in 2014 (PARIS21 2016a, 23–24). This is a lower-bound estimate that does not include

the many activities in which the World Bank produces, shares, or uses data as an input to or by-product of other work (for example, analytical work that helps countries analyze and interpret data, or impact evaluations that collect surveys).

Relatively few countries absorbed most donor support for data—the top 25 recipients received more than 60 percent of support. Furthermore, countries with the lowest statistical capacity do not always receive the most assistance. The World Bank's long-term statistical capacity development support is also concentrated on a few countries, leaving others with a minimal level of assistance. Twenty-six countries obtained World Bank assistance with a loan or grant of more than $2 million dedicated entirely to data. In other countries, the World Bank privileged direct support to targeted data collection or sharing needs. Box 3.1, Figure 3.1, and appendix B provide more details on the World Bank's portfolio of activity.

The World Bank has been most effective when partnering with other donors to support statistical capacity building, but it too rarely does so. The portfolio review found that only 28 of the 201 data projects reviewed involved other development partners.[2] In surveys conducted for the Statistics for Results Facility evaluation, national statistical offices (NSOs) expressed concern that development partners continue to have weak harmonization (Ngo and Flatt 2014). As in other sectors, the advantages to countries of multi-partner data support include more coordinated and harmonized approaches and pooled funding. Although the World Bank is governments' preferred partner to work on data issues in certain regions (especially Africa), it has less leverage to influence reforms in

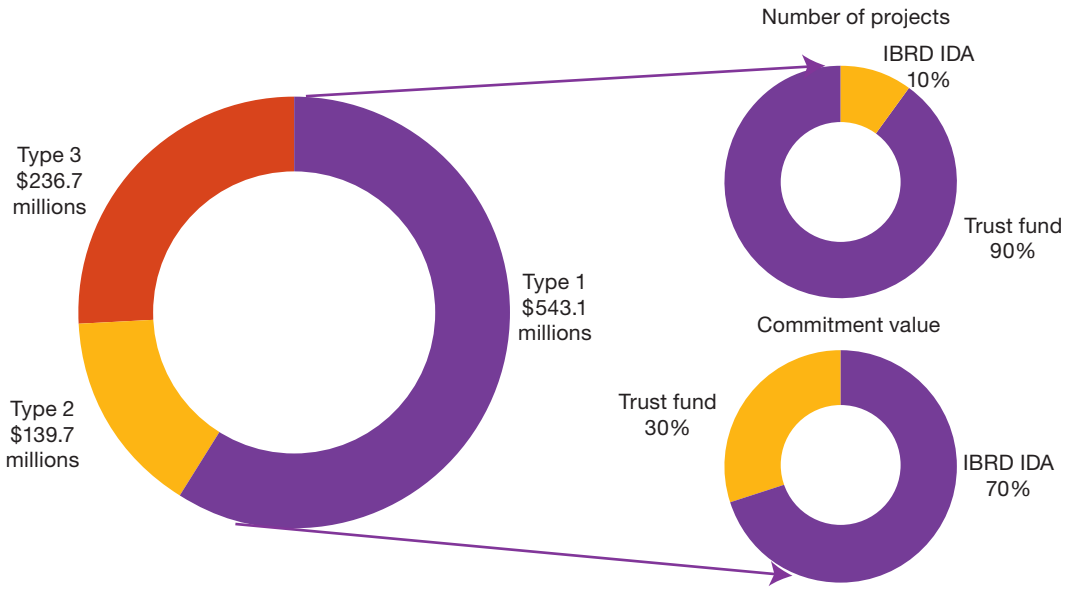
### Box 3.1 | The World Bank Portfolio of Projects on Data for Development

IEG identified 225 World Bank projects that supported countries' data production, sharing, or use between 2005 and 2015. IEG classified these projects into the following three categories:

- Type 1: The entire project supported data
- Type 2: At least one entire component supported data
- Type 3: The project supported relevant data activities, but the project components were not specifically data related.

Statistical capacity–building initiatives are type 1 projects. The World Bank relies heavily on multidonor partnerships to invest global resources in national statistical systems. Among the 139 type 1 projects, 125 were trust funded (these represented 30 percent of total data commitments).

FIGURE 3.1 | Overview of World Bank Financing Commitments




*Note:* Type 1 projects supported data activities entirely; type 2 projects had at least one component that supported data entirely (the commitment value for only the data component is included); and type 3 projects supported relevant data activities, but the project components were not specifically data related. Only the commitment value for data is included. This report understates the commitment value of these projects because IEG excluded development policy financing, as the amounts could not be reliably estimated. The IBRD IDA category combines the data for both sources of funding. IBRD = International Bank for Reconstruction and Development; IDA = International Development Association.

others. About 44 percent of the number of data commitments and 45 percent of the value of data commitments were to African countries.

Pooled-funding mechanisms are particularly effective to ensure donor alignment and government ownership. However, this is the exception in World Bank lending. In Kenya and Rwanda, the World Bank followed the United Kingdom in joining an established basket-fund modality and had a critical role in the Joint Government-Development Partners Steering Committee. This mechanism was a channel for communication, setting priorities, and upholding professional standards among the participating agencies. All parties interviewed during the case study depicted it as a key factor in explaining the fast development of Rwanda's NSO capacity. Conversely, in the Democratic Republic of Congo, the World Bank did not join an existing pooled-funding mechanism established by the United Nations Population Fund to support the census.

Partners and clients most appreciated the World Bank's capacity to combine statistical expertise—providing credibility to the statistical information generated by country systems—and the managerial expertise to lead on large investments. These comparative advantages emerged clearly in all case studies. The World Bank's contribution to reestablishing trust in the Peruvian statistical system is highly significant in this regard. Official poverty estimates were unavailable in Peru between 2004 and

2007, which triggered a loss of credibility of the NSO. The authorities requested World Bank technical assistance to improve methodologies and help restore public trust. Instead of providing only traditional technical assistance, the World Bank established an external advisory committee made up of poverty experts from the public sector, academia, and international organizations to agree on the best way to produce comparable poverty estimates. The NSO was able to issue comparable poverty figures for all years from 2001 on, public trust was restored, and several data initiatives resulted from this experience.

## Direct Support to Data Collection and Sharing

Direct financing of data collection activities is the most widespread form of World Bank support to data production, and 56 percent of the projects reviewed involved support for collecting data. Production of household survey data received the most attention in a number of projects (20 percent). The World Bank Group Data Council identified five priority areas that will be the focus of World Bank engagement going forward.[3] About 40 percent of projects included support for collecting data in at least one of these five priority areas. Although this form of direct support quickly triggers visible impact, it fails to address systemic issues that hinder countries' long-term production capacity. Capacity-building initiatives can better address these issues. The World Bank used two approaches depending on the configuration of countries' needs, existing capacity, and funding availability.

In responding to a broadening data agenda that recognizes the importance of data sharing and use, World Bank support has widened to encompass a broader array of activities such as improving data dissemination and open data initiatives. Of the 201 projects reviewed, 68 projects provided direct support for increasing public access to development data—for example, through open data portals, training on publishing microdata, and technical assistance for dissemination policy.

## Building Capacity with Institutional Reforms and Technical Strengthening

Under the auspices of global partnerships, the World Bank has contributed to testing and adopting a sectorwide statistical capacity–building approach anchored on the design, funding, and implementation of national strategies for the development of statistics (NSDS). Existing evaluations (Willoughby 2008; World Bank 2014; UN 2016; UNFPA 2015) point to the approach's high relevance to country needs and to encouraging progress (though slow) in improving data production. Evidence is still thin on the impact of specific statistical capacity–building components and on the most adequate sequencing.

The larger statistical capacity–building projects the World Bank managed have typically had the following components: institutional development and legal reform, human resource capacity development, development of statistical systems and databases, data collection and dissemination, and support to physical infrastructure and equipment. Only few countries benefited from this full package. Most projects—particularly those funded by the Trust Fund for Statistical Capacity Building (TFSCB)—had the resources to cover only one or two of those components.

In the overall portfolio, 50 percent of the projects supported the strengthening of client institutional capacity with various degrees of success. In particular, the World Bank supported reforms that seek to enhance NSOs' autonomy and stature, ensuring that they are independent of a parent agency and their management is not vested into a governing body (typically a board of directors). Making NSOs autonomous helps shield official statistics from political interference and improves the organization's effectiveness and control over staff resources (Kiregyera 2015). Autonomous NSOs are also more respected, attract public confidence, and raise the profile of statistics in the country, as shown in the examples in box 3.2.

To improve the quality of data produced by client countries, World Bank financing set priorities for human capacity strengthening, especially through training for NSO staff—the most common

### Box 3.2 | The Significance of National Statistical Office Autonomy: Two Contrasting Cases

Kenya: In August 2010, the government of Kenya acted on the recommendation of its autonomous national statistical office (NSO) and rejected census results submitted by eight northeastern districts because the population figures were inflated and unsupported by documented trends of births, deaths, and migration. The eight districts' leaders promptly filed a lawsuit with the high court. After four years, a five-judge panel agreed with the NSO. The bureau was free to declare its figures official statistics eligible for use in public policy, including determining how to divide the national revenue among the 47 counties.

The decision gave the NSO much-needed credibility. If the appellate judges had ruled in favor of the eight districts, the NSO would have faced years of uphill struggle trying to nurture a nascent institution while repairing the damage to its reputation.

Ukraine: Under the 1993 law, the NSO reported directly to the Cabinet of Ministers of Ukraine, had its own budget, and enjoyed operational autonomy. However, the government reorganization of 2013 put the NSO under the supervision of the Ministry of Economic Development and Trade. The NSO lost the autonomy it previously enjoyed, along with much of its professional independence. Under this arrangement, the NSO submits its work program to its parent agency, which can approve or reject the line items. Resources are insufficient to cover physical infrastructure maintenance.

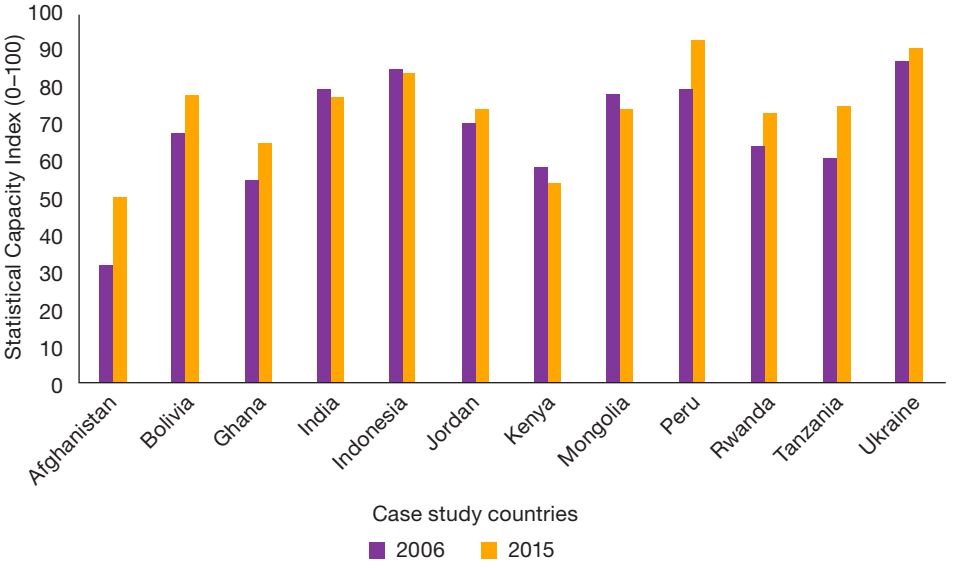
A new draft law on statistics is now in preparation. It includes changes to give the NSO greater operational autonomy and professional independence by returning to the reporting structure that was in place before 2013 and establishing an advisory body of data producers and users.

form of support in 88 percent of the reviewed projects in the portfolio. Furthermore, 40 percent of the reviewed portfolio sought to improve statistical methods, standards, and classifications. The interviewed NSO staff particularly appreciated the World Bank's support for the adoption of internationally accepted standards in data collection and the transfer of best practices in projection for economic statistics, an area somewhat neglected by other donors.

IEG reviewed the available project completion documents for 75 of 146 closed World Bank operations to assess the results of World Bank support for data activities. IEG rated the extent of results achieved for each dimension of statistical capacity building on a scale of 0 to 3 (with 0 representing no documented results and 3 representing a high degree of results achievement).[4] Strengthening legal frameworks and building human capacity are two dimensions with well-documented positive achievements, though efficiency could improve. Issues often surfaced regarding per diem for trainees, selection of trainees, and the cost-effectiveness and sustainability of training. The World Bank could have used the technical expertise of specialized institutes better, such as the East Africa Statistical Training Center based in Dar es Salaam. And higher salaries in the private sector can make it hard for NSOs to retain trained staff, yet support to NSOs on human resource management was rare.

Data reliability, timeliness, and quality control improved in client countries where the World Bank intervened with a comprehensive package of activities and large funding, as illustrated by the evolution of the Statistical Capacity Indicator (SCI) in case study countries (figure 3.2).[5] However, this progress is not attributable to World Bank interventions alone because the SCI also increased elsewhere, but improvements in national statistical systems' fundamentals associated with World Bank interventions likely translated into improved data production capacity also beyond the SCI metrics.

FIGURE 3.2 | Statistical Capacity Improvement in Case Study Countries




*Source:* World Bank data.

Development and implementation of NSDS has been the cornerstone of the World Bank's statistical capacity building, and most projects have used the NSDS as their operative backbone. Until recently, the World Bank was a main funder of PARIS21, which spearheaded the NSDS. The World Bank also implemented a large number of TFSCB initiatives centered on developing or operationalizing NSDS.

As of January 2016, 58 of 77 IDA countries have implemented an NSDS, are now designing an NSDS, or are awaiting the implementation of an NSDS. An additional 14 countries are in the process of planning an NSDS (PARIS21 2016b, 2). The growing number of countries implementing NSDSs is promising because an NSDS is a powerful framework for building capacity and mainstreaming statistics; they also promote donor alignment (PARIS21 2015b). Unlike plans in some other sectors, NSOs have mostly owned NSDS and used them to coordinate donor support. In India, state governments are developing their own NSDS with World Bank support. The feedback from interviews and surveys is largely positive on the usefulness of NSDS and the effectiveness of the World Bank in supporting them. However, while providing a common framework for cross-sectoral data collection, NSDS are not sufficient to ensure effective coordination between the NSO and line ministries, which remains weak in many countries. In addition, in countries that do not benefit from substantial funding from the World Bank and its partners, NSDS implementation can stall because of low capacity and lack of resources (PARIS21 2016c). Closing both the multicountry Statistical Capacity Building Program (STATCAP) and the Statistics for Results Facility Catalytic Fund, and the decision to stop funding PARIS21, threaten future progress.

## Statistical Capacity Building in Fragile States

Data gaps are often dire in countries affected by fragility, conflict, and violence. The community of experts working on fragility is divided on whether it is a good idea to set up formal statistical systems—an inherently slow process—or whether this step should be advanced through technology and alternative data sources. Meanwhile, the World Bank has undertaken statistical capacity–building activities in almost all countries in a fragile situation. These are mostly small trust-funded activities targeting specific data collection (for example, support to the household budget survey in the Republic of Yemen or to Kosovo's judicial statistics), or just-in-time support to the NSO (for example, support to Lebanon's statistical master plan).

The World Bank also planned large projects in several fragile countries, committing $14 million to Afghanistan to strengthen the country's statistical system, $11 million in the Democratic Republic of Congo, and $9 million in South Sudan. In Sudan, the World Bank supported the fifth population census with a $34.4 million grant. The population headcount was instrumental in defining power sharing between North and South and the territorial organization of the new state of South Sudan.

Although the World Bank has had some success in the face of adverse conditions, it designed projects that were too complicated in Afghanistan, Iraq, and Sudan. In Afghanistan, a series of events in 2013 that were outside the statistical office's control, coordination challenges with the twinning partner, the political situation and security issues, and inadequate design slowed project

implementation and led to the cancelation of two-thirds of the funds. In contexts where institutions and capabilities are the weakest, the World Bank and its partners need to adapt their standard model and deploy specific expertise to fit these special circumstances.

## Success Factors in Statistical Capacity-Building Initiatives

In the fast-moving, tech-heavy world of the data revolution, statistical capacity–building initiatives are reputedly slow-paced, unwieldy, and somewhat archaic endeavors that could somehow be bypassed through investments in smart devices and big data analytics. This reasoning is misguided because technological solutions cannot be useful without the right institution and proper skills (the core of statistical capacity building). World Bank–supported statistical capacity initiatives have had high transaction costs and have been slow to show results. However, this is characteristic of this type of intervention, which seeks change at the system level. Shaping institutions requires building trust, which takes time, perseverance, and soft skills.

The World Bank has learned through the years how to design and implement statistical capacity initiatives that improve data production, and it should fully apply the lessons it learned in a larger number of countries. Success factors include gaining government's trust and using its leverage through formal mechanisms such as the systematic country diagnostic (SCD) and country partnership framework (CPF) and the NSDS. Other factors include continuous policy dialogue and technical assistance at multiple levels, engaging for the long term (eight to 10 years according to the case studies), and using the right instrument mix.

### Fostering Trust and Ownership

The World Bank's effectiveness in statistical capacity building depends on staff's ability to combine technical expertise and soft skills and to stay informed of political developments. Many World Bank staff, especially those based in country offices, provide valuable day-to-day support and dialogue. The in-country statisticians funded through Statistics for Results or some STATCAP projects helped ensure that countries sustain the gains made in statistical capacity–building projects. Building relationships is far more difficult when task team leader or in-country statistician turnover is high, or when project supervision is entirely from headquarters.

The IEG's evaluation of World Bank Group country engagement (World Bank 2017) found uneven attention to data issues in SCDs and CPFs: "Many SCDs identified knowledge gaps to improve the evidence base for future policy making; this was a useful input for the analytical agendas in the CPFs. Data gaps also inevitably meant that some SCDs suffered from weaknesses in their analysis of current circumstances and future needs for achieving the twin goals. It is therefore important that SCDs identify knowledge gaps and data limitations, and that CPFs aim to close gaps and improve data quality." From the perspective of this evaluation, one may add the need to ensure links to the NSDS and policy dialogue.

