

## CHAPTER 2

# How to Do Government Analytics

## Lessons from the *Handbook*

*Daniel Rogger and Christian Schuster*

### SUMMARY

How can practitioners and researchers undertake government analytics effectively? This chapter summarizes lessons from *The Government Analytics Handbook*. The chapter begins by introducing a public administration production function, which illustrates how different data sources, such as procurement data or public servant survey data, shed light on different parts of the machinery of government. The chapter then highlights lessons to keep in mind when undertaking any measurement and analysis of government administration. In the chapter-by-chapter summary that follows, lessons on how to generate and analyze a range of core data sources for government analytics are presented. These are data from administrative records (such as human resources and payroll, budget, procurement, administrative cases, task completion and projection completion); data from surveys of public servants; and external assessments of government, such as household or citizen surveys. The chapter concludes by showcasing how different data sources can be integrated to understand core challenges in the administration of government, such as personnel management.

### ANALYTICS IN PRACTICE

- This *Handbook* presents a wealth of approaches and data sources available to governments to improve the analytics of the machinery of government and identify evidence-based improvements. Many of these approaches rely on data that governments already collect as part of their day-to-day operations.
- By conceiving of government's electronic records as data in themselves, existing government records, and in particular the vast troves of data now being produced, can be repurposed as a means of diagnosing and strengthening government administration.
- Government analytics can be undertaken with at least three types of data: administrative data collected or published by government entities (such as payroll data); surveys of public servants; and external assessments (such as household surveys or anthropological assessments).

- To organize the various data sources assessed throughout the *Handbook*, this chapter introduces a public administration production function. A production function relates input factors of production through processes (such as management practices) to the output of an organization, and their eventual outcomes.
- Some data sources are better suited to assessing inputs into public administration, such as payroll data assessing the costs of different personnel. Some data sources are better suited to assessing the processes, practices, and cultures that convert inputs into outputs, such as surveys of public servants that assess how they are being managed. And some data sources are better suited to assessing the outputs and outcomes of public administration, such as citizen satisfaction surveys. What type of data source is appropriate for analytics depends on what aspect of public administration the analyst is seeking to diagnose and improve. Data sources can also be powerfully combined to understand overarching themes in how government is functioning, such as corruption.
- Frontier government analytics would integrate the analytics of these distinct data sources. It would generate them at a scale sufficient to inform the decision-making of individual managers. And it would make them easily accessible to those managers across government organizations and departments. For instance, dashboards integrating data sources and updating in real time would provide managers with comparisons for their staffing issues, process quality, the perceived quality of management practices, and so on. They could track outputs and outcomes, from task completion and case productivity to external assessments from citizens. Comparative data would allow them to benchmark themselves against other government organizations, or where appropriate, other countries. Managers would be capable of understanding the limitations and strengths of different analytics. The result would be a transformational change toward leveraging data to strengthen public administration.

## INTRODUCTION: AN UNPRECEDENTED EXPLOSION OF DATA

A common perception of government is that it is a place that no one truly understands: an incomprehensible maze, filled with a procedural fog. Actually, public administration is brimming with information. The everyday business of government has generated an abundance of records on who does what where. Each of these records has its origin in a contemporary administrative need of public officials. With the right approach, these records also provide a window into the workings of the administrations that created them—a way to clear the fog and provide a map through the maze.

For example, as a social security officer decides how to organize their work, they are making decisions that will affect the delivery of cases they are working on. These choices are reflected in the time stamps that accompany their cases, the degree to which they follow procedure, and the feedback they receive from those they serve. On its own, the record of their work is interesting, but combined with the records of their colleagues from their own organization and across the country it becomes a means to understanding the functioning of a public sector organization. Do all the officers working for a particular manager perform well on some aspects of their work but less so on others? Are some offices able to regularly process retirement benefits faster than others? The fog of uncertainty about how social security officers perform across their community can be cleared by turning records into analytics of the public service.

The data trail of government records like these has expanded greatly recently. The move toward digital government has multiplied the scale of electronic transactions between public administrations and citizens. From citizens paying taxes electronically to registering online for driver's licenses, every interaction is now recorded electronically in government data systems. Digital government has also multiplied electronic transactions within public administrations. Payrolls, procurement records, and budget disbursement records are some examples. Digital government has also facilitated the collection of other data, for instance by enabling governments to survey their employees at scale online. All this makes data to understand public administration easier to access than ever before.

This trove of data is critically important to the productive functioning of government, much in the same way as it has improved firm performance (Bakshi, Bravo-Biosca, and Mateos-Garcia 2014). “Without high-quality data providing the right information on the right things at the right time, designing, monitoring and evaluating effective policies becomes almost impossible,” a United Nations report notes (UN 2014, 2).

While the use of data for public *policy* has exploded, far less emphasis has been placed on how government records might be repurposed to understand how effectively the administrative machinery of government is functioning. By conceiving of government’s electronic records as data in themselves, existing government records can be seen as a means of diagnosing and strengthening government administration.

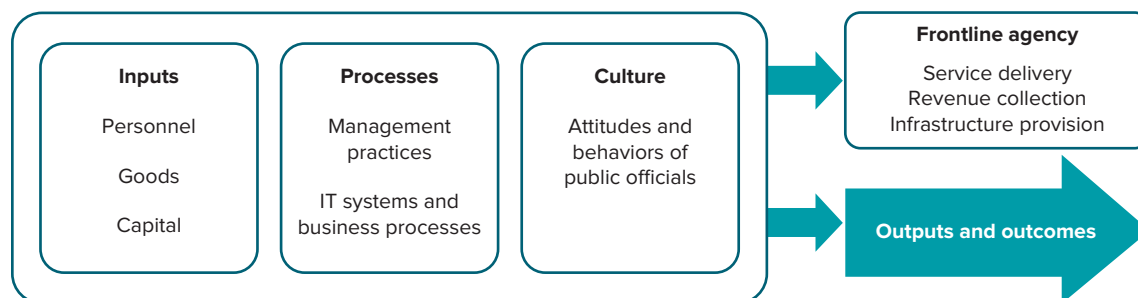
As discussed in chapter 1, public administration presents particular measurement challenges and opportunities. This *Handbook* focuses on sources and methods of measurement to meet those challenges, and on the need to create a system of measurements that is fit for public administration, rather than repurposed from other settings. Although current government records are an excellent foundation for analysis, their suitability for generating knowledge requires work. Once analytics becomes a goal, the range and nature of the public record is likely to change, to optimize government data for both the task at hand and the broader assessment of government functioning. This chapter summarizes lessons to that end—and on how to do government analytics—from across the *Handbook*.

## WHAT GOVERNMENT ANALYTICS CAN ANALYZE: UNDERTAKING ANALYTICS ALONG A PUBLIC ADMINISTRATION PRODUCTION FUNCTION

Government analytics refers to the use of data to diagnose and improve the machinery of government, or public administration. This chapter introduces a public administration production function to provide an overarching framework to organize the various data sources assessed in different chapters.<sup>1</sup> A production function relates input factors of production to the output of deliverables of an organization, and their eventual outcomes. The productivity of an organization thus depends on the quality and quantity of outputs relative to inputs. Figure 2.1 visualizes the different components of our production function for public administration (Meyer-Sahling et al. 2021; World Bank Group 2019). While many core elements coincide with typical private sector production functions (Mas-Colell, Whinston, and Green 1995), the functioning of government administration has been characterized as distinct from that of private firms due to the multiplicity of principals, the ambiguity of tasks, and the presence of principals with political incentives, among other features.

In public administration, inputs include personnel (public employees), goods (such as computers), and capital (such as office space). Outputs refer to, first, the deliverables produced by public

**FIGURE 2.1** The Public Administration Production Function



Source: Adapted from World Bank Group 2019.  
Note: IT = information technology.

administration organizations themselves. For instance, a ministry of finance might issue public sector debt at a certain interest rate. Further, public administration organizations produce outputs (activities) that enable frontline agencies in the public sector—such as hospitals, schools, or police forces—to deliver services and goods to citizens. The outcomes in these examples are better health, education, or public safety, respectively. To fund the outputs, a ministry of finance may oversee budgets that frontline agencies then disburse to deliver their services.

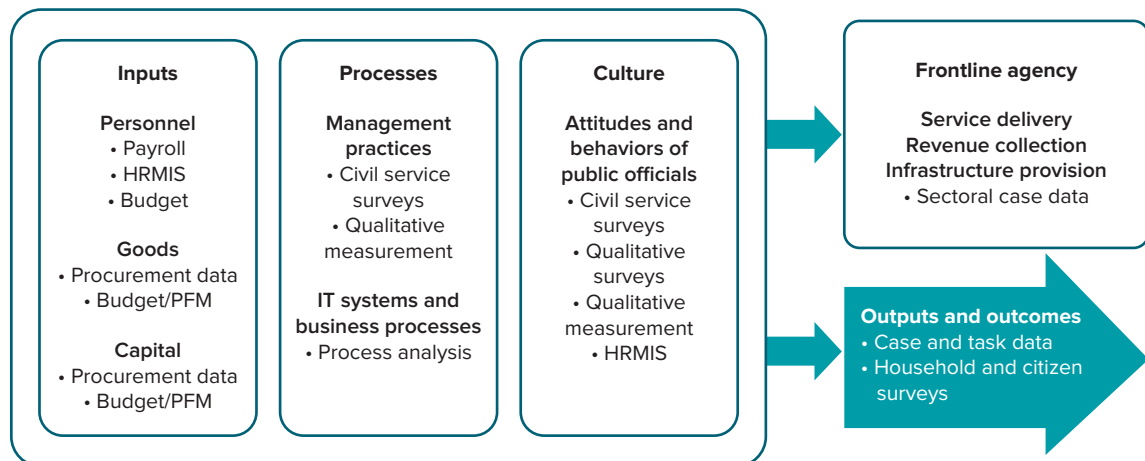
How do public administrations convert inputs (such as personnel) into outputs and outcomes? In our production function, this conversion is enabled by policies (organizational objectives and work procedures), systems, and management practices, and mediated by norms and behaviors inside public administration. For instance, a ministry of finance may have a policy in place to review a budget for an organization by a certain date. A team lead inside the ministry then manages employees to ensure the task is completed well and on time—such as through effective performance management practices. Those practices and organizational policies shape the norms and behaviors of the ministry’s employees—such as their motivation to work hard—which in turn then allows the ministry to produce outputs (such as a budget review).<sup>2</sup>

By utilizing different data sources and different methods, government analytics can shed light on all parts of this production function and identify bottlenecks, whether overpriced input goods, ghost workers on the payroll, high staff turnover, or slow processing of administrative cases, to name just a few. Contemplating government analytics along the production function enables analysts to diagnose public administration challenges holistically, and to understand how different data and approaches to government analytics relate.

To illustrate, figure 2.2 maps a number of different data sources analyzed in various chapters to their respective components in the production function. Several types of administrative data have particular strengths in diagnosing inputs into the public administration production function. For instance, payroll data and human resources management information system (HRMIS) data can help governments understand personnel as an input into the public administration production function, such as whether pay of public servants is competitive and fiscally sustainable, or whether staffing levels are adequate (see chapters 9 and 10). Budget data and procurement data can help governments understand spending on goods and capital as inputs into public administration—for instance, whether public administrations acquire similar goods cost-effectively across organizations in the public administration (see chapters 11 and 12).

