

ANALYSIS

Modernising the National Statistical System: Transforming Pakistan's Future

Professor Dr. Zahid Asghar

School of Economics
Quaid-i-Azam University, Islamabad

1. Introduction

In an age marked by the digital era, rapidly changing technological progress, and data-driven governance, modernizing a country's statistical system is no longer optional— it is a strategic requirement. With a population of nearly 250 million facing many socio-economic challenges, the need for accurate, timely, and relevant data cannot be overstated. Intelligent government is based on intelligent public policymaking, which in turn is based on the right and timely data. According to Genesis, there will be a paradigm shift in wars, and data centres will be built in bunkers. Capturing territories may be less desirable, but data centres and other digital infrastructures may become the real target. In short, data is the basic fuel in the AI era.¹

Modernizing the national statistical system is very important for providing the necessary statistics. These statistics help in making smart public policies, which are essential for a smart government. Government and policymakers need updated statistics to make informed decisions and create effective policies. Businesses and industries use statistics to make decisions, innovate, and stay competitive. Researchers and academics rely on accurate data to conduct studies and support evidence-based policies. International organizations require up-to-date statistics to collaborate with other countries and address global challenges. NGOs and civil society depend on reliable data to advocate for change and monitor progress.

Media and journalists can use accurate statistics to inform the public. The general public benefits in making informed decisions for their personal consumption, investment and other matters. Investors and financial institutions use numbers for assessment of economic conditions and investment opportunities. The healthcare sector relies on precise data to design public health policies, to improve patient outcomes and manage resources efficiently. Covid-19 has already highlighted importance of accurate statistics. Pakistan has been facing revenue shortfall. To overcome these issues, the government should accelerate digitalization efforts to enhance tax compliance and reduce leakages.

Expanding the use of e-filing systems and leveraging big data analytics can help identify underreported incomes and improve overall efficiency. With a population of nearly 250 million facing many socio-economic challenges, the need for accurate, timely, and relevant data cannot be overstated. In a rapidly changing technological progress, and data-driven governance, modernizing a country's statistical system is no longer optional— it is a strategic requirement. As Rosling et al. said

¹ Henry A Kissinger, Eric Schmidt, and Craig Mundie, Genesis (Little, Brown, 2024).

"Numbers alone can't explain the world, but the world can't be explained without numbers,"². Intelligent government is based on intelligent public policymaking, which is based on the right and timely data.

This document explores the challenges, opportunities, and actionable strategies for reforming Pakistan's national statistical system. Currently, the Pakistan Bureau of Statistics (PBS) and Provincial Bureaus of Statistics (PBOS) operate largely under pre-digital-era frameworks. Despite some work by the PBS using digital gadgets for collecting the Consumer Price Index (CPI), censuses, and other data, there has not been much progress in processing, automating, and providing data in usable formats. Long delays in releasing digital census results, survey data for poverty calculation, and the non-provision of data in machine-readable formats are examples in this regard.

Our statistical agencies are not following the global shift toward real-time and AI-driven data analytics. This shift is a threat that could marginalize these bureaus if they do not evolve rapidly. Business as usual is causing erosion of trust in official statistics in this digital and artificial intelligence era. Official statistics— long considered the cornerstone of public trust— risk losing credibility when data release is delayed, when advanced analytical methods remain unused, and when the public has access to information (not necessarily correct) from alternative sources. Discourse increasingly leans on alternative data sources. The main objectives of this article are to:

- Highlight the pressing challenges facing Pakistan's national statistical system, especially in light of the digital revolution and the emergence of generative AI tools.
- Propose a modernization framework and outline a strategy to ensure a smooth transition from archaic data mechanisms to robust, real-time, and interoperable data solutions.