### Continuity of Engagement and the Right Instrument Mix

Statistical capacity building takes time. The average length of larger statistical capacity initiatives (more than $2 million) is 5.5 years and can range from three to 11 years. The World Bank lacks a readily

available instrument that allows long-term engagement of the kind needed for statistical capacity building. Realizing that it takes more time to achieve the intended transformation, the World Bank often resorted to various options for prolonging engagement, including additional financing or a second intervention. The World Bank supported three or more data-related interventions in 34 of 97 countries during the 10 years covered by the portfolio review. Considering the numerous analytical and knowledge services not captured in the portfolio, the number and diversity of data-related activities in any given country is even wider. Therefore, the question of the sequencing of operations becomes important.

The World Bank wisely used smaller grants to prepare for larger and more long-term lending, and to ensure continuity of engagement. Statistical capacity building is an area in which trust funds have aligned remarkably well with other core World Bank activities. In Indonesia, for example, the World Bank supported many data-related activities financed with trust funds or nonlending technical assistance ranging from informal advice to conducting special surveys, developing and maintaining useful databases, sharing tools (for example, ADePT, Survey Solutions, computer-assisted personal interviewing, and microdata library support), convening knowledge networks, and releasing publications that help socialize Indonesian data. An evaluation of these grants noted that they had a considerably larger effect than might be expected from their modest amount. The grant-funded activities also created trust and paved the way for a large lending operation to modernize the NSO.

The selected financing modality determines the length and nature of engagement. By commitment volume, investment lending projects accounted for 82 percent of the data for development portfolio, followed by policy lending (17 percent) and Program for Results financing (1 percent). Modalities that allow for long-term and hands-on engagement, by combining investment lending and technical assistance, were preferable. The World Bank chose a stand-alone development policy financing (DPF) in India for $107 million (the only such example in the portfolio) based on the realization that supervising a lending operation in multiple states would have been impractical. However, the DPF instrument did not provide the means for maintaining necessary World Bank engagement and technical assistance with the states.

## Mobilizing Domestic Resources and Strengthening Administrative Data

It is essential for countries to mobilize domestic resources for their statistical systems. Using donor funding for a core government function such as statistics may provide the resources needed for collecting a survey or build capacity in the short-to-medium term, but is not sustainable in the long term. The literature and interviews with staff and partners make it clear that many governments are not necessarily inclined to dedicate enough domestic budget for national statistical systems. A related issue is that governments rarely raise the status of statisticians by establishing a separate profession in the civil service with salaries and career paths to attract and retain the right candidates. Consequently, NSOs face difficulties in managing their own human resources and lose qualified staff to the private sector, civil society, and international organizations, or rely on project per diem allowances to maintain staffs.

The World Bank should use its leverage and lending instruments more effectively to ensure that data-related activities are adequately funded, including through domestic resources. The World Bank's approach should demonstrate the value of using different forms of data, promote evidence-informed decision making, and raise data issues in country policy dialogue more systematically. Survey respondents believe that mobilizing funding for development data should be among the World Bank's top priorities (51 percent of staff and 64 percent of stakeholders indicated so). The support to the *World Bank Group Strategic Actions Program for Addressing Development Data Gaps* by the IDA18 Replenishment participants also opens the door to leveraging IDA as a funding source to supplement domestic resources and trust funds.

The World Bank has concentrated its support to NSOs so far, with 71 percent of type 1 projects (entirely dedicated to supporting data) targeting the NSO primarily or solely.[6] However, an undue focus on NSOs to the neglect of the national statistical *system* would be a missed opportunity; the capacity of other parts of the national statistical system must also be improved. Government counterparts interviewed in all case studies consistently emphasized that data used to inform policy making, service delivery, and monitoring and evaluation, needs to be disaggregated enough to meaningfully represent the local level, and it must be available regularly. Surveys can rarely meet these needs. National statistical systems have struggled to keep up with the growing demand for data from the global community, and there is concern that the numerous Sustainable Development Goal indicators will stay unmeasured. Administrative data, which are typically collected by line ministries and subnational governments, can potentially bridge this gap. One of the priorities in the *World Bank Group Strategic Actions Program for Addressing Development Data Gaps* is for Civil Registration and Vital Statistics, which is based on administrative records.

In many sectors, however, data quality in administrative data systems is weak and data are little used. Service providers often collect administrative data, frequently without any independent monitoring, which raises questions of data integrity. While NSOs are technical organizations staffed by professional statisticians and governed by international statistical principles and standards, there is much greater variance in processes and systems for data collection and production across line ministries, and capacities are often weaker at lower levels of government. Efforts to build administrative systems should take stock of the landscape and factor in the cost and time that will be needed. More than half of the respondents of IEG's three surveys indicated that they use household surveys as one of the primary data sources, yet less than one-third of respondents in each survey indicated similar use of administrative data (figure 3.3).

A recent report by Development Gateway provides detailed insights on constraints and progress underpinning Tanzania's health administrative data systems (Bhatia and others 2016). Although a new web-based health information management system has led to better coordination of data collection in the health sector since its rollout in 2013, many rural clinics still cannot access the system. Facility staffs continue to collect and manage data on paper and could spend as much as 25–30 percent of their time filling out reporting forms, typically near reporting deadlines. Furthermore, remote facilities often struggle with getting data to district offices. Administrative data completeness, quality, and timeliness suffer as a result.

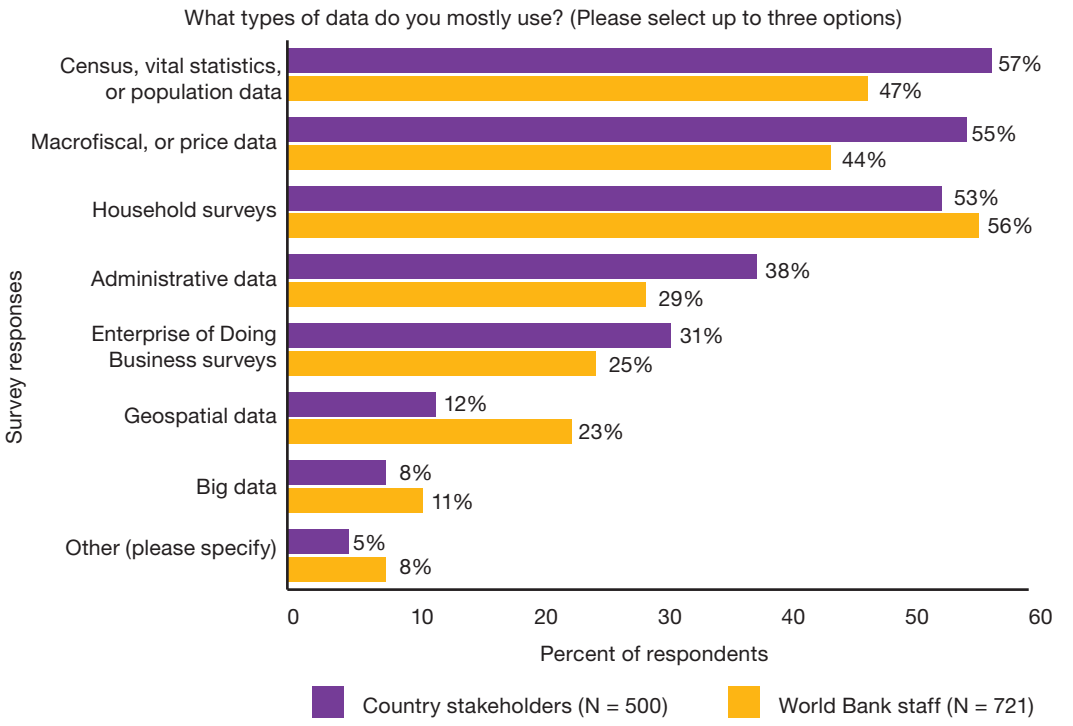
Development partners aligned well in building NSO capacity to produce data, but efforts to build administrative data systems are dispersed and donor-centric. Officials described donors' tendency to build sector management information systems to fit their own monitoring and evaluation needs instead of the countries' systemic data needs, causing a proliferation of fragmented databases across various parts of governments. One person interviewed called this trend "the monitoring and evaluation curse." While World Bank support to strengthen administrative data systems takes place across global practices, primarily as components of other sectoral interventions, this support should be better tracked and coordinated both in the World Bank and within the Government (across the NSO, line ministries, and sub-national levels.) Any support to administrative data systems provided through cross-sectoral engagements should also be tracked. This would be critical to achieving a shared digital infrastructure for data which avoids duplication and maximizes synergies.

An exception seems to be the coordinated efforts to build management information systems in social protection, an area in which the World Bank provides leadership. In Rwanda, for example, DFID, the United Nations Children's Fund, and the World Bank (through its DPF series) have supported the ministry of local governments in its ambitious attempt to create an integrated management and evaluation information system linking more than eight social protection programs. This endeavor requires hands-on support, and DFID is funding three in-country advisers embedded fully within the government. This example illustrates the need to explore different capacity development approaches to cost-effectively support data producers in line ministries and local governments where capacity is often weaker and more heterogeneous.

Similarly, in Peru, the World Bank effectively supported major reforms in the social sectors and in the country's data systems by combining a five-year programmatic advisory service with a subsequent DPF series. The programmatic advisory service generated useful data and supported dialogue with the incoming authorities, and extensive dissemination helped create consensus around social sector reforms. It paved the way for the subsequent DPF series, which supported accountability frameworks in health, nutrition, and education with several prior actions focused on improving data production and dissemination. Civil registration and vital statistics (CRVS) systems urgently need support, as shown by their priority status on several countries' NSDS.

The World Bank Group *Strategic Action Program for Addressing Development Data Gaps* makes CRVS one of its three priorities for the near future. The program recognizes that "Robust CRVS systems, together with national identity management systems, form the foundation of all sectors and pillars of the economy and contribute to the World Bank Group twin goals of poverty eradication and boosting shared prosperity" (World Bank 2015, 16). Improving CRVS also requires addressing systemic undercoverage of groups that are particularly difficult to reach through household surveys (for example, refugees, migrants, and people in bonded labor). The commitment to enhance support to CRVS has been integrated in the IDA 18 replenishment, as one of three data-related indicators in its results measurement system. In the portfolio reviewed, 18 projects provided support to population statistics based on census or civil registration, mostly through sectorwide statistical capacity–building initiatives. IEG identified only a few targeted efforts, such as a multisector demographic support

FIGURE 3.3 | Administrative Data Are Still Underused




*Source:* Structured surveys conducted for this evaluation.

project in Niger or a TFSCB grant of $250,000 in Peru, which helped design a new system for improving vital statistics production and record keeping. The commitment to improve CRVS requires a dramatic increase in the level of World Bank support.

## Conclusion

The World Bank has worked with many country clients to improve their data quantity and quality, increase technology use, make data and microdata freely available publicly, and improve data analysis and use. By building on its comparative advantages—trust of country counterparts, sought-after technical expertise, convening power, and funding ability—the World Bank can design and deliver ambitious capacity-building initiatives.

There is still a long way to go to build effective national statistical systems that can track progress across a broad spectrum of development objectives (PARIS21 2016b; Serajuddin and others 2015). To ensure that client countries escape a scenario of low data supply and use, and continue on a trajectory of improved data production, sharing, and use, the World Bank should consider taking the following steps:

- Strengthening domestic and international long-term funding for data and statistical capacity building
- Making data more central in policy dialogue, promote evidence-based decision making, and demonstrate the value of using data
- Moving toward a data capacity–building model that reaches beyond the NSO's boundaries to other parts of the national statistical system
- Scaling up support for administrative data systems in collaboration across global practices and with other development partners, and aligned with country priorities.

[1] Of 225 projects in the IEG portfolio, 146 are closed projects, and completion documents were available for 75 of those. IEG validated an Implementation Completion and Results Report Review (ICRR) in only 39 projects, and only a small portion of those were projects dedicated entirely to data and statistical capacity building. Furthermore, considering the high number of trust fund grants in the portfolio and the sparse reporting on their results, the assessment of results achieved by closed projects was limited.

[2] Of the 225 data-related projects identified, only 201 projects had enough documentation for IEG to review.

[3] The initial five priority areas are: household surveys, population statistics based on census and civil registration, national accounts, price statistics, and labor and job statistics. It is expected that more areas will be added.

4 IEG reviewed three main types of completion documents: Implementation Completion and Results Reports prepared by the implementing team at project closure, ICRRs prepared by IEG, and Implementation Completion Memorandums for trust fund grants. Specific results are presented in appendix B (Table B.9).

[5] The Statistical Capacity Indicator (SCI) is based on a diagnostic framework assessing methodology, data sources, and periodicity and timeliness. The SCI scores countries against 25 criteria in these areas using publicly available information and calculates the overall score as the simple average of the three area scores. Background work for this evaluation concludes that the SCI is a recognized, well-accepted tool for assessing statistical capacity, and it has both strengths and weaknesses. For example, it reflects statistical outputs more than statistical capacity from an institutional or governance perspective, and it takes no account of data quality. Because of the binary nature of many of its components, it can display large swings from year to year for a particular country. Therefore, this evaluation makes only limited use of the SCI.

[6] Although the evaluation covered the World Bank's support to national statistical offices (NSOs) in more depth, case studies also covered data support in line ministries and the issue of coordination within the national statistical system, though more superficially than support to NSOs.

## References


Bhatia, Vinisha, Susan Stout, Brian Baldwin, and Dustin Homer. 2016. *Results Data Initiative: Findings from Tanzania*. Washington, DC: Development Gateway.

IEG (Independent Evaluation Group). Forthcoming. *World Bank Group Country Engagement: An Early Stage Assessment of SCD and CPF Process and Implementation.* Washington, DC: World Bank.

Kiregyera, Ben. 2015. *The Emerging Data Revolution in Africa: Strengthening the Statistics, Policy, and Decision-Making Chain*. Stellencosch, South Africa: SUN MeDIA.

Ngo, Brian T., and Andrew J. Flatt. 2014. "World Bank Statistics for Results Facility—Catalytic Fund (SRF-CF): Evaluation Report of the Pilot Phase." Report 88366, World Bank, Washington, DC.

PARIS21 (Partnership in Statistics for Development in the 21st Century). 2015. "How to Build Statistical Capacity." Knowledge Series: Capacity Building, PARIS21 Secretariat, Paris.

———. 2016a. *Partner Report on Support to Statistics: PRESS 2016*. Paris: PARIS21 Secretariat.

———. 2016b. *National Strategies for the Development of Statistics Progress Report 2016.* Paris: PARIS21 Secretariat.

———. 2016c. *PARIS21 National Strategies for the Development of Statistics: Progress Report 2016.* Paris: PARIS21 Secretariat.

Serajuddin, Umar, Hiroki Uematsu, Christina Wieser, Nobuo Yoshida, and Andrew Dabalen. 2015. "Data Deprivation: Another Deprivation to End." Policy Research Working Paper No. WPS 7252, World Bank, Washington, DC.

UN (United Nations). 2016. *Independent System-Wide Evaluation of Operational Activities for Development.* New York: UN.

UNFPA (United Nations Population Fund). 2016. *Evaluation of UNFPA Support to Population and Housing Census Data to Inform Decision-Making and Policy Formulation 2005–2014.* New York: Evaluation Office of UNFPA.

Willoughby, Christopher. 2008. *Overview of Evaluations of Large-Scale Statistical Capacity Building Initiatives*. Paris: PARIS21 Secretariat.

World Bank. 2015. *World Bank Group Strategic Actions Program for Addressing Development Data Gaps: 2016–2025.* Washington, DC: World Bank.

———. 2017. *World Bank Group Country Engagement: An Early-Stage Assessment of the Systematic Country Diagnostic and Country Partnership Framework Process and Implementation.* Independent Evaluation Group. Washington, DC: World Bank.

# 4

# Toward a User-Centered Data Culture

## highlights

1. The World Bank has supported open data portals and practices, championed open government policies, and influenced several countries to share data and microdata publicly. However, the World Bank has not used its leverage fully in client countries that are reluctant to openly share development data.

2. The World Bank has far paid far less attention to promoting government and citizen data use so far, and therefore success is limited.

3. The World Bank has an opportunity to draw on insights offered by new tools, such as behavioral science and big data analytics, to understand the decision makers' motivations and encourage them to use data. The rise in demand for data to monitor the Sustainable Development Goals, increased interest in performance-based budgeting among some governments, and the surge of citizen surveys of service quality also provide opportunities to promote a more inclusive, user-centered data culture.

**SUPPORT FOR NATIONAL STATISTICAL SYSTEMS** has enhanced data production more than it promoted in-country data sharing and use. As recent reports show, this applies to data development partners in general, not just the World Bank (UN 2014; PARIS21 2015). Focusing on the World Bank's contribution specifically, only 68 of the 201 projects reviewed for this evaluation included support for increasing public access to development data. IEG's structured survey of World Bank staff and country stakeholders and the interviews with development partners found that these groups perceive that support to in-country data production has been more effective than support to data sharing and use (figure 4.1).

Three types of data sharing seem to be important. First, there is data financed by the World Bank which must be shared with the World Bank. Second, there is a set of essential data financed from domestic or other sources which countries would do well to share with the World Bank allowing it to report on aggregate statistics. Third, there is country level data which if more openly shared with the country's public could improve transparency, accountability, and evidence-based policy making. The World Bank made a significant contribution to data sharing in some countries by promoting an open data agenda using a combination of legal reforms and technical updates to make official data and

FIGURE 4.1 | Perceptions of the Effectiveness of World Bank Data Support


*Source:* IEG structured survey 2016.

*Note:* In estimating the percentages, IEG excluded "Do not know" and "No opinion" responses from the denominator.

microdata more accessible. It provided many countries with technical assistance to develop access and dissemination policies that are in line with the UN Fundamental Principles of Official Statistics and the African Charter on Statistics. The World Bank also helped upgrade the websites of national statistical offices (NSOs) as well as open data portals that increase user access to data. However, the World Bank has not used its leverage fully with governments that have been reluctant to share data.