Government analytics can also shed light on the processes and practices that convert inputs into outputs and outcomes. Surveys of public servants and qualitative measurement have particular strengths

**FIGURE 2.2 Mapping Different Government Analytics Data in the Public Administration Production Function**



Source: Original figure for this publication adapted from World Bank Group 2019.

Note: HRMIS = human resources management information system; IT = information technology; PFM = public financial management.

at diagnosing management practices. Management quality is fundamentally experienced by employees and a result from the interaction between managers and employees. Surveys can, for instance, ask public servants how they perceive the leadership of their superior or the quality of their performance feedback (see chapter 18). Government analytics can also shed light on the quality of processes inside public administration, such as whether these processes adhere to government procedure or meet deadlines (see chapter 13).

Whether practices and processes effectively turn inputs into outputs and outcomes is, as noted, mediated by the norms, attitudes, and behaviors of public administrators. Surveys of public servants and qualitative measurement are standard practice in many governments to evaluate this component of public administration production—for instance, to understand how engaged, committed, and ethical in their behavior public administrators are (see, for example, chapter 18). HRMIS data often complement rich survey data by providing insights into specific behaviors of public employees that are digitally recorded, such as whether public servants leave the organization, work overtime, or take sick leave (see chapter 9).

Last, public administrations produce outputs and outcomes both of their own (such as a ministry of finance issuing debt), and to enable outputs and outcomes of frontline providers. The productivity of frontline, service delivery agencies such as hospitals, schools, and police forces has been extensively measured, not least as direct contact with citizens enables more direct measurement of service delivery outcomes (such as patient outcomes in hospitals or learning outcomes in schools) (see chapter 29).

This *Handbook*, instead, focuses on the analytics of *administrative* outputs and outcomes. Administrative case data are one important source for measurement in such contexts. Such data are often routinely collected by organizations (for instance, the number of tax or social security cases processed) and can be repurposed by organizations to measure outputs and outcomes (such as the amount of tax revenue raised), and thus gauge productivity (see chapters 14 and 15). Beyond administrative data, surveying households and citizens (such as by asking citizens about their trust in public administration organizations) can be an important data source to understand outcomes of public administration (see chapter 28). What will be of most significance to measure and analyze will depend on a specific organizational setting and topic of interest to decision-makers.

The various chapters of this *Handbook* provide insights into how to use these different data sources to understand how government is functioning and improve the management of public administration. For a brief overview, table 2A.1 in annex 2A at the end of this chapter presents a chapter-by-chapter mapping of data sources to the topics covered in chapters 10 to 30 of the *Handbook*, to help readers pick and choose the chapters most of interest to them. The remainder of this chapter thus summarizes their lessons on how to do government analytics well.

## PART 2: FOUNDATIONAL THEMES IN GOVERNMENT ANALYTICS

The second part of the *Handbook* focuses on four cross-cutting challenges in the measurement and analysis of public administration: how to ensure that government analytics are undertaken (1) with due recognition and management of the risks involved (from managing political pressure to protecting key indicators from manipulation); (2) in a way that ensures analytical practices accord with current best practice in social science; (3) in an ethical manner; and (4) with measures to determine whether the government analytics generated are, in fact, used.

### Measuring What Matters: Principles for a Balanced Data Suite That Prioritizes Problem Solving and Learning

Chapter 4, by Kate Bridges and Michael Woolcock, tackles the first of these four challenges. The chapter notes that overreliance on quantitative government analytics comes with risks. For example, as with any performance targeting, hitting easy-to-measure targets risks becoming a false standard of broader success. Overemphasis on measurement also risks neglect of other important forms of knowledge—such as

qualitative work or practical knowledge—and thus may curtail a deeper understanding of key public administration problems. Moreover, political pressures, if undetected, can lead to falsification of data—for instance, to cover up problems such as corruption—and thus undermine the integrity of government analytics. Left unattended, these risks may mean that government analytics may curtail the problem-solving and implementation capabilities of public sector organizations, rather than strengthening them.

In light of these risks, chapter 4 offers four principles for government analytics based on a balanced data suite that strengthens the diagnosis and solving of problems in public administration:

1. Identify and manage the organizational capacity and power relations that shape data management: for instance, by defining and communicating to all staff professional standards for collecting, curating, analyzing, and interpreting government analytics data.
2. Focus quantitative measures of success on those aspects that are close to the problem: for instance, by targeting measurement of public administration problems prioritized by government.
3. Embrace a role for qualitative data, especially for those aspects that require in-depth, context-specific knowledge.
4. Protect space for judgment, discretion, and deliberation in those (many) decision-making domains that inherently cannot be quantified.

### **Practical Tools for Effective Measurement and Analytics**

Chapter 5, by Maria Ruth Jones and Benjamin Daniels, turns to a second cross-cutting challenge: how to employ the best practices of modern social science when utilizing statistical tools for government analytics. The chapter discusses important underlying statistical principles for government analytics to ensure that government analytics are credible, including the transparency of analysis and reproducibility of results. Producing analysis that accords with best practice requires considering the full life cycle of data work, such that each stage of handling data can be designed to support the next stages. The chapter introduces the suite of tools and resources made available for free by the World Bank's Development Impact Evaluation (DIME) Department to support the achievement of best-practice statistical analysis, such as research-cycle frameworks, extensive training tools, detailed archives of process and technical guidance, and a collaborative approach to data and analytics. The chapter also links to online tools, such as an online training hub, available to help implement these practices.

### **The Ethics of Measuring Public Administration**

Chapter 6, by Annabelle Wittels, discusses a third cross-cutting challenge in government analytics: how to undertake ethical measurement and analysis in public administration. While guides for the ethical collection of data on citizens exist, there is a dearth of practical guides on the ethics of government analytics, particularly with respect to data collection by governments on their own employees. Chapter 6 introduces a heuristic to encourage ethical government analytics, which balances three, at times competing, ethical considerations: (1) an individual dimension that encompasses demands by public employees for dignity and privacy; (2) a group dimension that relates to allowing for voice and dissent of public employees; and (3) a public-facing dimension that ensures that analytics enable public administrators to deliver on public sector values—accountability, productivity, and innovation. These three considerations can be in tension. For instance, data diagnostics on public personnel can inform better management practices and enable greater productivity but impinge on the privacy of employees, whose data are required for the diagnostic.

To guide practitioners, chapter 6 presents a 10-point framework. For instance, the ethics of government analytics requires consideration of the granularity of data required for the diagnostic. Are data that identify individuals required for the diagnostic, or could group-level or anonymous individual-level data enable a



similar diagnosis? In a survey of public servants, do demographic questions need to identify the exact age of a respondent in an organization (which risks identifying the respondent), or are broad age bands (with a lower risk of identification) sufficient? As a second example, ethical government analytics requires consideration of who has a say in what gets measured. In particular, are those who are being measured consulted in advance and given an opportunity to provide inputs? Questions in the framework like these can guide ethical government analytics.

### Measuring and Encouraging Performance Information Use in Government

Chapter 7, by Donald Moynihan, discusses efforts to address a fourth challenge in government analytics: measuring whether government analytics measures are, in fact, being used. If costly data collection or analysis is undertaken on public administration but the resulting analytics are not used to improve public administration, government analytics harms rather than advances better government. Chapter 7 illustrates how survey and administrative data can be drawn on to measure use of government analytics data (or, to use the term in chapter 7, performance information). In particular, the chapter recounts the experience of the US Government Accountability Office (GAO), which periodically surveys public employees on their use of performance information. The chapter describes the survey measures deployed to this end. The chapter also discusses administrative data that governments can utilize, such as data that track the use of dashboards displaying government analytics data. The use of government analytics can thus be subjected to the same evaluative measurement rigor that government analytics applies to public administration.

### Understanding Corruption through Government Analytics

Chapter 8, by James Anderson, David S. Bernstein, Galileu Kim, Francesca Recanatini, and Christian Schuster, illustrates how the approaches presented in this *Handbook* can be combined and leveraged to holistically diagnose a major obstacle to more effective public administrations. Chapter 8 is discussed further at the end of this chapter.



## PART 3: GOVERNMENT ANALYTICS USING ADMINISTRATIVE DATA

The nine chapters in part 3 discuss the range of administrative data sources that governments can draw on to undertake government analytics. Each main type of administrative data is covered in a different chapter. Chapters 9 and 14 contextualize these discussions, showcasing how to create underlying data infrastructures (chapter 9) and how to combine data sources to measure the performance of public organizations with multidimensional missions (chapter 14).

### Creating Data Infrastructures for Government Analytics

Chapter 9, by Khuram Farooq and Galileu Kim, focuses on how to create data infrastructures—or management information systems (MIS)—that are well suited for government analytics. Using the case of human resources management information systems (HRMIS), the chapter provides a road map to guide the development of data infrastructures that enable government analytics. The road map emphasizes the importance of first getting in place high-quality foundational data modules in the MIS, and only then transitioning to developing more complex analytical modules. In the case of HRMIS, for instance, ensuring high-quality foundational modules including basic information on personnel and payroll compliance should take precedence over more advanced modules such as talent management and analytics. Without quality foundations,

other modules will produce imprecise or inaccurate analytics. Analytical modules that include dashboards and reports require that foundational modules are set in place and their data are accurate.

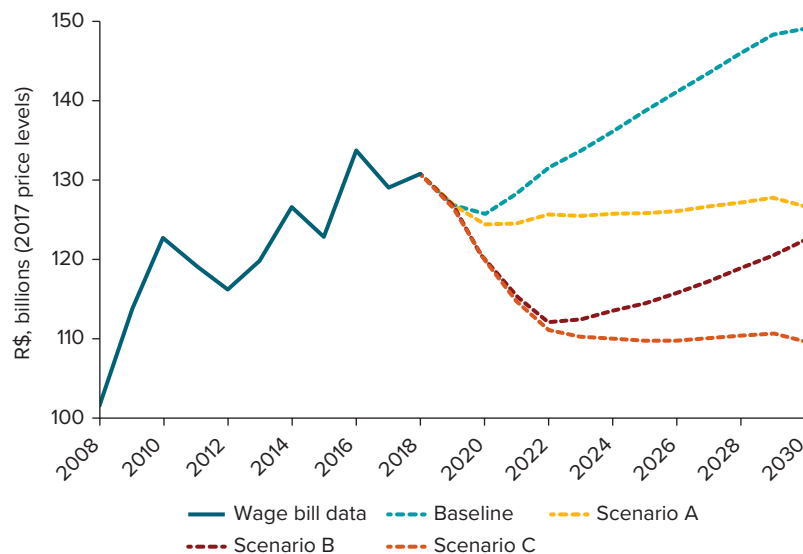
The road map thus emphasizes a sequential approach to creating data infrastructures, coupled with repeated testing of new infrastructures, accessible technical support for users of new MIS systems, tracking of usage, and building in-house capacity to maintain the system. To illustrate potential applications, the chapter is complemented by three case studies of analytical transformations in HRMIS systems in Luxembourg, Brazil, and the United States. Ludwig Balmer, Marc Blau, and Danielle Bossaert discuss how the government of Luxembourg introduced a Business Intelligence Center for human resources (HR) analytics, which transformed HR operations. Luciana Andrade, Galileu Kim, and Matheus Soldi Hardt showcase the development of a system to detect irregularities in Brazil's federal payroll, using machine learning. Robin Klevins and Camille Hoover illustrate how a US federal government agency developed a simple but effective dashboard to extract insights from a federal public employee engagement survey.