I shall deliberate about the nature of modern data— structured, unstructured, private, and public— to point out why these new forms of data matter for Pakistan. I also highlight major obstacles to modernization based on interviews conducted with bureaus and other officials dealing with statistics and other government sections, such as the lack of skilled human resources and fragmented institutional mandates. I discuss how generative AI, nowcasting, and integrated data platforms can help Pakistan harness the full potential of its vast data resources. Finally, I conclude that a well-funded, skilled human resource base (in Machine Learning, Big Data Analytics, LLMs), a national data policy, and a continuously evolving statistical system are the foundations of an intelligent government.

2. The Evolving Data Landscape

2.1 Real-Time Data and Speed

By 2030, it is projected that over 30% of global data will be generated, processed, and consumed in real-time. This data will come from a broad spectrum of sources: e-commerce transactions, financial services, consumer goods analytics, GPS navigation, live traffic monitoring, and more.

² Hans Rosling, Ola Rosling, and Anna Rosling Rönnlund, *Factfulness : Ten Reasons We're Wrong about the World - and Why Things Are Better than You Think* (London: Sceptre, 2019).

“A new beauty has been added to the splendour of the world— the beauty of speed.” (Filippo Tommaso Marinetti, the Italian poet)

In today’s digital environment, people expect quick and urgent responses to any issues they face. Taking actions based on fast-emerging information has become a necessity for governments, policymakers, economists, and businesses. Real-time data can inform “nowcasting”— short-term, data-driven economic forecasting that allows governments to anticipate rapid shifts in economic conditions. Countries that have integrated real-time data analytics into their official statistics are better positioned to make agile, evidence-based decisions.

2.2 Wealth of Public and Private Data

The 21st-century data ecosystem is characterized by an unprecedented proliferation of both public and private data, both structured and unstructured. Where official statistics once depended heavily on large surveys and censuses released after long intervals, modern systems can tap into continuous data feeds from payment processors (e.g., Mastercard), social media platforms, online marketplaces, and more.

In Pakistan, however, the custodians of official statistics— PBS and PBOS— have yet to fully leverage these opportunities. Interviews with bureau staff reveal that much of the effort still goes into manual data processing, repetitive surveys, and administrative tasks that do not exploit advanced analytical techniques. While official statistics remain central to public discourse, they increasingly face skepticism. This is partly because perceptions based on alternative sources of data (mainly through social media and other digital platforms) have more impact than the official data released with delays. Many people form opinions based on the alternative data available before the official releases come out.

2.3 Traditional Mechanisms vs. Modern Requirements

PBS remains relatively strong in some traditional domains— conducting censuses, household surveys, and measuring GDP— but the global data landscape is shifting toward real-time feeds, automated dashboards, and machine-readable databases. Without embracing these innovations, the slow turnaround of statistics risks making official numbers less relevant.

A telling example is the 2023 digital census in Pakistan, whose complete results have yet to be publicly released. Data provided on the PBS website is not in machine-readable format, which implies that users have very limited opportunities to analyze the data. This delay undermines public trust and raises questions about data authenticity and the transparency of official procedures.

3. The Need for Modernisation

3.1 Data-Driven Decision-Making for an Intelligent Government

Census data, GDP calculations, and consumer price indices are very important but emergence of real-time data, artificial intelligence (AI), and machine learning (ML) has transformed the data landscape. However, our statistical system remains entrenched in pre-digital mechanisms and is at

risk of marginalization in the modern digital era. The modern world demands data-driven public policy. Establishing a modern statistical system is vital to ensure that each of these stakeholders can access, analyze, and interpret data effectively. If Pakistan's official statistics do not keep pace with today's data revolution, policymakers and the broader public risk making critical decisions based on incomplete or outdated information.

3.2 The Digital Revolution and Generative AI

Generative AI— driven by Large Language Models (LLMs) offers new opportunities to work with data. APIs for easy data access and the use of open-source software should be practiced in bureaus.

Tools like “StatGPT” or “ImpactAI” can be customized by the bureaus as per their own requirements. One can use generative AI to detect anomalies in real-time data feeds could significantly enhance the timeliness and reliability of economic indicators. However, to benefit from these innovations, there is a need for skilled workforce, open data formats, and a cultural shift that government organisations and statistical bureaus must urgently embrace.