The World Bank could do much better on data use. Statistical capacity building objectives often included serving data users' needs, but IEG's review found that the World Bank could do more to promote enhanced data use strategically, for example, by understanding the different kinds of data users and their needs and motivations, and including both government and nongovernment data users in the design of its projects. Only 27 of the 201 projects reviewed for this evaluation supported activities to build capacity for data use. Weak data literacy, limited internet and smartphone connectivity, and in some cases resistance by interest groups impeded progress on data use. Staff also reported in interviews that when data use is an explicit project objective, it is difficult to prove its achievement. However, data are valuable only if they are used. Finding evidence of data use requires carefully tracing all its influences on decision making or resource allocation, and determining the extent to which the particular World Bank project contributed to them. This is challenging, though not impossible, and it must be undertaken given that the outcome of interest is data use for sound decision making and resource allocation.

The World Bank has a well-established approach to building data producers' capacity, but it has not yet formulated a conceptual model to consider ways of assessing user capacity. Increasing data production, data production capacity, and data quality will not be sustainable without consistent demand. The Data Council has yet to address the need to fundamentally rethink how to develop a more inclusive, user-centered data culture. The Open Data Readiness Assessment methodology, a rapid diagnostic tool to assess the demand for open data and the capabilities of diverse user groups, should be explored as a possible starting point. This evaluation looked for a theory of change or other framework for understanding how data and other support might foster data use, but it did not find either. The literature review points to a gap in theory and empirical studies into the causal relationship between data and decisions.[1]

The World Bank's approach to fostering data demand and use has, in practice, revolved around data-driven research and analysis, global data portals, benchmarking exercises, encouraging data sharing, and open data and government initiatives. However, these are not sufficient to foster data use, especially beyond academia. To effectively use data, practitioners emphasize the importance of starting with the question to be answered instead of the data itself (World Bank and SecondMuse 2014), and of grasping the political economy of data use and nonuse.

There are several reasons why decision makers may not demand or use data: the available data may not be relevant to their goals; data relevant to their goals are not available; they do not know how to analyze and use the data (low data literacy); they do not trust the data (poor data quality); or they find the available data politically inconvenient. On the supply side, data visualization and new technological platforms can help to increase the accessibility and usability of data. On the demand

side, demonstration-by-example, training, and investments by the Bank in improving data literacy – showing governments the value of specific data types to address specific goals and building their capacity for data analytics to be able to draw actionable insights from data – can help increase uptake. However, staff interviews indicated that the World Bank's country directors often do not use the World Bank's fullest leverage to foster countries' interest in data, and could do more to promote greater data demand and use if they were themselves committed.

One factor, rejection of politically inconvenient data, can be widespread and is the hardest to address. World Bank staff needs to understand the reasons for decision makers' lack of interest in data and develop approaches to change their behavior by changing their incentives. One approach is to motivate action by broader groups of stakeholders (private sector, other parts of government, legislators, and civil society) through data and analysis on particular issues. Better data sharing and accessibility can lead to public scrutiny and debate of government policy and stimulate data demand and use.

Another approach is to publicize data that compare the performance of government programs and agencies across jurisdictions, which can often gain attention and use. Comparison with other countries (or subnational units or agencies) seems to produce a spirit of competition in government leaders or possibly embarrassment or envy. Authorities in several Latin American countries were startled to see how poorly their students scored in international comparative tests of educational learning outcomes, and this spurred them into more vigorous reform efforts. The World Bank could explore the use of benchmarking exercises or comparative indicators to nudge client countries toward evidence-based policymaking.

Development data sometimes reveals politically inconvenient truths that decision makers act upon only after broader political change occurs. Shifts in the distribution of power could empower new decision makers with new goals or priorities, leading to a greater appetite for data use. *World Development Report 2017: Governance and the Law* stresses that even though history partly determines the distribution of power in society, it can still change when elites reach agreements to restrict their own power, when citizens engage (through voting, political organization, and public deliberation), and when donors support rules that strengthen reform coalitions (World Bank 2017).

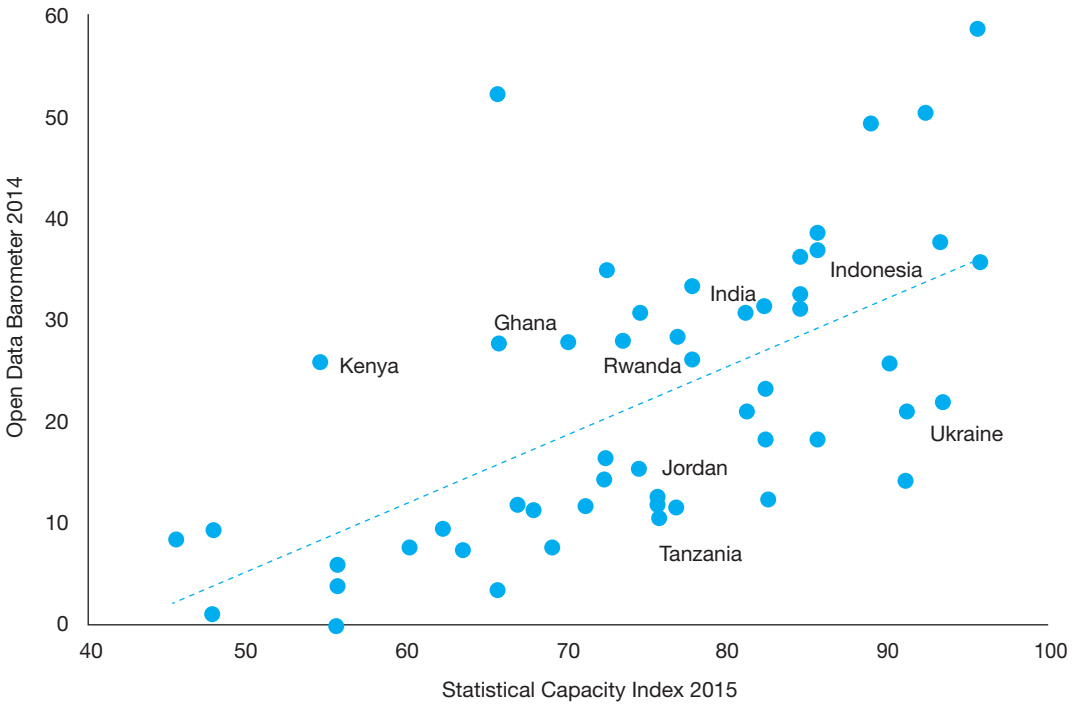
Evaluations have noted weaknesses in the promotion of data use for a long time, and not just in the World Bank's programs. The Partnership in Statistics for Development in the 21st Century (PARIS21) was established partly to bring together policy makers, data users, and statisticians. An inventory of evaluations of different statistical capacity programs concluded that these programs had little impact on the use of statistics in countries (Willoughby 2008). An evaluation conducted by the European Commission (2007) covering 30 projects from 1996 to 2005 concluded that few projects tackled the contribution of statistics to evidence-based decision making. Another study in 2009 concluded, "While support to the production of statistics has increased, the link between production and use in-country is still far too weak" (OPM 2009, 5). Part of the problem is that policy makers, data users, and statistical systems managers each see the world differently and mechanisms to connect them are lacking.

## Open Government Policies Support Data Sharing

The World Bank tended to be effective in promoting data sharing in countries where it also successfully helped strengthen NSOs. Countries with sound statistical capacity are more likely to endorse open data policies and the release of microdata and metadata (figure 4.2). NSOs in a number of countries increasingly provide statistical calendars with expected release dates, and they strive to respect the deadlines.

The World Bank helped countries build national capacity in microdata preservation, analysis, and dissemination through its support to PARIS21 and direct technical assistance to its country clients. This involved establishing national data archives and implementing the Accelerated Data Program (ADP). The ADP has provided training to more than 2,000 staff from 150 national organizations in about 70 countries on microdata anonymization, documentation, archiving, and sharing. ADP increasingly includes outreach to microdata users and training for them. All case study countries for this evaluation—except Ukraine—made progress on data sharing, helped by regular diagnostic reports prepared by ADP.

FIGURE 4.2 | Positive Relationship Between Statistical Capacity and Data Openness




*Source:* Based on the Statistical Capacity Indicator and Open Data Barometer.

IEG's country case studies and interviews found mixed progress on data sharing and open data policies and that the World Bank has been most effective in countries where governments were already committed to sharing data. The World Bank has occasionally raised data issues at high levels of policy dialogue, and has sought to influence countries to share data and microdata publicly. However, where that did not work, World Bank management has been unwilling to make open data access a prerequisite for financing of statistical capacity building—or anything else. Realistically, it is hard to hold the World Bank's program hostage to the performance of this one item on the broad menu of development interventions, but the World Bank does need to ensure that it uses its fullest leverage to help foster data sharing. And the World Bank must ensure that it has access to data produced with its financial support.

The World Bank engaged closely in Indonesia on opening up government records. Indonesia is one of the eight founding members of the Open Government Partnership in 2011. The NSO's publications are freely available online, the office posts its annual budget online for public view, and each agency that produces statistics gives notice of future release dates. World Bank teams have used training and technical assistance to help several ministries use and interpret data. However, open data competes with other priorities, the country's Freedom of Information Law is only partially implemented, and some ministries continue to release data in formats that are not machine-readable. Jakarta province government leads the way on open data, while the Ministry of Finance launched a fiscal transparency portal in 2016 to share budget data; the practice of making data publicly available is uneven across agencies. Many local governments are unable to produce data on a regular schedule.

After a slow start, Kenya (another case study country) is now one of the more advanced countries in Africa in open access to official data. Statistical techniques were improved substantially with support from the World Bank. Improvements included updating the base years for most data sets, better data validation, and bulletins informing users about revisions. However, the project monitoring and evaluation neglected data use even though this was an explicit feature of the project results framework. World Bank supervision missions emphasized the delivery of outputs more than progress toward outcomes, data accessibility, and outreach to users. Opportunity still exists for making data easier to select and download, clarifying the terms of use, and providing more complete metadata.

Rwanda has made much progress toward improved data access and dissemination with support from the World Bank and PARIS21 through the ADP. The Statistics for Results project, launched in 2012, emphasized dissemination and services for users and supported the NSO to update its website, provide more complete metadata, digitize statistical information, and develop an electronic national data archive to allow users to access microdata. The ongoing Public Sector Governance Program for Results is supporting the government to open some administrative data collected by one of the main data-producing line ministries. Progress has been slow to date, as this effort requires implementing quality control protocols and sensitizing various layers of government to the value of open data.

The World Bank's efforts to promote open data in Ukraine through seminars and outreach events had little government support initially. Online availability, machine readability, ease of download, ability to

filter, clarity of definitions, or quality of metadata received no priority. IEG's case study found that an April 2015 law on public access to information and open data is yet to be widely implemented or even understood. The government still perceives anonymized data as confidential, does not conduct user satisfaction surveys, and little discussion takes place on user needs or obstacles to data access. Those interviewed reported so-called "confidential data" are only "available on the black market for a fee." Although the World Bank made a major contribution to Ukraine's data production, it has not been effective in addressing key constraints on data access and use even though it was an explicit part of its objectives.

## Websites for Data Sharing and Use

The World Bank has been particularly effective in helping NSOs develop websites and data portals, as in Ghana and Rwanda. In Tanzania, data users took part in the design of the website, and demand for statistics is now stronger from ministries and development partners. However, releasing information still suffers delays, and traditional publications take priority over digital data.

The World Bank was equally effective in Peru where it worked with the NSO to develop a website that offers free microdata and metadata downloads from 35 sources, including censuses and surveys. Under the auspices of the new Ministry of Social Development and Inclusion, the World Bank established a digital data repository and supporting web platform to collate administrative data on education, health, finance, citizen registration, housing, and sanitation. Users can freely download data, cross-tabulate variables, and generate basic reports. IEG's country case study found that uptake has been faster by the private sector than by universities.

The data format is important on official websites. In India, researchers told IEG that officials publish survey reports in PDF format, which makes data tables inseparable from lengthy descriptive material. Data cannot be downloaded for analysis and reuse. Research institutions and government do not discuss improving data exchange and usability.

## Making Data Use Inclusive and Empowering

The World Bank made substantial progress in promoting open data policies and web-based data access in some countries, and it should now redouble its efforts to increase the number of people using data to help shape the development agenda. Enabling officials in central government to understand, analyze, and communicate data is part of the solution—data literacy at this level is indispensable. However, the bigger challenge lies beyond central government: equipping local administrations, universities, the media, and civil society to be more discerning data users so they can hold government accountable and improve service quality. Survey respondents rated the World Bank low on fostering in-country demand for data. Only 33 percent of World Bank staff and 45 percent of country stakeholders rated the World Bank as highly effective or effective on this dimension. Despite this low rating of effectiveness, neither of the two groups surveyed included

"generating country-level demand for data" among their top choice of areas needing strengthening going forward (appendix C). However, many university teachers and researchers among the survey respondents urged the World Bank to make outreach more effective.

Citizens are more likely to be empowered when governments establish public forums for participation. In Peru, World Bank technical assistance to the new Ministry of Development and Social Inclusion is helping develop channels for user feedback. It will improve knowledge management, information, and communication through the implementation of an integrated information platform that includes data from different programs, thus helping to embed a culture of data use and results orientation. Elsewhere, the World Bank had setbacks in its efforts to promote public forums to make data production and use more inclusive. In Ghana, after a brief period of existence between 2004 and 2006, the National Advisory Committee of Producers and Users of Statistics was discontinued for lack of funding. Its reinstitution was inserted into a new Statistics Bill which was supposed to be passed by the end of 2012, but remained pending after the December 2016 session.

The World Bank encouraged the use of surveys to measure user satisfaction with statistics. Surveys are another aspect of an inclusive, user-centered data culture and are now standard part of World Bank technical assistance. User surveys in Rwanda and Tajikistan, for example, point to increased satisfaction with official statistics.

## Improving Subnational Data

Nurturing a data culture at the local level needs more attention. People interviewed in many countries told IEG that they want to know how their region, city, or community is doing relative to others in the same country. Decision makers in central and local governments need this information to set priorities and compel action. In Rwanda, for example, growing demand for district-level data means that samples need to be representative below the national level, and regular surveys are essential to inform local planning and service delivery. Surveys that are representative at the district level are valuable, but only routine administrative systems can provide the type and frequency of data needed to meet most local needs. This data collection is done by teachers and nurses, but they have little incentive to do so because the data are channeled to central authorities without any attempt to use them to inform local decisions and without providing access for local administrative staff.

Indonesia's NSO cannot keep up with local governments' growing demand for technical assistance. In Tanzania, poorly trained local government staffs give low priority to data production. The poor quality of the data produced by provincial administrations is a source of frustration and an impediment to their use. Local municipalities in Peru do little to track service delivery, making monitoring and evaluation difficult. The lack of coordination between the different levels of government means that textbooks and vaccines do not always reach the children who need them. Interviewees in India told IEG that state and district officials need training in data analysis and presentation. Preparation of the 2010 Statistical Strengthening Project involved close engagement with state authorities in 16 of India's states and included discussion of data accessibility. Increasing

user awareness at the state level is a project objective, but there is no corresponding indicator in the results framework, and IEG found no evidence of user-producer dialogue at the state level.

## Performance Management Frameworks

The more citizens hold their governments accountable, the greater the demand and use for data will be for measuring government performance against indicators and targets. One way to make government agencies more accountable and more efficient is to widely publicize data about their achievements and shortfalls, and then adjust funds delivered in the next budget cycle to reward strong performers. Peru adopted performance-based budgeting in 2006 with World Bank support. The number of programs covered has steadily increased since then, and much of the budget now ties to performance indicators. Several line ministries worked with the World Bank to develop indicators. So far, the World Bank has been more effective at proposing metrics than suggesting ways to integrate the various ministries' rapidly growing data sets properly.

Since 2006, all public institutions in Rwanda have been required to sign performance management contracts with the president of the republic. Independent evaluators annually assess progress toward agreed targets. This is a data-intensive exercise that collects information from the localities and the lowest levels of government. As one IEG interviewee noted, hard evidence of results is required at annual meetings, and anecdotal reports are no longer enough. Each district has its own scorecard and is expected to review the targets' relevance, the effort needed to reach them, and the quality of information needed to report on achievements. This system of cascading performance contracts (called *Imihigo*) has increased the demand for data. A high-level statistician observed, "When they start using data, people become addicted, they want more and more" (IPAR 2015). Going a step further, the World Bank and DFID have recently adopted performance-based financing instruments that trigger disbursements with evidence of data use.

Overambitious performance targets can encourage data falsification. Kenya abolished school fees and gave local authorities resources that put more children through primary school. The administrative data promptly showed a rapid increase in enrollment that data from the Demographic and Health Survey did not support (Sandefur and Glassman 2015). Performance contracts must be validated independently, and they need to be embedded in a results-based culture that has data users who are sufficiently literate and committed to holding government accountable.

## Nurturing a User-Centered Data Culture

The World Bank has often been effective at supporting data sharing in the countries in which it engaged NSOs. Much depends on countries' willingness to share data, and several countries still refuse to share data, for political reasons or because of quality concerns or from a reluctance to lose a revenue source. The World Bank has not used its leverage fully to influence additional countries to share data.

The next step is to foster reciprocity between the multiple agencies that produce and share data and the equally diverse data users, creating a user-centered data culture. This goal is broad and diffuse. Creating a user-centered data culture in the poorest countries faces several obstacles. Internet use is limited, universities are weak, and such countries lack a vibrant research community that demands data for its studies. Fostering a user-centered data culture would require the World Bank to locate and recognize the receptiveness among different groups and institutions within individual countries (media, universities, and subnational governments). The World Bank can help nurture an ecosystem of data use by working with local governments, civil society institutions, the media, and academia, using approaches tailored to the needs of different users.

[1] Exceptions include, for example, Massett and others (2013).

## References


IPAR (Institute of Policy Analysis and Research). 2015. *Imihigo Evaluation FY 2014/2015.* Brighton, U.K.: Institute of Development Studies.

Massett, Edoardo, Marie Gaarder, Penelope Beynon, and Christelle Chapoy. 2013. "What is the Impact of a Policy Brief? Results of an Experiment in Research Dissemination." *Journal of Development Effectiveness* 5 (1): 50–63.