The quality of analytics arising from any MIS is critically dependant on the quality of the underlying data, and the analysts' understanding of their origins and limitations. The rest of part 3 discusses how to ensure quality measurement and analytics with a range of administrative data sources that measure aspects of the production function of public administration.

### Government Analytics Using Human Resources and Payroll Data

In chapter 10, Rafael Alves de Albuquerque Tavares, Daniel Ortega Nieto, and Eleanor Florence Woodhouse illustrate how to use payroll and HRMIS data for government analytics. Using a series of examples from Latin American countries, the chapter underscores how the analytics of such data can enable more evidence-based decisions around both fiscal planning and personnel policy. For instance, payroll data can be drawn on to better understand the likely future dynamics of wages and retirement based on the modeling of individual career trajectories. Figure 2.3 illustrates how analytical use of payroll data allowed Brazilian policy makers to simulate the financial implications of a set of policies related to pay and employment. The difference between the wage bill costs under the least and most expensive options amounted to nearly 50 percent

**FIGURE 2.3** Wage Bill Projection and Policy Scenarios, Brazil, 2008–30



Source: Original figure for this publication (see chapter 10).



of total wage bill expenditures. Such a granular lens on employment trends enables governments to better plan for the longer-term fiscal implications of public employment and understand its drivers.

The analytics of payroll and HRMIS data can also enable governments to improve personnel policy. For instance, data on workforce allocation across different regions of a country (such as those of a tax administration) matched to data on the number of service users (such as taxpayers) in each region can help governments understand under- or overstaffing and improve workforce allocation by prioritizing new recruitment in understaffed regional offices. Payroll data also enable governments to measure turnover of employees working in different organizations or for different managers, helping governments pinpoint retention problems and their origins. By comparing pay for similar positions across government institutions, payroll analysis identifies potential salary inequities. Chapter 10 recommends centralizing payroll and HR data collection systems to render such data accessible and provides a road map to this end that complements the discussions and case studies in chapter 9.

### **Government Analytics Using Expenditure Data**

Chapter 11, by Moritz Piatti-Fünfkirchen, James Brumby, and Ali Hashim, discusses government analytics using expenditure data. Budget and government expenditure data are already widely used by governments to understand whether resources are used for budgeted priorities; whether spending is sustainable, efficient, effective, and equitable; and which government transactions (typically, large-value ones) might have high fiduciary risks. The World Bank, for instance, has a long-standing practice of Public Expenditure Reviews. Such reviews are often accompanied by benefit incidence analysis to orient spending toward those in need, by linking data on spending distribution with who receive services. The chapter briefly reviews these well-established uses of expenditure government analytics, and then delves into an aspect missing in much expenditure analytics: how to ensure high-quality data for expenditure analytics.

The chapter underscores the need to periodically review government expenditure microdata for five features: (1) data provenance and integrity; (2) comprehensiveness; (3) usefulness; (4) consistency; and (5) stability. This requires a prior, clear definition of what counts as a government expenditure, as well as a clear understanding and documentation of how transactions across spending items in government are created, what control protocols they are subject to, how this information is stored, and how microdata are aggregated for analysis (such as by classifying government transactions by function, to be able to analyze aggregate spending data by function). The chapter provides practitioners with seven questions to include in expenditure analytics to ensure that the data are high quality, and the resulting analytics are as informative as possible to improve government spending.

### **Government Analytics Using Procurement Data**

Chapter 12, by Serena Cociolo, Sushmita Samaddar, and Mihaly Fazekas, discusses how to use procurement records as data for government analytics. The digitization of national public procurement systems across the world has multiplied opportunities for procurement data analytics. Such analytics allow governments to strategically monitor procurement markets and trends, to improve the procurement and contracting process through data-driven policy making—for instance, by identifying overpriced goods or corruption risks in procurement—and to assess the potential trade-offs of distinct procurement strategies or reforms. The chapter explores the range of procurement indicators that can serve these purposes. For instance, indicators to measure the economy and efficiency of procurement include the time needed for contracting and the final price paid. Indicators to proxy transparency and integrity include the share of single bidder tenders and the share of excluded bids. Indicators to measure competition include the number of bidders. Indicators of inclusiveness and sustainability include the share of bids coming from small and medium enterprises.

When e-procurement systems are integrated with other e-government systems—such as systems generating firm registries or tax data—analytics can go even further: for instance, by allowing governments

to detect potential family relations (and thus collusion risk) between owners of firms bidding for government contracts and procurement officials. Chapter 12 also showcases how governments can use interactive dashboards to track, analyze, and display key procurement indicators through customizable and user-friendly visualizations. All this requires that procuring entities record procurement transactions consistently, that such records are then centralized, and that (automated) data quality checks and periodic data audits ensure the data are accurate.

### **Government Analytics Using Data on the Quality of Administrative Processes**

The eventual value of the core inputs to the effective functioning of government (personnel, goods, and capital) is determined by how they are processed and managed. Chapter 13—by Jane Adjabeng, Eugenia Adomako-Gyasi, Moses Akrofi, Maxwell Ampofo, Margherita Fornasari, Ignatius Geegbae, Allan Kasapa, Jennifer Ljungqvist, Wilson Metronao Amevor, Felix Nyarko Ampong, Josiah Okyere Gyimah, Daniel Rogger, Nicholas Sampah, and Martin Williams—presents approaches to assessing the quality of *administrative* processes. Applying proper processes and procedure to a project, file, or case is core to the work of public administrators. The chapter presents a range of indicators to this end, such as the share of processes undertaken by public administrators that are timely with respect to deadlines, adhere to government procedure, and are logical in flow. Such data can be collected automatically as part of digitized government work, or manually by assessors employed to judge the quality of process in the physical records of projects, files, or cases. Chapter 13 showcases two applications of this approach. The example from Liberia highlights adherence to new processes for performance appraisal. The example from Ghana highlights the quality of process in core office duties, such as project planning, budgeting, and monitoring.

### **Government Analytics Using Customs Data**

Chapter 14, by Alice Duhaut, provides an overview of government analytics using customs data. Customs agencies typically have three core objectives: facilitating trade, collecting revenue, and ensuring the security and safety of the goods entering or exiting the country. As in many other government agencies with multidimensional missions, attaining one objective (such as greater safety of traded goods) can come at the expense of another (such as facilitating trade). Incomplete measurement of objectives risks encouraging attainment of measured objectives while unknowingly impairing other objectives. This puts a premium on effective measurement of all dimensions of a customs mission, which often requires triangulating different data sources. Chapter 14 showcases how this can be done, deriving indicators for trade facilitation (such as costs of the process, particularly in terms of delays); revenue collection (such as trade volume and revenue collected based on the assessed value); and safety (such as number of goods in infraction seized). Collecting these indicators requires integrating multiple data sources. The chapter thus discusses several data sources. These include the Automated System for Customs Data, used by 100 countries and designed by the United Nations Conference on Trade and Development (UNCTAD), which captures items declared, excise, and duties; as well as complementary data sources, such as time-release studies and GPS data on the time spent at borders. The chapter also illustrates how such data can be used not only for risk management *ex ante*, but also to assess customs performance *ex post*.

### **Government Analytics Using Administrative Case Data**

Chapter 15, by Michael Carlos Best, Alessandra Fenizia, and Adnan Qadir Khan, provides insights into the analytics of administrative case data in government more broadly. A case file is typically a collection of records regarding, for instance, an application for government payments (such as social security) or access to services (such as government-sponsored child care); to obtain licenses and permits; or to bid on a government contract. The chapter draws on three example types of administrative cases: social security programs,

tax collection, and public procurement. In all three examples, governments routinely collect case data as part of their day-to-day operations. The chapter shows how these case data can be repurposed to construct objective measures of performance. These include measures to benchmark productivity across, say, regional tax offices, such as the average time to complete similar tax cases or the number of cases completed per officer. Map 2.1 presents an index of the productivity of social security case processing in Italy. It illustrates how substantial the variability in government productivity can be across a single country, with some offices taking 2.5 times as long as others to process similar cases. Case data also enable governments to understand differences in quality, such as comparing the share of tax cases leading to appeals by taxpayers or being identified as erroneous by audits.

Chapter 15 also emphasizes the importance of accounting for the complexity of each case. For example, a social security claim that clearly meets the requirements of regulation is less complicated to process than a case in which there are ambiguities in eligibility and external validation is required. The chapter provides guidance on how to adjust case data for complexity. When accounting for complexity and quality, the chapter concludes that the case data governments already collect provide a wealth of performance information to make such adjustments.

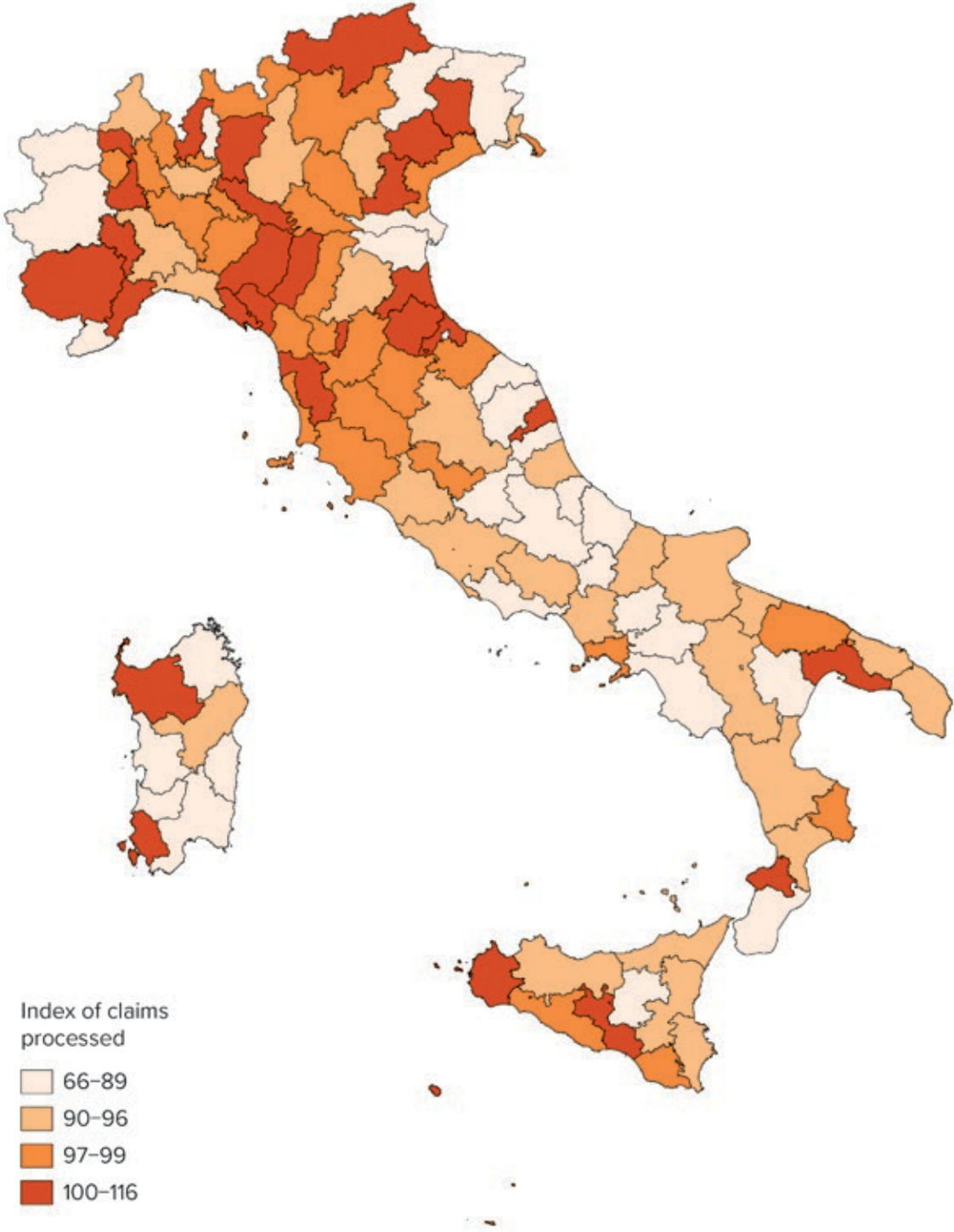
### Government Analytics Using Machine Learning

