



USAID
FROM THE AMERICAN PEOPLE



EXECUTIVE SUMMARY

Avoiding Data Graveyards:

Insights from Data Producers & Users in Three Countries

Editors

Samantha Custer and Tanya Sethi



Executive Summary

Government, development partner, and civil society leaders make decisions every day about how to allocate, monitor and evaluate development assistance. Which communities should receive scarce resources? Which programs generate the highest returns on investment and should be flagged for expansion? Are development projects being delivered on time and achieving their desired results?

Policymakers and practitioners can theoretically draw from more data sources in a variety of formats than ever before to answer these questions. But will they choose to do so? Those who collect data and produce evidence are often far removed from those who ultimately influence and make decisions. Technocratic ideals of evidence-informed policymaking and data-driven decision-making are easily undercut by individual prerogatives, organizational imperatives, and ecosystem-wide blind spots.

In other words, the data revolution for sustainable development could succeed in building a supply of better data, but may falter if there is insufficient demand for its use. In writing this report, the authors shed light on how leaders make decisions about development assistance in three countries and the role of data and evidence in those choices.

In 2016, researchers from the AidData Center for Development Policy interviewed nearly 200 decision-makers and those that advise them in Honduras, Timor-Leste, and Senegal. Central government officials, development partner representatives based in country, and leaders of civil society organizations (CSOs) shared their experiences in producing and using data to target development projects, monitor progress, and evaluate results.

Specifically, the report answers three questions:

- Who produces development data and statistics, for what purposes, and for whom?
- What are the technical and political constraints for decision-makers to use development data in their work?

- What can funders and producers do differently to encourage use of data and evidence in decision-making?

In the remainder of this brief, we summarize the main findings from the report using a theory of change adapted from a Custer et al. (2016) study, From Pork to Performance: Open Government and Program Performance Tracking in the Philippines. The theory of change presents the causal logic of getting from data to impact (e.g., improved development outcomes) as the interaction of four C's: content, channel, choice, and consequence. In this study, we exclusively look at the first three C's as building blocks for data use. On this basis, we identify nine barriers and corresponding operating principles for funders and producers to make demand-driven investments in the next generation of development data and statistics.

Theory of Change

Governments and organizations disclose data on development resources and results (content), disseminate this information to users online or offline (channel), whereby citizens and officials take action individually or collectively (choice) to improve the country's performance on achieving sustainable development for all (consequences).

Content:

Is development data fit-for-purpose?

Government officials, development partner representatives, and CSO leaders across the three countries reported that development data is often: (1) insufficiently granular; (2) of questionable accuracy; and (3) fragmented across numerous organizational and agency silos. We discuss each barrier below and identify principles for funders and producers to overcome these content constraints to encourage greater uptake of development data and statistics.

Granularity

Barrier: *Decision-makers want more data that is disaggregated by sector, geography, and demography*

There was high demand for more data on development infrastructure (e.g., clinics, schools) and results (e.g., student performance, unemployment figures) that is disaggregated by district (or relevant subnational unit). When asked about high-value data sources, interviewees most frequently mentioned geo-referenced and sector-specific administrative data produced by line ministries, and surveys and censuses from national statistical offices. Yet, demand for this information outstrips the capacity of the three countries to produce it consistently. The production of highly granular sector-specific data relies on providers at the point of service delivery, who often have constrained data skills to report information accurately, and limited visibility on its value if they do.

Principle: *Funders should invest in regular collection and disclosure of geo-referenced survey / census data*

Funders wanting to make demand-driven investments in data and statistics should invest in the capacity of national statistics organizations and other actors to conduct censuses and surveys in sectors such as education, health, and agriculture with greater frequency. These investments will generate real dividends only if producers geo-reference and disaggregate census and survey data by key demographic attributes (e.g., age, sex, disability

status). Funders and official data producers must also expand coverage of censuses and household surveys to include under-represented groups.