4. Challenges in Collecting and Disseminating Economic Statistics

During interviews with staff of bureaus, it was found that months pass between data collection and release, diminishing the value of these statistics for immediate policy interventions. Moreover, any use of modern digital tools often remains limited to pilot initiatives or donor-funded projects that end once external support ends.

4.1 Lack of Skilled Human Resources

One of the most cited issues is the scarcity of professionals trained in data science, big data analytics, ML, and AI. More troublingly, the organizational culture tends to sideline employees who propose new technologies and methods. According to interviews with both current and former staff, senior managers frequently resist innovation, preferring the familiarity of administrative routines. Talented graduates who join the bureaus often leave after finding few growth opportunities and little appreciation for modern skill sets.

Furthermore, some employees in allied organizations (e.g., those dealing with education, health, or labor statistics) complain that PBS does not facilitate the exchange of sampling frames, leading to duplication of surveys and an overall lack of coordination. Provincial bureaus also lament the limited and irregular training opportunities. Sessions are often scheduled haphazardly, leaving staff members without a cohesive learning path. Training opportunities are not provided based on individual needs but rather on who can be spared for training. Moreover, if a training program consists of several components spread over time, different participants in each component can undermine the whole purpose of the training.

4.2 Bureaus Dual Role

Persistent delays in releasing critical surveys or census results lead many stakeholders to rely on third-party or even anecdotal data. For instance, repeated postponements in making the 2023 digital census data publicly available have fuelled scepticism, with some questioning whether granular details will ever fully see the light of day.

This issue is compounded by the fact that PBS sometimes doubles as both a producer and an in-house user of data, delaying public release in favour of conducting its own analysis. A senior demographer mentioned in an interview that this practice is an attempt to create confusion over whether to use Bureau analysis or Bureau-released data. Bureaus release data in PDF format while keeping machine-readable data to themselves, putting common users at a disadvantage when trying to reproduce results. Bureaus have limited analytical capacity and should perhaps restrict their role to being data producers.

4.3 Fragmented Data Ecosystem

*There's a saying, "Water, water everywhere, but not a drop to drink," from Samuel Taylor Coleridge's poem *"The Rime of the Ancient Mariner."* It describes a sailor surrounded by water but unable to drink— a paradox of abundance without usability. We face a similar "Data Paradox" today. We have plenty of data, but not the right kind to make informed decisions. Despite an enormous amount of information, the public sector lacks proper indicators for evidence-based policymaking. Generative AI is based on right kind of data.³*

This is a data paradox: on one hand, there is a large amount of data, but on the other, relevant data are missing. The fragmentation within Pakistan's data landscape is considerable. PBS collects certain data, while PBOS gathers overlapping or parallel data at the provincial level, often without coordination or standardized methodologies. The result is duplication, inconsistencies, and missed opportunities for integration. One of the PBOS Director Generals mentioned in a seminar that PBS and PBOS have little coordination on most issues.

A 2011 Act intended to grant PBS greater autonomy has not been fully implemented. Bureau heads are often compromised when it comes to releasing statistics that have political economy dimensions— such as poverty, national accounts, and other important statistics.

4.4 Lack of Digitization and Machine-Readable Formats

Despite being in a digital age, large portions of Pakistan's administrative and historical data remain locked in paper-based or PDF formats. This severely limits their utility in modern analytics. Even when data are digitized, they are rarely offered in machine-readable formats (CSV, JSON, RDF, XML). This poses a barrier for researchers and analysts who want to integrate official statistics into dashboards, interactive tools, or advanced models.

4.5 Inadequate Financial and Institutional Support

Over successive administrations, financial allocations to PBS and PBOS have been insufficient for the scale of modernization required. Consequently, hardware, software, and human capital investments lag behind. Many modernization initiatives stay restricted to isolated projects rather than scaled-up, system-wide reforms.

