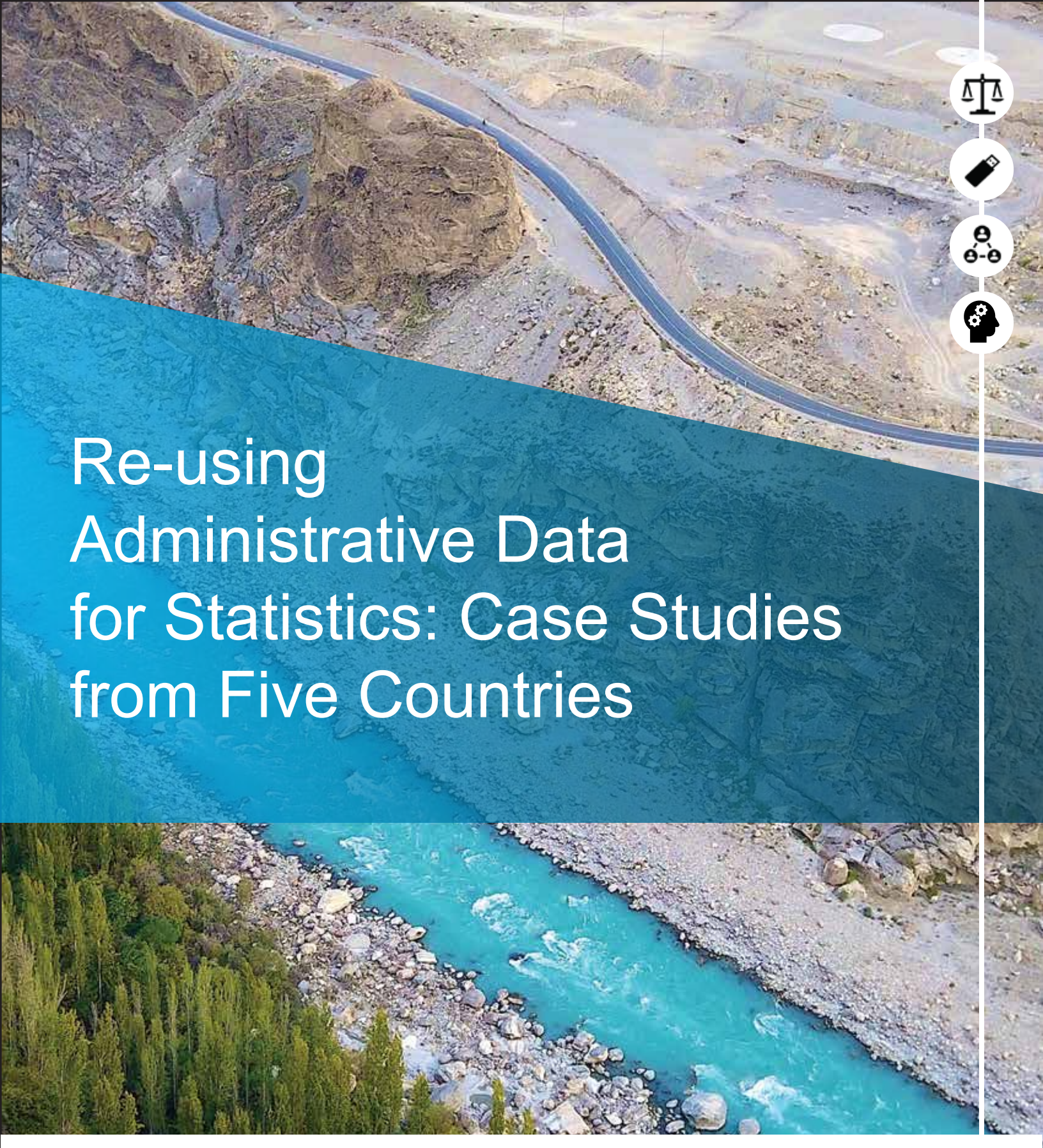


# Re-using Administrative Data for Statistics: Case Studies from Five Countries



LSE Department of International Development

United Nations | DESA Statistics

Global Partnership for Sustainable Development Data

# Table of Contents

# Acknowledgments

*Re-using Administrative Data for Statistics: Case Studies From Five Countries* is commissioned by the United Nations Department of Economic and Social Affairs (UNDESA), Statistics Division Collaborative on Administrative Data.

We would like to thank the following individuals for their unwavering guidance throughout the development of this report:

Viebeke Oestreich Nielsen, UNDESA
Karen Bett, Global Partnership for Sustainable Development Data (GPSDD)
Tiziana Leone, London School of Economics and Political Science (LSE)

We are also grateful to the respondents from Chile, Denmark, Ghana, Kenya, and Pakistan who agreed to participate in our research. Lastly, we would like to extend our appreciation to all those who supported our research and connected us with key informants, including:

Federico Segui, Instituto Nacional de Estadística (INE)
Omurbek Ibraev, National Statistics Committee of the Kyrgyz Republic (NSC)
Ximena Clark, National Bureau of Economic Research (NBER)
Faryal Ahmed, United Nations Department of Economic and Social Affairs (UNDESA)
Afsaneh Yazdani, United Nations Economic and Social Commission for Asia and the Pacific
Rasmus Larsson, Statistics Sweden (SCB)
Lars Thygesen, Thygesen Statistics Consulting
Paula Darville, United Nations Resident Coordinator Office Chile
Robin Choudhury, International Development Cooperation