OPM (2009) Evaluation of the Implementation of the Paris Declaration: Thematic Study—Support to Statistical Capacity Building, Synthesis Report." London: DFID.

PARIS21 (Partnership in Statistics for Development in the 21st Century). 2015. *A Road Map for a Country-led Data Revolution*. Paris: OECD.

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

UN (United Nations). 2014. *A World that Counts: Mobilising the Data Revolution for Sustainable Development*. New York: UN.

Willoughby, Christopher. 2008. *Overview of Evaluations of Large-Scale Statistical Capacity Building Initiatives*, Paris: OECD.

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

# 5

# Implications of Big Data for the World Bank

## highlights

1. A lack of an understanding of when and how big data can complement traditional data in answering key development questions and a lack of corporate agreements to ensure World Bank access to such data have hindered efforts to use big data.

2. The World Bank has not systematically analyzed big data's potential contributions and pitfalls for addressing its mission.

3. The World Bank's ad hoc approach to big data is unlikely to work well for scaling-up and institutionalizing big data, though initially it helped in facilitating small-scale exploration and experimentation.

**ERIC SCHMIDT,** the former CEO of Google, said in 2010, "There was five exabytes of information created between the dawn of civilization through 2003, but that much information is now created every two days, and the pace is increasing" (Einav and Levin 2013, 1). Much of this information growth stems from the rise of big data (sometimes called new data or non-traditional data). Big data refers typically to extremely large data sets created through satellite or geospatial imagery, remote sensing, Global Positioning System (GPS) tracking, computer search engines, social media, crowdsourcing, online payments, call detail records, smartphones, and the Internet of Things. Volume, velocity, and variety characterize such data (Hilbert 2013). Extracting patterns, trends, and associations from these data sets through computational analytics can provide a wide range of real-time information about people, often much faster and at lower cost than was previously possible (Oroz 2016).

## World Bank Support for Geospatial and Other Forms of Big Data

In the mid-1990s, the World Bank realized the potential of geospatial data and tried to build its own capacity to analyze such data. However, these initiatives did not flourish because staff was skeptical and the potential of such data was unproven. A 2014 campaign to champion big data innovations at the World Bank uncovered several issues regarding lack of access to certain types of big data, data science expertise, storage and computational capacity, guidance on handling privacy, opportunities for peer-to-peer learning, and platforms and norms for sharing data and software (World Bank 2016). These gaps have prevented the World Bank from combining big data with traditional data and could represent a missed opportunity.

As of November 2016, the World Bank estimates the number of World Bank–supported projects involving big data at more than 60. Of these, 14 projects had won an innovation challenge in 2014 that had attracted 131 entries. The ongoing big data projects are in sectors as diverse as agriculture, transport, urban development, energy, environment, employment, economic productivity, financial inclusion, governance, property rights, and natural disasters. Only two of these projects are lending operations; the rest are advisory and analytic services. Most are mapped to the Development Economics Vice Presidency and the Office of the Sustainable Development Chief Economist, followed by the Energy and Extractives Global Practice; the Social, Urban, Rural, and Resilience Global Practice; and the Transport and Information and Communications Technology Global Practice. Most projects are cross-regional.

Specific innovations using big data in the World Bank include the following:

- Cities in the Philippines are minimizing traffic congestion by observing the flows of vehicles with cellphone GPS data

- Drones are being used in Albania and Kosovo to help map land boundaries and secure property rights
- Satellite imagery is being used to determine the maize yield in Ugandan farms.
- Rural poverty in Sri Lanka is being estimated using satellite imagery of building density and roof material
- Cities in Latin America are using satellite images to identify slums, roads, and commercial areas
- Citizens in the Philippines are using crowd-sourced photos, maps, and satellite imagery to monitor road infrastructure projects.

Several of the ongoing operations are in the piloting or incubation phase, and it is too early to assess their effectiveness. However, the sheer numbers now of big data projects (more than 60) show greater World Bank willingness to explore big data's potential in helping to solve development problems. Big data is not a complete substitute for traditional data. Because the World Bank's big data projects are getting started and the new data sources are still unproven, it is essential that big data be complemented by or validated with traditional data. For example, satellite images of building density and roof material—proxies for poverty—need to be validated with household surveys and census data on poverty. Furthermore, turning satellite images into accurate crop yield estimates requires training the computer model with actual crop-cutting data from farm visits. Box 5.1 describes ways of combining big data and traditional data to answer development questions.

In IEG's structured survey, about 50 percent of country stakeholders and World Bank staff agreed that the World Bank has been highly effective or effective regarding global data innovations, including big data (appendix C). Asked to choose among areas of strategic thrust for the World Bank going forward, respondents gave a relatively low priority to global data innovation. Only 32 percent of World Bank staff and 37 percent of country stakeholders included this among their five preferred areas for the World Bank's strategic thrust, in contrast to the world's successful global companies such as Amazon, Google, Netflix, and LinkedIn that are using big data to deliver extraordinary results, for example, by using such data to understand client preferences and to find the best way to respond to those preferences (Marr 2016).

Big data initiatives at the World Bank so far have been ad hoc, driven by individual initiative instead of a coordinated institutional approach. According to World Bank staff interviewed by IEG, this has resulted in inadequate quality control and lack of knowledge of who to approach for advice on using big data.

The World Bank currently spreads responsibility for geospatial and big data work across three separate units, and the division of labor is unclear. The three units are the front office of the Senior Vice President of Operations, the Sustainable Development Practice Group Vice Presidency (which also houses the Geospatial Operations Support Team), and the Analytics and Geospatial Working Group under the Data Council. And there are two communities of practice and two working groups related to big data (for a total of four). The rationale for this arrangement needs more thought. Overlapping responsibilities and a lack of strong coordination can result in inefficiencies.

### Box 5.1 | Combining Big Data and Traditional Data: Two Examples

**Easing Urban Congestion with Smartphones in the Philippines.** The traditional method of collecting traffic congestion data in the Philippines uses travel time surveys. Two local contractors with a clipboard and stopwatch drive in a car and manually measure the time it takes to drive between intersections, repeating these measurements for a month. The average of the results determines typical traffic speeds. This is a slow and costly process.

The World Bank developed Open Traffic, a platform that provides an alternative to this process. A partnership with the taxi-hailing app Grab gives Open Traffic access to real-time, anonymized GPS data from hundreds of thousands of taxis. These big data have the same use as the travel time surveys, but there is much more data gathered in real-time at almost no cost. Open Traffic does not completely replace travel time surveys, however. The World Bank has been using the travel time surveys to validate the new approach. Furthermore, transport planning still requires traditional surveys to obtain data for more granular analyses of different vehicle types, such as studying the flows of motorcycles.

**Securing Property Rights with Drones in the Balkans.** Cadastral maps in the Balkans are usually produced at the national level through a costly and time-consuming process. An orthophoto (aerial photo corrected for distortion in the same way as a map) is a key part of a cadastral map traditionally produced for the whole country from satellite imagery or from manned airplanes. The World Bank has been using drones, or unmanned aerial vehicles, since 2013 in Albania and Kosovo to produce high-resolution orthophotos of specific towns and villages. Instead of waiting years for a new national orthophoto to update a particular town's cadastre, drones can produce a new orthophoto in just days. Drone imagery combined with other new technology, such as open source software to record property rights and platforms to manage cadastral data, offers a cost-effective methodology to secure property rights. National orthophotos are still the best way to achieve large-scale cadastral mapping, but drones provide a fit-for-purpose mapping approach when a specific town needs updated aerial imagery and cadastral maps.

The World Bank's data science staff is spread across the organization, and in units dealing with disaster management, information technology, environment, and other issues, and it can be hard for others to find them or know what they are doing. The World Bank's human resources data show that as of November 2016, 18 staff members have the title of data scientist in various grades, half of those were direct hires into the data scientist title, and the other half had their titles converted to data

scientist. IEG interviewed World Bank staff working on geospatial data who pointed out that along with training staff in big data and ensuring that geographers, statisticians, economists, and World Bank staff from other disciplines work together, the World Bank also needs to undertake strategic hiring of data scientists, GIS experts, modeling professionals, and other experts in big data analytics. Such experts should also help other staff grasp the intricacies of big data.

Furthermore, the World Bank has sent mixed signals about big data's priority. The *Strategic Actions Program for Addressing Development Data Gaps* (World Bank 2015) does not contain any proposals or actions specifically on big data (World Bank 2015). However, a recent reform of geospatial information at the World Bank aims to make it a sophisticated consumer of geospatial and big data analytics. The note outlines the Geospatial Operational Support Team's mandate as "The creation, brokering, or scaling of institutional public goods with significant utility in development lending, specifically around three core areas: (i) efficient spatial data management; (ii) knowledge capture and dissemination; and (iii) procurement support." The agenda this mandate implies deserves wide circulation and discussion within the World Bank to ensure buy-in.

Interviews and case studies for this evaluation suggest that the World Bank's ad hoc approach to big data might have helped initially by facilitating small-scale exploration and experimentation. However, it is unlikely to work well for scaling-up and institutionalizing the World Bank's big data work if a decision is taken to do so. Scaling-up big data use at the World Bank will require clearly defining responsibility among units, avoiding overlapping mandates, and ensuring the necessary data science expertise on relevant forms of big data (from geospatial to social media) in answering key development questions. The World Bank recognizes the importance of satellite imagery and other forms of big data in situations of fragility, conflict, and violence, especially given the lack of security in the field and low government capacity or interest in conducting household or other surveys in remote and marginalized areas, but the use of big data has not yet been institutionalized in those situations.

## Future Big Data Use by the World Bank

A major challenge so far has been the lack of a widely-shared understanding and appreciation among World Bank staff of when and how big data can complement traditional data in answering key development questions. Staff interviewed for the Strategic Needs Assessment for the World Bank Big Data Analytics Program saw an important role for the World Bank in this area and said that the World Bank should be "An innovative leader in the use of big data to improve the well-being of the poor" (Vital Wave 2015).

Although big data analytics can be outsourced to specialized firms, the World Bank still needs in-house data science expertise to examine proposals and quality control deliverables. Many staff interviewed by IEG were opposed to wholesale outsourcing to external firms (or even to data scientists in other parts of the World Bank) because data science expertise needs to be complemented by subject matter and country expertise, and the latter resides in specific World Bank operational units. A review undertaken by the World Bank found that the majority of geospatial

analytics tasks have been outsourced on an ad hoc basis, with little to no coordination between project expenditures, creating significant leakages of data, loss of expertise, higher average costs, and little institutional memory.

The World Bank does not have a central repository or systematic cataloging for big data sets obtained by different parts of the organization. Consequently, staff in different areas can both put effort into obtaining the same data, which is inefficient.

Other organizations have recently built up their big data analytics capacity. For example, the U.S. Office of Science and Technology Policy has a chief data scientist, the Organisation for Economic Co-operation and Development (OECD) has a Futures Unit, and the United Kingdom has the Foresight initiative. Policy Horizons Canada is a foresight and knowledge organization in the Canadian government, and the UN's Global Pulse Labs work on new approaches to using big data for development (box 5.2). These organizations' experiences might hold lessons for the World Bank.

Big data can be extremely beneficial in approaching problems from new angles, but without proper management and analysis, it can also cause big errors. In developing its capacity for big data analytics, the World Bank will need to ensure that it prepares adequately for big data's analytical, ethical, governance, privacy, and exclusion challenges and pitfalls (box 5.3).

Important unresolved questions surround access to big data. The World Bank tried unsuccessfully to acquire call data records during the 2014–16 Ebola outbreak in West Africa. What kinds of corporate agreements or data-sharing partnerships should the World Bank establish with big data producers (such as the U.S. National Aeronautics and Space Administration, the U.S. National Oceanic and Atmospheric Administration, Uber, Verizon Communications, Facebook, and Twitter)? How will the World Bank safeguard privacy concerns, and what protocols will it follow when sharing big data with governments? How will the World Bank ensure the ethical use of big data by itself and country clients?

### Box 5.2 | The UN Global Pulse Labs

UN Global Pulse Labs are pioneering new ways to use big data to pursue development goals, aiming to show how new sources of digital data and emerging technologies can help understand what is happening to vulnerable populations. The headquarters lab is in New York, and other labs are in Jakarta and Kampala. Pulse Labs design projects with UN agencies and public sector institutions that provide sectoral expertise, and with private sector or academic partners that often provide access to data or analytical and engineering tools. Research projects include food security, humanitarian logistics, economic well-being, gender discrimination, and health. Host countries must be willing to share lessons, experiences, and findings with labs in other countries.

*Source:* Oroz 2016.

### Box 5.3 | Big Data Key Challenges and Pitfalls

**Analytical Challenges.** Big data can be biased. For example, social media users are a subset of the population (generally young people living in cities), and information drawn from them is not representative of the population at large. Big data is often incomplete. A researcher might analyze how often topics appear in tweets, but if Twitter uses its editorial rights and removes all tweets that contain content it deems inappropriate, the analysis will be skewed. Big data is often not available in a standardized format, which makes it more difficult to process than traditional data. Big data can be misinterpreted. Mobile phone data might suggest that workers spend more time with their colleagues than their spouses, but this does not necessarily mean that colleagues are more important than spouses. O'Neil (2016) showed how algorithms can produce disastrously wrong results if they use imperfect proxies for what cannot be directly measured, and become what she calls "weapons of math destruction."

**Ethical Challenges.** Combining big data sets might offer new insights, but it can also violate privacy because some of the information is tagged with the user's identity. It is not easy to ensure that users give informed consent. Private companies may thwart government efforts to serve the public good in this regard, and governments could use big data to suppress opponents or discriminate against some groups.

**Inclusiveness Challenges.** Access to big data is often only available for a fee, and this might be beyond many organizations' financial capacity. Big data also requires technical and analytical processing capacity that poorer countries lack, and their access and technical support is likely to be limited. This opens a new digital divide.

*Source:* Boyd and Crawford 2012; Hilbert 2013; Shirky 2016.

The World Bank needs to decide on the extent and nature of its support to country clients in building their capacity for big data. Typical statistical capacity–building projects have focused on building clients' capacity to produce traditional forms of data. Although these initiatives are still highly relevant, the World Bank should consider when, where, and how it should also support the development of clients' big data capacity. Big data are often faster, less costly, more reliable, more frequent, and more disaggregated than traditional data, and could represent the future for many types of use, such as geographic targeting.

The World Bank now needs to examine its own experience and that of other relevant organizations with the usefulness of big data in complementing traditional data. Based on what it learns, the World Bank should implement coordinated actions to ensure that sufficient, advanced big data analytics underpin its own decisions, and that it provides effective support to country clients for big data use.

Those actions are likely to include the following:

- Ensuring outreach to World Bank staff and country clients to support their understanding and use of big data
- Ensuring that geographers, statisticians, economists, and World Bank staff from other disciplines work together, and recruiting adequate data science experts to strengthen both the World Bank's own work and to improve its support to countries to raise their awareness of and appetite for big data use
- Considering when and where it makes sense to grow NSOs' big data capacity
- Fostering data-sharing partnerships between the World Bank and public and private big data producers
- Ensuring systematic cataloging of big data obtained by different parts of the World Bank and creating a centralized repository for it
- Implementing ethical, governance, and privacy safeguards for big data use by the World Bank and its country clients.

## References


Boyd, Danah, and Kate Crawford. 2012. "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication, and Society* 15 (5): 662–79.

Einav, Liran, and Jonathan D. Levin. 2013. "The Data Revolution and Economic Analysis." NBER Working Paper 19035, Cambridge, MA, National Bureau of Economic Research.

Gourlay, S., Kilic T., and Lobell, D. 2017. "Could the Debate be Over? Errors in Farmer-Reported Production and Their Implications for Inverse Scale-Productivity Relationship in Uganda." Paper Presented at the 2017 Centre for the Study of African Economies (CSAE) Conference, Oxford, UK.

Hilbert, Martin. 2013. "Big Data for Development: From Information to Knowledge Societies." University of California, Davis, January 15, 2013. https://ssrn.com/abstract=2205145.

Marr, Bernard. 2016. *Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results*. West Sussex, U.K.: Wiley.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

Oroz, Miguel Luengo. 2016. "10 Big Data Science Challenges Facing Humanitarian Organizations." *United Nations Global Pulse Blog* (blog), November 25. http://unglobalpulse.org/news/10-big-data-science-challenges-facing-humanitarian-organizations.

Shirky, Clay. 2016. "Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy, by Cathy O'Neil." *New York Times Book Review*, October 3. http://www.nytimes.com/2016/10/09/books/review/weapons-of-math-destruction-cathy-oneil-and-more.html?smprod=nytcore-ipad&smid=nytcore-ipad-share.

Vital Wave. 2015. *Strategic Needs Assessment for the World Bank Big Data Analytics Program*. Palo Alto, CA: Vital Wave.

World Bank. 2015. *World Bank Group Strategic Actions Program for Addressing Development Data Gaps: 2016–2025*. Washington, DC: World Bank.

——. 2016a. *Big Data Innovation Challenge: Pioneering Approaches to Data-Driven Development.* Washington, DC: World Bank.

# 6

# Conclusions and Recommendations

**THE WORLD BANK** has a strong reputation on development data. It effectively supported many individual countries' data needs and supported data as a global public good. Major gaps in data quantity, quality, and availability remain, and no country is anywhere close to collecting all 230 Sustainable Development Goals (SDGs) monitoring indicators. Although the World Bank was effective in supporting data production in many countries and encouraging some countries to share data, support for data use by governments and citizens lagged. The World Bank has been a leader on development data for global audiences, and it now needs to assess and adjust its approach where needed and meet recent commitments for a stronger effort on data. *The Forward Look* (World Bank 2016) expresses these commitments, and envisions an expanded role for the World Bank in addressing global public goods as part of IDA18 and in the corporate goal that commits the World Bank Group to ensuring a household survey in IDA and blend countries at least every three years.[1]

Management of the World Bank Group's institutions has recently signaled its intent to step up support for data production and clarified what types of data will be given priority. The creation of the World Bank Group Data Council and Development Data Council (until recently Development Data Directors group) and its associated working groups established an internal framework for governance and coordination. The Data Council formulated goals and priorities for the World Bank's work in data and put forward specific, ambitious costed proposals for expansions in CRVS, price, survey, and geospatial data collection. The World Bank also created a theme code for data that will help track and manage the World Bank's portfolio on data going forward. Although these plans appear to align broadly with country needs and World Bank technical strengths, they should not displace an emphasis on strengthening long-term statistical capacity.