Accuracy

Barrier: *Data sources are viewed as incomplete or out of date for both technical and political reasons*

Concerns regarding the timeliness, accuracy, and validity of government-produced data substantially dampen broader use. While most often associated with government data, these challenges extend beyond the public sector. For example, development partners often do not allocate sufficient attention to reporting on their activities in a timely manner. Interviewees cited several technical and political reasons that compound the risk of inaccurate data. These include publication delays, episodic data collection, limited data management capacity at the subnational level, and gaps in coverage. Front-line service providers are often under-resourced, under-trained, and under-motivated to submit complete and accurate reports. Interviewees also noted that powerful incentives exist for public servants to make the official numbers look good.

Principle: *Producers should engage citizens, companies, and CSOs as data co-creators and validators*

While data producers understandably face trade-offs in terms of timeliness versus accuracy, the sole burden of verifying the accuracy of data need not fall on the producer alone. Government agencies could proactively enlist the support of citizens, civil society groups, and development partners to help catch errors or augment official data sources. This would require government producers to put a spotlight on potential limitations, gaps, and deficiencies in the data. Since the world of data is not immune to politics, this recommendation is likely most feasible in contexts where high-level government champions are able to create an authorizing environment for agencies to openly share data and collaborate with end users. Without this political cover, inertia will be difficult to overcome.

Integration

Barrier: *Disconnected initiatives make integration, interoperability, and triangulation of data sources difficult*

Decision-makers need to see the full service delivery chain from upstream budgets and intermediate outputs to downstream outcomes. Nonetheless, fragmentation is often the norm: disconnected data points are siloed between disparate information management systems maintained by different agencies and collected using non-standardized methods. These challenges extend beyond the government. Development partners and CSOs expend substantial resources to produce development data that meets their specific needs. These independent data collection exercises often use different methodologies and generate datasets that may not be publicly available, making comparisons difficult. Several interviewees noted the substantial risk of ad hoc data collection efforts that are seldom coordinated with each other, even within departments of the same organization.

Principle: *Funders should align incentives of data producers in favor of integration and interoperability*

Governments and their development partners should tie financing with the need to demonstrate greater integration, interoperability, and openness in data collection, management, and publication practices. Funders should make it more difficult for government agencies and organizations to undertake duplicative data collection activities or maintain siloed data management systems. They should also align their funding with national statistical development strategies. This requires funders to take the long view, eschewing one-off data extraction exercises that would allow them to quickly get data for their own purposes and instead investing with an eye towards sustainability, so that domestic actors have the capacity to produce and share data that meet their needs.

Channel:

Can prospective users easily find, access, and use the information they need?

Interviewees across the three countries reported that dissemination channels miss the mark when: (1) there is uneven access to information; (2) different levels of connectivity exacerbate a digital divide; and (3) data is invisible such that people will not find it and use it. We discuss each barrier below and identify principles for funders and producers to overcome these channel constraints and encourage development data use.

Access

Barrier: *Absent Freedom of Information (FOI) legislation, access depends upon tenuous voluntary disclosure*

In countries without FOI laws, such as Timor-Leste and Senegal, visibility on public sector activities depends on the government's willingness to open up this data, or on the strength of one's network to secure it via back channels. This reliance on voluntary disclosure creates uneven access that disproportionately affects less well-connected groups and individuals. Even in countries like Honduras that have adopted FOI legislation, access to development data is only as strong as the government's commitment to enforce compliance. Development partners and CSOs conduct independent data collection exercises and maintain proprietary databases that have valuable information on development projects and indicators, but which typically do not fall under the jurisdiction of FOI law. Government officials sometimes view these groups as bypassing them to collect information without sharing their results.

Principle: *Producers should open data by default, and funders and advocates should help make access sustainable*

Funders of data and statistics capacity should make their investments contingent upon producer willingness to adhere to open data principles and ensure sustainable access to that information through supporting FOI legislation. To address real concerns

regarding the costs of data collection and the risks of increased public scrutiny, funders could help producers develop sustainable business models and data management practices that cover costs and reduce their liability. Meanwhile, funders of transparency and open data initiatives could incentivize producers to disclose datasets that are in highest demand and advise them on implementing open data principles in practice.