# List of Acronyms

| | |
|---|---|
| BoS | Bureaux of Statistics |
| CPR | Civil Registration System |
| CSO | Central Statistical Office |
| EASD | East African Statistical Department |
| EU | European Union |
| FBS | Federal Bureau of Statistics |
| FGD | Focused Group Discussion |
| GPSDD | Global Partnership for Sustainable Development Data |
| GSSP | Gender Sector Statistics Plan |
| GSS | Ghana Statistical Services |
| IBRD | International Bank for Reconstruction and Development |
| IMF | International Monetary Fund |
| INE | The National Statistics Institute of Chile |
| IRB | Internal Review Board |
| KI | Key Informant |
| KNBS | Kenya National Bureau of Statistics |
| KSDS | Kenya Strategy for the Development of Statistics |
| LMIC | Low- and Middle-Income Countries |
| LSE | London School of Economics and Political Science |
| MDAs | Ministries, Departments, and Agencies |
| MOUs | Memorandums of Understanding |
| NSO | National Statistics Office |
| OECD | Organisation for Economic Co-operation and Development |
| PBS | Pakistan Bureau of Statistics |
| SDG | Sustainable Development Goals |
| SoP | Standards of Practice |
| SSC | Strategic Sector Cooperation |
| ToR | Terms of Reference |
| UNDESA | United Nations Department of Economic and Social Affairs |
| UNICEF | United Nations Children's Fund |
| UNSD | United Nations Statistics Division |

# Explanatory notes

## Administrative Data

Administrative data refers to information collected and stored by administrative systems, including public sector ministries, departments, and agencies (MDAs) (Connelly et al., 2016). According to Woollard (2014), this type of data is collected for record-keeping essential to service delivery. For instance, these data can be extracted from activities concerning taxation, housing, health, and education.

## Ministries, Departments and Agencies (MDAs)

This report uses an acronym for ministries, departments, and agencies (MDAs) to account for differences in terminology for governmental entities across countries. MDAs are grouped together in this report because they can serve as providers of administrative data for national statistics offices (NSOs).

## National Statistics Offices (NSOs)

Each country included in this report has a national statistics office (NSO) responsible for producing national statistics informed by a variety of sources, including administrative data from MDAs. While NSO titles and corresponding acronyms differ across countries, the acronym "NSO" is used throughout this report to refer to the main entity responsible for producing national statistics in a given country.

# Report Highlights

## Overview

The report on *Re-using Administrative Data for Statistics: Case Studies From Five Countries*, prepared by post-graduate student consultants at the London School of Economics and Political Science (LSE) in collaboration with the United Nations Department of Economic and Social Affairs (UNDESA), examines challenges and opportunities for enhanced administrative data collaboration between National Statistics Offices (NSOs) and government Ministries, Departments and Agencies (MDAs).

The report was produced in response to the increasing interest in administrative data combined with the amplified imperativeness of leveraging these data for policymaking during the COVID-19 pandemic. It is motivated by existing literature on the benefits of administrative data for improving the efficiency and accuracy of developing national statistics and evidence-based policies. The report focuses on five case studies – Chile, Denmark, Ghana, Kenya, and Pakistan – by presenting country profiles and illustrating insights of local experts on existing data sharing practices.

Primary qualitative data was collected through interviews with NSO statisticians and MDA administrators from the five countries of interest. A total of 21 respondents were interviewed, with 12 representing NSOs and 9 representing MDAs. Thematic analysis was used to develop a coding system for the transcribed interviews and to organise the coded content by theme.

## What are the key findings?

Four themes were developed from the transcribed and coded interviews to account for the main factors determining the extent of administrative data re-use across the five countries. Figure 1 portrays the main findings under the following themes: **legislation**, **data systems**, **communication, collaboration, and partnerships**, and **awareness and attitudes**.