Producing data is a core government function, and governments cannot achieve good governance with bad data. However, countries do not always understand data's value well, domestic funding for statistics is still low in many places, and several governments are reluctant to borrow from the International Bank for Reconstruction and Development (IBRD) or International Development Association (IDA) for data. Countries' ownership and financing of data, while not necessarily within the World Bank's control, are crucial for measuring progress on the twin goals. The World Bank's current approach to advocating for data (often by demonstrating how specific pieces of data analysis can generate solutions to policy problems) has merit, but is insufficient. The Data Council has proposed principles for funding data production and vital registration systems in IDA-eligible countries based on blending donor funding with an increasing share of domestic funding over time. However, the suggested funding principles (gradually increasing domestic financing) are not binding or enforceable. The World Bank's systematic country diagnostics (SCDs) identify data gaps unevenly, and its Country Partnership Frameworks (CPFs) and country policy dialogue do not consistently include data issues.

No mechanism exists today for medium- to long-term financing of data, even though funding needs for data are significant.[2] Trust funds for statistical capacity building were central to past successes, positioning the World Bank as a premier global funder and coordinator of data and allowing it to engage also in countries that were reluctant to borrow for data. However, relying on trust funds creates uncertainty and dependency on a small number of donors, and it hinders long-term planning, which affects even some of the most high-profile initiatives, such as the Living Standards Measurement Study and PovCalNet. Furthermore, the Statistical Capacity Building Program (STATCAP) is now completed. The envisioned expansion in data production, the ability to track the twin goals and key SDGs, and the sustainability of past gains in statistical capacity in some (mostly lower-income) countries are at stake.

The World Bank, its global partners, and client governments should join forces in setting up and implementing a multipronged mechanism to ensure adequate long-term funding for development data. Blending of domestic and international support could be a guiding principle. This mechanism should result in greater funding predictability, less ad hoc donor support tied to collection of specific surveys, and a gradual increase in shares of domestic financing for data aligned with countries' fiscal strength. The World Bank should also consider doing more to incorporate development data support and issues of data funding consistently into its engagement and dialogue in client countries. SCDs should more consistently pinpoint data gaps.

Some countries produce data that they have little capacity to analyze and use; some even receive World Bank support for collecting data that they do not share with the public or even occasionally with the World Bank. This is unreasonable. For reasons that are not entirely clear, the World Bank has not used its leverage fully to gain access to essential data or to promote open data sharing. Support from the international community to data production, as a rule, should be conditional on countries agreeing to share data (suitably anonymized) openly and promptly.

The World Bank should do more to influence countries toward greater data use. IEG identified several good practices in the World Bank, but no framework or approach for ensuring that interventions in

different sectors align to fostering a user-centered data culture. The *Strategic Actions Program for Addressing Development Data Gaps* (World Bank 2015) has limited discussion of data sharing, and almost no discussion of data use. Staff has mixed views on the merit of pursuing a data use objective and, in practice, often pays limited attention to it. The World Bank and other global actors should develop a framework for marshaling the disparate interventions to encourage developing countries toward greater data demand and evidence-driven policy making.

Some of the factors underlying data use or non-use by government decision makers, especially those related to government ability, can be addressed through communication, policy dialogue, and training. The World Bank can invest in showing governments the value of specific data types to address specific goals and build their data analytics capacity so they can better draw actionable insights from data. Open data initiatives can also generate pressure on government to act. The World Bank could also use comparative data on the performance of government programs and agencies across jurisdictions to change incentives and spur action.

The World Bank has no clear goals for its contributions to the global statistical system and data partnerships. An overarching vision should articulate goals for engagements in global data partnerships and how the World Bank can help maintain coherent global data efforts with clear roles for the numerous agencies and partnerships active in data. This coherence existed with a well-defined partnership architecture in the past, but the present reality is a more fragmented landscape with unclear funding.

The World Bank could do more to pursue long-term data goals and foster connections across different data-related activities in country programs. Except for the relatively small number of core projects dedicated to statistical capacity, much of the World Bank's support for data production, sharing, and use occurs as a by-product of other work. Data efforts are often task-focused and rarely work toward a common purpose related to data. As stated by the external panel review of the Development Economics Vice Presidency (DEC), "Data are seen as a by-product of other activities rather than a resource in their own right, with a lack of a coherent data infrastructure, and data that cannot be integrated and reused" (Besley and others 2015). Country dialogue gives uneven priority to development data, and interviews show a shared sense that the World Bank could and should do even better in its client-facing data work.[3]

Could the World Bank organize its data work better? Data resides in all global practices and is concentrated in the Poverty Global Practice (which handles most statistical capacity–building projects) and DEC (which handles many global and corporate responsibilities). Because of its decentralized structure and entrepreneurial staff, it is hard to manage a cross-cutting topic like data in the World Bank. At the corporate level, it is unknown whether the Data Council can emerge as an effective governing body or its different working groups can coordinate on technical issues. The Data Council has not resolved internal budget issues. Some observers have informally suggested an advisory data committee (a body or council) with external representation, broader responsibility, and stronger powers. It would involve an eminent group of data users and producers in governance and budgeting for the World Bank's own data work. The World Bank sometimes uses such advisory

bodies with external representation, but they rarely have real decision-making power or budgetary authority. Given that staff is already well-connected to global data actors, the value added of a new body is questionable.

Regarding big data, the World Bank currently spreads responsibilities across three World Bank units, which can result in inefficiencies from overlapping responsibilities and a lack of strong coordination. Despite many pilot initiatives, a common understanding is lacking of big data's potential and pitfalls in answering development questions, and internal capacity is weak. The World Bank often pursues innovation through bottom-up initiatives. The challenge is how to scale big data and other data innovations and how to ensure sustainability. After establishing big data's potential and pitfalls for development, the World Bank will need to decide how to strengthen and consolidate its skills and efforts in this area. Other needs include a framework for managing privacy, ethical, and other risks.

Based on the evidence, the evaluation offers five recommendations.

Recommendation 1: Implement goals and priorities reflecting the findings of this evaluation with regard to the World Bank's support to global data and global partnerships, country data capacity, and a user-centered data culture.

Steps to be considered by World Bank Management could include:

- articulating goals and priorities;
- specifying accountabilities for the implementation of new and existing goals and priorities; and
- ensuring sufficient management oversight so that the new and existing goals and priorities are implemented.

Recommendation 2: Mobilize and deliver additional support to countries' statistical systems, using a more comprehensive model of statistical capacity building that also factors in needs and opportunities to strengthen administrative data systems.

Recommendation 3: Step up engagements with global partners and client governments on long-term funding for development data.

Steps to be considered by World Bank Management could include:

- requiring CPFs to explicitly indicate how the SCD-identified knowledge and data gaps, which are most relevant to CPF objectives, will be addressed;[4]
- elevating attention to funding for data in the policy dialogue with client governments; and
- initiating high-level discussions on establishing a global umbrella mechanism for long-term financing of data.

Recommendation 4: Scale up promotion of data sharing and data use.

Steps to be considered by Bank Management could include:

- ensuring that all data financed by the World Bank are shared with the World Bank;

- developing and using a list of essential data items that countries are expected to share with the World Bank;
- incentivizing governments to more openly share data with the public, for example, by more prominently using a ranking of countries on open data performance; and
- scaling-up promotion of government and citizen demand for data and the voice of data users in the kinds of data that are produced.

**Recommendation 5:** Implement coordinated actions so that World Bank operations benefit from big data's insights and clients receive appropriate support for big data use.

Steps to be considered by World Bank Management could include:

- reviewing opportunities to scale up the use of big data for development;
- specifying accountabilities for implementation of the coordinated actions; and
- ensuring sufficient management oversight so that the coordinated actions are implemented.

---

[1] The commitment requires funding estimated at $148 million every three years.

[2] According to one estimate, for example, "The estimated cost of an expanded program of surveys and censuses and improvements in administrative data systems for 77 IDA-eligible countries over the SDG [Sustainable Development Goal] period is $17.0 billion to $17.7 billion. Total expenditures by IBRD countries to produce SDG indicators are expected to be $26.5 billion to $27.6 billion. Total aid needed to support the production of Tier I and II indicators for the SDGs is expected to be $635 million to $685 million a year over the period of 2016 to 2030" (GPSDD 2016).

[3] An evaluation of data in the IMF has a similar conclusion. "Efforts are under way in this regard…but these efforts are, as previous attempts, piecemeal without a clear comprehensive strategy which recognizes data as an institutional strategic asset, not just a consumption good for economists" (IEO 2016, 1).

[4] This step echoes recommendation 3 in the IEG report, *World Bank Group Country Engagement: An Early-Stage Assessment of the Systematic Country Diagnostic and Country Partnership Framework* (World Bank 2017).

## References


Besley, Timothy, Peter Henry, Christina Paxson, and Christopher Udry. 2015. *Evaluation Panel Review of DEC. A Report to the Chief Economist and Senior Vice President.* World Bank, Washington, DC.

Glassman, Amanda, Alex Ezeh, Kate McQueston, Jessica Brinton, and Jenny Ottenhoff. 2014. *Delivering on the Data Revolution in Sub-Saharan Africa: Final Report of the Data for African Development Working Group.* Washington, DC: Center for Global Development.

GPSDD (Global Partnership for Sustainable Development Data). 2016. "The State of Development Data Funding 2016." Open Data Watch. http://opendatawatch.com/the-state-of-development-data-2016/.

IEO (Independent Evaluation Office, International Monetary Fund). 2016. *Behind the Scenes with Data at the IMF: An IEO Evaluation*. Washington, DC: IMF.

World Bank. 2015. *World Bank Group Strategic Actions Program for Addressing Development Data Gaps: 2016–2025.* Washington, DC: World Bank.

——. 2016. “Forward Look: A Vision for the World Bank Group In 2030.” Report to Governors at Annual Meetings 2016, World Bank, Washington, DC.

——. 2017. *World Bank Group Country Engagement: An Early-Stage Assessment of the Systematic Country Diagnostic and Country Partnership Framework Process and Implementation.* Independent Evaluation Group. Washington, DC: World Bank.

APPENDIXES

# Data for Development

## An Evaluation of World Bank Support for Data and Statistical Capacity

IEG
Independent Evaluation Group

WORLD BANK GROUP
World Bank • IFC • MIGA

WHAT WORKS

# Appendix A. Methodological Approach

## Evaluation Questions

1. The evaluation's objective was to assess how effectively the World Bank has supported the production, sharing, and use of development data, and to suggest ways to improve. This overarching objective inspired four lines of inquiry that guided the collection and analysis of data and the framing of findings and recommendations (box A.1).

### Box A.1. Four Lines of Inquiry that Guided the Evaluation

- Has the World Bank contributed effectively to data for the global public good and data partnerships?
- How effectively has the World Bank helped countries strengthen data production?
- How effectively has the World Bank promoted data sharing and use in countries?
- Is the World Bank keeping up with technological innovations, particularly those relating to big data?

## Overarching Principles

2. Three central principles motivated the evaluation design: multilevel analysis, theory-based evaluation, and mixed methods. First, the evaluation adopted a multilevel perspective because the assessments covered both the global and national dimensions of World Bank support to data production, sharing, and use. Second, the evaluation was grounded in a theory of change—a reconstruction of how the changes sought by the World Bank's support to global partnerships and national statistical systems were expected to improve data production, sharing, and use. IEG reconstructed the theory of change through an iterative process and validated it with key stakeholders. Third, the evaluation followed a mixed-methods approach combining a range of methods for data collection and analysis, and applied systematic triangulation to ensure the robustness of the findings.

## Evaluation Components

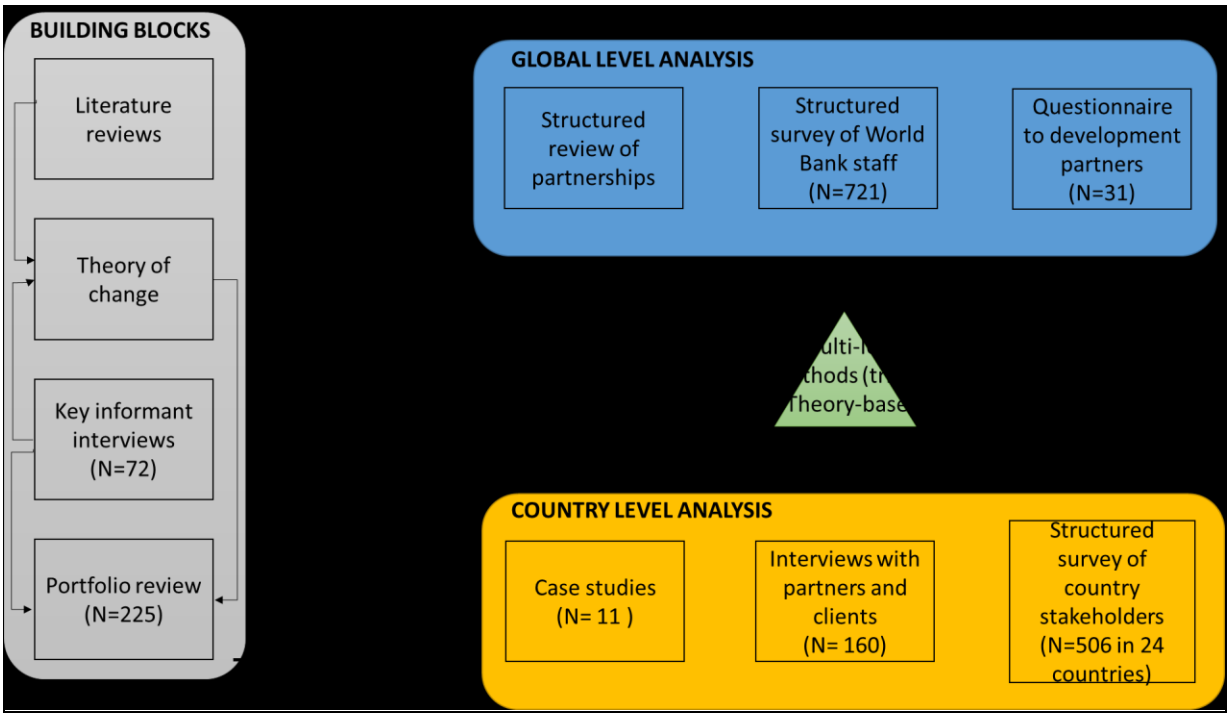
3. Table A.1 lists the evaluation components, and figure A.1 shows their articulation within the overall evaluation design. The next two sections provide more details on each component.

## Table A.1. Evaluation Components

| Evaluation Component | Description |
|---|---|
| Literature reviews and background papers | Structured review of the academic, evaluation, and gray literature on data production, sharing, and use; background papers on topics such as the World Bank's contribution to gender statistics |
| Portfolio review | Systematic desk review and assessment of 225 core projects across 95 countries |
| Interviews with World Bank staff | Semistructured interviews with 72 World Bank staff |
| Reconstruction of a theory of change | Reconstruction of how the desired changes sought by the World Bank to data production, sharing, and use were expected to happen |
| Systematic review of partnership programs | Review of major global partnership on data, including synthesis of existing evaluative evidence |
| Structured survey of World Bank staff | Survey addressed to staff across the World Bank focused on the organization's effectiveness in promoting data production, sharing, and use, and on the factors hindering or enhancing its effectiveness |
| Questionnaire administered to development partners | Questionnaire seeking development partners' views on the World Bank's comparative advantage and its role as a global partner on data |
| Structured survey of country stakeholders | Survey fielded in 24 client countries with more than four World Bank engagements related to data and development, asking for feedback on the World Bank's effectiveness in promoting data production, sharing, and use and on the factors hindering or enhancing its effectiveness |
| Case studies of the World Bank's role in supporting national statistical systems in data production, sharing, and use | In-depth analysis of the World Bank's support to country clients' statistical systems through field-based case studies (India, Indonesia, Peru, Rwanda, Tanzania, and Ukraine), PPARs (Ghana and Kenya), and desk-based case studies (Afghanistan, Bolivia, and Jordan) |

*Note*: PPAR = Project Performance Assessment Report.

### Figure A.1. Methodological Design




## Ensuring the Validity of Findings

4. IEG took several steps to guarantee a consistent approach across evaluation team members—for example, using a case study template and interview protocols to ensure a common framework and evaluative lens across studies. Similarly, IEG secured interrater reliability across team members charged with coding interview transcripts.

5. Furthermore, the team applied triangulation at multiple levels, first by crosschecking evidence sources within a given methodological component. Within case studies, for example, the team compared and contrasted evidence from interviews with national statistical offices (NSOs), development partners, and World Bank staff on the same topic. Second, the team applied triangulation across evaluation components—for example, cross-validating findings from case studies with findings from surveys and portfolio analysis.

6. The evaluation team also applied external validation mechanisms at various intervals during the evaluation process. For example, the team identified the portfolio of core activities through an iterative process in dialogue with the Development Data Group. Five peer reviewers provided feedback at the beginning, during, and end of the evaluation process. Finally, the team organized workshops with a panel of key stakeholders at the beginning of the evaluation process to validate the scope and the approach, and at the end to ensure the relevance and feasibility of the evaluation recommendations.

## Limitations

7. Notwithstanding these steps, the team documented several limitations to the evaluation design that broadly fell into two categories. The first set of limitations is the result of conscious choices about scope, and the second set is limitations due to other methodological and data availability reasons. Limitations in scope include the following:

- The team made a necessary trade-off between breadth and depth of analysis, covering some themes in detail and others more superficially, including the World Bank's role in supporting data systems in fragile and conflict-affected states
- The evaluation scope was deliberately outward looking and paid limited attention to internal coordination issues covered in previous evaluative work
- The team purposefully centered the case study selection model on countries in which the World Bank had a significant data engagement to gauge effects at the system level. Cases selected included at least one core project such as a major statistical capacity–building initiative. This purposive sampling of countries is not representative of the total population of countries in which the World Bank is active
- The evaluation focused particularly on capacity-building initiatives that supported NSOs and went into less depth on support to line ministries, partly because statistical support to line ministries is difficult to identify in project documents. The evaluation found it impossible to develop a comprehensive picture of the extent of sectoral data support apart from that emerging from the review of type 2 and type 3 projects. The evaluation therefore somewhat superficially covered support to line ministries through its portfolio review, structured surveys that also targeted staff in line ministries, and case studies covering projects supporting line ministries' data production and sharing.