4.6 Bureaus' Staff Opinion: Resistance to Innovation

Interviews conducted during seminars, workshops, and face-to-face discussions confirm the difficulties faced by PBS and PBOS staff. Key points include:

³ Nitin Seth, *Mastering the Data Paradox* (Penguin Random House India Private Limited, 2024).

- i. **Culture of Routine:** Tasks that require minimal skill dominate. There is little incentive for continuous learners or those who wish to apply cutting-edge methods (e.g., big data, ML, or LLMs).
- ii. **Talent Drain:** Qualified employees often exit within a few years because the work environment does not reward innovation.
- iii. **Fear of Transparency:** Senior staff may feel threatened by advanced analytical methods that could expose discrepancies in long-standing workflows.
- iv. **Low Coordination:** PBOS directors and other agencies, such as the Pakistan Institute of Education, report difficulty obtaining updates on many issues from the PBS.
- v. **Delayed Releases:** PBS tends to hold onto census or survey data for its own preliminary analyses, postponing the public availability of these datasets. This stalling often leads to suspicion and criticism.
- vi. **Poor Data Formats:** Data sets are released, if at all, in messy PDFs, making them nearly impossible to integrate into advanced analytical software without labor-intensive conversion steps.

5. A Framework for Modernization

As the above challenges demonstrate, the current system is unable to meet the fast-evolving demands of a digital economy. Yet, the opportunities are immense. Below is a suggested modernization framework:

5.1 National Data Policy

A clear, overarching National Data Policy is essential. This policy must define:

- **Objectives:** What the country intends to achieve with its data— improved governance, SDGs, economic growth, social services, etc.
- **Scope:** Which agencies (including PBS, PBOS, SBP, FBR, and others) are covered and how they coordinate data-sharing.
- **Standards and Protocols:** Setting guidelines for data quality, collection methods, metadata documentation, and machine-readable releases.
- **Governance:** Roles and responsibilities, especially concerning who owns the data, who can access it, and under what conditions.
- **Data Repositories:** A central platform at the national and provincial levels where repositories of all data are available and accessible through APIs.

Countries like Estonia and South Korea provide instructive examples of how robust policies can facilitate digital transformation and e-governance, streamlining public services and strengthening trust in official data.

5.2 Nowcasting and Real-Time Data

Forecasting has always served a useful purpose based on past data collected over long periods. But now, real-time data has made it possible to analyze the economy in real time. Incorporating real-time data from digital payment systems, retail platforms, or transportation networks can help policymakers stay ahead of economic fluctuations. Known as “nowcasting,” these short-term forecasts can complement existing surveys and censuses, filling the gap between official releases.

When integrated carefully, nowcasting provides governments with the flexibility to implement timely interventions, especially in times of crisis such as natural disasters or sudden macroeconomic shocks.

5.3 Data Integration and Synthetic Data

Linking census data with survey-based microdata, administrative records, and other large-scale data sets (potentially augmented by synthetic data to preserve privacy) can yield better, more nuanced insights. Synthetic data generation, specifically, allows for data sharing and advanced analytics without compromising individual confidentiality. Such methods can significantly enhance research opportunities and utilities while upholding ethical standards. To achieve this objective, bureaus need staff well equipped with ML, AI, LLMs, and Big Data skills.

5.4 Automating Reports and Parameterized Outputs

Bureau documentation often lacks recent trends. Automation is critical for reducing human error, ensuring consistency, and freeing staff to focus on higher-level analytical tasks. Reproducible data pipelines and parameterized reporting can save significant time, mitigate errors, and facilitate quick updates. By employing software that automates routine tasks—such as data cleaning and the generation of standard tabulations—PBS, PBOS, and other organizations can keep their outputs current with minimal manual intervention.