***Figure 1: Key findings***

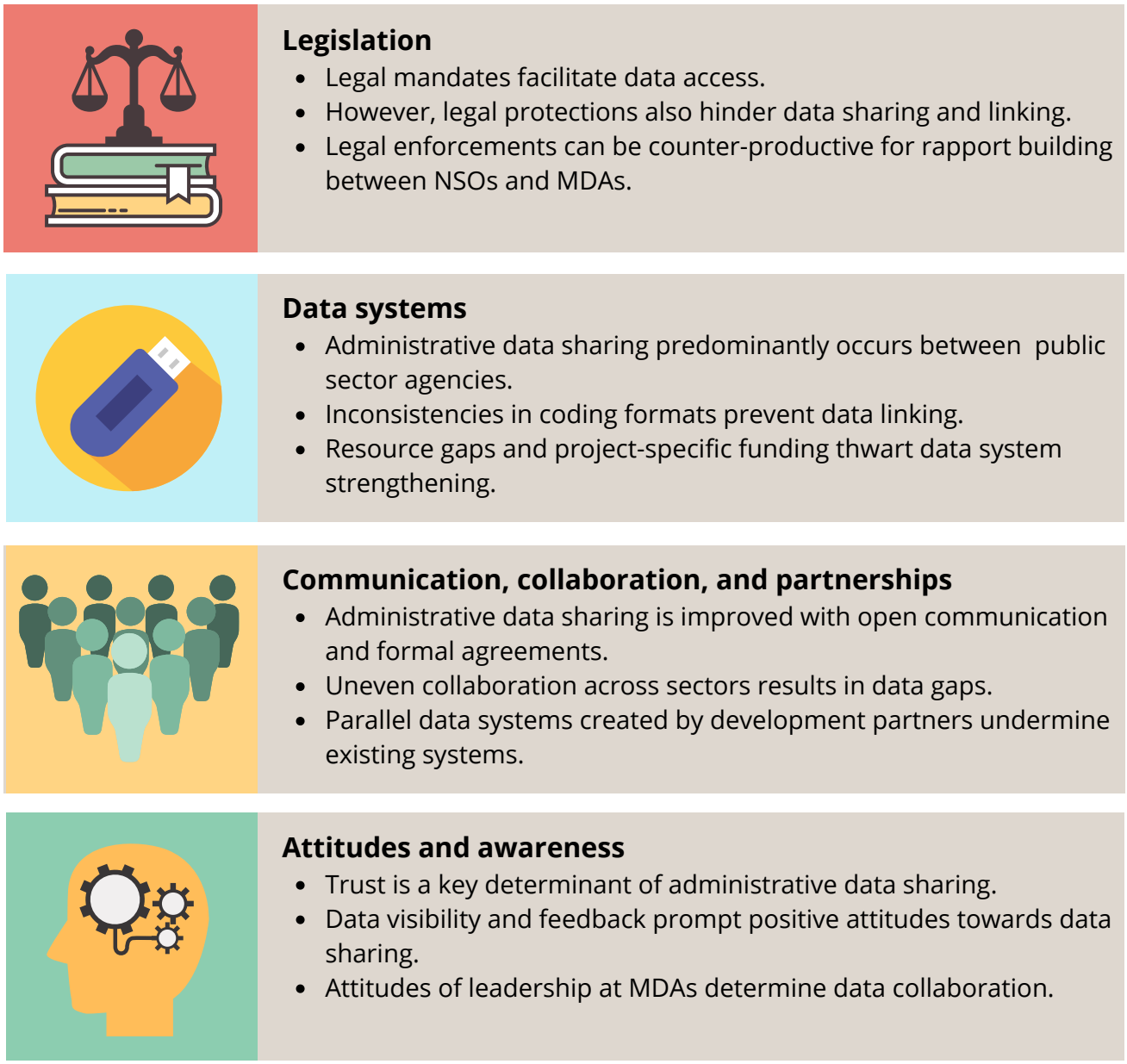


***Figure 2: Key recommendations***

## INFORMATION

**Increase public and government awareness of the benefits of re-using administrative data for statistics**

Utilise social media and other interactive platforms for citizen-led accountability.

Organise briefings with government officials to encourage investment in data system strengthening.

Engage in political discourse to frame administrative data as a resource for decision-making.

Facilitate and advocate for research using administrative data.

## GUIDANCE

**Promote legislative reforms and develop standards of practice (SOPs) to boost data sharing**

Lobby for sufficient resource allocation for administrative data collection and access.

Produce and publish SOPs for data privacy and cybersecurity.

Establish common identifiers to facilitate data linking and support standardised data formatting.

## COORDINATION

**Bolster dialogue and relationships with MDAs and development partners**

Establish a taskforce for multi-sectoral coordination with defined terms of references including:

a. Identifying region-specific examples of data system successes to promote progress.

b. Exchanging strategies and tools for data collection and sharing.

c. Building accountability structures to enforce data sharing and adherence to SOPs.

Produce guidelines for international donors collecting data to align with existing data systems and contribute to data sharing flows.

## What are the key recommendations?

NSOs can play a vital role in improving the interest and recognition of administrative data re-use for statistics due to its collective commitment and influence on data processes within the selected countries. Therefore, this report proposes the need for a Strategic Framework initiated by the NSO to advance progress in data quality and availability. Figure 2 illustrates the recommended objectives under three pillars – namely, **Information**, **Guidance**, and **Coordination** – to ultimately enhance data sharing between the NSO and MDAs.

## Considerations

While a centralized body, like the NSO, can be conducive in accelerating development of data systems, a consolidated multi-sectoral and multi-lateral response is required to achieve sufficient progress. Various actors involved in data practices and policymaking including development partners and MDAs are urged to commit to data system strengthening.

# 1.Introduction

Administrative data present unique opportunities for guiding policy change and improving government efficiency. Over the past several decades, the number of studies based on administrative datasets has increased in diverse fields from economics to medicine (Enayati and von Schrader, 2016; Mazzali and Duca, 2015; Sanhueza, 2014; Bharadwaj, Løken and Neilson, 2013; Nagler, 2013; Dinkelman and Martínez A., 2014). While there has been growing interest in the re-use of administrative data for research and evidence-based policy implementation, emerging challenges such as the COVID-19 pandemic have increased the urgency of efforts seeking to expand administrative data sharing.

There are myriad benefits to leveraging administrative data for statistical purposes. Studies have shown that administrative data can be more cost-effective and more representative of large, heterogeneous populations than traditional survey methods. For example, re-using administrative datasets can increase the efficiency of research, particularly longitudinal studies, by reducing the time and cost of data collection (Enayati and von Schrader, 2016; Holman *et al.*, 2008; Hand 2018). Additionally, while surveys can be limited by small sample sizes, low levels of attrition, and non-response bias, administrative data offer larger sample sizes and, in some cases, more universal coverage of populations of interest (Gavrielov-Yusim and Friger 2014; Card et al. 2010). As a result, administrative data can not only increase statistical power, but can also make it possible for researchers to gain historical and up-to-date information on events or groups that are difficult to capture with traditional survey methods (Enayati and von Schrader 2016; Boden and Rees 2010; Harron *et al.*, 2017; Figlio, Karbownik, and Salvanes 2017). In light of these potential advantages, increasing administrative data sharing to inform statistics is a promising strategy for enhancing evidence-based policymaking and increasing government efficiency (Allard *et al.*, 2018).