Chapter 16, by Sandeep Bhupatiraju, Daniel Chen, Slava Jankin, Galileu Kim, Maximilian Kupi, and Manuel Ramos Maqueda, shifts the focus to a different data source—text-as-data—and a different methodological approach, the use of machine learning (ML) and artificial intelligence (AI) for government analytics. Machine learning is fundamentally a methodological approach: it defines a performance indicator and trains an algorithm to improve this indicator, using the data collected. Such data can include text, allowing government to classify quantitatively the “big data” of texts it produces in records or communications. As a result, ML can be applied in a range of government analytics domains, from detecting payroll fraud to understanding bias in welfare appeal decisions, to name a few. The chapter illustrates the use of ML and AI for government analytics in the case of the judiciary. In the justice system, the increasing digitization of legal documents and court sentences, and the development of new techniques in natural language processing, enable analytics to improve judicial decision-making. India, for instance, has 27 million pending court cases; the sheer number of cases precludes manual classification of (often inconsistent) records and legal texts. ML algorithms can be trained to classify such records. This enables, for instance, analytics of bias and discrimination in courts (such as where judges with certain characteristics are associated with certain judicial outcomes in similar cases), or evaluations of how judicial reforms shape judicial productivity and bias. The chapter also describes the enabling environment for ML application—including how to build ML human capital and data infrastructure; the ethical considerations to keep in mind; and the importance of collaboration between ML engineers, domain experts, and the agencies that will use the technology to develop effective ML-based government analytics.

### Government Analytics Using Data on Task and Project Completion

Chapter 17, by Imran Rasul, Daniel Rogger, Martin Williams, and Eleanor Florence Woodhouse, discusses government analytics using task completion data. Much government work consists of the completion of tasks, from creating major reports to undertaking training programs and building infrastructure. A task completion approach allows for the investigation of which units and organizations are most likely to initiate, make progress on, and complete tasks—particularly where organizations complete similar tasks (such as preparing budgets). A task completion approach is particularly important to understand performance in administrative organizations in which the productivity data discussed in previous chapters (such as case data or frontline service delivery indicators) are not available.

**MAP 2.1** Variations in Productivity of Processing Social Security Cases, Subregions of Italy



Source: Fenizia 2022, using Italian Social Security Agency data.

Note: The key refers to the number of social security claims of a particular type that are processed by an office in a particular time period divided by the full-time equivalent of workers of that office during that time.

Data for task completion can be extracted from a variety of sources. In Ghana, for instance, government units complete quarterly progress reports on all activities, which can be digitized and repurposed into task completion data. In the United Kingdom, the National Audit Office completes major project reports that assess the progress of large infrastructure projects against corresponding planning documents. By subsetting—that is, culling—these data to similar tasks undertaken by all government units, units can be benchmarked on the average time it takes them to complete tasks and the share of uncompleted tasks, for instance. Matching these data to other data about these units—for instance, on budget disbursements or management practices—can help governments understand why some government units are more effective at completing tasks and identify drivers to improve productivity in lagging government units. Chapter 17 underscores the importance of—and provides guidance for—classifying task characteristics correctly (such as in terms of their complexity) to ensure that cross-unit benchmarking is valid.

### Cross-Cutting Insights from Part 3

The chapters in part 3 reveal some cross-cutting insights to keep in mind when undertaking government analytics using administrative data. To begin with, high-quality administrative data are not a foregone conclusion. The infrastructure for—and thus quality of—the underlying measurement and resulting data is paramount and requires careful consideration. The first challenge is data coverage. A central payroll system, for instance, might cover only part of the public employment workforce (chapter 10); and a central financial management information system might cover only part of government expenditures (chapter 11). A second challenge is data validity. To illustrate, case data in tax or social security are often not recorded or designed to measure performance. Careful thought is needed to repurpose such data for performance measurement. Self-reported data—as in the case of some task completion data—may also suffer from inaccurate or manipulated data entry by the units being evaluated, putting a premium on independent, third-party data collection or validation (chapter 7). A third challenge for performance data in particular is completeness relative to the mission of an organization. As chapter 14 notes, missions and goals of public sector organizations are typically multidimensional (and at times contradictory). This requires the triangulation of multiple data sources to ensure performance is measured holistically, so that government analytics does not incentivize the attainment of one goal of an organization at the expense of another (chapter 4).

The chapters also emphasize the human capital requirements for government analytics using administrative data. Creating the information technology (IT) systems that underlie centralized data infrastructures requires IT and data science skills (chapters 9 and 16). Processing records into usable data, analyzing the data, and making the data accessible for analysis and management improvements (such as through dashboards) similarly require data science and visualization skills. In some governments, advanced data science skills (such as for machine learning) might be in short supply. Upskilling data scientists in government, or creating data science teams for government analytics, is thus important to make the most of government analytics opportunities.

Lastly, the chapters also emphasize the importance of data accessibility for decision-makers. Creating dashboards to allow users to explore key insights from the data—for instance, on procurement indicators or HR—facilitates such accessibility. These can be complemented by briefings and slide decks with key diagnostic findings and insights and data storytelling for senior policy makers to take action. In other words, government analytics data do not speak for themselves, but need to be made understandable to support government action.

## PART 4: GOVERNMENT ANALYTICS USING PUBLIC SERVANT SURVEYS

Part 4 focuses on a single data source for government analytics: surveys of public servants. There are three reasons for dedicating an entire section to one data source.



First, surveys of public servants are one of the most widely used data sources for government analytics. The review of the global landscape in chapter 18 finds that the number of governments implementing governmentwide surveys of public servants repeatedly every year or two has increased continuously over the last two decades. At least nine governments of member countries of the Organisation for Economic Co-operation and Development (OECD) were conducting annual or biannual surveys as of 2021, and many others are surveying their public servants on a more ad hoc basis.

Second, surveys of public servants can be costly in terms of staff time in a way that repurposing administrative data is not. Such surveys often sample a census of (that is, all) government employees. The staff time cost of completing the US federal government employee survey reaches US\$30 million annually (as cited in chapter 20). It is therefore important to design such surveys to be as efficient and effective as possible.

Third and more important, many key features of public administration production cannot be measured efficiently through other data (such as administrative data or citizen surveys). For example, understanding how public servants are managed, their motivations, and their behaviors are all internal to the official's lived experience, yet matter for public sector productivity. Public employees' motivations are difficult to observe outside of their own expressions of their motives. Thus, self-reporting through surveys becomes the primary means of measurement for many aspects of the public administration production function, and serves as a lever for improving public sector productivity.

For all these reasons, effectively designing, implementing, and making the most of surveys of public servants for government improvement is crucial.

## Surveys of Public Servants: The Global Landscape

In chapter 18, Ayesha Khurshid and Christian Schuster review current government practices in surveys of public servants. The chapter finds that governments undertake such surveys with relatively similar objectives, and thus most governments tend to measure similar concepts in their surveys. These include, on the one hand, measures of core employee attitudes correlated with productivity, such as job satisfaction and engagement, commitment to the organization, and intent to remain working for the organization. On the other hand, governments measure a relatively consistent set of management practices as antecedents of these employee attitudes, such as the quality of leadership, performance management, and training.

Yet chapter 18 finds that governments differ in how they design, implement, and report on surveys of public servants. For instance, the wording of survey questions differs, even when similar concepts are being measured. Approaches to sampling public servants differ, as do survey modes or approaches to dealing with nonresponse. Governments also differ widely in how they report survey results: for instance, in terms of what kind of benchmarks are reported or what levels of hierarchy inside organizations are measured and provided with results reports. Given that governments undertake surveys with similar objectives, why is there such diversity in how they design, implement, and report on surveys?

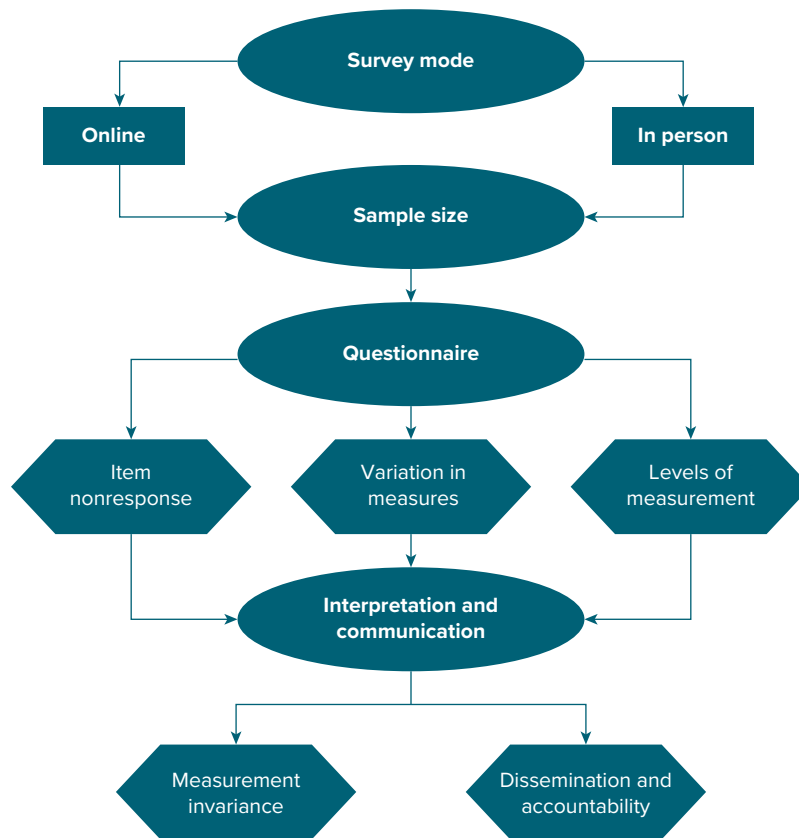
The answer, arguably, lies in part in the limited evidence available to governments that could guide choices about design, implementation, and reporting of surveys of public servants. The chapters in part 4 thus provide novel empirical evidence to enable governments and practitioners to make more evidence-based choices in response to some of these and other methodological questions in public servant surveys.