Other limitations include the following:

- The team faced several challenges in identifying the core portfolio of data activities. The World Bank typically uses a sector coding system to account for its projects and operations, but it did not have a dedicated code to identify its data projects or projects with a data-related component until June 30, 2016. Therefore, the team estimated the total amount of World Bank commitments using various information sources and identification criteria (appendix B)
- A large range of data-related activities funded by trust funds had important information gaps. Similar information gaps existed for advisory services and analytics, where the World Bank does not document its progress as systematically as it does for investment projects
- The evaluation had to rely on proxy measures and perception data to study the topic of data use—a core concept of the evaluation—because it is particularly difficult to conceptualize, observe, and ultimately measure, and it has important construct validity

issues. Another challenge was the relative lack of research on data use compared with the abundance of literature on data production and sharing

- Survey respondents did not represent the overall population for several reasons. The primary reason is the development methods of the survey frames that purposively targeted countries in which the World Bank had more than four data engagements and where the official or operative language was English, French, or Spanish. This purposive sampling and the surveys' response rate have implications for the generalization of findings.

## Building-Block Studies

8. Four of the 10 methodological components were building blocks for the rest of the evaluation and informed further methodological choices, selection strategy, and the substance of the evaluation findings. They included in-depth literature reviews, a reconstruction of the theory of change, a systematic portfolio review, and interviews with a large number of people both inside and outside the World Bank.

### Literature Reviews

9. The team conducted two types of literature reviews during the evaluation process. The first study analyzed and synthesized a diverse and extensive set of academic, evaluation, and gray literature to understand the current state of development data—from production to use—across the development spectrum. The study sought to understand the key accomplishments across the entire statistical value chain, from identifying user needs to data collection, archiving, analysis, dissemination, and eventual use. It used the four lines of inquiry listed in box A.1 as a guide to identify sources and synthesize existing evidence.

10. A second, more targeted literature review sought to address a particularly complex question: Is there a tension between global monitoring and national policy-relevant data? IEG commissioned the review to Morten Jerven, a world-renowned expert on the political economy of data and statistics. The review synthesized the latest theoretical and empirical literature to establish patterns of the effect of international and donor data priorities on national statistical capacity. It also examined the phenomenon of the large increase in national strategies for the development of statistics and what this means for national priorities.

### Theory of Change

11. The evaluation drew on a theory-based evaluation approach, systematically and iteratively reconstructing and recalibrating the theory of change underlying the World Bank's support to client countries' statistical systems. Data production is understood relatively well, so the theory of change focuses on reconstructing the causal chains and processes underlying data dissemination and use. The idea of information polity—defined as "stakeholders, data sources,

data resources information flows, and governance relationships involved in the provision and use of government-held and nongovernmental data sources"—is an important construct in understanding the factors and processes that underlie data use (Helbig and others 2012).

12. Several recent evaluations of donors' support to countries' statistical system have adopted a theory of change approach (for example, UN 2016; UNFPA 2015). This evaluation used their frameworks, which were validated empirically, as a starting point and modified them to reflect the specificity of World Bank support. A theory of change underlying complex systems is necessarily a simplification. It relies on stylized relationships and makes cognizant decisions about the systems' boundaries and the enabling conditions to which the evaluation paid close attention. The theory of change also informed the design of data collection methods. IEG conducted the process of reconstructing the theory of change by integrating insights from the literature review, the portfolio analysis, and evidence from case studies. More specifically IEG relied on the following elements:

- A review of existing empirical research and evaluations on data production, open data, and data use
- A review of the documents produced under the auspices of the Data Council
- A review of documents for data for development projects—for example, project appraisal documents, Implementation Completion and Results reports (ICRs), and evaluations
- Stakeholder consultation and validation of the model.

13. Figure 1.1 in chapter 1 shows the World Bank's various forms of support to country clients' statistical systems, on both the supply side and demand side. The green-shaded boxes represent the World Bank's various forms of support. Referring to figure 1.1, the World Bank sought to bring change to its country clients' statistical systems in the following ways:

- Providing indirect support through its global-level work (top of the figure)
- Strengthening the capacity of national stakeholders to produce, analyze, and share data (left side of figure)
- Fostering an ecosystem for enhanced data use (right side of figure).

## Portfolio Review

14. The portfolio review exercise involved a systematic review of relevant project documents for the data projects portfolio. This initial portfolio included 291 projects approved during the FY06–15 period.[1] The portfolio review's main goal was to establish the extent of World Bank support for data activities and the nature of activities supported through the World Bank's lending to client countries. Given the broad nature of the initial portfolio, the review also sought to identify a core portfolio of projects by eliminating projects with no relevance to data for development.

15. The identification of a relevant portfolio focused exclusively on IBRD and IDA lending (including development policy financing), recipient-executed trust fund grants, and World Bank analytic work approved during FY06–15. The review excluded World Bank–executed grants, terminated or dropped projects, or projects still in the pipeline.

16. Given the lack of a harmonized system for tracking World Bank support for development data activities, IEG constructed the portfolio by compiling an extensive list of projects from different sources (discussed in the next sections) and excluding false positives based on a manual review of project documents.

17. The first step involved selecting all projects approved under the World Bank's different statistical capacity–building programs and trust funds. IEG retrieved a partial list of the World Bank's statistical capacity–building programs from the World Bank website.[2] The team identified a list of projects based on data available on the initiatives' websites, including the Statistical Capacity Building Program (STATCAP), the Multi-Donor Trust Fund to Support Statistical Capacity Building in Eastern Europe and CIS Countries (known as ECASTAT), and the Statistics for Results Facility Catalytic Fund. IEG obtained the commitments approved under the Trust Fund for Statistical Capacity Building separately from the World Bank's Business Intelligence database.

18. IEG supplemented the project lists with project data compiled and provided by the World Bank's Development Data Group, which maintains a list of all World Bank development data activities led by the group's project staff.[3]

19. The preliminary portfolio also included all the World Bank activities (loans, grants, and Advisory Services and Analytics) assigned the theme code 22 (economic statistics, modeling, and forecasting). Although the World Bank lacks a theme or sector flag to identify data-related activities, IEG felt that this particular code is a close approximation. The subsequent manual review of project documents eliminated several false positives.

20. IEG used a keyword search of project titles to identify projects whose names indicated support for development data activities. The keywords used included statistical capacity building, devstat, stats, survey, and census.

21. Keyword search of relevant databases: The final step in the portfolio selection process involved keyword searches of the prior actions database and components database.[4] The search included the following keywords, among others: data, statistical, open government, statistics, websites, open data, civil registry, living standards measurement, census, and survey.

22. IEG identified relevant advisory services and analytics activities based on only theme codes and project title searches.

23. The World Bank Development Data Group then validated the final portfolio.

24. Of the 291 projects in the initial portfolio, IEG excluded 66 projects from the final portfolio because they were not relevant to data for development. The core portfolio of 225 interventions included 201 commitments to 95 countries and 24 commitments to country groupings such as the Organisation of Eastern Caribbean States, Andean countries, Western Africa, Pacific Islands, and the like.

25. IEG developed a template for systematic data extraction in line with the evaluation questions. The team systematically mined the information contained in the project documents (project appraisal documents, concept notes, and ICRs and Implementation Completion Results Reviews (ICRRs) when available) and created an Excel database to record the extracted information and proceed with data aggregation. The team then generated simple frequency statistics.

### Key Informant Interviews

26. Between June and September of 2016, the evaluation team conducted 76 semistructured interviews with World Bank staff and external experts on a range of topics related to the World Bank's role in promoting data for development. IEG conducted the following interviews:

- 21 interviews with World Bank staff early in the evaluation process to identify prominent World Bank initiatives related to data production, sharing, or use
- 43 interviews with senior-level World Bank staff in the global practices, cross-cutting solutions areas, and World Bank regions selected to obtain cross-cutting perspectives
- seven interviews with World Bank staff involved in the day-to-day business operations of the World Bank's main data partnerships or trust funds
- five interviews with external informants.

27. The evaluation team took detailed, written notes for each interview and systematically coded and analyzed those using content-analysis software (NVivo) to derive themes and key messages from the interviews that could be triangulated with each other and with other information sources (notably survey responses and in-depth case studies).

## Assessing the World Bank's Contribution to Data for Development at the Global Level

### Systematic Analysis of Data Partnerships and the World Bank's Contribution to Gender Data

28. IEG performed a systematic analysis of formal partnerships on the main evaluation questions about data production, sharing, and use when applicable. Specifically, the review's objectives were as follows:

- Describe the main partnerships in which the World Bank engaged for advancing data for development and summarize findings and recommendations from existing evaluations
- Describe the partnership's stated and perceived purpose from a member's perspective
- Assess the extent to which the partnership is meeting the stakeholders' and beneficiaries' needs from a member's perspective
- Describe the member's contributions to the partnership from a member's perspective
- Describe the outcomes associated with the partnerships' data production, dissemination, and use from the member's perspective
- Identify key success factors for effective partnerships from the member's perspective
- Identify perceived barriers or other factors that limited the effectiveness of the partnership from a member's perspective
- Describe the partnership's relevance in the context of the Sustainable Development Goals from a member's perspective

29. IEG based the assessment on three primary data sources: partnership documents and websites, available evaluations of partnerships (five formal evaluations were available out of the 10 partnerships reviewed), and interviews with World Bank and partners' staff and stakeholders involved in the partnerships. The data collection methods consisted of a systematic review of documents, interviews, and a questionnaire distributed to select partners. Table A.2 presents the main information sources for each partnership.

### Table A.2. Evaluation and Annual Reports by Partnerships Reviewed

| Partnership | Evaluation Report | Annual Reports |
|---|---|---|
| Marrakech Action Plan for Statistics 2004 | Yes | Yes |
| Statistics for Results Facility Catalytic Fund | Yes | Yes |
| PARIS21 | Yes | Yes |
| Trust Fund for Statistical Capacity Building | Yes | Yes |
| International Comparison Program | Yes | Yes |
| Living Standards Measurement Study | No | No |
| Open Data for Development | No | Yes |
| Global Findex Database | No | Yes |
| Living Standards Measurement Study Integrated Surveys on Agriculture | No | Yes |
| Global Partnership for Sustainable Development Data | No | No |

*Note*: PARIS21 = Partnership in Statistics for Development in the 21st Century.

## Questionnaire Administered to Development Partners

30. The evaluation team consulted with the World Bank staff in charge of managing data partnerships and conducted desk-based research to compile a list of development partners involved in global data partnerships in which the World Bank is an active member. IEG sent a questionnaire to 135 partners and reached 123 people (SurveyMonkey reported three opt-outs and nine bounce-backs). The team administered the questionnaire between October 4 and November 11, 2016 and sent eight e-mail reminders (table A.3).

### Table A.3. Coverage of the Development Partner Surveys and Structured Questionnaire

| Concept | Value | Description |
|---|---|---|
| Target population | 135 | Number of development partners involved in World Bank data partnerships (survey sent to 135 partners) |
| Population reached | 123 | Survey had nine bounce backs and three opt-outs with a reachable population of 123 people |
| Respondents (response rate = 25.2%) | 31 | Number of partners that responded to the survey (25.2% of the target population) |

## Structured Survey of World Bank Staff

31. To build the survey frame, the evaluation team obtained from the World Bank's human resource department a list of staff in all global practices, regions, cross-cutting solutions areas, DEC, and the Office of the President's Special Envoy whose grade level is GF and above. The team excluded staff who were not in a direct operational or research role, yielding a total population of 4,500 staff members. The team randomly assigned staff members to one of two surveys that IEG conducted concurrently (one survey for this evaluation and one survey for the evaluation of shared prosperity). The survey's sample size was 2,420 staff members, 52 of whom had previously opted out of SurveyMonkey surveys, bringing the total number of respondents to 2,369. IEG administered the anonymous, confidential survey questionnaire through SurveyMonkey between October 4 and November 11, 2016. Nonrespondents received 11 reminders, and a random sample of nonrespondents received phone reminders during the last two weeks of the survey.[5] Furthermore, the Director of the IEG Human Development and Economic Management Department sent an e-mail to a random sample of 60 directors asking them to encourage their staff to take the survey (table A.4).

### Table A.4. Coverage of World Bank Staff Structured Survey

| Concept | Value | Description |
|---|---|---|
| Target population | 4,500 | Number of World Bank staff in the survey frame |
| Random sample | 2,420 | Number of World Bank staff randomly sampled to receive the survey (sent to 2,369 World Bank staff) |
| Sample reached | 2,369 | 52 opted out of SurveyMonkey surveys |

| | | |
|---|---|---|
| Respondents (response rate = 30.18%) | 721 | Number of World Bank staff who responded to the survey (30.18% of the target population) |

32. The evaluation team applied descriptive statistics (a description of sample frequencies) and goodness of fit tests to the two main surveys (the next section describes other surveys), and used sample frequencies and crosstab analyses to assess the results of conditional filters (for example, sorting respondents by particular characteristics).

## Assessing the World Bank's Contribution to Data for Development at the Country Level

### Case Studies

33. The evaluation conducted 11 case studies of the World Bank's role in supporting national statistical systems in data production, sharing, and use. Case studies are in-depth analyses of specific configurations intersecting a specific World Bank support setting, country context, and modality of support. These specific configurations give rise to a constellation of factors that influence the outcome of statistical capacity–building initiatives. The objective for each case study was to assess the extent of success of World Bank interventions and understand the specific constellation of factors that account for particular outcomes. IEG conducted case studies in countries where World Bank support was sufficiently large that effects were observable at the system level. Taken together, the cases illustrated the range of World Bank support modalities for data production, sharing, and use and a sufficiently diverse set of contexts.

34. Table A.5 summarizes the case selection and the type of study. The case study countries were selected purposefully, based on the following criteria:

- World Bank financial and technical support to data: the sample included countries receiving a medium to high level of support[6]
- Statistical Capacity Initiative: the sample had to include countries where the World Bank funded at least one statistical capacity–building initiative
- Diversity of the World Bank's support modalities: the sample had to include countries that also had other data-related projects and Advisory Services and Analytics as identified in the preliminary portfolio
- Timing of World Bank support: the main statistical capacity–building initiative had to be active or recently closed
- Diversity of contexts: the sample had to include statistical systems in all World Bank regions and a mix of large and small countries, with deliberate oversampling of IDA countries.

### Table A.5. Case Studies Selected

| World Bank Regions | Case | Type of Study |
|---|---|---|
| Africa | Ghana | PPAR |
| | Kenya | PPAR |
| | Rwanda | Field-based case study |
| | Tanzania | Field-based case study |
| East Asia and Pacific | Indonesia | Field-based case study |
| Europe and Central Asia | Ukraine | Field-based case study and PPAR |
| Latin America and the Caribbean | Bolivia | Desk-based case study |
| | Peru | Field-based case study |
| Middle East and North Africa | Jordan | Desk-based case study |
| South Asia | Afghanistan | Desk-based case study |
| | India | Field-based case study |

*Note*: PPAR = project performance assessment report.

35. Each case study involved an extensive desk-based review of project documents, an analysis of pertinent indicators, and a review of existing empirical evidence on the country's statistical system (for example, a review of case studies conducted by other partners, self-evaluation material, and diagnostic studies conducted by the Partnership in Statistics for Development in the 21st Century). IEG also conducted interviews with World Bank staff in charge of the portfolio of interventions in each country (40 interviews across case studies). A team composed of at least one IEG evaluation expert and a statistics expert also conducted an in-depth field visit in eight of 11 cases. The team conducted interviews and roundtable discussions with data users and producers, including country client officials, staff in partner agencies, and researchers (160 interviews across case studies). Three of the eight field-based case studies were project performance assessment reports.

36. The evaluation team wrote a case narrative for each country using an established template. Subsequently, the team systematically analyzed, compared, contrasted, and synthesized the evidence emerging from the 11 case narratives using a qualitative analysis software (NVivo). The team then coded individual cases along the dimensions of the theory of change and fed the synthetic evidence into the evaluation report.

## Structured Survey of Country Stakeholders

**37.** The evaluation team compiled a list of 24 countries (box A.2) where the World Bank had more than four data engagements and where the official or operative language was English, French, or Spanish. This purposive sampling has implications for the generalization of findings. For each country, the evaluation team asked country offices to identify target respondents in five stakeholders group (government agencies, civil society organizations, partners, academia, and the private sector). For government agency stakeholders, the team targeted only director-

level officials. The total target survey population was 2,371, but 446 bounce-backs and individuals who opted out reduced the total survey population to 1,925. The survey was anonymous and administered in English, Spanish, and French through Snap Survey Software between October 4 and November 11, 2016. The team sent five reminders to all targeted participants and phone reminders to a random sample in the last two weeks of the survey period.[7]

**Table A.6. Coverage of Stakeholders in Client Countries**

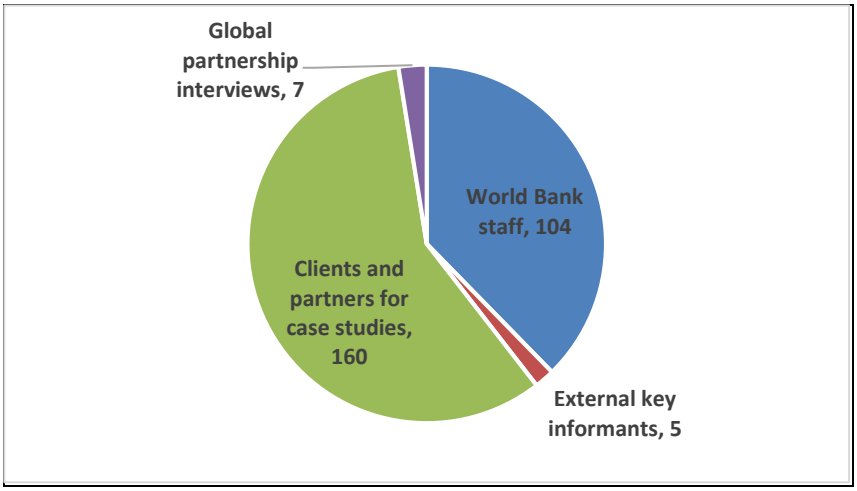
| Concept | Value |
|---|---|
| Target population | 2,371 |
| Reached population | 1,925 |
| Coverage (response rate = 26.52%) | 506 |

**Box A.1. Countries Taking Part in the Survey**

| | |
|---|---|
| Algeria | Namibia |
| Burkina Faso | Niger |
| Congo, Dem. Rep. | Nigeria |
| Dominican Republic | Pakistan |
| Ethiopia | Philippines |
| India | Rwanda |
| Jordan | Senegal |
| Kenya | Sierra Leone |
| Madagascar | South Africa |
| Malaysia | Tanzania |
| Maldives | Tunisia |
| Morocco | Zambia |

38. Figure A.3 shows the distribution of interviews conducted throughout the evaluation across methodological components. IEG consulted 276 people through interviews.