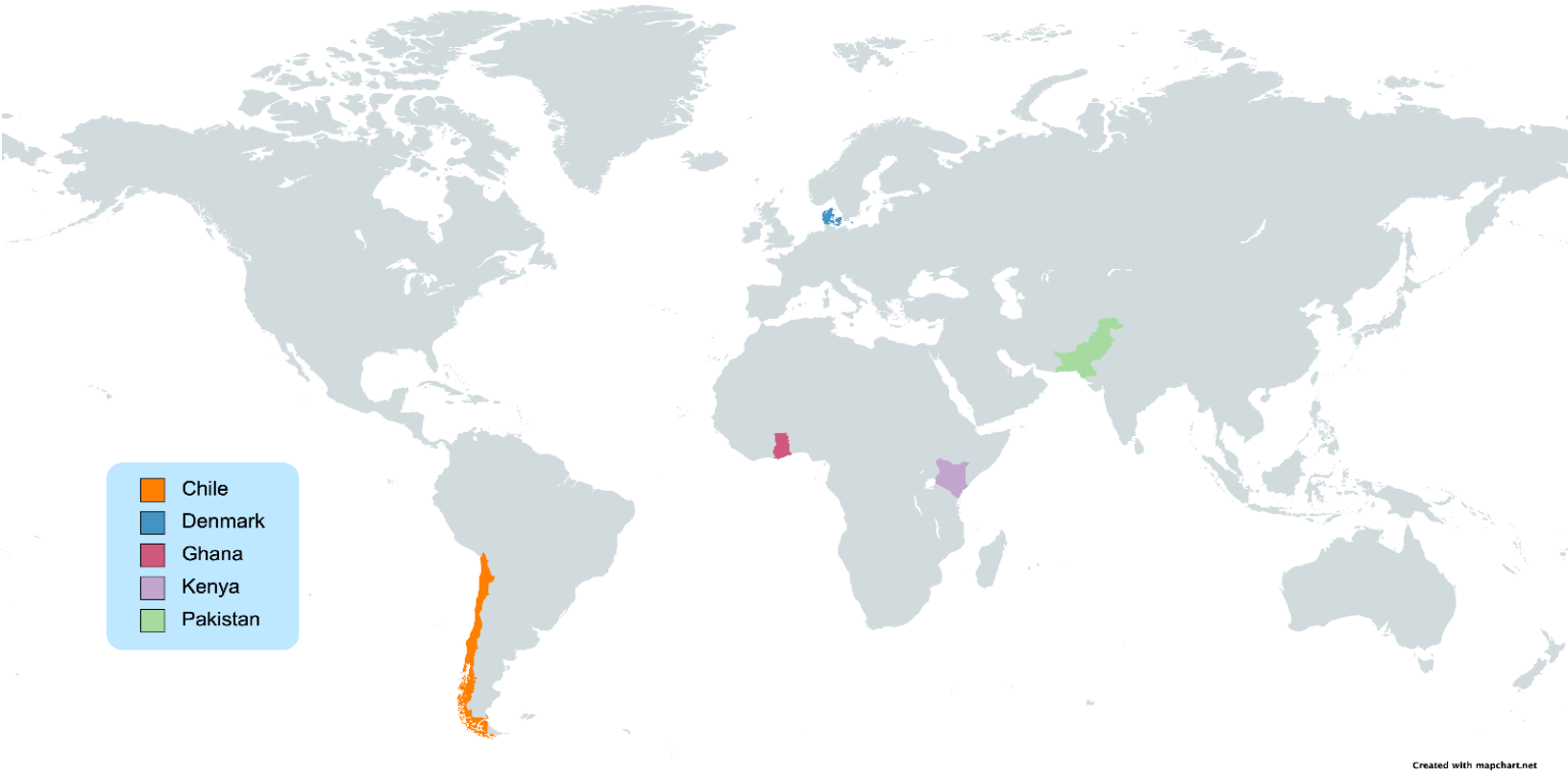
However, despite these benefits, administrative data remains an under-utilized resource in many countries. A variety of factors can pose barriers to administrative data sharing and effective re-use of administrative data for statistics, including decentralized data systems, lack of legislative guidance, and resource gaps. For example, in the absence of standardized identifiers, linking data from decentralized administrative sources can pose major technical challenges and limit statistical applications (Harron, Goldstein, and Dibben 2015; Jutte, Roos and Brownell, 2011; Harron *et al.*, 2017; Allard *et al.*, 2018). Additionally, data privacy laws can prevent researchers from accessing administrative data, and lack of trust in government and statistical authorities can limit political support for increased administrative data sharing (Schnell, 2014; Goroff, Polonetsky, and Tene 2018; Hand *et al.* 2018; Petrila 2018; Sexton *et al.,* 2017).

While there is a growing body of literature on administrative data, limited research has focused on administrative data re-use in low- and middle- income countries (LMICs). The potential for administrative data to enhance the quality of statistics while reducing costs could be especially beneficial for LMICs, which tend to face more gaps in data availability than high-income countries (Khalid, Sharma and Dubey, 2020). However, additional barriers, such as financial constraints, technical limitations, and political reluctance, can further complicate efforts to increase administrative data sharing in some LMICs (Kumar et al., 2018). More research is needed to better understand the barriers to – and the benefits of – administrative data sharing across different LMIC contexts.

This report contributes to existing literature by drawing on key informant interviews to understand factors impacting administrative data sharing between government agencies and National Statistics Offices (NSOs) in different country contexts. Namely, the report includes case studies on Chile, Denmark, Ghana, Kenya, and Pakistan. Primary qualitative data was collected from each country through key informant interviews with statisticians and administrators from each country. The report presents the perspectives of local experts on the roles of legislation, data systems, attitudes, and partnerships in shaping administrative data sharing. Following the results is a discussion of how these respective factors interact to create barriers or offer solutions to effective re-use of administrative data. The report concludes with recommendations for the UN Statistics Division Collaborative on Use of Administrative Data for Statistics.

# 2. Country profiles

*Figure 3: Map of countries of focus*




## 2.1 CHILE

Chile is a high-income country situated alongside the western seaboard of South America with a population of 18.7 million (WBG, 2018). As of 2010, Chile became the first Latin American country to join the Organisation for Economic Co-operation and Development (OECD). Since then, Chile has committed to improving its statistical capacity and enhancing the usage of big data.

The National Statistics Institute of Chile (INE) is a public agency in charge of disseminating statistics. Since its inception in 1843, INE has been responsible for the general census of population and housing. In 1927, INE became an executive branch of the Ministry of Economy that collects, produces, and publishes official demographic statistics. Throughout the institutional development of nationwide statistics, the legal framework for statistical legislation has not been modified since its publication. However, there is ongoing discussion in congress to mandate the law. Updates to the role of INE and other Chilean institutions in terms of access to information have so far ensured stable progress in meeting OECD's standard of practices. In

particular, INE's ability to request administrative data as an autonomous authority has been grounded in the mandate from the Statistics Act 2008.

