

The Problems of Government Analysts in the Developing World*

Chiara Casanova, Towera Moyo, Daniel Rogger

Flavia Sacco Capurro and Christian Schuster[†]

June 29, 2026

Abstract

A substantial chunk of analysis on the effectiveness of government and public policy in developing countries is undertaken inside public administrations by government analysts in public institutions. Yet little is understood about the nature of this analysis nor the challenges to its effective undertaking. This paper reports on a global call for analytical projects and a related survey that sketches out the analytical landscape within governments across low- and lower-middle-income countries. We document the questions government analysts seek to answer, the data and methods they employ, and the constraints they face in producing evidence for decision-making. Government analysts address a wide range of operational and strategic challenges, often relying on administrative data and applied empirical methods. However, analytical work is constrained by limitations in data access and quality, insufficient time and staffing, fragmented institutional incentives, and weak mechanisms for translating findings into policy action. Building an improved ecosystem for analytics inside government would strengthen the evidence base on which public policy is defined and implemented.

Keywords: bureaucrats; public administration; economic analysis

*We gratefully acknowledge financial support from the World Bank's Development Impact Fund, Bureaucracy Lab and Governance Global Practice. We are grateful to Aidan Coville for helpful comments. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

[†]Casanova, Moyo, Rogger (corresponding author) and Sacco Capurro: The World Bank, 1818 H St NW, Washington DC, USA; Schuster: University of Oxford

1 Introduction

While the central role of bureaucracy in shaping development outcomes is increasingly well recognized [Acemoglu and Robinson, 2012; Besley et al., 2022], surprisingly little is known about the analytical work conducted within bureaucracies themselves. A fast-growing literature on bureaucracy and development looks at the role of governments as consumers of evidence Cairney [2016]; Parkhurst [2017]; Hjort et al. [2021]. This literature has, among others, underscored the importance of evidence-based policy-making for development outcomes [Banerjee and Duflo, 2011; Vivalt, Coville and KC, 2025]. However, little is known about the role of governments as producers rather than consumers of evidence to inform policymaking. Yet, a substantial share of the analysis that informs policy in developing countries is produced not by external experts, but by government analysts working inside public institutions.

Consider the following: there are approximately 500,000 local governments in the world, 183 national governments with myriad sectors, and various regional entities. Each of these organisations tends to have, at a minimum, budget analysts assessing spending of organisations relative to budget allocations. By implication, the number of government analysts conducting research and analytics in-house dwarves the number of academic researchers studying the same topics by a large margin. Adding any other analytics undertaken in government multiplies the estimate of the number of analysts in government further.

Analysis undertaken within public administration is, perhaps unsurprisingly, the main source of insights for decision-making in public administration [Weiss, 1979, 1980; Armstrong et al., 2014; Newman, Cherney and Head, 2017; Koga et al., 2022; Cantarelli, Belardinelli and Bellé, 2023]. Rogger and Somani [2023] interview Ethiopian public servants on the sources of their information for decision-making and find that only 2.12% of officials selected external sources/media as their top source of information. Similarly, Koga et al. [2024] summarize their study of evidence-use in Brazilian public administrations by arguing that, “scientific knowledge is only used to support internal sources (e.g., reports, legal or control recommendations, etc.) and sporadically used by bureaucrats in general”. AidData’s ‘Decoding Data Use’ project [Masaki et al., 2017] interviewed public officials from 126 low- and middle-income countries and find that the two most important features of evidence is that it “reflects the local context”, and “draws upon data or analysis produced by the government”.¹

While the production of academic research and donor-funded studies on development are extensively scrutinized and debated, little attention has been paid to the analytical work done by public servants inside government — the body of analysis on which many real-time public policy decisions are made. The canonical view of developing country public administrations - reflected in large-scale investments by development institu-

¹The reasons for this lack of use of external sources of analytics is poorly understood, but Vivalt, Coville and KC [2025] highlight the preference public officials have for ‘locally-generated’ knowledge.

tions - is that government analysts in lower-income countries lack the human capital, statistical training and technical skills to conduct sophisticated analysis [OECD, 2020; Mejía-Guerra et al., 2023]. There is little systematic evidence supporting the notion that analytics are, in fact, scarce in public administration or that skills gaps are causing this scarcity.

This paper addresses this gap through a survey of almost 5,000 government analysts from countries in the Global South seeking World Bank support for analytics projects. The survey included submissions of government analytics projects participants were working on. The survey underscores, in the first place, that respondents employ a rich array of methods in their analytics projects — including machine learning, natural language processing, geospatial analysis, and predictive modeling — which challenge notions that sophisticated analytics are precluded by human capital shortcomings inside developing country governments. The survey also provides evidence that government analysts are working with internal records that can directly inform program design and administrative functioning, underscoring that government analysts can often undertake analytical work that is more tightly linked to local problems than external researchers.

Survey findings also underscore that binding constraints for better analytics in government are as much organizational as related to human capital: inter-agency cooperation, data sharing, and funding for analytical infrastructure rank as the dominant reported barriers. While the survey underscores there is space for strengthening the quality and skills of government analysts, their existing skills are not being catalyzed as they could be. Yet, research has largely neglected the question of how to improve the organizational ecosystem for research and analytics inside government by government analysts.

The paper continues as follows: Section 2 overviews the data collection approach; Section 3 outlines the analytical projects that public servants report working on in their submissions to the World Bank call; Section 4 covers the core barriers they report to being able to undertake analytical work more broadly; and, Section 5 concludes.

2 Data

The survey data on which this paper is based was collected as part of a World Bank call to public servants for the provision of assistance with ‘government analytics’, or analytical work related to the functioning of government and related policy. This call was framed as offering public servants support to undertake this analysis through a ‘government analytics fellowship’ that included both training on related analytical methods in Washington DC and a six-month series of online mentoring and learning sessions that aimed to assist the officials in completing their analysis.

The call was targeted to client countries of the World Bank, primarily low- and middle-

income countries eligible for support through the International Development Association (IDA) and the International Bank for Reconstruction and Development (IBRD). While the call was focused on central government administrative entities, there was no restriction on applications being received from any tier of government. Eligible applicants were expected to hold mid-level management or senior analyst positions (or equivalent), with sufficient authority to design and implement a government analytics project.²

The call received 2244 responses in 2024 and 2606 responses in 2025. We will analyze the data from both rounds of the call jointly since the results from the two rounds are qualitatively similar. Applications were received from around the world, with a strong representation from African and South Asian countries. However, we restrict our analysis in all sections to those countries classified by the World Bank as low or lower-middle income, which represent 85% of the total pool of survey responses, dropping the 421 responses from those countries classified as upper middle or high income. This enables our analysis to focus squarely on the challenges of analysts in the developing world according to standard classifications. The distribution of remaining applicants that we focus our analysis on is described in Figure 1.

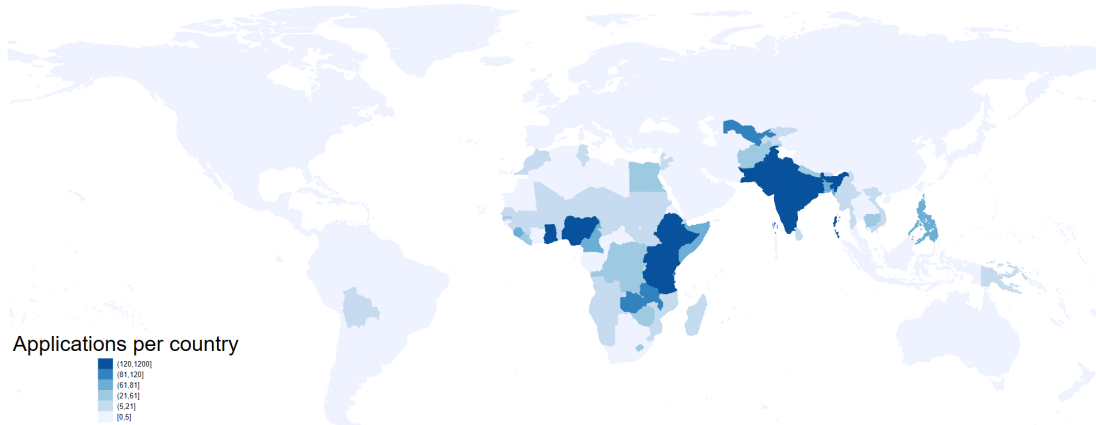


Figure 1: Total applicants per country.