The decision tree pictured in figure 2.4 structures the choices facing governments in surveying public servants. This decision tree should *not* be read as a linear set of steps: there are interdependencies between choices. For instance, how many respondents need to be sampled depends on the expected variation in survey measures, which in turn is a function of questionnaire design.

Nonetheless, a common first choice concerns a survey mode: Are surveys conducted online, on paper, in person, by phone, or through a combination of these modes? Governments then need to determine the appropriate survey population, and an approach to sampling respondents, including determining the desired sample size given the purpose of the survey. Subsequently, questionnaires need to be designed. While measures may vary across concepts, several general concerns apply across them. For instance, how can measures be designed so that public servants are willing to answer questions (thus avoiding item nonresponse)?



**FIGURE 2.4** Decision Tree for Surveys of Public Servants



Source: Original figure for this publication.

How can measures be designed that vary sufficiently, so that comparisons between organizations or groups of public servants on these measures become meaningful? And should survey measures inquire about the individual experience of public servants themselves or ask them about their perceptions of practices in the organization as a whole?

Finally, governments need to decide how to interpret and report results. For instance, can responses from different groups—such as public servants in different countries, organizations inside a country, or demographic groups inside a country—be meaningfully compared? Or might concepts such as job engagement mean different things to different public servants (even when answering the same question), so benchmarking is not valid? Once decisions about who to benchmark are made, how can results be reported effectively to enable action? For instance, how are survey results presented to decision-makers, and who receives results? How is capacity built to enable decision-makers to take action based on survey results, and how are they incentivized to do so? The chapters in part 4 provide novel evidence on each of these key questions.

### Determining Survey Modes and Response Rates: Do Public Officials Respond Differently to Online and In-Person Surveys?

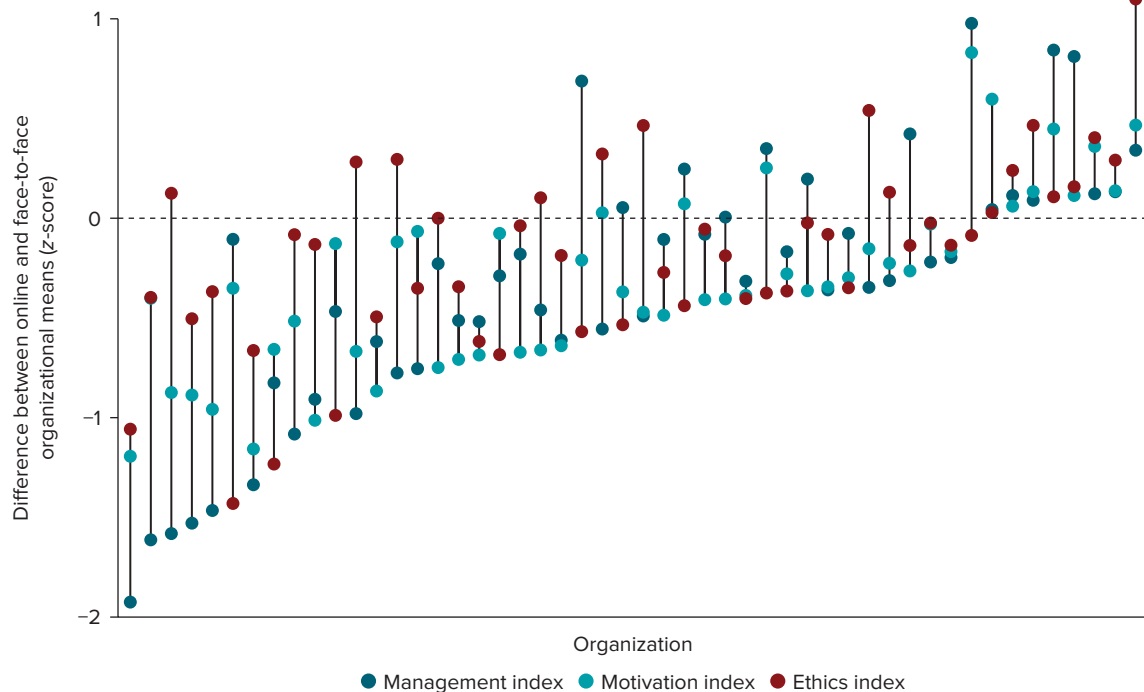
Chapter 19, by Xu Han, Camille Parker, Daniel Rogger, and Christian Schuster, assesses the first methodological choice in the decision tree: which enumeration method or survey mode to choose. This matters because different survey modes may come with different response biases to questions and different overall response rates. In OECD governments, surveys of public servants are typically conducted online, though not exclusively so. All nine government surveys reviewed in chapter 18 are implemented online, although,

to enhance accessibility (for instance, for staff who have difficulty accessing or completing an online survey), Colombia, Switzerland, the United Kingdom, and a few agencies in Australia also offer their survey in a paper format, while New Zealand offers its survey through paper and telephone upon request. The advantage of a predominantly online approach to surveying public servants across governments is clear: it reduces costs and may reduce biases, such as those induced by respondents' notions of socially desirable answers when faced with an in-person or phone interviewer.

Online surveys, however, also tend to have lower response rates than other survey modes, such as in-person surveys. For instance, the last US Federal Employee Viewpoint Survey had a response rate of 44 percent. This raises a concern that the data resulting from an online survey are not a valid representation of the population—in the case of the United States, the entire federal public administration. In some instances, these concerns about the validity of online surveys of public servants become severe. In the United Kingdom, the validity and quality of the underlying data of the Civil Service People Survey was questioned in a parliamentary inquiry, in part motivated by low response rates in some government organizations (UK Parliament 2022).

To what extent are low response rates in online surveys a concern (thus putting a premium on survey modes with higher response rates, such as in-person surveys)? To find out, chapter 19 presents evidence from a randomized control trial that compares face-to-face and online survey responses from Romanian public servants. The face-to-face surveys had consistently high response rates across Romanian government organizations, while the response rates for the online surveys varied across organizations, as is typical in other governments. The results suggest that these diverging survey modes do not substantially affect aggregate estimates at the national level. They do, however, affect the comparability of findings across organizations. Figure 2.5, reproduced from chapter 19, shows how big of a difference the mode of survey makes for indexes of survey topics. For some organizations, the impact of the survey mode is substantial.

**FIGURE 2.5 Average Difference between Survey Modes for Different Topics across Romanian Government Organizations**



Source: Original figure for this publication (see chapter 19).

Note: The figure shows, by organization in the Romanian government, the difference in the management index, motivation index, and ethics index between online and face-to-face survey respondents.

Basic organizational and demographic characteristics explain little of the variation in these effects. In other words, survey weights are not effective in addressing these effects.

Governments that offer varying survey modes should thus be careful when comparing the scores of organizations if some implement the survey primarily online while others implement it primarily based on pen and paper. Rankings of organizations in such instances do not appear to be valid. Nonetheless, chapter 19 does not find evidence to suggest that the (lower-response) online survey mode biases national-level inferences from the survey of public servants in Romania. More research is required to confirm the external validity of this finding in other countries.

In a second step, governments need to define survey populations and their sampling approach. Who the appropriate survey population is, of course, depends on the government's measurement objectives. The global review in chapter 18 shows that the survey population generally consists of central government civil servants, albeit with variations in the extent to which public sector organizations and types of employee contracts outside the (legally defined) civil service are also covered—for instance, in other branches of government or frontline services. To cite just one example, for the United Kingdom's government employee survey, all public servants from 101 agencies are eligible, excluding the Northern Ireland Civil Service, the National Health Service (NHS) (which conducts its own survey), and frontline officials (such as police officers and teachers) (UK Cabinet Office 2022).

Who, then, should be sampled within this survey population? As chapter 18 shows, approaches to sampling vary across governments, ranging from census to random, ad hoc, and stratified sampling. Australia, Canada, Ireland, New Zealand, and the United Kingdom adopt a census approach where all eligible public sector employees may participate in the survey. Switzerland and the United States use (stratified) randomized sampling approaches for most years but conduct a census every few years. Colombia's public servant survey uses a mixed approach: for larger organizations, a stratified sampling approach is used, while for smaller organizations a census is taken to protect anonymity. The Republic of Korea adopts a sampling approach for all annual surveys.

### **Determining Sample Sizes: How Many Public Officials Should Be Surveyed?**

As noted in chapter 20, by Robert Lipinski, Daniel Rogger, Christian Schuster, and Annabelle Wittels, determining the appropriate sample of a public administration survey is often a trade-off between increasing the precision of survey estimates through greater sample sizes and the high costs of surveying a larger number of civil servants. Greater precision enables both more precise benchmarking (such as between organizations) and survey result reports at lower levels of hierarchy in an organization. Less precision reduces the staff time lost responding to the survey.

How can this trade-off be resolved? Chapter 20 shows that, ultimately, survey administrators must decide on the sample size based on the type of inferences they want the survey to yield and the staff time they can claim for the survey. By employing Monte Carlo simulations on survey data from Chile, Guatemala, Romania, and the United States, chapter 20 shows that governmentwide averages can be reliably derived using sample sizes considerably smaller than those used by governments currently. On the other hand, detecting differences between demographic groups (such as gender and rank) and, in particular, ranking individual public administration organizations precisely requires larger sample sizes than are collected in many existing surveys.

These results underscore, on the one hand, the importance of not overinterpreting the substantive significance of small differences between organizations in public servant survey results (or individual ranks of organizations in results). On the other hand, the results emphasize that governments should determine sample sizes based on the type of inferences and benchmarking exercises they wish to make with the data. Existing governmental surveys of public servants do not seem to be based on such a data-driven approach to sampling. Chapter 20 addresses this gap and offers an online sampling tool to enable such sampling.

In a third step, surveys of public servants require the definition of a questionnaire. Part 4 sheds light on three cross-cutting dimensions of questionnaire design: How can measures be designed that vary sufficiently

so that comparisons between organizations or groups of public servants become meaningful? How can measures be designed so that public servants are willing to answer? And should survey measures inquire about the individual experience of public servants or instead ask them about their perceptions of practices in the organization as a whole?

### **Designing Survey Questionnaires: Which Survey Measures Vary and for Whom?**

A first prerequisite for effective question design is variation: Survey measures should provide a sufficient degree of discriminating variation across respondents to be useful—or, in other words, sufficient variation to understand differences between key comparators, such as organizations or demographic groups. Without discriminating variation across organizations, demographic groups, or countries, survey measures cannot inform governments about strengths and areas for improvement. With this in mind, chapter 21, by Robert Lipinski, Daniel Rogger, Christian Schuster, and Annabelle Wittels, assesses variation in a set of typical indicators derived from data sets of public service surveys from 10 administrations in Africa, Asia, Europe, North America, and South America.

The results show that measures related to personal characteristics such as motivation do not vary as much as those relating to management practices such as leadership. When respondents are asked to assess practices of others, such as their superior or their organization, survey responses exhibit significant discriminant variation across organizations and groups. By contrast, when respondents are asked to self-assess whether they possess desirable characteristics such as work motivation, survey responses across countries tend to be heavily skewed toward favorable answers, and variation is so compressed that meaningful differences between organizations or demographic groups are difficult to detect. Standard measures for desirable attitudes, such as motivation, may therefore need to be redesigned in surveys of public servants to better discriminate between values at the top end of indexes.