In practice, however, requests for data relating to personal information from administrative registers are often met with challenges due to conflict in institutional legal framework and INE's mandate. Concerns over confidentiality under Article 29 and Article 30 of the governance of Law No.17.374 have challenged the data access for INE. In addition, a constitutional change in 2008 to access public information, under Law No.20285, enforces the administrative bodies to display the collected information publicly (Nacional, 2008).

With the increasing popularity of using administrative data in social science academia, the present COVID-19 pandemic is bringing a renewed focus within INE on the development of explicitly modelled administrative records to support their use for the country's policies. Beyond the systematic update of the sampling framework for more effective surveillance, integrating administrative records has become one of the critical priorities for INE.

## 2.2 DENMARK

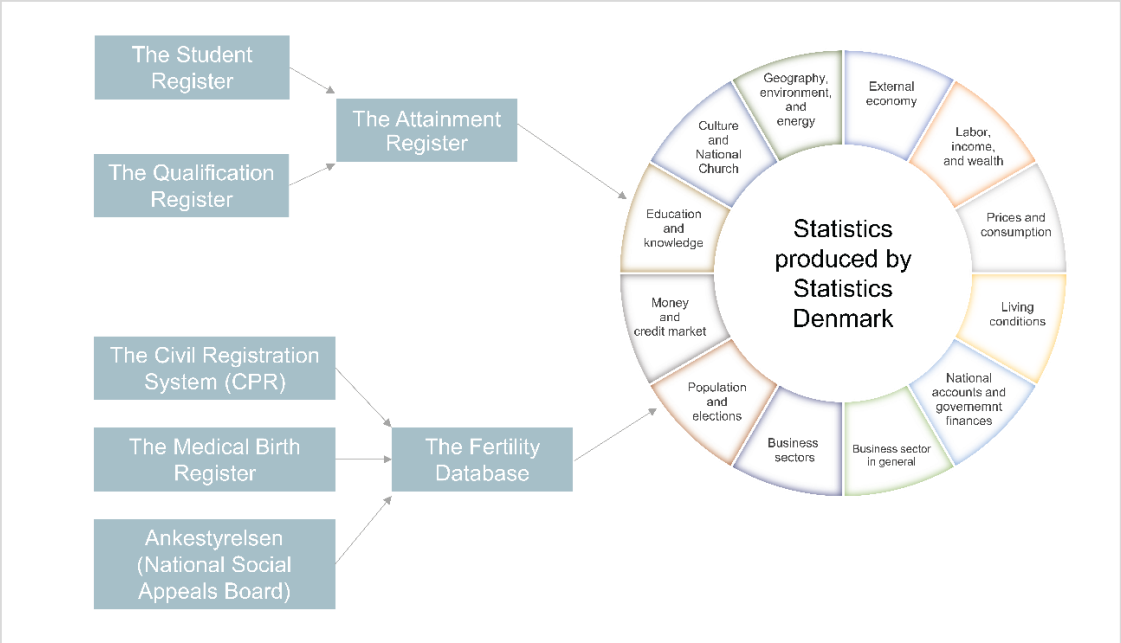
Denmark is a high-income Nordic country bordered by Germany and the North Sea with a population of 5.8 million (WBG, 2019a). As early as 1645, all Danish pastors were required by law to keep parish registers including information on births, baptisms, marriages, and deaths in their parish. In 1924, local municipal registers were established to collect data on all municipality residents, expanding administrative data capacity. In 1968, the Danish Civil Registration System (CPR), which includes data on every legal Danish resident, was created to centralize the collection of administrative data. The CPR remains a key source of administrative data for statistical purposes and has been updated and expanded periodically since its establishment.

Statistics Denmark, created in 1850 under the name Statistical Bureau, is the central authority for producing Danish statistics. The Act of Statistics Denmark of 1966 defines Statistics Denmark as an independent institution, outlines its responsibilities, and mandates its access to data sources. Most statistics produced by Statistics Denmark rely on administrative data from Danish registers, including the CPR as well as registries for businesses, buildings, and other subjects. Figure 4 shows examples of the data flow from registers. Information from different registers can be linked easily because observations have universally unique identifiers, such as a unique personal identification number, or CPR number, for every individual. Some statistics produced by Statistics Denmark, such as statistics regarding quality of life, supplement registry data with survey data.

Legislation including The EU Regulation on European Statistics, the EU Data Security Directive, the Danish Act on Processing of Personal Data, the Danish Public Administration Act, the Danish

Access to Public Administration Files Act, and the Act on Statistics Denmark undergird Statistics Denmark's data confidentiality policy. Statistics Denmark also assures data quality in accordance with the Code of Practice for European Statistics. Public awareness of and attitudes toward Statistics Denmark are regularly evaluated with public image surveys (Denmark Statistics, *nd*).

*Figure 4: Examples of data flows to Statistics Denmark*




## 2.3 GHANA

Ghana is a lower middle-income country with an estimated population size of 30.4 million (WBG, 2019c). It is bordered by Burkina Faso, Togo, the Atlantic Ocean, and Cote d'Ivoire. Ghana's earliest efforts at collecting and disseminating data began in 1891 when it conducted its first census, in the then Gold Coast. In 1948, the Office of the Government Statistician was created, and was later expanded and renamed the Central Bureau of Statistics (CBS) in 1961. By 1985, the Ghana Statistical Service (GSS) was established as an independent body, reporting directly to the Office of the President, under Statistical Service Law (PNDCL 135). However, to enhance the statistical legal framework of the country, the Statistical Service Act of 2019 (Act 1003) which recognizes the GSS as the country's central statistics producing and coordinating institution, replaced the Statistical Service Law of 1985. The GSS has 10 regional and over 100 district offices, and is mandated to conduct censuses, surveys and compile socio-economic data critical for the

management and growth of the country and development of the private sector (Ghana Statistical Services, 2020).