### Figure A.3. Interviews Conducted During the Evaluation




## References

Helbig, Natalie, Anthony M. Cresswell, G. Brian Burke, and Luis Luna-Reyes. 2012. "The Dynamics of Opening Government Data: A White Paper." The Center for Technology in Government, University at Albany, SUNY (State University of New York), Albany, NY.

United Nations, 2016. "Evaluation of the Contribution of the United Nations Development System to Strengthening National Capacities for Statistical Analysis and Data Collection to Support the Achievement of the Millennium Development Goals (MDGs) and Other Internationally-Agree, New York, NY: https://www.unjiu.org/en/reports-notes/JIU%20Products/JIU_REP_2016_5_English.pdf

---

[1] Given the lack of a harmonized system for tracking World Bank support for development data activities, IEG constructed the portfolio of 291 projects through a process of triangulating data from different sources. The Approach Paper describes the process and criteria IEG used to select the 291 projects.

[2] This partial list of World Bank statistical capacity programs included only the programs managed by the Statistical Development and Partnership Team of the Development Data Group. IEG excluded other existing programs in other World Bank departments. Please see the list at http://www.worldbank.org/en/data/statistical-capacity-building/trust-fund-for-statistical-capacity-building.

[3] The IEG team asked for similar lists from other World Bank departments and global practices but learned that other global practices do not systematically track this data.

[4] The World Bank prior actions database is a consolidated database on all prior actions associated with development policy operations approved since 1980. The database is maintained by the World Bank's

Operations Policy and Country Services Vice Presidency. The project components database contains information on the projects' components of investment loans approved since 1997. IEG created the database and currently maintains it.

[5] SurveyMonkey has a special feature to send reminders only to nonrespondents while keeping responses anonymous.

[6] The sample included countries with a medium or high level of World Bank engagement because it did not make sense to allocate time and resources to conducting in-depth, in-country case studies in countries where the World Bank had minimal data engagement or none at all.

[7] Snap Survey Software does not allow sending reminders only to nonrespondents because the survey was anonymous and IEG did not collect the respondents' personally identifiable information, such as e-mail addresses.

# Appendix B. Portfolio Review of World Bank Data for Development Lending Commitments

1. The portfolio review exercise involved a systematic review of relevant project documents for the data projects portfolio IEG identified. This initial portfolio was composed of 291 projects approved during the period FY06–15.[1] The exercise intended to establish the extent of World Bank support for data activities and the nature of activities supported through World Bank lending to client countries. Furthermore, given the broad nature of the initial portfolio, the review also sought to identify a core portfolio of projects by eliminating projects with no relevance to data for development.

## Identified Portfolio

2. The document review process yielded 225 core projects across 95 countries.[2] Financing by IBRD or IDA accounted for a greater share of financing by value (80 percent) though there were more trust-funded grants (table B.1). Estimating the value of policy-based financing for data activities was difficult because the project documents do not specify these amounts. Consequently, this report understates overall commitment amounts because IEG excluded the DPF contributions.

### Table B.1. Data for Development Projects by Product Line

| Product line | Number of Projects | Percent of Projects by Number | Commitment Value (US$, millions) | Percent of Projects by Value |
|---|---|---|---|---|
| IBRD/IDA | 92 | 41 | 739.0 | 80 |
| Trust fund grants | 133 | 59 | 180.5 | 20 |
| Total | 225 | 100 | 919.4 | 100 |

*Note*: This report understates the commitment value because IEG excluded the amounts from development policy financing, which could not be accurately estimated.

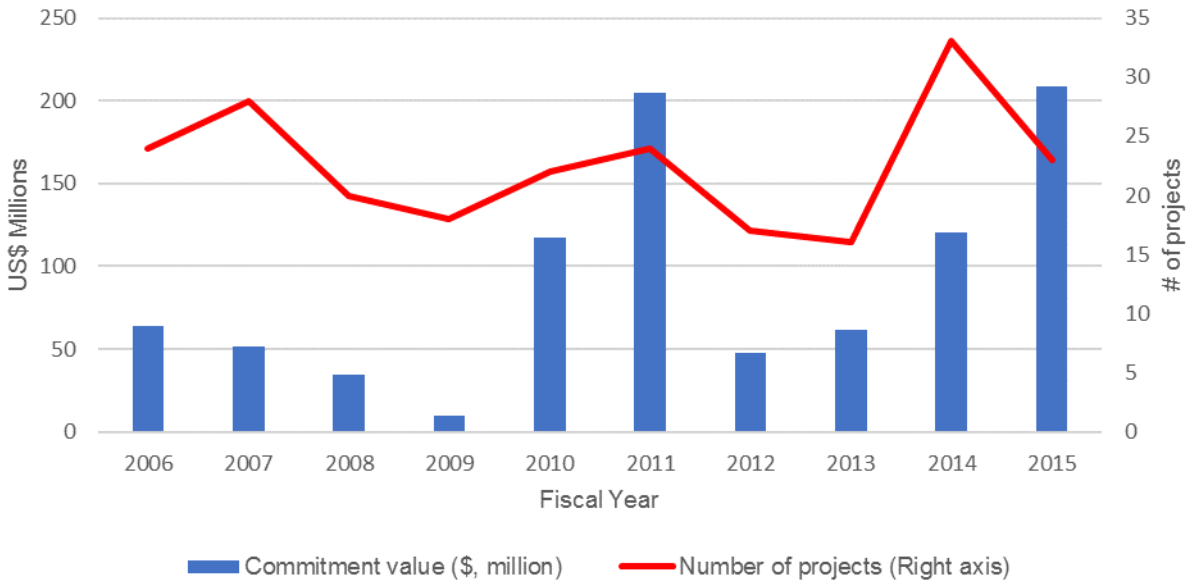
3. Investment lending was the main form of support for data activities. By number, investment lending projects accounted for 82 percent of the data for development portfolio, followed by adjustment at 17 percent, and Program for Results projects at 1 percent (table B.2).

### Table B.2. Data for Development Projects by Instrument Type

| Instrument Type | Number of Projects | Percent of Projects |
|---|---|---|
| Policy operations | 38 | 17 |
| Investment (lending and grants) | 184 | 82 |
| Program for results | 3 | 1 |
| Total | 225 | 100 |

4. World Bank commitments for data activities averaged about $90 million per year and increased in the latter half of the evaluation period (FY06–15). Average annual commitments during the 10-year period were about $90 million, with sharp increases in FY11 and FY15.[3] Declining commitments characterized the first part of the evaluation period—commitments fell from about $64 million in FY06 to $10 million in FY09. However, lending for data for development activities has increased steadily since 2012 from about $48 million to $209 million in 2015 (figure B.1). Consequently, the average annual commitments have more than doubled from $55 million during the FY06–10 period to $129 million during the FY11–15 period. By number, data for development projects did not show any sustained trend during the period, averaging about 23 projects each year.

**Figure B.1. Number of Projects and Commitments by Fiscal Year**




*Source*: World Bank Business Warehouse (database).

5. The Africa Region accounted for the largest share of data for development commitments in both value and number. About 44 percent of the number of commitments and 45 percent of the value was committed to countries in the Africa Region (table B.3). Overall, the top five recipients of World Bank support for data activities included Bolivia, India, Indonesia, Nigeria, and Rwanda.

### Table B.3. Data for Development Projects by Region

| Region | Number of Projects | Projects by Number (%) | Commitment Value (US$, millions) | Projects by Value (%) |
|---|---|---|---|---|
| Africa | 100 | 44 | 410.9 | 45 |
| East Asia and Pacific | 22 | 10 | 100.7 | 11 |
| Europe and Central Asia | 28 | 12 | 66.9 | 7 |
| Latin America and the Caribbean | 40 | 18 | 112.0 | 12 |
| Middle East and North Africa | 19 | 8 | 53.7 | 6 |
| Other | 1 | 0 | 0.2 | 0 |
| South Asia | 15 | 7 | 175.0 | 19 |
| Total | 225 | 100 | 919.4 | 100 |

## Types of Data Coverage

6. As part of this exercise, IEG classified the data for development projects into three categories based on the extent of support for data activities. Type 1 projects (stand-alone projects) used the entire loan or grant amount to support data activities and were about 62 percent of the data for development portfolio by number and 59 percent by value (table B.4).

### Table B.4. Data for Development Projects by Project Type

| Project Type | Number of Projects | Projects by Number (%) | Commitment Value (US$, millions) | Projects by Value (%) |
|---|---|---|---|---|
| 1 | 139 | 62 | 543.1 | 59 |
| 2 | 23 | 10 | 139.7 | 15 |
| 3 | 63 | 28 | 236.7 | 26 |
| Total | 225 | 100 | 919.4 | 100 |

*Note*: Type 1: The entire project supported data; Type 2: At least one entire component supported data; Type 3: The project supported relevant data activities, but the project components were not specifically data related.

7. Type 2 projects had *at least one* entire component that supported data for development and were 10 percent of the portfolio by number and 15 percent by value. Type 3 projects supported relevant data for development activities, but the projects and components were not specifically data related. These projects accounted for 28 percent of the number and 26 percent of the portfolio value.[4]

8. Of the 225 data for development projects, only 201 projects had enough documentation for IEG to review.

## World Bank Support for Identifying Data Gaps

9. The portfolio review tried to assess the extent to which World Bank–supported activities were informed by previously identified data gaps or were developed in response to them. The review found only a few cases in which the project documents explicitly referred to the existence of previously identified data gaps. Only 20 projects explicitly discussed the existence of data gaps or activities undertaken to diagnose existing data gaps. However, in several cases the documents referred to a previously undertaken national strategy for the development of statistics (NSDS) or noted the project's support for preparing the strategy.[5] Of the 201 projects reviewed, 41 projects (20 percent) either supported preparation of an NSDS or mentioned the existence of a previously undertaken NSDS.

## World Bank Support for Strengthening Production of Specific Data Types

10. The World Bank targeted support toward specific data types in about 56 percent of the reviewed projects. The type of activities typically supported included improving subject matter methodologies, capacity building aimed at enhancing skills for the production of the specific data types, and collecting the relevant data, especially through surveys. About 40 percent of projects included support for strengthening data production in at least one of these five priority areas. Production of household survey data received the most attention in number of projects.

**Table B.5. World Bank Support for Strengthening Data Production in Priority Areas**

| Priority Area | Number of Projects | Percent of Reviewed Projects |
|---|---|---|
| Household surveys | 40 | 20 |
| National accounts | 37 | 18 |
| Price statistics | 24 | 12 |
| Population statistics based on census and civil registration | 18 | 9 |
| Labor and job statistics | 11 | 5 |

*Note*: Some projects supported more than one priority area.

11. The World Bank also supported data production strengthening for other key types of data, including statistics in education, health, gender, tourism, and agriculture. Of the 201 projects reviewed, 78 projects (39 percent) supported the strengthening of these data types.

## World Bank Support for Data Collection Activities

12. The portfolio review found that World Bank financing was used in several cases to support actual data collection—household surveys, business surveys, population census, agriculture census, and surveys of private sector activity. Fifty-six percent of the projects reviewed involved support for collecting data. World Bank–financed surveys included, among

others, pilot surveys for agriculture statistics, health facility surveys, school censuses, and establishment surveys. For example, the FY14 Armenia agriculture census project used grant funds to undertake the following activities related to data collection:

- Training field staff (enumerators, registrars, supervisors, coordinators, and census area managers)
- Providing methodological assistance and supervising the fieldwork during the census
- Preparing census documents, including printing questionnaires, instructions, and publicity materials, and providing stationary necessary to conduct the pilot agricultural census
- Conducting fieldwork and interviewing respondents.

## Project Execution of Data Activities

13. National statistical offices (NSOs) were the primary implementing agencies in World Bank–supported data activities, especially for stand-alone projects. NSOs implemented the data activities in about 53 percent of the projects reviewed and 73 percent of the type 1 projects. For projects aimed at strengthening the production of specific sectoral data, the statistics departments of the responsible ministry were typically the implementing agencies. Other implementing agencies for data activities included the ministries of finance, research institutions, and regional institutions such as AFRISTAT (table B.6).

### Table B.6. Data for Development Project Implementing Agencies

| | Number of Projects | | | | Percent of Projects | | | |
|---|---|---|---|---|---|---|---|---|
| Implementing Agency | Type 1 | Type 2 | Type 3 | Total | Type 1 | Type 2 | Type 3 | Total |
| National statistical office | 83 | 10 | 14 | 107 | 72 | 43 | 22 | 53 |
| Ministry of finance or planning | 6 | 6 | 24 | 36 | 5 | 26 | 38 | 18 |
| Regional organization | 14 | 0 | 2 | 16 | 12 | 0 | 3 | 8 |
| Other ministries | 3 | 3 | 7 | 13 | 3 | 13 | 11 | 6 |
| Ministry of education | 0 | 3 | 8 | 11 | 0 | 13 | 13 | 5 |
| Research institute | 9 | 0 | 0 | 9 | 8 | 0 | 0 | 4 |
| Ministry of health | 0 | 1 | 8 | 9 | 0 | 4 | 13 | 4 |
| Total | 115 | 23 | 63 | 201 | 100 | 100 | 100 | 100 |

## Beneficiaries of Data for Development Projects

14. Project documents seldom specified the beneficiaries of World Bank support for data activities. Reviewed documents rarely specified even the intended project beneficiaries, and when they did, they did not specify the beneficiaries of the data activities separately. This could be because a large number of projects are trust-funded projects, which do not specify the intended project beneficiaries (unlike IBRD/IDA projects). However, when specified, the main

beneficiaries of data for development projects included NSOs, government ministries, departments and agencies, policy makers and decision makers, research institutions, international development partners, and the public.

## World Bank Support for Strengthening Client Capacity

15. The portfolio review also assessed the World Bank's contribution to strengthening client capacity to produce, disseminate, and use data. The review considered three key dimensions of client capacity: institutional capacity, legal and regulatory capacity, and human capacity.

16. Financing for institutional strengthening was an important form of World Bank support. Almost half of the projects reviewed (46 percent) supported strengthening client institutional capacity (table B.7) and delivered support to about 60 countries and three regional subgroupings. World Bank–supported institutional strengthening activities included organizational restructuring of NSOs, specifying formal coordination mechanisms between and among data producers and users, and strengthening sectoral offices. For example, the FY07 Statistical Capacity Building Program (STATCAP) project in Kenya supported the organizational restructuring of the Central Bureau of Statistics and its transformation from a government department in the ministry of finance to a new, semiautonomous agency. In the Kyrgyz Republic, a 2009 Trust Fund for Statistical Capacity Building grant provided technical assistance for defining the interaction between the National Statistical Committee and relevant line ministries and statistical agencies. In South Sudan, World Bank support was used to assess the NSO's organizational structure and implement changes to align the organization with its corporate objectives.

### Table B.7. World Bank Support for Strengthening Institutional Capacity

| | Number of Projects | | | Percent of Projects | | |
|---|---|---|---|---|---|---|
| Project Type | No | Yes | Total | No | Yes | Total |
| 1 | 57 | 58 | 115 | 50 | 50 | 100 |
| 2 | 7 | 16 | 23 | 30 | 70 | 100 |
| 3 | 44 | 19 | 63 | 70 | 30 | 100 |
| Total | 108 | 93 | 201 | 54 | 46 | 100 |

*Note*: Type 1: The entire project supported data; Type 2: At least one entire component supported data; Type 3: The project supported relevant data activities, but the project components were not specifically data related.

17. Support for strengthening the legal framework for data activities was less frequent than other forms of capacity strengthening, but it was still an important form of World Bank support. Only 20 percent of the reviewed projects involved support for strengthening the legal framework. In countries lacking a legal framework for statistical activities, the World Bank aimed to support the enactment of laws and regulations to guide these activities. In countries with existing legal frameworks, World Bank support was used to assess the adequacy of the

laws and regulations and when necessary support revision of the regulatory and procedural framework for government statistics. The World Bank also supported the revision of existing statistics legislation to give NSOs more professional and technical independence and to strengthen the accountability of official statistics producers. For example, the Strengthening the National Statistical System of Mongolia Project in FY09 financed a review of the law on statistics to better incorporate UN Fundamental Principles of Official Statistics and the Mongolian Code of Practice. The Strengthening the National Statistical System of Kazakhstan Project in FY11 supported the development of regulations and bylaws to improve interagency cooperation on statistical activities, and the revision of agreements between statistical agencies to ensure efficient interaction and information exchange. One of the prior actions of the FY11 Poverty Reduction Support Credit 5 DPO to Senegal was submission to parliament of amendments to the statistics law that mandated greater open access to primary data.

18. Support for strengthening human capacity was the largest form of capacity-strengthening financing. Seventy-one percent of the reviewed projects supported activities to strengthen human capacity to produce and use data (table B.8). Activities to strengthen human capacity included, among others, assessing competencies and training needs, supporting key staff participation in various training programs, and strengthening cooperation with universities to develop a training curriculum for statistics and to educate trainers. For example, the FY14 Comoros Statistics project supported the establishment and operation of a statistics training school at a university.