To characterize the corpus of projects submitted by government analysts through this process, we utilized the BERTopic topic modeling technique to systematically organize and group project descriptions into coherent themes.³ This methodology facilitated

²In reporting how they heard about the fellowship, applicants responded that the top channels were social media (LinkedIn, Twitter/X, etc.; 38%), colleagues or word of mouth (36%), the World Bank website or mailing list (36%), and internal channels like government mailing lists, etc. (19%). Numbers add up to more than 100% as respondents could tick more than one option.

³BERTopic is a topic modeling technique that uses document embeddings to create dense clusters of semantically similar texts and generates interpretable topic descriptions by selecting the most representative words for each cluster. More details are available at <https://maartengr.github.io/BERTopic/index.html>.

the identification of the primary areas of analytical work represented within the corpus and allowed for systematic summarization for subsequent analysis. Furthermore, we used the GPT-5 model through the OpenAI API in an automated prompt engineering pipeline to extract structured metadata from the themes generated by BERTopic, specifically focusing on topic labels, methodologies, data types, and concise summaries to support more in-depth analysis. This approach ensured a consistent and reproducible procedure for acquiring this additional information.⁴

To better understand the administrative problems public servants face in implementing their analytical projects, we administered a follow-up survey to all applicants. The instrument aimed to capture individual-level challenges in daily tasks, as well as broader organizational barriers to the use of data and evidence in decision making. The survey covered aspects such as operational and institutional constraints to task completion, the presence and strength of an analytical culture inside organizations, and perceived organizational capacity for government analytics. 58% of total applicants responded to this survey, providing insights into both micro-level obstacles and macro-level factors that shape the feasibility and impact of analytical work in the public sector.

The sample of analysts on which the paper is based are a self-selected group applying to a program to support public servants to undertake analytical projects.⁵ As such, the sample is unlikely to be representative of all government employees, or even all government analysts. Rather, it captures a self-selected group with unusually high levels of motivation, interest in evidence, and willingness to invest effort in analytical work. While this limits the external validity of estimates intended to characterize the average public servant, it makes the sample particularly informative for understanding the role of analysis within government. These individuals are likely to constitute the leading edge of analytical capacity in their respective institutions: they are the officials most actively seeking to use data and evidence to solve policy and operational problems.

Development outcomes are often disproportionately influenced by a relatively small number of entrepreneurial bureaucrats who initiate reforms, champion evidence use, and introduce new analytical practices within government [Mintrom, 1997; Teodoro, 2011; Andrews, Pritchett and Woolcock, 2017]. Understanding the motivations, activities, and constraints of this group is therefore of significant interest. Rather than describing the average government employee, the sample sheds light on the subset of officials most likely to drive evidence-informed policymaking and organizational learning.⁶

Our evidence is also largely self-reported and perception-based; measures such as the

⁴Further details on the processing of the underlying application data is provided in the appendix.

⁵Relatively few of our sample of respondents are from the same organization, implying we can abstract from clustering concerns in terms of projects or problems faced in undertaking analytical work.

⁶We may not be picking up the full extent of defense-related analyses given security concerns, despite the sector being a major component of public expenditure in many developing countries.

share of colleagues using data, or the salience of particular barriers. These may be shaped by social desirability and by respondents' vantage point within their organisation, and we lack objective or administrative measures against which to validate them. Relatedly, we cannot identify the causal effect of any given constraint on analytical output, nor link analytical activity to downstream policy or development outcomes. We see these limitations less as threats to the central claims than as markers of where longitudinal, and outcome-linked data on government analysts would be most valuable. We hope that future research will take up this challenge.

3 Data Analyses by Public Servants in the Developing World

Our survey results provide insights into both the extent and nature of data analytics inside public sector organisations - or what the 'analytical problems' of government analysts in the developing world are. How widespread is data analytics? Which topics are addressed? And which data sources and methods are used?

3.1 How widespread is data analytics practice in public sector organisations in developing countries?

To capture how widespread actual analytics practice is, we asked respondents whether "My organization uses data and evidence to achieve its goals and targets". 74% of respondents agree with this statement, indicating relatively broad use of analytics in the public sectors of developing countries as a whole. Focusing on the proportion of individuals engaged in bridging analytics and policy, we also asked, "What percentage of your colleagues within your organization would you say have used government data to improve the functioning of your administration?" The respondents estimated that, on average, 40% of the colleagues within their organization have used government data to improve the functioning of their administration. Thus, three-quarters of the organizations represented in our sample are generally undertaking analytics for improved government functioning, with four of ten staff engaged in that endeavor at some point in their work.

Responses varied widely across countries. Figure 2 presents averages of responses to this question for those countries that we received applications from stacked vertically around the corresponding regional average as grey dots. We see that most regions have a relatively high degree of variation, with standout cases both on the lower and upper ends of the distribution. A natural benchmark is the United States Federal Government, arguably the most advanced public service in the world. How do these regional and country-level averages from developing countries compare with the extent

of analytics in a frontier government? U.S. Government Accountability Office [2021] provides evidence on the extent of evidence-based policymaking there, with the caveat of a slightly differently phrased survey question. Across federal public service, 72.6% of officials believe that “My agency’s top leadership demonstrates a strong commitment to using performance information to guide decision-making” (responding ‘to a very great’, ‘great’ or ‘moderate’ extent); similar to the regional means seen in Figure 2. Yet, this average masks variation across agencies, with 39% of US officials stating they see no or only a small commitment to analytics across the US Federal Government, corresponding to a distribution of organizational-averages not dissimilar to that observed in our data.

These indicators provide suggestive evidence that government officials in the Global South are no less likely to believe that their administrations are evidence based than at the global frontier. Rather, we observe similar levels of variation in those perceptions across the world.