### **Designing Survey Questionnaires: To What Types of Survey Questions Do Public Servants Not Respond?**

Chapter 22, by Robert Lipinski, Daniel Rogger, and Christian Schuster, shows that surveys of public servants differ sharply in the extent to which respondents skip or refuse to respond to questions. So-called item nonresponse can affect the legitimacy and quality of public servant survey data. Survey results may be biased, for instance, if those least satisfied with their jobs are also most prone to skipping survey questions. Understanding why public servants respond to some survey questions but not others is thus important.

The chapter offers a conceptual framework and empirical evidence to further this understanding. Drawing on other survey methodology research, the chapter theorizes that public servants are less likely to respond to questions that are too complex (because they are unable to answer them) or sensitive (because they are unwilling to respond). Coding the complexity and sensitivity of public servant survey questions in Guatemala, Romania, and the United States, chapter 22 finds one indicator of complexity to be the most robust predictor of item nonresponse across countries: respondents' lack of familiarity with the information or topic examined by a survey question. By contrast, other indicators of complexity or sensitivity or machine-coded algorithms of textual complexity do not predict item nonresponse. The implication for survey design is clear: Avoid questions that require public servants to speculate about topics with which they are less familiar.

### **Designing Survey Questionnaires: Should Surveys Ask about Public Servants' Perceptions of Their Organization or Their Individual Experience?**

A third prerequisite for effective questionnaire design is valid measurement of organizational aggregates, such as which surveyed organization has the highest level of job satisfaction among its employees or

which organization has the highest quality of leadership practices of superiors. This raises the issue of whether respondents should be asked about their perceptions of organizational practice overall (so-called organizational referents) or whether questions should ask about the respondent's own experience, such as the quality of their superior's leadership practices or their own job satisfaction (so-called individual referents). In chapter 23, Kim Sass Mikkelsen and Camille Mercedes Parker examine this question using survey experiments with public servants in Guatemala and Romania. The survey experiments randomly assign public servants to respond to questions about a topic with phrasing using either an organizational referent or an individual referent.

The chapter finds that, while there are no strong conceptual grounds to prefer either organizational or individual referents—both have advantages and disadvantages—the choice matters to responses and alters response means (such as the average job satisfaction of employees in an organization). Organizational questions are particularly useful when questions are very sensitive (such as on corruption) because respondents may skew their response more strongly toward a socially desirable response option when asked about their own individual experience rather than practices in the organization. Individual questions are particularly useful when the attitudes or practices being measured are rare in the organization. In such cases, many respondents may lack the information to accurately assess the prevalence of a practice in the organization, risking that they rely instead on unrepresentative information or stories, for instance, rather than actual organizational characteristics. In short, whether survey questions should ask about the individual's own experience or the individual's perception of organizational practice depends on the characteristics of the question and the organization it seeks to measure.

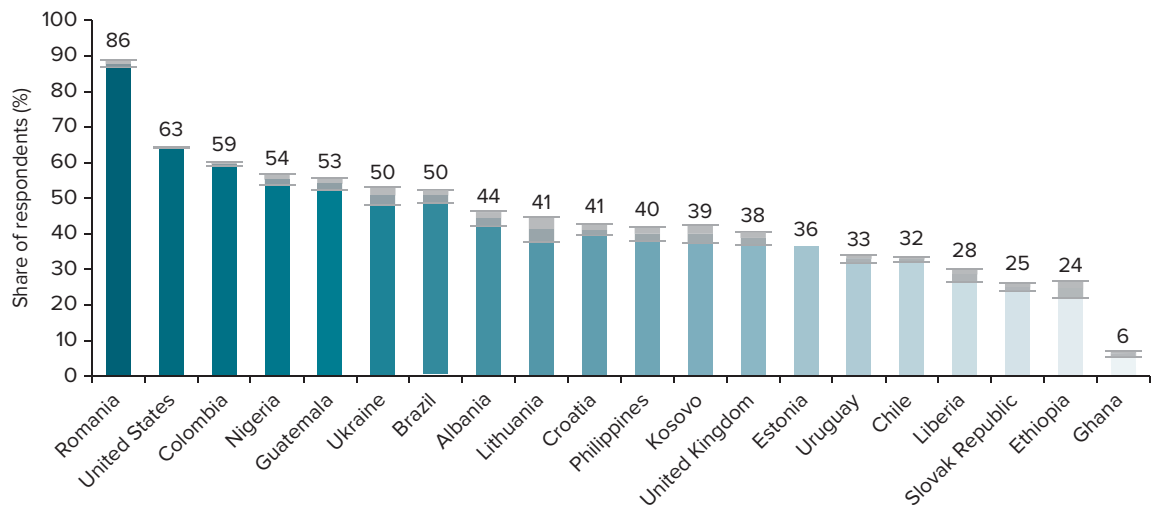
### Interpreting Survey Findings: Can Survey Results Be Compared across Organizations and Countries?

Chapters 24 to 26 turn to the interpretation and reporting of survey results. In chapter 24, Robert Lipinski, Jan-Hinrik Meyer-Sahling, Kim Sass Mikkelsen, and Christian Schuster focus on interpretation and in particular the question of benchmarking: Can survey results be compared across organizations and countries? This matters because survey results can rarely be understood in a void. Rather, they require benchmarks and points of reference. If, for instance, 80 percent of public servants are satisfied with their jobs, should a manager interpret this as a high or low level? Historical comparisons provide a sense of dynamics over time, but not a sense of degree. The availability of similar statistics from comparator organizations in the public sector or other countries is a potentially valuable complement to a manager's own results. However, such benchmarking requires that survey questions measure the same concept in the same way, making meaningful comparisons possible. Even when questions are phrased in the exact same way, however, the validity of comparison is not obvious. For multiple reasons, including work environment, adaptive expectations, and cultural factors, different people might understand the same question in distinct ways and adjust their answers accordingly. This might make survey results incomparable not only across countries but also across different groups of public servants within a national public administration.

To assess this concern empirically, chapter 24 investigates to what extent the same survey questions measure the same concept similarly—that is, questions are measurement invariant—using questions related to “transformational leadership” and data from seven public service surveys from Europe, Latin America, and South Asia. The chapter finds support for so-called scalar invariance: the topic (in this case, means of transformational leadership) can be compared *within* countries across organizations and demographic groups (the chapter authors test for gender and educational levels). Across countries, the chapter finds tentative evidence for scalar invariance, and stronger evidence when countries are grouped by regions and income.

The findings—although tentative and requiring further confirmatory evidence from other settings—thus underscore the utility of global benchmarking of countries in surveys of public servants.<sup>3</sup> As chapter 18 explains, the Global Survey of Public Servants offers a tool to harmonize questions across governments. In conjunction with the freely accessible Global Survey Indicators dashboard, the Global Survey of Public

**FIGURE 2.6 Share of Public Servants Satisfied with Their Pay and/or Total Benefits, Various Countries**



Source: Fukuyama et al. 2022, figure 9.

Note: Years of measurement vary by country. Colors denote the extent of job satisfaction, with darker shades signifying greater job satisfaction. The gray vertical bars denote 95% confidence intervals.

Servants thus enables governments to understand strengths and areas for development of their public administration in global comparative terms. For example, the Global Survey provides comparative data on pay satisfaction from public administrations around the world. As can be seen from figure 2.6, this varies greatly across countries.

Once decisions are made about whom to benchmark against, consideration turns to how to report and disseminate survey results—that is, how to make the most of survey results. Chapters 25 and 26 provide two complementary perspectives on this challenge.

### **Making the Most of Public Servant Survey Results: Lessons from Six Governments**

Chapter 25—by Christian Schuster, Annabelle Wittels, Nathan Borgelt, Horacio Coral, Matt Kerlogue, Conall Mac Michael, Alejandro Ramos, Nicole Steele, and David Widlake—presents a self-assessment tool that lays out low-cost actions governments can take to support evidence-based reforms based on the insights from public servant surveys. The chapter applies this tool to governments of six countries (Australia, Canada, Colombia, Ireland, the United Kingdom, and the United States) to assess the comprehensiveness of their ecosystem to turn survey results into management improvements.

The self-assessment tool focuses on three main components of an ecosystem to turn survey results into management improvements: (1) information, (2) capacity, and (3) incentives to take action. For the first component (information), public servant survey results can improve public administration by providing tailored survey results to four main types of users: the government as a whole; individual public sector organizations; individual units or departments within a public sector organization; and the public, including public sector unions. Results reporting should identify key takeaways about the strengths and weaknesses of particular organizations and enable users to explore aggregate survey results in a customized manner, such as through dashboards.

For the second component (capacity to take action), reporting of results can become more effective when it includes (automated) recommendations to users—such as managers of units or organizations—on how to best address survey findings, as well as action plans for users to develop their own actions. Where more



resources are available, tailored technical assistance—or a human resources management (HRM) consultancy, provided either by a central HR unit or an external provider—can further help managers turn survey findings into improvements.

For the third component (incentives to take action), accountability mechanisms are key. For instance, governments can introduce central oversight of actions taken in response to survey findings by government organizations and units; can construct and publicize “best place to work” in government indexes to foster external oversight; or measure employee perceptions of the extent to which government organizations take action in response to survey findings.

Applying this self-assessment framework to the six governments, chapter 25 finds that many governments could undertake a range of additional low-cost actions to enhance the benefits they derive from public servant surveys to improve public administration.

## Using Survey Findings for Public Action: The Experience of the US Federal Government

Chapter 26—by Camille Hoover, Robin Klevins, Rosemary Miller, Maria Raviele, Daniel Rogger, Robert Seidner, and Kimberly Wells—complements chapter 25 by delving into the experience of the United States, the country with the longest-standing governmentwide employee survey. The chapter emphasizes, first, the importance of considering action in response to survey results at the time the survey is being designed. In particular, questions should focus on topics that staff and senior leaders find most important to achieve their mission. Second, the chapter describes the critical architecture necessary in each public sector organization to translate survey results into management improvements. This includes, for instance, a technical expert in the organization capable of interpreting survey data; a strong relationship between that expert and a senior manager in the organization who acts as a “change champion”; and the development of a culture for initiatives for improvements informed by the survey.

The chapter also provides guidance on how to develop a culture of responsiveness to surveys of public servants. It emphasizes the importance of leaders in an organization being transparent in sharing and discussing the survey results with their workforce, codeveloping action plans with staff, and coproducing improvements in response to survey results.

Part 4 thus provides evidence-based good practice on a range of choices involved in designing, implementing, interpreting, and reporting on surveys of public servants. This evidence can inform actions by governments seeking to improve their existing regular governmentwide employee surveys, as is the case in many OECD countries. It can enable governments that have yet to introduce surveys of public servants to leapfrog to best practice from the start. Such governments are encouraged to consult the range of toolkits on the Global Survey of Public Servants toolkit site.<sup>4</sup>

## PART 5: GOVERNMENT ANALYTICS USING EXTERNAL ASSESSMENTS

Part 5 turns to select data sources available to undertake government analytics through external assessments: that is, assessments conducted on or by those outside government (rather than data on public servants or administrative data collected by government organizations themselves). Chapters 27 through 30 provide guidance on four external data sources: household survey data, citizen survey data, service delivery indicators, and anthropological methods. These data sources illustrate the possibility of government analytics through external assessments but do not cover the full range of microdata for external assessments. Enterprise surveys of businesses, for instance, can provide insights on topics such as bribery or government regulation (World Bank 2023). With that caveat in mind, the part 5 chapters provide important insights about how to do government analytics using external assessments.