Connectivity

Barrier: *Delivering data exclusively online hampers access due to connectivity and capacity constraints*

Limited Internet connectivity, language barriers, and low levels of data literacy and numeracy inhibit practical access to, and use of, development data. In multilingual Timor-Leste, for example, data is often available online in the official languages of Portuguese and/or English, but not in Tetum, the lingua franca of most citizens. Meanwhile, domestic policymakers and practitioners are predisposed to favor an oral culture of information sharing and are far less comfortable using complex online databases. Connectivity and capacity constraints in utilizing online sources of development data are not unique to Timor-Leste. If data producers rely exclusively on online dissemination strategies, this can inadvertently privilege access to information among the urban, educated, and well-to-do (who tend to be more digitally savvy) at the expense of other groups.

Principle: *Producers should partner with infomediaries to promote data and identify actionable insights*

To ensure their data products are being used, producers must (1) embrace a mandate beyond collecting, producing and publishing data, that includes translating raw data into actionable insights, and (2) craft an intentional dissemination strategy. Media, civil society, and think tanks, among others, could play a complementary role in communicating data via various online and offline channels. In brokering partnerships with these infomediaries producers could adopt a hybrid dissemination strategy recommended by Gigler et al. (2014) that blends “high-tech” (Internet, social media, mobile phones) and “low-tech” (e.g., radio, print, community meetings) communications.

Visibility

Barrier: *Prospective data users have low familiarity with many publicly available information sources*

Even when data is available and accessible, it may not be visible enough to prospective users for them to take advantage of this information in their work. Nonetheless, government, development partner, and CSO producers seldom have an intentional strategy to proactively communicate and promote the development data they publish via datasets, databases, and data portals. For example, only a third of government and development partner representatives interviewed in Honduras, and even fewer CSOs, were aware of publicly available sources of aid information. People will not use development data that they do not know exists, increasing the risk of data graveyards.

Principle: *Training can increase visibility, but funders and producers must systematically test what works*

If generic launch events and media coverage allow producers to go broad in advertising their data with a general audience, in theory, customized skills trainings and focused consultations are an opportunity to go deep with key user groups. AidData and its consortium partners are conducting field experiments in Timor-Leste and Honduras to assess whether and how participation in a training event changes usage patterns of the aid information management system among domestic policymakers and practitioners. Funders, data producers, and researchers should build upon this evidence base to test how users respond to different formulations of data training and statistical capacity programs to improve the returns on their investments.

Choice:

Will the perceived benefits of using data outweigh the costs?

Interviewees highlighted that they are less likely to use development data when: (1) information sources are seen as lacking credibility; (2) they can more easily access this information via informal networks; and (3) the perceived costs outweigh the benefits of making decisions based upon evidence rather than other

factors. We discuss each barrier below and identify principles for funders and producers to overcome choice constraints and encourage uptake of development data.

Credibility

Barrier: *Government data sources suffer from a trust deficit, and are perceived as out-of-date and unreliable*

Some interviewees were skeptical that government agencies have sufficient data management capacity to validate information and safeguard against inaccuracies, both inadvertent (e.g., mistakes in data entry) and intentional (e.g., tampering to skew reporting). Distrust in data could also be symptomatic of low public trust in government more broadly and political concerns regarding perverse incentives for officials to use data as propaganda. Related research by Masaki et al. (2016) identifies perceived credibility (or lack thereof) as a determinant of whether leaders in 126 countries used governance data. Prospective users care greatly about the reputation of a data source and the trustworthiness of the data producer in determining whether or not to put this information to use.

Principle: *Producers should be more transparent in their methods and quality assurance procedures*

Opaque methods for collection and validation do little to alleviate concerns that producers are manipulating data behind the scenes for their own political ends. Increasing public trust in development data will require not only technical fixes to improve the accuracy of the data, but also a political commitment to greater openness and transparency about what is being measured, how data is collected, and what procedures protect this information from tampering. If credibility is indeed a major barrier to use, data producers, both official and unofficial, need to become more transparent in documenting their assumptions, methods, and processes for managing data. Funders of data initiatives and senior leaders in governments and organizations should reward and reinforce the efforts of producers to become more transparent. In doing so, they will help create the conditions for more robust data use.