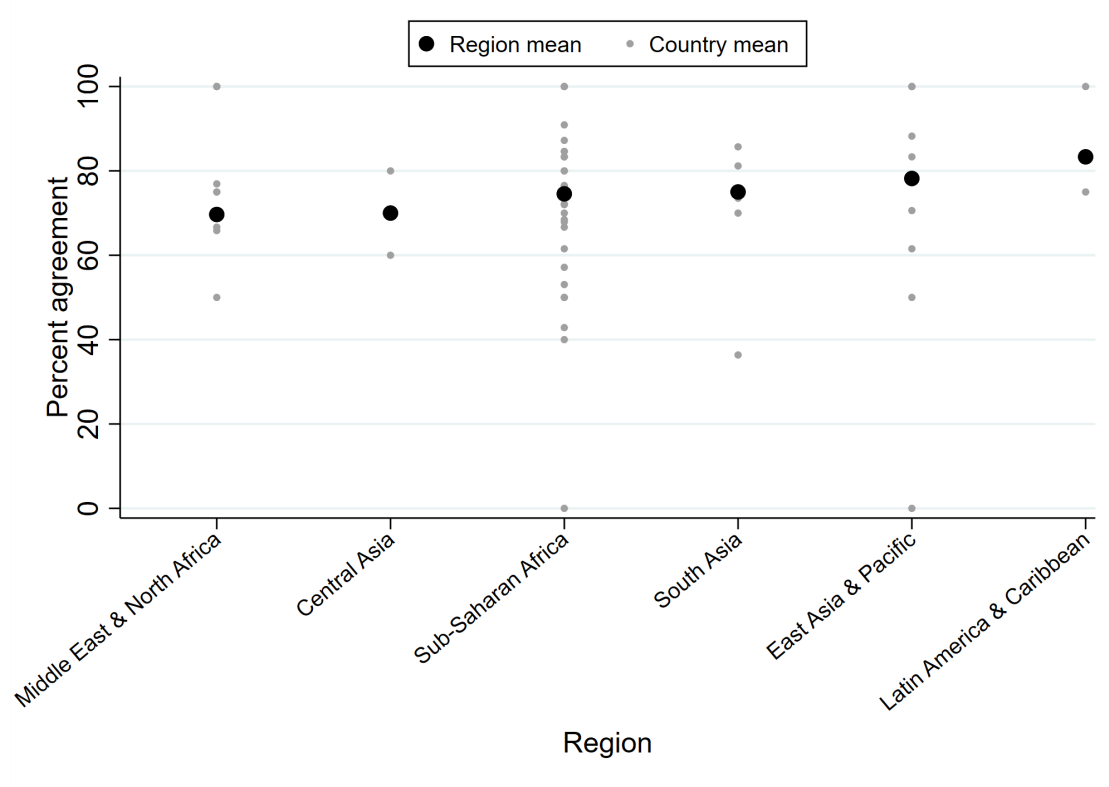


Figure 2: Percent of respondents agreeing with the statement “My organization uses data and evidence to achieve its goals and targets”

3.2 What analytical problems are public servants in the developing world working on?

Figure 3 shows the results of the topic modeling of descriptions of the analytical projects submitted to the government analytics fellowship.⁷ The top panel indicates the top five clusters of analysis. The most prevalent area of analysis, grouped as Topic 1 in Panel A and accounting for 28.6% of projects, relates to data integration and analytics for strengthening health care systems. For example, one project addresses fragmented health financing data in Bangladesh by “developing a unified health resource tracker that distinguishes domestic and donor funding streams across maternal health, tuberculosis, NCDs, and vaccination, and uses forecasting and geographic disparity mapping to identify funding gaps”.

The second largest group of projects, accounting for 22.0%, relates to financial flows, including tax and revenue systems, public finance, auditing, and corruption. One example addresses uncertainty around the performance and future value of Ghana’s e-Levy data infrastructure by “analyzing transaction data from electronic transfers to assess revenue performance and inform decisions about whether to retain the system even if the tax is eventually abolished”.

The third topic, accounting for 13.4% of projects, focuses on trade, export and procurement. An example is a project in Sierra Leone that addresses corruption risks and limited transparency in public procurement by “developing a digital platform that consolidates beneficial ownership disclosures, contract award data, and real-time monitoring tools”.

The fourth topic, accounting for 11.8% of projects, concerns agriculture, water, energy, and climate resilience. For example, one project addresses limited visibility into land degradation and climate vulnerability in Ethiopia by “integrating geospatial data, remote sensing, and predictive modeling to map land-use change, soil erosion risks, and deforestation”.

The fifth topic, accounting for 7.9% of projects, relates to education data systems and analytics for evidence-based policy and resource allocation in the sector. An example is a project in Kenya that addresses fragmented and inconsistent education records across 78 universities and thousands of schools by “centralizing data in a unified system with unique learner identifiers, enabling real-time tracking of enrollment, retention, and resource distribution to support equitable policy decisions.”

The bottom panel of Figure 3 outlines the next most prevalent topics for analysis in governments in the developing world, as reflected in the submitted projects. These

⁷Topics were identified using BERTopic, which groups semantically similar project descriptions using transformer-based embeddings and density-based clustering. Sector classifications were assigned to each topic using GPT-5-mini, based on the concatenated project descriptions within each cluster. Full methodological details are provided in the Technical Appendix. Documents not assigned to a named cluster (n=1,440) are excluded from the analysis.

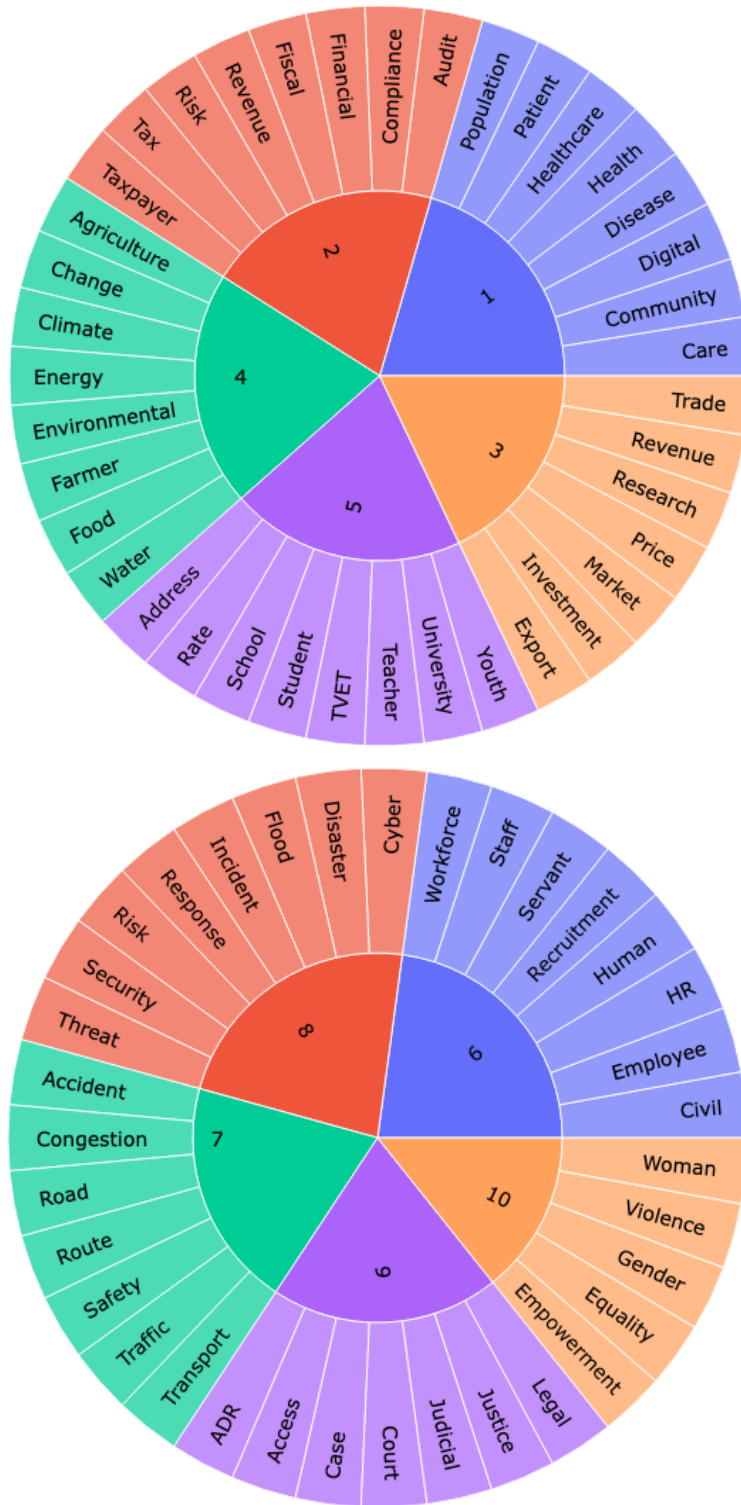


Figure 3: Topics of government analytics projects: first to fifth most prevalent topic groupings (top panel) and sixth to tenth most prevalent topic groupings (bottom panel)

relate to transportation and road safety, crime and disaster management, gender and equality studies, risk management, and public policy relating to businesses. While these ten topics are not an exhaustive list of sectors in which the public sector is frequently involved, the list covers many of the main areas, reflecting the breadth of coverage in the underlying projects.⁸

The identified topics might have a naturally corresponding sector, such as the case of the education and health projects. Yet a greater proportion of projects are system-wide assessments of sector budgets, revenues, or personnel. These correspond to key decision-points of centralized actors - how to generate revenues, distribute budgets, and manage or optimize expenditures. 35.7% of projects target the public administration as a whole. Additionally, 21.4% assess the financial and economic issues of the government as a whole.