Ghana, despite having a central agency for producing and coordinating statistics, operates a semi-decentralized system. Separate from the GSS, the Ghanaian Statistical structure comprises statistical units within individual ministries and regional administrative districts that produce, coordinate, and use administrative data for statistical purposes as it applies to their respective programmes. However, many of these sector ministries and districts lack robust statistical units capable of generating credible and reliable data in a timely manner, needed by the country to inform decision making and improve government efficiency (Ghana Statistical Services, 2020).

To strengthen the administrative data system and production of statistics in Ghana, the government of Ghana is collaborating with Statistic Denmark on a new Strategic Sector Cooperation (SSC) focused on building knowledge and expertise in the use of administrative data for statistics production in Ghana (Ministry of Foreign Affairs of Denmark, 2019).

*Figure 5: NSOs of countries of focus*

| COUNTRY | NSO TITLE | NSO SYMBOL |
|---|---|---|
| CHILE | NATIONAL STATISTICS INSTITUTE (INE) | INE Instituto Nacional de Estadísticas · Chile |
| DENMARK | STATISTICS DENMARK | STATISTICS DENMARK |
| GHANA | GHANA STATISTICAL SERVICE (GSS) | GHANA STATISTICAL SERVICE |
| KENYA | KENYA NATIONAL BUREAU OF STATISTICS (KNBS) | KNBS KENYA NATIONAL BUREAU OF STATISTICS Keeping you informed |
| PAKISTAN | PAKISTAN BUREAU OF STATISTICS (PBS) | PAKISTAN BUREAU OF STATISTICS Government of Pakistan |

## 2.4 KENYA

Kenya is a lower middle-income country in Eastern Africa bordered by South Sudan, Ethiopia, Somalia, Uganda, Tanzania, and the Indian Ocean with an estimated population size of about 53 million (WBG, 2019d). In 1948, the East African Statistical Department (EASD) was established to collect data and disseminate statistics for Kenya, Uganda and Tanganyika.

In 1956, the EASD created a separate statistical unit for Kenya. Kenya's Statistics Act of 1961 brought the statistical unit under Kenya's jurisdiction and integrated it into the Kenyan government's Economics and Statistics Division. In 1963, the Economics and Statistics Division was divided into two separate departments: a department for economic planning and a department for statistics, the latter of which became the Central Bureau of Statistics (CBS) in 1972.

Following low levels of survey data collection and statistics production by the CBS, the Statistics Act of 2006 established the Kenya National Bureau of Statistics (KNBS) to replace the CBS. According to the Statistics Act of 2006, KNBS is a semi-autonomous agency within the Ministry of State for Planning, National Development and Vision 2030 and is responsible for collecting, analysing and disseminating statistics in Kenya. KNBS, headquartered in Nairobi, is led by a Board of Directors and Director General and is divided into six main directorates: Strategy and Development, Population and Social Statistics, Macroeconomic Statistics, Production Statistics, Information Technology, and Finance and Administration. Statistics produced by KNBS relies mostly on survey data and is shared on a number of different platforms, including Kenya Open Data and the Kenya Data Portal.

To address persisting resources and coordination gaps, KNBS recently adopted the Kenya Strategy for the Development of Statistics (KSDS) and the Gender Sector Statistics Plan (GSSP). These plans offer strategic direction to enhance Kenya's statistical capacity by 2023, with a particular focus on measuring progress toward gender equality and other SDGs. (Kenya National Bureau of Statistics, *nd*)

## 2.5 PAKISTAN

Pakistan is a lower middle-income country in South Asia with an estimated population of 216. 5 million (WBG, 2019b). After the country gained independence, a Central Statistical Office (CSO), associated with the Economics Affairs Division, was established in 1950. Upon recommendation from the International Bank for Reconstruction and Development (IBRD), the CSO was realigned as a technical wing under the Statistics Division in 1971. In 1981, the CSO transitioned once again and was recognized as the Federal Bureau of Statistics (FBS). During this time, the Statistics Division consisted of the FBS, Population Census Organization (PCO), and the Agricultural Census

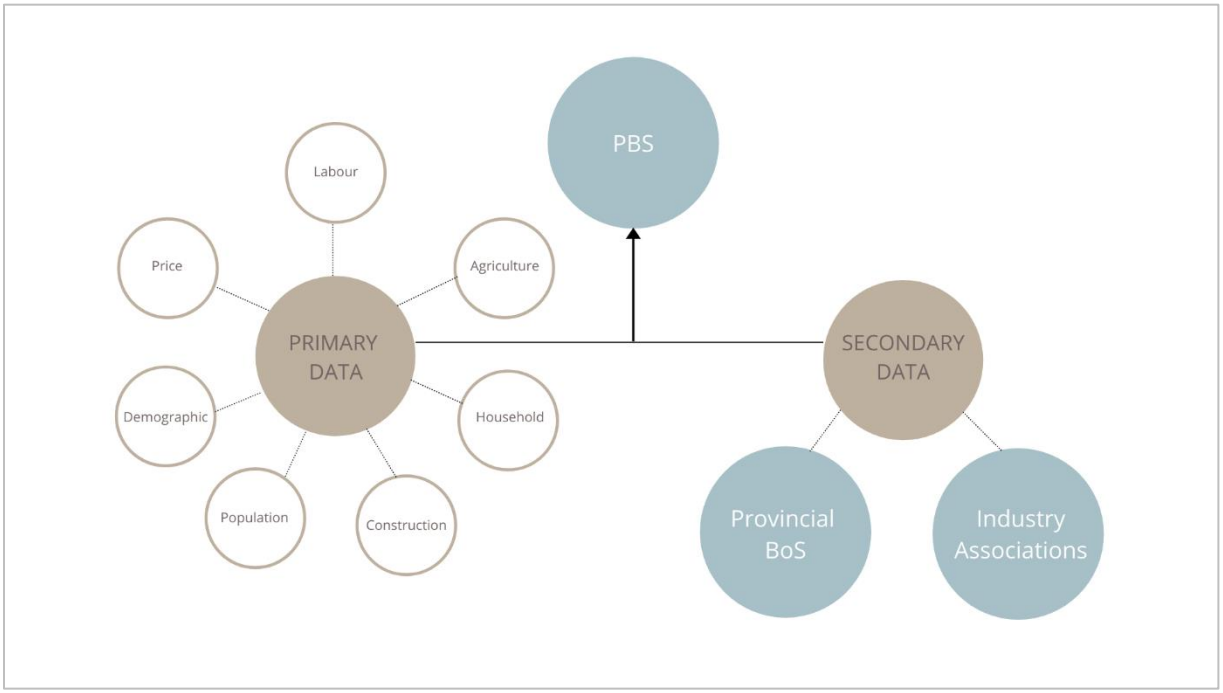
Organization. Today, the presiding statistical system known as the Pakistan Bureau of Statistics (PBS) - responsible for producing, gathering and circulating statistics - encompasses the FBS, the Population Census Organization, the Agriculture Census Organization, and the Statistics Division (Pakistan Bureau of Statistics, *nd*).