5.5 Geospatial and Global Data Linkages

Modern statistical organizations incorporate geospatial data to identify spatial patterns in poverty, health outcomes, and infrastructure gaps. Linking satellite imagery (e.g., NASA, NOAA) and other global datasets on climate or nightlights can provide local administrators with a richer view of regional disparities. The same approach can be used for national planning, guiding the placement of schools, health facilities, and roads more efficiently.

5.6 Legislation and Coordination

Improved coordination among all departments, including private and public bodies like PBS and PBOS, must be mandated by legislation or at least high-level policy directives. Fragmentation leads to duplication, inconsistent data definitions, and wasted resources. A coherent legal framework or well-enforced set of rules is needed to harmonize these bodies' mandates, data collection schedules, and reporting formats.

6. Role of Data Analytics and Generative AI in Policy-Making

6.1 Bridging the Data-Analytics Gap

Pakistan's public sector has vast repositories of text documents: policy directives, ministry records, judicial opinions, parliamentary proceedings, project plans (e.g., PG-1), and more. However, these often lie dormant as “garbage” because they are not digitized or analyzed. Introducing a robust text-data analytics framework powered by Natural Language Processing (NLP) can uncover hidden insights, detect policy inconsistencies, and inform evidence-based decision-making.

6.2 Embracing Generative AI

Generative AI, particularly LLMs, can automate many aspects of statistical work:

- **Code Translation:** Tools like a “SAS to R Code Assistant” reduce the time needed for adopting new software.
- **Enhanced Explanations and Summaries:** AI-generated narratives can improve the clarity of statistical reports, bridging the gap between technical experts and general audiences.
- **Efficiency Gains:** Repetitive tasks like data transformation or preliminary analysis can be handled by generative AI, allowing statisticians to focus on deeper analysis.

6.3 Building Human Resource Capabilities

However, harnessing these benefits demands personnel who understand both data science and prompt engineering. Training programs for PBS and PBOS staff must emphasize open-source software, advanced analytics, and generative AI. Pilot projects can serve as learning platforms, after which broader adoption strategies can be designed. Consistent engagement with the evolving field of AI will be key, as new tools and techniques are introduced at an accelerating pace.

6.4 Right not Easy Data

As Rosling et al. points out “Its not numbers but numbers which matter.”⁴ Rather than chasing “easy data,” bureaus must focus on “right data.” This helps ensure that the numbers being collected measure genuine outcomes and inform strategic choices. By integrating geospatial layers, household surveys, real-time data feeds, and textual analyses, policymakers can craft more nuanced, localized solutions to socio-economic challenges.

7. Way Forward

Digital payments, GIS and satellites technologies have become more in use, Bureaus can start using real-time data in its national statistics, helping the country respond faster to challenges like natural disasters or inflation. This demands close collaboration among PBS, SBP, FBR, health agencies, and international organizations. This teamwork will reduce duplication, share costs, and ensure systems work smoothly. Innovation labs at universities or institutions like PIDE can help test new data tools in a controlled environment. At the same time, sharing data more quickly and in user-friendly formats will help rebuild public trust in official statistics, reducing doubts about data manipulation and allowing people to focus on solving real policy issues. However, for this modernization to succeed, it must be a continuous effort. Political leaders and policymakers need to recognize the importance of an updated statistical system and provide consistent funding and support. Without this commitment, modernization efforts will remain incomplete and unsustainable.

8. Conclusion

It is a tough job to update national statistical system in digital era but its no more a choice but obligatory. One option is to keep using old, outdated methods, messy data systems, and losing public trust in governments. The other option is to adopt a modern, data-driven approach to governance that can handle the challenges of the 21st century. This decision will shape Pakistan’s future for decades. To achieve this, Pakistan needs strong political will, proper funding, skilled people, and involvement from all sectors. If done correctly, the statistical system could become the foundation for building a smart, efficient government.

⁴ Hans Rosling, Ola Rosling, and Anna Rosling Rönnlund, *Factfulness : Ten Reasons We’re Wrong about the World - and Why Things Are Better than You Think* (London: Sceptre, 2019).