Overall, our findings imply that analytics on a broad range of public policy problems are being undertaken by government analysts across the developing world. This is true for both system-wide assessments, as well as sectoral or thematic problems.

3.3 What data do government analysts use to undertake this analysis?

When asked in the survey about the primary sources of data most frequently relied on for analytical work, respondents overwhelmingly cited budget, expenditure, and other financial data, with 64% selecting this category. As elaborated in Figure 4, this was followed by specialized administrative data from within government (54%) and case processing data, task completion or other output data (42%).⁹ External or citizen-facing data sources were less commonly used: citizen surveys were selected by 25% of respondents, while administrative data from private sources was used by 26%. In 22% of projects, analysts aimed to integrate multiple sources of data to conduct their analyses. Among projects relying on multiple data sources, the most common combination involves payroll, budget, and procurement data, which provides analysts with a multidimensional view of how public resources are managed.

A similar picture emerges when we assess the data sources being used for the submitted analytics projects. Administrative data, such as budget and expenditure data, and measures of service delivery, which would be included in specialized administrative data, are among the most widely used data types. Some data types are particularly prominent in specific sector projects, such as household and labor force surveys being most intensively used in public administration and governance-related projects.

⁸In contrast to the dominant proportion of higher-income public expenditure it represents, social security spending is a very small component of most of the countries under study here [World Bank, n.d.].

⁹Respondents could choose up to three data sources as their primary ones, such that percentages can sum up to more than 100%.

Two conclusions follow. First, our findings highlight a preference for internally generated operational data over survey-based or external datasets. Second, analysts inside government often combine multiple types of data in their analysis. These conclusions underscore another reason for the importance of fostering effective analytics inside government: government analysts often have access to internal data sources for analytics to improve government - either through standalone analytics of these data sources or their integration with other data sources - which outside researchers cannot access or only access with significant friction. This implies a substantial potential value of government analytics undertaken by government analysts.

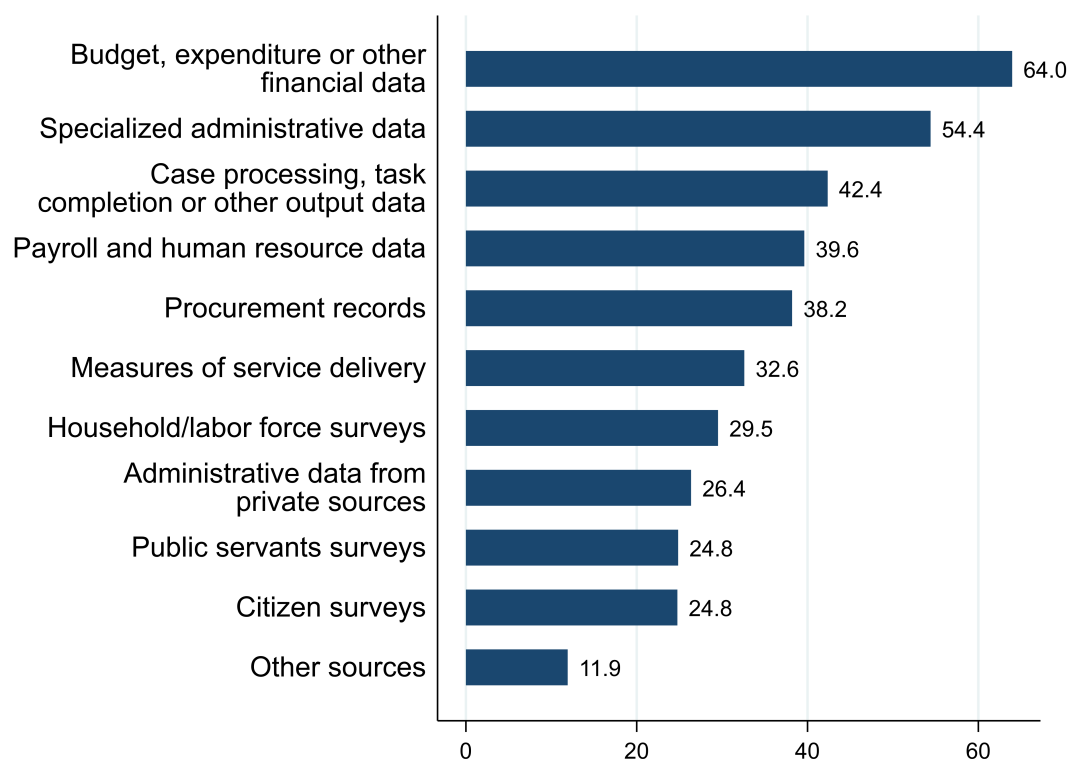


Figure 4: Main data sources used by government analytics projects (Percent)

3.4 What methods do government analysts use to undertake their analysis?

Government analytics projects draw on a wide range of analytical methods. Data visualization is the most commonly proposed method, appearing across all sectors. A prominent example of data visualization in government is the creation of dashboards of administrative and survey data for decision-making support. Predictive modeling is the second highest proposed method, with a substantial demands for predicting

future outcomes in the process of public policy planning. AI-related methods are observed being used across all the sectors we define, including anomaly detection, machine learning, and natural language processing.

Despite their prevalence in applied economics, statistical and correlational analysis such as regressions are less likely to be observed in government analytics projects than these other approaches. We observed very few randomized control trials, and these, along with any other impact evaluations, are a small minority in the body of analytical work undertaken by government analysts. Thus, government analysts are far more likely to be focusing on novel presentations of data than developing counterfactual assessments using it.