## Government Analytics Using Citizen Surveys: Lessons from the OECD Trust Survey

Chapter 28, by Monica Brezzi and Santiago González, examines government analytics using citizen surveys. These surveys can help governments shed light on certain outcomes in the public administration function. In particular, they can capture the outcomes of public governance as perceived and experienced by people, through nationally representative population samples. For instance, citizen surveys are used in many countries to measure satisfaction with widely used public services (such as tax administrations, schools, or hospitals).

As chapter 28 shows, they can also be used to understand broader government outcomes. In particular, the chapter illustrates the potential of such surveys for government analytics using the example of the OECD's Survey on the Drivers of Trust in Public Institutions (OECD Trust Survey). The survey measures trust of citizens in government and captures their expectations of and experiences with public institutions around key drivers of trust. Measures of trust and its drivers include the competence of public institutions—including access to public services and their quality and reliability—as well as the perceived values of public institutions, notably in terms of integrity, openness to involving citizens, and fairness in the treatment of citizens.

Chapter 28 showcases how governments have used evidence resulting from the survey to develop concrete actions to strengthen institutional trust. The chapter provides guidance for other governments wishing to apply this international benchmark on measuring trust in public institutions, following the OECD Guidelines on Measuring Trust.

## Government Analytics Using Measures of Service Delivery

In chapter 29, Kathryn Andrews, Galileu Kim, Halsey Rogers, Jigyasa Sharma, and Sergio Venegas Marin go beyond this book's core focus on public administration to the frontline of service delivery and introduce measures of service delivery (MSDs). Currently, MSDs are focused on education and health. Mirroring this *Handbook's* approach to conceptualize government analytics along the public administration production function from inputs to outcomes, MSDs provide objective measurements not only of service quality (such as absenteeism of medical doctors, and test results of students in school) but the entire process involved in delivering frontline public services (including input and process measures such as availability of medicine, and management practices in schools).

MSDs offer a granular view of the service delivery system, providing actionable insights on different parts of the delivery chain, from the physical infrastructure to the knowledge of frontline providers. These can be usefully viewed as outputs and “outcomes” of the administrative environment embedded in government agencies under which these frontline providers fall. Measurement of these different factors of production allows practitioners to map out, conceptually, how each part of the production chain is faring, and where improvements can be made, at the individual provider level as well as part of the wider production function for government outlined in figure 2.1. MSDs also provide action-oriented visualizations of these indicators, enabling practitioners to design their service delivery policies in an intuitive and evidence-based approach.

Chapter 29 provides a road map to interested practitioners to produce MSDs, from design, implementation, and analysis to dissemination. The chapter emphasizes that developing service delivery indicators requires considering and defining relevant dimensions of quality in a public service, along with relevant inputs, with a premium on defining indicators according to policy objectives and resource constraints. Akin to the discussions in chapters 25 and 26, chapter 29 also underscores the importance of linking MSDs directly to stakeholders who have the ability to enact change in the delivery system. Many of the steps involved in the analytics of core public administration are thus mirrored in the analytics of service delivery.

## Government Analytics Using Anthropological Methods

Finally, chapter 30, by Colin Hoag, Josiah Heyman, Kristin Asdal, Hilde Reinertsen, and Matthew Hull, returns to chapter 4's call for embedding qualitative studies in government analytics. It considers how government analytics can be undertaken through an anthropological approach, a powerful means of collecting qualitative data. Anthropologists are most commonly associated with immersive, ethnographic methods such as participatory observation. Chapter 30 applies that lens to studying public administration. As the chapter authors emphasize, "Anthropologists are motivated by an abiding concern with empirical rigor—a refusal to ignore any sort of data or to content oneself with a single view of such a multifarious thing as bureaucracy." Doing so risks overlooking factors that shape organizations.

Anthropological methods suggest that data collection should approach government analytics by engaging with the staff who are involved at every level of the organization, from senior officers to low-level staff and contractors, and across different demographic groups; studying everyday documents; and watching how officials interact. By observing every part of what public officials do at work in a holistic way, from their interactions in corridors and meetings to the protocols they observe in their relationships, the analyst undertakes the most holistic data collection strategy feasible.

Such an approach requires analysts to develop relationships with a variety of types of people in an organization and have open-ended conversations about their work and unrelated issues to understand their values and perspectives. It also requires analysts to engage in participant observation to capture activities that may be so routine they go unnoticed by public officials and are not self-reported in surveys. Moreover, it requires analysts to collect the widest practical range and amount of qualitative and quantitative data, even if such data cannot be easily standardized. Finally, it requires analysts to study not only data but also the interactions and microscopic decisions that affect the gap between stated policy goals and the actual work being carried out by public officials—for instance, by studying what public officials say and do, including the rationales they draw on for their decisions.

Chapter 30 thus emphasizes that government analytics can incorporate methods that provide insights into aspects of the public administration function that quantitative microdata cannot. It also brings the *Handbook* back to the first cross-cutting chapter (chapter 4), which emphasizes the importance of a holistic and "balanced data suite." Part of this "suite" is ensuring that problem analysis in public administration is holistic: that important parts of a problem are not neglected due to the absence of quantitative data and measurement. This, in turn, puts a premium on utilizing qualitative and anthropological methods to complement insights gleaned from microdata.

Holistic analytics benefit not only from triangulating different quantitative and qualitative methods, but also from triangulating and integrating the analytics of different components of the public administration production function. The chapter concludes with a challenge: How can the different approaches and data sources in the government analytics toolbox be integrated effectively to diagnose major challenges in public administration holistically?

## CONCLUSION: TOWARD A HOLISTIC ANALYTICS OF CORE CHALLENGES IN THE MACHINERY OF GOVERNMENT

How can practitioners take the distinct data sources detailed in the *Handbook* to the frontier? Frontier government analytics would integrate the analytics of the data sources described across the *Handbook* into standard government practice. It would generate them at a scale sufficient to inform the decision-making of individual managers. And it would make them easily accessible to those managers across government organizations and departments. For instance, dashboards integrating data sources and updating in real time would provide managers with comparisons for their staffing issues, process quality, the perceived quality of

management practices, and so on. They could keep tabs on outputs and outcomes, from task completion and case productivity to external assessments from citizens. Comparative data would allow them to benchmark themselves against other government organizations, or where appropriate, other countries. Managers would be capable of understanding the limitations and strengths of different analytics. The result would be a transformational change toward leveraging data to strengthen public administration.

Where should the journey toward this frontier begin? As a first step, the government analytics of the individual data sources explored in detail in various chapters can provide important insights into different components of the public administration production function. Once governments have analytical tools for several components in place (such as payroll diagnostics, and surveys of public servants), the possibilities for government analytics further expand. The integration of multiple government analytics data sources enables governments to diagnose and address major government challenges holistically (chapter 9). In some cases, such as customs (chapter 14), such integration is a vital part of measuring attainment of the organization's mission. In others, the insights that can be gleaned go beyond those available from individual data sources and thus enable a holistic perspective on public administration.

To illustrate, consider how the integration of government analytics data sources described in this *Handbook* can shed light on key challenges in public administration: corruption and personnel management.

Corruption is a multidimensional phenomenon, affecting public administration across its production function. As chapter 8 discusses, corruption can first be detected with input data. For example, payroll data can be drawn on to detect ghost workers. HRMIS data can be drawn on to detect instances of nepotism in recruitment, such as when family members with similar last names are hired. Procurement data can be drawn on to detect procurement fraud and collusion risks (such as when organizations grant contracts without competitive bidding). Expenditure data can be drawn on to detect off-budget spending at risk of embezzlement.

Second, corruption can be detected in the processes and practices that convert inputs into outputs and generate norms and behaviors in public administration. For instance, surveys of public servants can measure unethical leadership by superiors (such as pressures on subordinates to collude in corruption), as well as the ethical norms and integrity of public servants themselves (such as their perceptions of corruption of colleagues).

Third, corruption can be detected in output and outcome data (such as in tax audit case data, or through surveys of citizens querying them about bribery requests from public officials). Understanding at a granular level where corruption occurs—in a specific public administration production function, or in a particular department or organization—enables governments to identify comprehensive but tailored and evidence-based solutions to curb corruption.

Consider, as a second example, civil service management—and how the integration of analytics across the public administration production function can aid better personnel management in government. As chapters 9 and 10 emphasize, data analytics of payroll and HRMIS systems can diagnose personnel as an input into the production function. To cite two examples: pay equity can be assessed across groups (such as by gender) and institutions; and retention challenges can be pinpointed by comparing turnover rates for public servants at different ranks or in different regions of the country. Such data can be complemented by labor force and household survey data (chapter 27)—for instance, to understand whether public sector pay is competitive relative to the private sector, or whether the workforce allocation across the territory is appropriate given the respective number of service users. Other input data, such as on budgets, can help understand whether productivity problems might arise from the lack of complementary inputs to personnel in the production function.

Data on processes and practices can then shed light on whether government is effectively converting personnel inputs into outputs and outcomes. Surveys of public servants, as explored in part 4, can measure the experience of public servants with management practices, as well as the culture, attitudes, and behaviors in public administration that mediate the conversion of inputs into outputs. For instance, is leadership by line managers effective? Are public servants motivated to work hard? Anthropological and qualitative

methods, as discussed in chapter 30, can enrich and contextualize survey responses through participant observation and detailed case study work. As chapter 13 shows, government analytics can also assess processes converting inputs into outputs through administrative data—for instance, to assess whether institutions follow due procedure when evaluating the performance of public servants.

Last, data on outputs and outcomes can help diagnose and improve personnel management. For instance, as chapter 17 explains, data on task completion can help understand where in government public servants are effectively completing tasks and where they are not. For public sector organizations that complete cases, administrative case data, as explored in chapters 14 and 15, can help understand differences in the productivity of public servants more broadly, while surveys of citizens can help understand differential citizen satisfaction and trust with public services across institutions and regions in a country.

### Strengthening Public Sector Management through Government Analytics

At the center of government are the public officials who navigate the strategic and daily decisions that determine public policy and the effectiveness of public administration. The information public managers have, and the extent to which they use it for running public organizations, drives the efficacy of government. An effective system of government analytics empowers officials to manage government based on evidence of the administration's current realities.

Empowering public officials with government analytics can transform the basis for personnel management—and public administration more generally. It enables government decision-makers to complement their tacit and practical knowledge about public administration with evidence-based and data-informed insights to improve how the machinery of government is run.

Almost all officials manage, organize, or process in ways that could usefully be analyzed as an input to strengthening government functioning. Thus, communicating analytical insights to government officials at the time that they are making a decision, as discussed in chapters 25 and 26, is a critical link in making analytics effective. One approach is to support decision-making through a human resources management dashboard that brings together the distinct data sources a government has available to visualize and benchmark the strengths and areas for improvement in personnel management to decision-makers in each government institution. By implementing any of the approaches in this *Handbook*, government or its stakeholders are building a platform for managers and decision-makers to do more with the public sector's finite resources.