However, the implementation of the 18th Amendment in 2010, devolved critical responsibilities to provincial governments. Therefore, various provincial Bureaux of Statistics (BoS) play an important role in the collection and distribution of data. The BoS in the four provinces - namely, Sindh, Punjab, Balochistan, and Khyber Pakhtunkhwa - function as a division within the provincial Planning and Development Departments.

Further, the General Statistics (re-organization) Act of 2011, allocates authority to PBS to collect, analyse and publish statistics for the welfare and benefit of the people of Pakistan. This includes the right entrusted to the PBS to access data from public and private entities. Additionally, the Act imparts regulatory powers to PBS to monitor and evaluate statistics-related work across Pakistan (including facilitating education and research in statistics).

The PBS compiles data on socio-economic sectors using primary and secondary sources, including administrative records. These have been illustrated in Figure 6. Regular working groups operate within the PBS such as the National Accounts Committee, which include members from various government agencies and industry associations.

*Figure 6: Data flows to PBS*

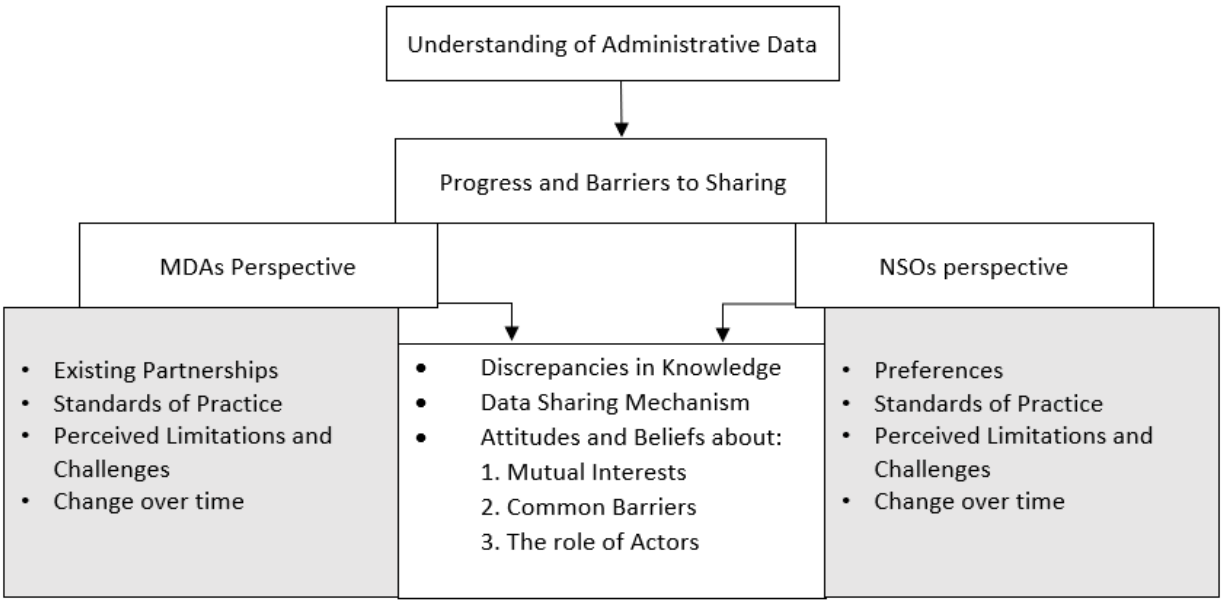


# 3. Methodology

## 3.1 RESEARCH DESIGN

This report is based on primary qualitative data collected through key informant and focus group interviews with respondents from Chile, Denmark, Ghana, Kenya, and Pakistan. These five countries were chosen in consultation with UNDSEA to allow for a broad analysis including countries with varying levels of administrative data use. Before beginning the interview process, two sets of interview questions were developed, one for respondents from NSOs and one for respondents from MDAs. The interview questions aimed to elicit detailed explanations and opinions of administrative data sharing. The framework for developing the questionnaires is presented in Figure 7. To ensure clarity, the interview questionnaires were tested in two pilot interviews with respondents from NSOs of countries not involved in the study. Feedback from the pilot interviews were used to finalize the interview questionnaires, which are included in the Appendix 2.

*Figure 7: Framework for questionnaires*




Eleven key informant and three focus group interviews were conducted between January 20, 2021 and March 5, 2021 with NSO statisticians and MDA administrators from the aforementioned countries. All interviews were conducted in English, with one interview facilitated by a translator. Additionally, with the exception of one written interview, the interviews were held on Zoom and lasted approximately 45 to 90 minutes each. Each Zoom interview was recorded and transcribed with permission of the respondent(s).