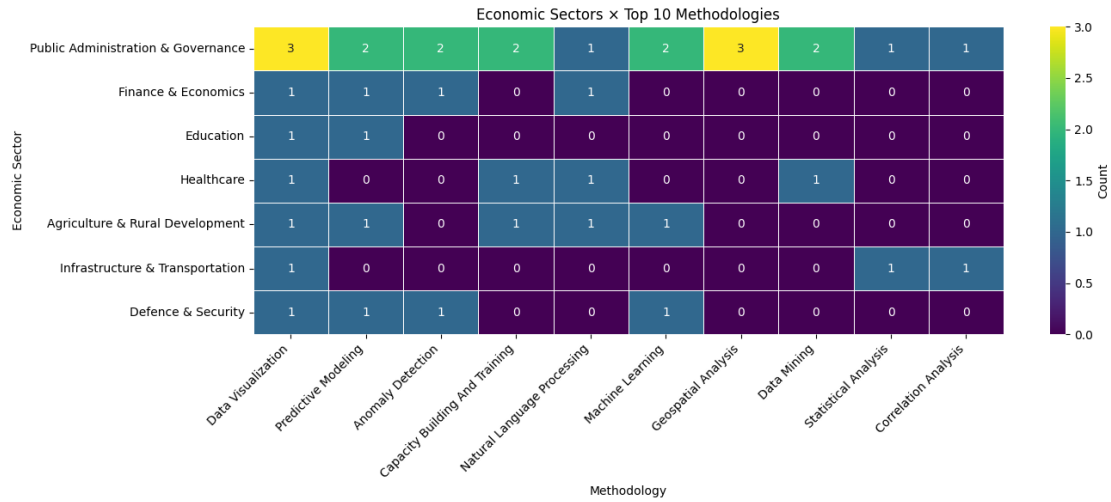
The broad patterns of mapping between sectors, data types, and methods are presented in Figure 5. Panel (a) describes the data types most prevalent in sectoral analytics, and panel (b) describes the most prevalent analytical methods. So in the first example described in the topic-modelling analysis above - focused on health financing in Bangladesh - the sector is health; the data type is budget data; and, the method is data visualization. Or an example of a project using specialized administrative data would include Uganda's project to use administrative microdata to assess corruption and overpayment in public service procurement by comparing service delivery costs against private sector benchmarks. An example of a data visualization project in public administration would be Malawi's project to develop interactive dashboards that visually represent progress toward the country's 2063 national vision, providing real-time updates on key indicators to track development goals and inform decision making.

In Figure 5, each cell represents the number of topic clusters within a given sector that are associated with a particular methodology or data type. For example, the value of 3 in the Public Administration and Governance row under data visualization reflects three distinct topic clusters in that sector use data visualization. All dashboard projects in public administration and governance would be one of the types of analysis. Similarly, the value of 5 in the Public Administration and Governance row under specialized administrative data reflects five topic clusters drawing on administrative records in that sector.

Figure 5 shows that Public Administration and Governance draws on the most types of data and is the most methodologically diverse sector. This is followed by finance analysis and then education assessments. Surveys of public servants are not observed being used in anything by analyses of public administration and finance. The overall picture beyond that sector is that there is substantial scope to expand the methodological diversity of much of government analysis.



(a) The data types most prevalent in sectoral analytics



(b) The analytical methods most prevalent in sectoral analytics

Heatmaps show the count of topics per sector associated with each methodology and data type respectively. Classifications were assigned using GPT-5-mini based on concatenated project descriptions within each topic cluster. Full methodological details are provided in the Technical Appendix.

Figure 5: Topics of government analytics projects

4 Barriers to Effective Data Analytics in Government in the Developing World

We now examine the institutional context in which government analytics is undertaken in developing-country public administrations. What are the principal challenges facing public-sector analysts as they conduct analytical work? To identify the key barriers

to effective government analytics, we draw on evidence from our survey of public analysts alongside broader assessments of public-sector working environments across the developing world.

Perhaps a natural first step is to understand the incentive environment for officials in developing country governments to undertake analytical work. A substantial majority of respondents to our survey (87%) agreed with the statement “The use of data and evidence in my work significantly impacts my career advancement”, suggesting that, contrary to common stereotypes, incentives for data use are well-aligned in government organisations. This reflects broader evidence that work performance in government matters for career advancement, a broad finding of ‘The Global Survey of Public Servants’ Schuster et al. [2023], the most comprehensive picture of public administration across the world. In a small number of cases, respondents outlined that they were insufficiently motivated to undertake their work; but this was a small minority of individuals.

Instead, respondents overwhelmingly point to weaknesses in the organisational ecosystem for data analytics - organisational, managerial and individual - as the major constraint for effectively completing their tasks. Figure 6 presents the most prevalent barriers to undertaking government analytics in the Global South. The primary obstacle, identified by 59% of respondents, is inadequate resources to undertake their work - ranging from computing and IT resources to the resources required to undertake primary data collection. This mirrors other recent work which has suggested that a key feature of government in the developing world is a lack of resources to undertake all aspects of public policy, including analytical work [Williams, 2026].

As found in (Santini et al. [2024]), dedicated funding remains one of the most persistent structural barriers to institutionalizing government analytics. Many organizations face significant challenges in securing consistent funding for analytical initiatives, because resources are limited and priorities often shift rapidly. This is not merely an abstract challenge; as one respondent puts it, there is an “Absence of budget head or provision allotted for government analytics within the recurrent expenditure scope”, meaning that even where the will to invest exists, the administrative architecture to do so systematically does not.

Beyond purely financial resources, organisations are frequently characterised by fractured data systems and related barriers to the undertaking of analytics that could improve government functioning. Half of respondents stated that they could not access internal administrative data to undertake the analytics they needed to. As one respondent stated, their major obstacle to more effectively undertaking their analytical work is a “lack of organization in data flow and bureaucratic obstacles to accessing the data.” Or as another respondent put it, “limited access to quality data, often fragmented across multiple departments, [makes] integration and analysis time-consuming.”

The challenges of limited resources are compounded by ineffective management of those



Figure 6: Challenges preventing from effectively completing tasks (Percent)

resources. 45% of respondents stated that ineffective management of the budgetary resources that do exist was a key issue for why they do not achieve their core duties. And 25% of the government analysts we interviewed argued that there was inadequate leadership for, or organization of, their work. The significance of ineffective management practices in public sector organizations across the developing world has now been documented widely [Rasul and Rogger, 2018; Rasul, Rogger and Williams, 2021; Bandiera et al., 2021; Schuster et al., 2023]. The lack of managerial support for analytical work that might guide more effective administrative activity is highlighted here as another consequence of poor management.

It may also be that among current public service managers, there is insufficient interest in building an analytical culture, despite the interest of a minority of analytics-focused officers. As one respondent summarised, “Challenges being faced within our government is that importance of data is not understood and its importance in decision making” and as another stated, “cultural resistance to change and skepticism towards data-driven decision-making processes can impede the adoption of analytics in government operations”. Such resistance creates ambiguity as to whether analytics are adding value to an organisation. As one respondent argued, “In most cases, executives just sit on data and both me and the rest of people in my organization never gets any

feedback on such data. Nobody knows whether the data was shared to the right people for decision making or not. Thus accessibility and sharing heavily affects analytical works.”

A lack of resources can be extended to more tacit forms of constraints. Roughly a fifth of analysts stated that they do not have sufficient information (26%) or guidance (19%) necessary to perform their tasks. Several respondents described working in environments marked by ambiguity and a lack of strategic direction. As one respondent noted, “The main obstacles hindering my organization’s ability are lack of direction, personal interests, and lack of knowledge. Management may have personal interest which stalls progress in gathering and analyzing data”, suggesting that beyond resource scarcity, unclear mandates and misaligned leadership priorities can significantly undermine the effective use of data within public sector organizations.

The survey also highlighted that time was a scarce commodity for some officials, echoing the phenomenon of bureaucratic overload - bureaucrats being heavily under-resourced relative to their responsibilities - documented in [Knill, Steinebach and Zink, 2025; Dodge et al., 2025]. 24% of the public servants we surveyed stated that ‘last-minute requests’ are a primary obstacle hindering their ability to complete their tasks effectively. As in Dasgupta, Keller and Siddiqi [2020], public analysts are regularly limited in their ability to undertake their normal duties due to tasks being assigned outside their core remit. As one respondent noted, “...last-minute decision-making undermines the value of prior analysis, as critical data and insights are overlooked in favour of quick fixes. This reactive approach limits the organization’s ability to plan strategically and achieve long-term goals”.