### Table B.8. World Bank Support for Strengthening Human Capacity

| | Number of Projects | | | Percent of Projects | | |
|---|---|---|---|---|---|---|
| Project Type | No | Yes | Total | No | Yes | Total |
| 1 | 14 | 101 | 115 | 12 | 88 | 100 |
| 2 | 6 | 17 | 23 | 26 | 74 | 100 |
| 3 | 39 | 24 | 63 | 62 | 38 | 100 |
| Total | 59 | 142 | 201 | 29 | 71 | 100 |

*Note*: Type 1: The entire project supported data; Type 2: At least one entire component supported data; Type 3: The project supported relevant data activities, but the project components were not specifically data related.

19. World Bank financing targeted the development of statistical methods, standards, and classifications to improve client countries' data quality. Eighty-one projects (40 percent of the reviewed portfolio) included support for these quality-enhancing activities, which included, among others, support for the adoption of internationally accepted standards and methodologies in data collection, compilation, and validation; improvement of questionnaire design and sampling frames; and improvement of sampling methods, population estimates, and projections. For example, the Additional Financing for Integrated Financial Management and Information System Project in The Gambia in FY14 provided training on the compilation and analysis of price and national accounts data.

20. To provide an enabling environment for data production, World Bank support to clients included financing for the acquisition of physical infrastructure, such as buildings and information technology (IT) equipment. Of the 201 projects reviewed, 79 provided this form of support. For example, the Tanzania Statistical Capacity Building Project in FY11 supported construction of new office buildings for the National Bureau of Statistics and the Office of the Chief Government Statistician. Other projects supported the acquisition of IT equipment, such as portable data assistants, computers, high-performance servers, and software.

## Support for Data Dissemination and Open Data

21. In several of its projects, the World Bank supported efforts to improve data producers' dissemination function and make development data more widely available to users. Of the 201 projects reviewed, 68 projects included support for increasing public access to development data. Activities financed under World Bank projects included the following, among others:

- Upgrades to NSO websites and the creation of open web portals to allow access to data users
- Publication of flagship statistical reports and documents produced by NSOs and other data producers
- Development of dissemination policies that respect the UN Fundamental Principles of Official Statistics and the African Charter on Statistics and technical assistance as needed.

22. For example, the FY12 Ghana Statistical Development Project supported the following activities relevant to data dissemination:

- Creation of a data dissemination and resource hub within the Ghana Statistical System
- Training provided to the respective line ministries on the communication and dissemination of statistics
- Improvements to the official national statistics website
- Development of a release calendar for national statistics
- Development of a policy for publications and dissemination.

## Partnerships

23. The portfolio review also sought to establish the extent and nature of partnership arrangements in World Bank data for development projects. The review found that only 28 of the 201 projects reviewed involved other development partners, including, among others, the U.K. Department for International Development (DFID), Sida, Statistics Norway, Statistics Korea, AusAID, Turkish International Cooperation Agency, and the European Union. The IMF was another key World Bank partner, especially in supporting clients to strengthen the

methodology for macroeconomic statistics. The partnership arrangements included cofinancing agreements such as with DFID under the FY07 Kenya STATCAP project, parallel financing of data activities such as surveys, and providing hands-on training and support. In several countries (including Kazakhstan, the Lao People's Democratic Republic, and Mongolia), World Bank projects supported the establishment of a twinning arrangement between the client NSO and a well-developed NSO (or a consortium of such offices). This approach was thought to promote greater knowledge transfer, and it was expected to reduce the transaction time and cost of implementation significantly, as well as the risk of improperly managing project funds.

## World Bank Support to Data Users

24. The portfolio review also considered the extent to which World Bank support enhanced clients' capacity to use the data produced. The review found that 27 projects (13.5 percent of reviewed projects) supported activities to build capacity for data use, such as user education workshops, training for media on how to use statistical information, and training and workshops to enhance data literacy among other data users.

## Results of World Bank Support for Data Activities

25. IEG reviewed the project completion documents for closed World Bank operations to assess the results of World Bank support for data activities.[6] Of the 225 projects in the IEG portfolio, 146 are closed, and completion documents are available for 75 of those. Considering the high number of trust fund grants in the portfolio and the sparse reporting on results for these grants, the assessment of results achieved by closed projects was limited. For the five dimensions of support shown in table B.9, the extent of results achieved for each dimension was rated on a scale of 0 to 3 (with 0 representing no documented results and 3 corresponding to a high degree of results achievement). The average project result score for strengthening data use was 1.7 compared with a higher score of 2.0 for building human capacity and an even higher score of 2.1 for strengthening the legal framework (table B.9).

### Table B.9. Average Results of World Bank Support for Data Activities

| Dimension | Strengthening the Legal Framework for Data Activities | Strengthening the Institutional Framework for Data Activities | Improving Data Access and Dissemination | Strengthening Data Use | Building Human Capacity |
|---|---|---|---|---|---|
| Projects with documented results | 20 | 34 | 34 | 12 | 47 |
| Average score | 2.15 | 1.9 | 1.9 | 1.75 | 2.04 |

*Note*: The rating scale is 0 (no results) to 3 (high degree of achievement).

More than half of the data for development projects validated by IEG received a satisfactory development outcome rating. However, the assigned outcome rating for type 1 and type 2 projects reflects the outcome of the whole

project, not just the project's data component. Overall, IEG validated 39 of the 146 closed projects (table B.10). Only seven of these were type 1 projects dedicated exclusively to data activities.

## Table B.10. IEG-Rated Data for Development Projects

| IEG Rating | Type 1 | Type 2 | Type 3 | Total |
|---|---|---|---|---|
| Highly satisfactory | | | 1 | 1 |
| Satisfactory | 3 | 1 | 7 | 11 |
| Moderately satisfactory | | 1 | 9 | 10 |
| Moderately unsatisfactory | 4 | 4 | 5 | 13 |
| Unsatisfactory | | 1 | 2 | 3 |
| Highly unsatisfactory | | | 1 | 1 |
| Total | 7 | 7 | 25 | 39 |

[1] Given the lack of a harmonized system for tracking World Bank support for development data activities, IEG constructed the portfolio of 291 projects through a process of triangulating data from different sources. The Approach Paper describes the process and criteria used to select the 291 projects.

[2] Of the initial portfolio of 291 projects, IEG excluded 66 projects from the final portfolio because they were not relevant to data for development. The core portfolio of 225 interventions included 201 commitments to 95 countries and 24 commitments to country groupings such as the Organisation of Eastern Caribbean States, Andean countries, West Africa, the Pacific Islands, and the like.

[3] In both FY11 and FY15, approval of a few large value projects caused the sharp rise in data for development commitments (projects such as the $65 million FY11 Strengthening Indonesian Statistics project and the $80 million component of the FY15 Nigeria Saving One Million Lives Initiative Program-for-Results Project, for example).

[4] This report understates the commitment value of type 3 projects because IEG excluded DPF amounts, which could not be reliably estimated. All excluded projects were type 3 projects.

[5] A national strategy for the development of statistics (NSDS) is expected to provide a country with a strategy for developing statistical capacity across the national statistical system. The preparation process for an NSDS is assumed to involve an assessment of existing data gaps and an implementation plan for closing these gaps.

[6] IEG reviewed three main types of completion documents: Implementation Completion and Results Reports prepared by the implementing team at project closure, Implementation Completion and Results Reviews prepared by IEG, and Implementation Completion Memorandums for trust fund grants.

# Appendix C. Survey Data Findings

1. The structured survey that IEG conducted for this evaluation shows that country stakeholders (academia, civil society, donor agencies, government, and the private sector) tend to rate the World Bank's effectiveness in promoting development data higher than the World Bank staff do.

2. In each of the two groups that IEG surveyed (World Bank staff and country stakeholders), slightly less than 70 percent rated the World Bank as highly effective or effective in making key data sets available globally. Between one-half and two-thirds of respondents in both groups rated the World Bank as highly effective or effective in developing standards and protocols to ensure global data quality. Respondents gave similarly high ratings of effectiveness to the World Bank's performance in supporting global data innovations (such as open data, systems for big data, or use of tablets and phones for surveys) and bringing development partners and governments together to discuss global data issues (table C.1).

3. Between 62 percent and 77 percent of country stakeholder subgroups rated the World Bank as highly effective or effective in making key data sets available globally (table C.2).

4. World Bank staff and country stakeholders that IEG surveyed agreed that the World Bank has been more effective at helping countries *produce* data than helping them to *share* or *use* data. On each of these dimensions, country stakeholders rated World Bank effectiveness higher than World Bank staff did (table C.3).

5. Only 19 percent of the staff rated the World Bank as highly effective or effective in helping countries adopt data innovations compared with 41 percent of country stakeholders. The staff's rating is higher than this on support for development of national statistical strategies—30 percent rated the World Bank as highly effective or effective compared with 41 percent of country stakeholders (table C.5). Respondents are more likely to agree that the World Bank puts its own data needs above those of its country clients than to agree that it gives priority to country needs (table C.6).

6. Asked to choose areas in which the World Bank's work needs strengthening, a larger proportion of respondents in each of the two survey groups (World Bank staff and country stakeholders) chose the category of prioritizing the use of development data in country-level policy dialogue. (table C.7).

7. Asked to reflect on World Bank priorities going forward, 59 percent of country stakeholders included "supporting countries in the production of development data" in their top five areas of strategic thrust, a higher proportion than for any other area. Fifty-three percent of World Bank staff chose support for country-level data production among their five preferred areas (table C.8).

8. Regarding support for data in client countries, country stakeholders rate the World Bank's effectiveness in production and sharing higher than its support for data use—this is the opposite for World Bank staff respondents, who rated the World Bank higher in data use than data sharing (table C.3).

9. Only 27 percent of World Bank staff rated the World Bank as highly effective or effective in promoting data use compared with 45 percent of country stakeholders (table C.3). Among the country stakeholder subgroups, the proportion of respondents rating the World Bank as highly effective or effective on data use ranges from 28 percent of the respondents in the donor agencies category to more than 50 percent in the private sector and government categories (table C.4).

10. Respondents gave a low rating to the World Bank's record in creating in-country demand for data. Only 27 percent of World Bank staff and 40 percent of country stakeholders rated the World Bank as highly effective or effective on this dimension (table C.5). Despite this low effectiveness rating, none of the three groups surveyed included "generating country-level demand for data" among their top choice of areas where, going forward, the World Bank needed strengthening.

11. Regarding support for global data innovations (including big data), 47 percent of the staff rated the World Bank as highly effective or effective compared with 45 percent of country stakeholders (table C.1). Among the country stakeholder subgroups, 53 percent of government respondents rated the World Bank as highly effective or effective in supporting global data innovations, followed by 44 percent of donor agencies, 43 percent of civil society, 42 percent of academics, and 35 percent of the private sector (table C.2).

12. Asked to reflect on the World Bank's priorities going forward, 54 percent of staff included '"making key data sets available globally'" in their top five areas of strategic thrust, a higher proportion than for any other area, while slightly less than 50 percent of country stakeholders chose "global availability of data sets" among their five preferred areas, giving more preference to "supporting countries in production of development data." Respondents gave a relatively low priority to global data innovation—only 32 percent of World Bank staff and 37 percent of country stakeholders included this among their five preferred areas for World Bank emphasis (table C.8).

13. The surveys paint a mixed picture of the priorities for future World Bank interventions on development data. Country stakeholders believe that the area in most need of strengthening is the World Bank's ability to mobilize funding for development data. World Bank staff gave high priority to both funding and inclusion of development data in the country-level policy dialogue, attached less importance to strengthening the understanding of in-country political economy issues, and gave even lower priority to generating country-level demand for development data (table C.7).

Appendix C
Survey Data Findings

## Questions about the Global Level

### Table C.1. Comparison of Responses across Two Survey Instruments

| How effective has the World Bank been in supporting development data at the global level in the following areas: Percent replying "highly effective" plus percent replying "effective" | World Bank Staff (%) N = 655 | Country Stakeholders (%) N = 496 |
|---|---|---|
| Making key data sets available globally | 66 | 67 |
| Developing standards and protocols to ensure global data quality | 46 | 65 |
| Supporting global data innovations such as open data, systems for big data, or use of tablets/phones for surveys | 47 | 45 |
| Bringing development partners and governments together to discuss global data issues | 36 | 46 |

*Source*: IEG's structured surveys of country stakeholders and World Bank staff.
*Note*: N = number of respondents for each survey per question.

### Table C.2. Comparison of Country Stakeholder Responses on World Bank's Support to Global Data (by Respondent Type)

| How effective has the World Bank been in supporting development data at the global level in the following areas: Percent replying "highly effective" plus percent replying "effective" | Country Stakeholders (percent) | | | | | |
|---|---|---|---|---|---|---|
| | Academia | Civil society | Donor agency | Government | Private sector | Other |
| Making key data sets available globally | 68 | 68 | 58 | 71 | 62 | 59 |
| Developing standards and protocols to ensure global data quality | 49 | 58 | 35 | 58 | 52 | 52 |
| Supporting global data innovations such as open data, systems for big data, or use of tablets/phones for surveys | 42 | 43 | 44 | 53 | 35 | 31 |
| Bringing development partners and governments together to discuss global data issues | 41 | 51 | 16 | 57 | 37 | 38 |

*Source*: IEG's structured surveys of country stakeholders and World Bank staff.

## Questions about the National Level

### Table C.3. Comparative Data among Two Survey Instruments on Support to Countries

| How effective has the World Bank been in supporting countries in the following areas: Percent replying "highly effective" plus percent replying "effective" | World Bank Staff (%) N = 644 | Country Stakeholders (%) N = 501 |
|---|---|---|
| Production of development data | 36 | 53 |
| Sharing of development data | 23 | 49 |
| Use of development data | 27 | 45 |

*Source*: IEG's structured surveys of country stakeholders and World Bank staff.
*Note*: N = number of respondents for each survey per question.

### Table C.4. Cross-Tabulated Data on Effectiveness of World Bank Support to Countries (by Stakeholder Group)

| How effective has the World Bank been in supporting countries in the following areas: Percent replying "highly effective" and "effective" | Country Stakeholders (percent) | | | | | |
|---|---|---|---|---|---|---|
| | Academia | Civil Society | Donor agency | Government | Private sector | Other |
| Production of development data | 46 | 50 | 49 | 61 | 56 | 37 |
| Sharing of development data | 44 | 42 | 44 | 58 | 61 | 30 |
| Use of development data | 41 | 43 | 28 | 53 | 52 | 30 |

*Source*: IEG's structured surveys of country stakeholders and World Bank staff.
*Note*: N = number of respondents for each survey per question.

## Table C.5.

| How effective has the World Bank been at the country level in the following areas: Percent replying "highly effective" plus percent replying "effective" | World Bank Staff N = 646 | Country Stakeholders = 498 | N |
|---|---|---|---|
| Creating in-country demand for data | 27 | 40 | |
| Helping client countries to adopt data innovations | 19 | 41 | |
| Supporting the development of national statistical strategies | 30 | 41 | |

*Source*: IEG's structured surveys of country stakeholders and World Bank staff.
*Note*: N = number of respondents for each survey per question.

## Table C.6.

| Which of the following statements do you agree with regarding the World Bank's past priorities? (Select only one option) | World Bank Staff (%) N = 657 | Country Stakeholders (%) N = 506 |
|---|---|---|
| The World Bank has prioritized the data needs of its country clients over its own data needs | 7 | 7 |
| The World Bank has prioritized its own data needs over the data needs of its country clients | 36 | 31 |
| The World Bank has considered its own data needs and the data needs of its country clients equally | 23 | 40 |

*Source*: IEG's structured surveys of country stakeholders and World Bank staff.

## Questions about Priorities for the Future

### Table C.7.

| For the World Bank to achieve optimal results in its support of development data going forward, which of the following areas should it strengthen? (Select up to three) | World Bank Staff N = 657 | Country Stakeholders N = 497 |
|---|---|---|
| The World Bank's ability to mobilize funding for development data | 50 | 64 |
| The quality of the World Bank's technical knowledge on data issues | 39 | 36 |
| The priority that the World Bank gives to development data at the global level | 26 | 24 |
| The priority that the World Bank gives to development data in its country-level policy dialogue | 55 | 54 |
| The World Bank's understanding of in-country political economy issues surrounding development data | 36 | 47 |
| The World Bank's focus on generating country-level demand for development data | 36 | 42 |

*Source*: IEG's structured surveys of country stakeholders and World Bank staff.
*Note*: N = number of respondents for each survey per question.

### Table C.8. Comparative Data on World Bank's Strategic Thrust Going Forward

| Which of the following areas should be the strategic thrust of the World Bank's support for development data going forward? (Select up to 5) Percent of each group that included each area among its five choices. | World Bank Staff (percent) N=657 | Country Stakeholders (percent) N=499 |
|---|---|---|
| Making key data sets available globally | 54 | 49 |
| Developing standards and protocols to ensure global data quality | 50 | 43 |
| Bringing development partners and governments together to discuss global data issues | 31 | 41 |

| **Which of the following areas should be the strategic thrust of the World Bank's support for development data going forward? (Select up to 5) Percent of each group that included each area among its five choices.** | **World Bank Staff (percent) N=657** | **Country Stakeholders (percent) N=499** |
|---|---|---|
| Supporting global data innovations such as open data, systems for big data, or use of tablets/phones for surveys | 32 | 37 |
| Supporting countries in the production of development data | 51 | 58 |
| Supporting countries in the sharing of development data | 28 | 28 |
| Supporting countries in the use of development data | 40 | 36 |
| Supporting in-country capacity development over the longer term for data production | 46 | 50 |
| Supporting in-country capacity development over the longer term for data sharing | 20 | 21 |
| Supporting in-country capacity development over the longer term for data use | 31 | 27 |
| Creating in-country demand for data | 20 | 17 |
| Helping client countries to adopt data innovations | 19 | 21 |
| Supporting the development of national statistical strategies | 16 | 26 |

*Source*: IEG's structured surveys of country stakeholders and World Bank staff.


**WORLD BANK GROUP**
World Bank • IFC • MIGA

The World Bank
1818 H Street NW
Washington, DC 20433