Even with 30 chapters, this book does not cover all potential data sources and approaches to government analytics. There will always be new analytical opportunities on the horizon. Chapter 3 discusses how government can continuously prepare for a transition to new analytical approaches. Above all, building the culture that binds public administrators and public administration together requires a commitment from senior management and individual officials to use evidence about their own administration in their decision-making. We invite all public servants and related stakeholders to capitalize on the insights summarized in this chapter and this *Handbook* and push forward the quality of management of their administration. The quality of government around the world will be shaped by how its decision-makers leverage data to strengthen their administration.



## ANNEX 2A MAPPING GOVERNMENT ANALYTICS DATA SOURCES TO CHAPTERS IN THE HANDBOOK

**TABLE 2A.1 Mapping Government Analytics Data Sources to Chapters 10–30**

Data source	Chap.	Examples of uses of analytics	Examples of indicators
<i>Part 3. Administrative data</i>			
Payroll and HRMIS	10	<ul style="list-style-type: none"> <li>Examine fiscal planning and sustainability of the wage bill.</li> <li>Allocate the workforce across government departments and territories.</li> <li>Set pay.</li> </ul>	<ul style="list-style-type: none"> <li>Wage bill by sector and years.</li> <li>Distribution of civil servants by rank.</li> <li>Turnover of civil servants.</li> <li>Pay inequity between ministries.</li> <li>Retirement projections.</li> </ul>
Expenditure	11	<ul style="list-style-type: none"> <li>Assess efficiency, equity, and effectiveness of government spending.</li> <li>Assess whether government spending corresponds to budget priorities.</li> <li>Identify large transactions with high fiduciary risk.</li> </ul>	<ul style="list-style-type: none"> <li>Budget execution rates.</li> <li>Share of government expenditures covered by the FMIS.</li> <li>Share of total expenditures by transaction value.</li> <li>Overdue accounts payable (payment arrears).</li> </ul>
Procurement	12	<ul style="list-style-type: none"> <li>Monitor procurement markets and trends.</li> <li>Improve procurement and contracting (for instance, by identifying goods organizations overpay for, or organizations with high corruption risks in procurement).</li> <li>Assess effects of distinct procurement strategies or reforms.</li> </ul>	<ul style="list-style-type: none"> <li>Time needed for contracting.</li> <li>Number of bidders.</li> <li>Share of contracts with single bidder.</li> <li>Final price paid for a good or service.</li> <li>Share of contracts with time overruns.</li> <li>Share of bidders that are small and medium enterprises.</li> </ul>
Administrative processes	13	<ul style="list-style-type: none"> <li>Assess the speed and quality of administrative back-office processes and process implementation (for instance, for project planning, budget monitoring, or performance appraisals).</li> </ul>	<ul style="list-style-type: none"> <li>Adherence of administrator's process work to accepted government procedure.</li> <li>Timeliness of administrator's process work with respect to deadlines.</li> </ul>
Customs	14	<ul style="list-style-type: none"> <li>Assess customs revenue collection.</li> <li>Assess trade facilitation (flow of goods) through customs across borders.</li> <li>Assess whether customs safeguards safety of goods and protects people (for instance, prevents dangerous goods from crossing).</li> </ul>	<ul style="list-style-type: none"> <li>Time delays in customs clearances.</li> <li>Cost of customs process.</li> <li>Total customs revenue collected.</li> <li>Number of goods in infraction seized in customs.</li> </ul>
Case data	15	<ul style="list-style-type: none"> <li>Assess productivity of organizations and individuals processing cases (such as tax claims).</li> <li>Assess the quality of case processing.</li> <li>Identify "best performing" offices and transfer best practices to other offices.</li> </ul>	<ul style="list-style-type: none"> <li>Total number of cases completed.</li> <li>Average time to complete one case.</li> <li>Error rate in completing cases.</li> <li>Timeliness of case completion.</li> <li>Share of cases with complaints.</li> </ul>

(continues on next page)

**TABLE 2A.1 Mapping Government Analytics Data Sources to Chapters 10–30 (continued)**

Data source	Chap.	Examples of uses of analytics	Examples of indicators
Text-as-data (machine learning)	16	<ul style="list-style-type: none"> <li>Analyze “big data” and convert unstructured text from government records into data for analytics.</li> <li>Examples of big data include payroll disbursements to civil servants or tax filings by citizens and firms.</li> <li>Examples of text data include court rulings, procurement tenders, or policy documents.</li> </ul>	<ul style="list-style-type: none"> <li>Risk score for a procurement tender or payroll disbursement based on a predictive algorithm.</li> <li>Bias score in court ruling, based on textual analysis of wording in the document (sexist or racial-profiling terms).</li> </ul>
Task and project completion data	17	<ul style="list-style-type: none"> <li>Examine the frequency of completion of tasks on schedule (such as delivery of training program, preparation of budget).</li> <li>Understand quality of administrative task completion.</li> <li>Understand drivers of differences in task completion across government units.</li> </ul>	<ul style="list-style-type: none"> <li>Number of tasks completed.</li> <li>Share of tasks completed on time.</li> <li>Share of unfinished tasks.</li> <li>Share of tasks completed as planned.</li> </ul>
<i>Part 4. Surveys of public servants</i>			
Civil service survey data	18–26	<ul style="list-style-type: none"> <li>Assess the quality of management practices (such as performance management, recruitment).</li> <li>Assess norms, attitudes, and cultures in public administration (such as work motivation).</li> </ul>	<ul style="list-style-type: none"> <li>Share of public servants hired through merit examinations.</li> <li>Share public servants wishing to leave the public sector in the next year.</li> <li>Share of public servants favorably evaluating the leadership practices of their superior.</li> </ul>
<i>Part 5. External assessments</i>			
Household survey data	27	<ul style="list-style-type: none"> <li>Assess wage differences between the public and the private sector.</li> <li>Identify regions with lower public sector wage premiums.</li> <li>Assess gender representation in the public sector.</li> </ul>	<ul style="list-style-type: none"> <li>Average public sector wage premium.</li> <li>Share of women in public employment.</li> <li>Share of public employment in total employment.</li> </ul>
Citizen survey data	28	<ul style="list-style-type: none"> <li>Assess satisfaction of citizens with public services.</li> <li>Assess trust of citizens in government institutions.</li> <li>Assess interactions of citizens with public administration (such as bribery requests).</li> </ul>	<ul style="list-style-type: none"> <li>Share of citizens satisfied with health services.</li> <li>Share of citizens trusting the civil service.</li> <li>Share of citizens asked for a bribe by a government official in the last year.</li> </ul>
Service delivery indicators	29	<ul style="list-style-type: none"> <li>Assess the quality of education and health service delivery.</li> <li>Assess facility-level characteristics (such as teacher presence).</li> </ul>	<ul style="list-style-type: none"> <li>Share of days during which teachers are not present in school.</li> <li>Availability of resources in hospital to treat patients.</li> </ul>
Anthropological analytics	30	<ul style="list-style-type: none"> <li>Observe everyday practices in public administration, to capture routine but unnoticed parts of administration.</li> <li>Observe social engagement to understand formal and informal rules, relationships, and interactions between public servants.</li> </ul>	<ul style="list-style-type: none"> <li>Holistic participant observation of everyday life inside particular public administrations.</li> <li>Understand how public servants interpret broader policy goals.</li> </ul>

Source: Original table for this publication.

Note: FMIS = financial management information system; HRMIS = human resources management information system.

## NOTES

1. In contrast to the relatively coherent consensus of functions of private sector production (Mas-Colell, Whinston, and Green 1995), no consensus has formed around an integrated model of a production function for public administration. This is due in part to the limited use of microlevel data on the workings of different components of public administration—the very gap that motivated this *Handbook*.
2. Whether inputs effectively convert into outputs is also moderated by exogenous factors, such as the political environment. This *Handbook* does not discuss the analysis of microdata to assess these exogenous factors. Instead, readers are encouraged to consult the many existing excellent resources to understand the exogenous and political environment of public administrations (see, for example, Moore 1995).
3. Read in conjunction with other research cited in chapter 24, this conclusion holds particularly for questions that are more factual and less culturally specific. For instance, questions on specific management practices (such as around the presence of certain performance evaluation practices) can be more plausibly benchmarked across countries without measurement invariance concerns than attitudinal questions (such as on how engaged employees are).
4. The Global Survey of Public Servants was cofounded by the two editors of this *Handbook*, along with a range of practitioners and academic colleagues. The toolkits on the website build upon and incorporate much of the evidence reviewed in this *Handbook*. See <https://www.globalsurveyofpublicservants.org>.

## REFERENCES

- Bakshi, H., A. Bravo-Biosca, and J. Mateos-Garcia. 2014. “The Analytical Firm: Estimating the Effect of Data and Online Analytics on Firm Performance.” Nesta Working Paper 14/05, Nesta, London.
- Fenzia, A. 2022. “Managers and Productivity in the Public Sector.” *Econometrica* 90 (3): 1063–84.
- Fukuyama, F., D. Rogger, Z. Husnain, K. Bersch, D. Mistree, C. Schuster, K. Sass Mikkelsen, K. Kay, and J.-H. Meyer-Sahling. 2022. Global Survey of Public Servants. <https://www.globalsurveyofpublicservants.org/>.
- Mas-Colell, A., M. D. Whinston, and J. R. Green. 1995. *Microeconomic Theory*. Oxford, UK: Oxford University Press.
- Meyer-Sahling, J., D. Mistree, K. Mikkelsen, K. Bersch, F. Fukayama, K. Kay, C. Schuster, Z. Hasnain, and D. Rogger. 2021. *The Global Survey of Public Servants: Approach and Conceptual Framework*. [https://www.globalsurveyofpublicservants.org/\\_files/ugd/acee03\\_91fc43ed92774eb88704717d6c7c80a2.pdf](https://www.globalsurveyofpublicservants.org/_files/ugd/acee03_91fc43ed92774eb88704717d6c7c80a2.pdf).
- Moore, M. 1995. *Creating Public Value: Strategic Management in Government*. Cambridge, MA: Harvard University Press.
- UK Cabinet Office. 2022. “UK Civil Service People Survey: Technical Summary.” <https://www.gov.uk/government/publications/civil-service-people-survey-2022-results/civil-service-people-survey-2022-technical-guide>.
- UK Parliament. 2022. “MPs to Probe Quality of Civil Service People Survey.” <https://committees.parliament.uk/committee/327/public-administration-and-constitutional-affairs-committee/news/173010/mps-to-probe-quality-of-civil-service-people-survey/>.
- UN (United Nations). 2014. *A World That Counts: Mobilising the Data Revolution for Sustainable Development*. Report prepared at the request of the United Nations Secretary-General by the Independent Expert Advisory Group on a Data Revolution for Sustainable Development. <https://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf>.
- World Bank. 2023. Enterprise Surveys. <https://www.enterprisesurveys.org/en/enterprisesurveys>.
- World Bank Group. 2019. “Innovating Bureaucracy for a More Capable Government.” <https://openknowledge.worldbank.org/handle/10986/31284?show=full>.