These results echo those outlined in Schuster et al. [2023] as key constraints of working in developing country administrations more generally. For example, the Global Survey of Public Servants measure of whether public servants have adequate work resources to perform their job effectively is markedly lower in lower and middle income countries (for instance, 48% in Kosovo, 62% in Brazil) compared to higher-income countries (for instance, 84% in Romania, 82% in Canada). Moreover, only 52% of public servants in Brazil and 54% in Kosovo report that their physical work conditions enable them to perform their jobs effectively, compared to, for instance, 74% in Austria, 83% in Romania, and 87% in Estonia. Analogous to households in low-income communities contending with multiple resource constraints, public servants in developing-country administrations face a diverse set of resource limitations that impede the execution of their daily responsibilities, including the analytical tasks that inform public policy decisions.

Beyond the organisational and managerial constraints to completing tasks, the most frequently identified challenge to undertaking analytical work specifically within developing country public services are a lack of skills and staff specializing in analytical

methods.¹⁰ In total, 61% of respondents stated that analytical work was hindered by a lack of competence in relevant methods. For example, one respondent stated that a major bottleneck to great analytics was a “shortage of qualified data analysts and scientists with the expertise to effectively analyze government data”. Another respondent decried “a significant percentage of the staff have no requisite data analysis skills to analyze the administrative data at their disposal” while a further respondent noted the “lack of skilled personnel with expertise in data analytics and evidence-based policymaking; and insufficient funding for advanced analytical tools and training”.¹¹

Hiring and retaining skilled data analysts within public administration remains a significant challenge. As one respondent noted “Government operations are bound by strict rules, which limit flexibility in hiring staff. In addition, funding constraints make it difficult to attract and retain the best analysts, posing a significant challenge within the government ecosystem.” These rigidities, often rooted in civil service structures, standardized pay scales, and lengthy recruitment processes, limit the ability of organizations to respond quickly to emerging analytical needs [Evans and Rauch, 1999; Finan, Olken and Pande, 2017; Schuster et al., 2023]. Competitive salaries and dynamic work environments offered by the private sector further exacerbate the problem, drawing talent away from the public sector. Even when governments do manage to hire capable analysts, retention can be undermined by bureaucratic silos, a lack of professional development opportunities, and limited exposure to advanced data tools or projects (Santini et al. [2024]).

Although some problems are systematically more likely to arise, the intensity and combination of constraints differ substantially across countries and institutions, reinforcing the need for setting-specific approaches to support government analysts. Figure 7 shows the heterogeneity of the challenges across countries. While staffing issues, data accessibility, and data quality are the most common constraints (each affecting more than half of countries), there is considerable variation in the extent to which individual countries and institutions report them. In contrast, legal issues and data security appear less prevalent on average, but again with notable dispersion.¹²

¹⁰This statistic is based on the survey question on top bottlenecks to analytical work: “What do you consider to be the top three most significant challenges hindering analytical work within your organization?”

¹¹How are public administrations trying to respond to these challenges? Governments are increasingly putting in place training courses that train public analysts in better using administrative data. 26% of the respondents to our survey were trained in government data usage over the past 12 months. Much of this - as echoed in both our survey and Santini et al. [2024] - relates to administrative or audit tasks rather than more causal or structural government analytics. Yet it includes foundational elements that could underpin analytical work; such as a familiarity with the variable fields available in common data sets. Yet that leaves 74% of government analysts who have not received any training on analytics in the past year. As one respondent notes, “Staff members are hardly trained on such issues. I personally have never attended any training which could help with such works. Such lack of skills makes it hard for workers to recognise the importance of data analytics and utilization.”

¹²Interestingly, we find little variation in the responses articulated by management-level officials versus more junior ‘technical staff’. This implies that the challenges to undertaking analysis and

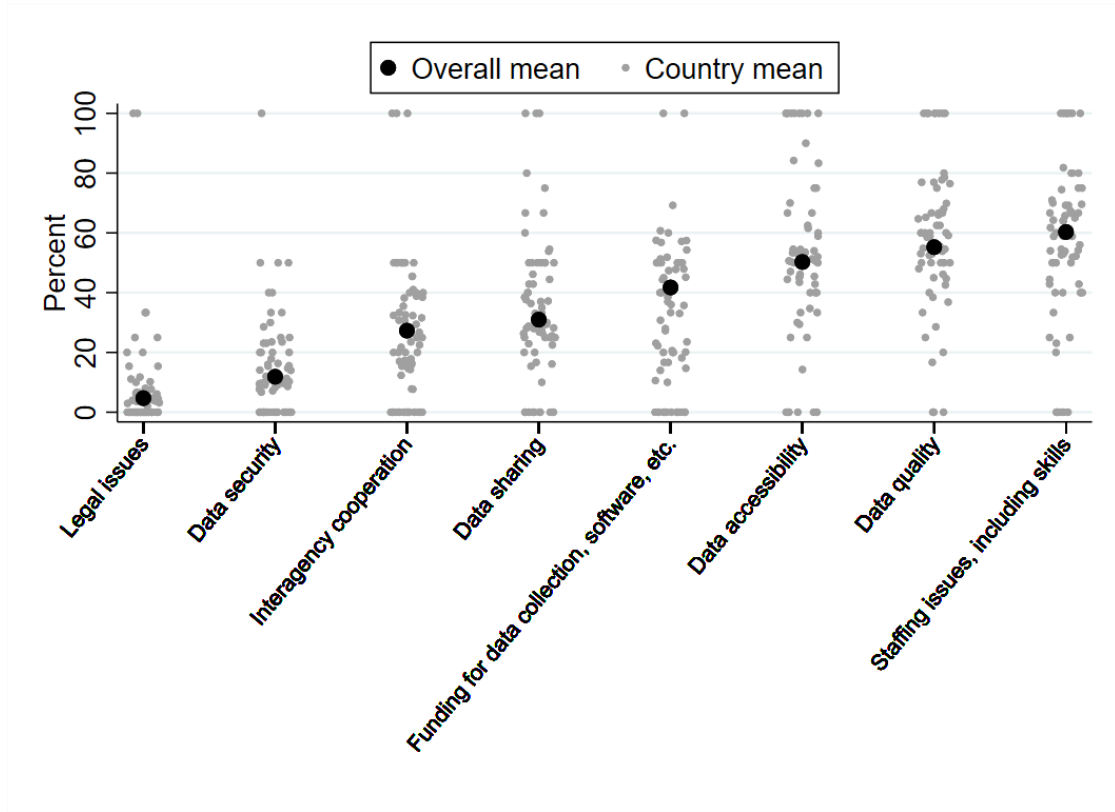


Figure 7: Challenges hindering analytical work within organizations

Perhaps the most natural benchmark to compare our survey results to are public officials in upper-middle income countries. Santini et al. [2024] assesses the ‘ecosystem’ for undertaking analytical work in Latin America and the Caribbean using expert surveys of management information systems in 20 countries in the region. The ordering of constraints to analytics found in those administrations is similar to that expressed by the analysts in the survey reported on here. Although governments in the region acknowledge the importance of analytical skills and in some cases offer training programs for analytics, only 12% of assessed countries have structured career tracks for data analysts, and only 25% of them have proficiency evaluations on data analytics for public servants. Moreover, only 33% of the interviewed experts reported funding for analytical projects. Governments in upper-middle income countries also face challenges in harnessing their data infrastructure as digitalization levels of government systems are often incomplete, preventing them from fully unlocking the transformative potential of their systems or catalyzing the value of the investments they have already made. These comparisons suggest that many limitations on moving public administration toward evidence-based policy making are still substantive at the next tier of

using it to inform public policy are common throughout the hierarchy of public services in developing countries.

development.