Networks

Barrier: *Many leaders source information from those they know from habit and lack of viable alternatives*

Even when official sources are available, interviewees sometimes prefer to source information from their professional and personal contacts, rather than proprietary or publicly available databases. This creates asymmetries: more senior individuals interviewed reported an easier time sourcing data from those in their networks, as compared with less well-connected junior staff. Development partners and government officials generally were more optimistic about their ability to access data than were their counterparts in civil society organizations. One driver of this preference for networked intelligence is a strong demand for qualitative lessons learned and more up-to-date information than is often available in databases maintained by official data producers.

Principle: *Data producers must seek to enhance, not replace informal knowledge sharing via networks*

Leaders have a natural tendency to source information from those they know due to the confluence of habit, preference, and lack of perceived alternatives. Data producers should support these robust human information networks, not replace or compete with them. They could proactively enlist the support of a sector working group or an association of CSOs to adopt centralized (and open) data systems in a mutually beneficial alliance. Producers could collaborate with network members as third-party validators to ensure that their databases are capturing up-to-date information or to give their seal of approval for the data's accuracy. As official producers grow in their data management capacity, they could provide critical assistance to networks with new tools, methods, and approaches to organizing qualitative insights that are often unstructured, subjective, and context-specific.

Incentives

Barrier: *Data is only one part of the decision-making calculus and is often crowded out by other factors*

Some interviewees reported using data in the context of siting new projects (e.g., feasibility studies) or assessing project results at closure. However, even at

these stages of the project cycle, organizations and government agencies may select project locations based on ease of access, available infrastructure, and existing networks. Political pressures or the desire to maintain good working relationships can also undercut willingness to frankly discuss (or publicize) lackluster results in implementation, thereby impeding organizational learning and corrective action. Societal norms and individual biases, such as an unfavorable view of international experts, can also decrease the perceived value of drawing upon data and evidence throughout policymaking processes or project cycles.

Principle: *Funders and producers should “work with the grain” to align incentives for data-driven decisions (Levy, 2014)*

If funders and producers of development data want to create the conditions for evidence-informed policymaking, they must crowd in, rather than short-circuit the interest of political actors in favor of using data as they allocate resources, target projects, and evaluate the performance of development programs. Monetary and non-monetary rewards should be employed to increase the perceived benefits for policymakers and practitioners to use data. Funders and producers could work with senior leaders in government agencies and organizations to tie performance bonuses or merit increases to behaviors that exhibit effective use of data to inform decision-making. Senior leaders could also prioritize access to training, recognition, and advancement opportunities for these decision-makers.

Citation:

Custer, S. and T. Sethi. (Eds.) (2017) *Avoiding Data Graveyards: Insights from Data Producers and Users in Three Countries. Executive Summary*. Williamsburg, VA: AidData at the College of William & Mary.

References:

Custer, S., Rahemtulla, H., Kaiser, K., & van den Brink, R. (2016). *From Pork to Performance: Open Government and Program Performance Tracking in the Philippines*. World Bank and AidData. July 2016. Retrieved from: <http://documents.worldbank.org/curated/en/174491467602317699/Pork-to-performance-open-government-and-program-performance-tracking-in-the-Philippines-phase-two>

Gigler, B.S., Custer, S., Bailur, S., Dodds, E., Asad, S., & Gagieva-Petrova, E. (2014). *Closing the Feedback Loop: Can Technology Amplify Citizen Voices?* In B.-S. Gigler, and S. Bailur (Eds.), *Closing the Feedback Loop: Can Technology Bridge the Accountability Gap?* World Bank.

Levy, B. (2014). *Working with the Grain: Integrating Governance and Growth in Development Strategies*. Oxford University Press: New York, NY. Kindle edition.

Masaki, T., Sethi, T., & Custer, S. (2016). *In the Eye of the Beholder: When is governance data “good enough”?* Williamsburg, VA. AidData at the College of William & Mary and the Governance Data Alliance.

Funded by USAID's Global Development Lab, the **AidData Center for Development Policy** provides geospatial data and tools that enable the global development community to more effectively target, coordinate, and evaluate aid. The Center is a consortium of five partners: the College of William and Mary, Development Gateway, Brigham Young University, the University of Texas at Austin, and Esri.