5 Discussion

This paper contributes to a growing literature on bureaucracy and development by shifting attention from the consumers of evidence inside government to its producers. While research has increasingly documented the importance of state capacity, bureaucratic management, and evidence-informed policymaking, remarkably little is known about the officials who generate much of the analysis on which governments rely. Yet, the evidence presented here suggests that government analysts constitute a substantial and largely overlooked analytical workforce operating inside public administrations across the developing world.

Three findings stand out. First, government analysts are undertaking a wide range of sophisticated analytical activities. The projects submitted through the fellowship span sectors as diverse as health, public finance, education, agriculture, procurement, and public management, and frequently employ methods that extend well beyond descriptive reporting. These findings challenge the common assumption that analytical work inside developing-country governments is scarce or technically rudimentary.

Second, government analysts often possess a unique comparative advantage relative to external researchers. Their work is grounded in administrative data, operational systems, and implementation challenges that are often inaccessible to researchers outside government. Much of the analysis identified in this paper seeks to improve the functioning of public administration itself, suggesting that a substantial share of development-relevant evidence is generated from within the state rather than supplied from outside it.

Third, the primary barriers to stronger government analytics appear to be organizational rather than purely technical. Analysts report constraints related to resources, data accessibility, inter-agency coordination, staffing, leadership support, and competing demands on their time. While skills gaps remain important [Mejía-Guerra et al., 2023], the results suggest that governments often possess more analytical talent than they are currently able to deploy effectively. The challenge is therefore not only building analytical capability, but creating institutional environments in which that capability can be productively used.

These findings point toward a broader reconsideration of where evidence for development comes from. Much of the literature on evidence-based policymaking implicitly treats governments as consumers of externally generated knowledge [Weiss, 1979; Cairney, 2016; Hjort et al., 2021; Vivalt and Coville, 2023]. By contrast, the results presented here suggest that governments themselves are major producers of evidence. Improving development outcomes may therefore depend as much on strengthening the

internal analytical systems of government as on expanding the external supply of academic research, evaluations, and technical assistance.

This observation opens a substantial research agenda. We know relatively little about the organizational determinants of effective government analytics. Which institutional arrangements best support analytical work? How should governments organize data access, analytical career paths, and incentives for evidence use? What management practices enable analysts to translate data into operational improvements? Which forms of training generate the greatest returns? Under what conditions does analytical work by internal government analysts influence policy decisions rather than remaining confined to reports and presentations? The answers to these questions remain largely unknown despite their potentially large implications for state effectiveness.

The emergence of artificial intelligence further increases the importance of these questions. Many of the tasks undertaken by government analysts—data cleaning, coding, summarization, forecasting, document analysis, and the production of analytical outputs—are likely to become substantially cheaper and faster [Noy and Zhang, 2023; Korinek, 2023; Brynjolfsson, Li and Raymond, 2025]. AI therefore has the potential to dramatically expand the analytical capacity of governments, particularly in resource-constrained settings. However, the results of this paper suggest that realizing these gains will depend less on access to AI itself than on the institutional foundations upon which AI must operate.

Artificial intelligence cannot compensate for inaccessible administrative data, fragmented information systems, weak incentives, inadequate staffing, or poor management. Indeed, many of the barriers identified by respondents represent precisely the conditions under which AI systems struggle to generate value [Brynjolfsson, Rock and Syverson, 2021; World Bank, 2016]. Governments that lack usable data, analytical workflows, and personnel capable of interpreting evidence are unlikely to benefit substantially from advances in AI, regardless of the sophistication of the underlying technologies. Conversely, governments that invest in these foundations may find that AI dramatically increases the productivity and reach of their existing analytical workforce.

The central message of this paper is therefore optimistic. Across the developing world, government analysts are already generating evidence to address pressing public problems. Analytical capacity thus exists within government. The challenge is whether governments can create the conditions under which that capacity can flourish. As analytical technologies continue to improve, the returns to getting these fundamentals right are likely to grow substantially. If governments succeed in doing so, the next generation of development gains may come not only from better research about government from academia and other external partners, but from better analysis conducted within government itself.

References

- Acemoglu, Daron, and James A Robinson.** 2012. *Why nations fail: The origins of power, prosperity, and poverty*. Crown Books.
- Andrews, Matt, Lant Pritchett, and Michael Woolcock.** 2017. *Building State Capability: Evidence, Analysis, Action*. Oxford:Oxford University Press.
- Armstrong, Rebecca, Elizabeth Waters, Barbara Crockett, and Helen Keleher.** 2014. “Understanding Evidence: A Statewide Survey to Explore Evidence-Informed Public Health Decision-Making in a Local Government Setting.” *Implementation Science*, 9(188).
- Bandiera, Oriana, Michael Carlos Best, Adnan Qadir Khan, and Andrea Prat.** 2021. “The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats.” *The Quarterly Journal of Economics*, 136(4): 2195–2242.
- Banerjee, Abhijit V., and Esther Duflo.** 2011. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. New York:PublicAffairs.
- Besley, Timothy, Robin Burgess, Adnan Khan, and Guo Xu.** 2022. “Bureaucracy and Development.” *Annual Review of Economics*, 14(1): 397–424.
- Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond.** 2025. “Generative AI at Work.” *The Quarterly Journal of Economics*, 140(2): 889–942.
- Brynjolfsson, Erik, Daniel Rock, and Chad Syverson.** 2021. “The Productivity J-Curve: How Intangibles Complement General Purpose Technologies.” *American Economic Journal: Macroeconomics*, 13(1): 333–372.
- Cairney, Paul.** 2016. *The Politics of Evidence-Based Policy Making*. Basingstoke:Palgrave Macmillan.
- Cantarelli, Paola, Paolo Belardinelli, and Nicola Bellé.** 2023. “Information Use in Public Administration and Policy Decision-Making: A Research Synthesis.” *Public Administration Review*, 83(5): 1177–1194.
- Dasgupta, Aditya, Thomas R. Keller, and Bilal Siddiqi.** 2020. “The political economy of bureaucratic overload: Evidence from rural development officials in India.” *American Political Science Review*, 114(4): 1316–1334.
- Dodge, Eric, Yusuf Neggers, Rohini Pande, and Charity M. Troyer Moore.** 2025. “From Delay to PayDay: Easing Bureaucrat Access to Implementation Information Strengthens Social Protection Delivery.” National Bureau of Economic Research 33756, Cambridge, MA.

- Evans, Peter, and James E. Rauch.** 1999. “Bureaucracy and Growth: A Cross-National Analysis of the Effects of “Weberian” State Structures on Economic Growth.” *American Sociological Review*, 64(5): 748–765.
- Finan, Frederico, Benjamin A. Olken, and Rohini Pande.** 2017. “The Personnel Economics of the Developing State.” In *Handbook of Economic Field Experiments, Volume 2.*, ed. Abhijit V. Banerjee and Esther Duflo, 467–514. North-Holland (Elsevier).
- Hjort, Jonas, Diana Moreira, Gautam Rao, and Juan Francisco Santini.** 2021. “How Research Affects Policy: Experimental Evidence from 2,150 Brazilian Municipalities.” *American Economic Review*, 111(5): 1442–1480.
- Knill, Christoph, Yves Steinebach, and Dionys Zink.** 2025. “Bureaucratic Overload and Organizational Policy Triage.” In *Triage Bureaucracy: The Organizational Challenge of Implementing Growing Policy Stocks*. Chapter 2. Cambridge:Cambridge University Press.
- Koga, Natália Massaco, Maria Rita Loureiro, Pedro Lucas de Moura Palotti, Rafael da Silva Lins, Bruno Gontyjo do Couto, and Suylan Nogueira Lima.** 2022. “Analysing the Information Sources Brazilian Bureaucrats Use as Evidence in Everyday Policymaking.” *Policy & Politics*, 50(4): 483–506.
- Koga, Natália Massaco, Pedro Lucas de Moura Palotti, Suylan de Almeida Midlej e Silva Nogueira Lima, and Pedro Cavalcante.** 2024. “Knowledge and Evidence Use by Street-Level Bureaucrats and Managers in the Brazilian Federal Bureaucracy.” *Policy and Society*, 43(4): 521–540.
- Korinek, Anton.** 2023. “Generative AI for Economic Research: Use Cases and Implications for Economists.” *Journal of Economic Literature*, 61(4): 1281–1317.
- Masaki, Takaaki, Samantha Custer, Anna Eskenazi, Andrew Stern, and Robert Latourell.** 2017. “Decoding Data Use: How Do Leaders Source Data and Use it to Accelerate Development?” AidData at the College of William & Mary Report, Williamsburg, VA.
- Mejía-Guerra, José Antonio, Christian Schuster, Magdalena Rojas Wettig, Kim Sass Mikkelsen, and Jan Meyer-Sahling.** 2023. “Making National Statistical Offices Work Better: Evidence from a Survey of 13,300 National Statistical Office (NSO) Employees in 14 Latin American and Caribbean Countries.” Inter-American Development Bank.
- Mintrom, Michael.** 1997. “Policy Entrepreneurs and the Diffusion of Innovation.” *American Journal of Political Science*, 41(3): 738–770.

- Newman, Joshua, Adrian Cherney, and Brian W. Head.** 2017. “Policy Capacity and Evidence-Based Policy in the Public Service.” *Public Management Review*, 19(2): 157–174.
- Noy, Shakked, and Whitney Zhang.** 2023. “Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence.” *Science*, 381(6654): 187–192.
- OECD.** 2020. “The Future of Work in the Public Service: Preparing the Workforce for Change in a Context of Uncertainty.” OECD Public Governance Directorate, Working Party on Public Employment and Management.
- Parkhurst, Justin.** 2017. *The Politics of Evidence: From Evidence-Based Policy to the Good Governance of Evidence*. Abingdon:Routledge.
- Rasul, Imran, and Daniel Rogger.** 2018. “Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service.” *The Economic Journal*, 128(608): 413–446.
- Rasul, Imran, Daniel Rogger, and Martin J Williams.** 2021. “Management, Organizational Performance, and Task Clarity: Evidence from Ghana’s Civil Service.” *Journal of Public Administration Research and Theory*, 31(2): 259–277.
- Rogger, Daniel, and Ravi Somani.** 2023. “Hierarchy and Information.” *Journal of Public Economics*, 219.
- Santini, Juan Francisco, Flavia Sacco Capurro, Daniel Rogger, Timothy Lundy, Galileu Kim, Jorge de León Miranda, Serena Cocciolo, and Chiara Casanova.** 2024. *Data for Better Governance: Building Government Analytics Ecosystems in Latin America and the Caribbean*. Washington, DC:World Bank. Published November 25, 2024.
- Schuster, Christian, Kim Sass Mikkelsen, Daniel Rogger, Francis Fukuyama, Zahid Hasnain, Dinsha Mistree, Jan Meyer-Sahling, Katherine Bersch, and Kerenssa Kay.** 2023. “The Global Survey of Public Servants: Evidence from 1,300,000 Public Servants in 1,300 Government Institutions in 23 Countries.” *Public Administration Review*, 83(4): 982–993.
- Teodoro, Manuel P.** 2011. *Bureaucratic Ambition: Careers, Motives, and the Innovative Administrator*. Baltimore:Johns Hopkins University Press.
- U.S. Government Accountability Office.** 2021. “Evidence-Based Policymaking: Survey Results Suggest Increased Use of Performance Information Across the Federal Government.” U.S. Government Accountability Office GAO-22-103910. Survey data can be downloaded at <https://files.gao.gov/special.pubs/gao-21-537sp/resultsall.htm>.
- Vivalt, Eva, Aidan Coville, and Sampada KC.** 2025. “Local knowledge, formal evidence, and policy decisions.” *Journal of Development Economics*, 174: 103425.

- Vivalt, Eva, and Aidan Coville.** 2023. “How Do Policymakers Update Their Beliefs?” *Journal of Development Economics*, 165: 103121.
- Weiss, Carol H.** 1979. “The Many Meanings of Research Utilization.” *Public Administration Review*, 39(5): 426–431.
- Weiss, Carol H.** 1980. “Knowledge Creep and Decision Accretion.” *Knowledge: Creation, Diffusion, Utilization*, 1(3): 381–404.
- Williams, Martin J.** 2026. *Reform as Process: Implementing Change in Public Bureaucracies*. New York:Columbia University Press.
- World Bank.** 2016. *World Development Report 2016: Digital Dividends*. Washington, DC:World Bank.
- World Bank.** n.d.. “Social Protection Expenditure (as % of GDP).” <https://www.worldbank.org/en/data/datatopics/aspire/indicator/social-expenditure>, Accessed: 2025-09-05.

Technical Appendix

To convert the raw application responses into a usable data set, we undertook data pre-processing, topic modelling, and post-processing. Our focus was on the text provided as proposals for analytical projects.

Pre-processing

Project descriptions were pre-processed to minimize noise and standardize the text prior to topic modeling. All text was converted to lowercase and country names were systematically removed to prevent geographic terms from disproportionately influencing topic formation. Standard English stopwords were eliminated and supplemented with a domain-specific list of common administrative terms, such as “project,” “program” and “service.” Words were then lemmatized to reduce them to their base form, thereby enhancing consistency across the corpus. During part-of-speech tagging, words were initially identified by their grammatical role (noun, verb, adjective, or adverb) to improve the accuracy of this process. Tokenization was applied on word boundaries, and records with empty or nearly empty cleaned text were excluded to ensure that only meaningful descriptions contributed to topic discovery. Country information was retained in a separate location field for reporting and summarization but was excluded from the text used for topic modeling.

Topic Modelling

Each refined project description was transformed into a semantic vector representation utilizing a transformer-based embedding model, specifically SentenceTransformer “all-MiniLM-L6-v2”. Subsequently, dimensionality reduction was conducted using Uniform Manifold Approximation and Projection (known as UMAP) to maintain semantic relationships within a lower-dimensional space. These reduced representations were then clustered using Hierarchical Density-Based Spatial Clustering of Applications with Noise, a density-based algorithm that groups semantically similar descriptions while permitting some to remain unassigned as noise. Outliers were reassigned based on their probabilities, and clusters were consolidated to yield a final set of fifteen coherent topics, representing the principal themes within the corpus.

Topic Representations

Topic representations were generated using BERTopic’s multi-aspect representation framework, which integrates part-of-speech filtering, keyword extraction, and maximal marginal relevance to produce interpretable descriptions of each cluster. BERTopic is a topic modeling technique that uses document embeddings to create dense clus-

ters of semantically similar texts and generates interpretable topic descriptions by selecting the most representative words for each cluster. More details are available at <https://maartengr.github.io/BERTopic/index.html>. Labels were generated using BERTopic’s built-in `generate_topic_labels` function and then refined to produce concise, human-readable labels. Keywords were filtered to remove country terms and domain-specific boilerplate, truncated by token length, and deduplicated using suffix normalization, noun lemmatization, and cosine similarity of sentence embeddings.

During metadata extraction, each topic’s concatenated documents were first truncated using a token counter (`tiktoken`) to ensure they fit within the model’s input limit. Prompts were structured to return results in a consistent, machine-readable format. The model’s responses were then parsed automatically, validated, and compiled into a per-topic summary table, ensuring that outputs were standardised and reproducible.