## 3.2 SAMPLE

Respondents were identified using snowball sampling and purposeful sampling. These sampling approaches were selected to increase the likelihood of response and ensure that respondents had relevant expertise (Robinson, 2014; Palinkas et al., 2015; Lavrakas, 2008). Potential respondents were limited to MDA administrators from health, education, and economic sectors and NSO statisticians knowledgeable about administrative data management, sharing, or re-use at the national level. Existing relationships between UNSD and NSOs were leveraged to recruit initial respondents. Additional respondents were identified by requesting relevant contacts from the initial respondents. A total of 21 respondents were interviewed across the five countries; 12 respondents represented NSOs and 9 respondents represented MDAs. Figure 8 summarizes the sample characteristics.

## 3.3 ANALYTIC APPROACH

Transcribed interview responses were evaluated using inductive thematic analysis. Thematic analysis is a common approach to analysing qualitative data in which the raw data is categorized into codes, which are then organized into broad themes (Braun and Clarke, 2006). This approach was selected because it allows for a complex, nuanced account of the data while also providing a well-structured framework for presenting results (Nowell et al., 2017; Rubin & Rubin, 2004). Rather than set predetermined themes, an inductive approach to thematic analysis derives a thematic structure from the data after it has been collected (Boyatzis, 1998). An inductive approach was selected for this report to allow flexibility for unanticipated insights and minimize researcher bias (Burnard et al., 2008).

## 3.4 ETHICS

This study was conducted in compliance with the ethical code and guidelines of the London School of Economics and Political Science (LSE). Ethical approval was obtained from the University's internal review board (IRB) prior to arranging interviews. During the interview process, each respondent provided informed consent. Additionally, data were kept confidential and stored without identifiable information.

## 3.5 LIMITATIONS

The selective approach to sampling combined with the time constraints of this report resulted in a low number of respondents, limiting the representativeness and generalizability of the findings. However, the in-depth nature of the interviews provided detailed data to inform the analysis, discussion, and recommendations.

The use of semi-structured, open-ended questions aimed to encourage rich discussion between interviewers and interviewees regarding the subject matter. However, the extent to which the questions addressed all key areas of interest was a concern. To address this limitation, the pilot

interview process explained above was used to produce a more robust and comprehensive survey instrument.

As a qualitative study, the findings are based on the perspectives, opinions, and beliefs of respondents. As such, it depends on their subjective understanding of the topic, limiting the extent to which the results can be generalized and replicated. However, the level of expertise of the respondents strengthens the reliability of the result and the qualitative approach offers valuable insight into the complex dynamics of data sharing.

*Figure 8: Summary of sample characteristics*

| NO. | CODE | QUESTIONNAIRE TYPE | INTERVIEW TYPE | DATE |
|---|---|---|---|---|
| 1 | PI 1 | NSO | KEY INFORMANT | 8 JAN 2021 |
| 2 | PI 2 | NSO | KEY INFORMANT | 15 JAN 2021 |
| 3 | KI 1 | MDA | KEY INFORMANT | 20 JAN 2021 |
| 4 | KI 2 | NSO | KEY INFORMANT | 26 JAN 2021 |
| 5 | KI 3 | NSO | KEY INFORMANT | 2 FEB 2021 |
| 6 | FGD 1: KI 4, 5, 6, 7 | NSO | FOCUS GROUP | 11 FEB 2021 |
| 7 | KI 8 | NSO | KEY INFORMANT | 15 FEB 2021 |
| 8 | KI 9 | MDA | KEY INFORMANT | 19 FEB 2021 |
| 9 | KI 10 | NSO | KEY INFORMANT | 17 FEB 2021 |
| 10 | FGD 2: KI 11, 12, 13 | MDA | FOCUS GROUP | 25 FEB 2021 |
| 11 | KI 14 | NSO | KEY INFORMANT | 3 MAR 2021 |
| 12 | KI 15 | MDA | KEY INFORMANT | 5 MAR 2021 |
| 13 | KI 16 | MDA | KEY INFORMANT | 9 FEB 2021 |
| 14 | KI 17 | MDA | KEY INFORMANT | 15 FEB 2021 |
| 15 | FGD 3: KI 18, 19, 20 | NSO | FOCUS GROUP | 22 FEB 2021 |
| 16 | KI 21 | MDA | KEY INFORMANT | 3 MAR 2021 |

# 4. Findings

Four main themes were generated from the interviews: (1) legislation, (2) data systems, (3) communication, collaboration, and partnership, and (4) awareness and attitudes. Similarities in responses were classified under distinct sub-categories within the broader themes.

## 4.1 LEGISLATION

### Effectiveness of Legal Support

Most respondents emphasized the role of legal mandates in facilitating administrative data access. Many respondents also indicated that legal reforms have driven changes in administrative data sharing over time.

*[...] every public authority is obliged to deliver data for the purpose of production of statistics; what we do is not based on interests, it's based on legislation (KI 2). We have a data sharing policy [...] that allows members of this country to access data from several institutions; the administrative data from ministries [....] used to be very difficult to access (KI 8). [...] but basically, there exists a generic obligation to give INE data (FGD1, KI 6). [...] we are willing to share but always in compliance with the law (KI 15).*

### Legal Challenges

a. Privacy

However, legislation can also complicate efforts to increase administrative data sharing. Some respondents expressed concern about the legality of data sharing, citing data privacy and security protections. For example, when datasets are not anonymised with common identifiers, legal protections can prevent data sharing and linking.

*We don't publicate this information because it is privacy (KI 1). We think that we have this legal restrictions, [...] the main concern for us is that we use the named data (KI 15). Anonymisation is one of the limitations to work with the data we cannot integrate the data we got (FGD1, KI 4). data protection and confidentiality [...] is an important aspect (KI 2). You find that because of security concerns, the data sharing is not as open as other ministries (KI 8).*

b. Priorities

Some key informants expressed interest in legal reforms that prioritize the needs of statisticians. Others indicated a need for greater clarity and full implementation of existing laws.

*Sharing is not a priority now, but we can picture someday that the law goes in (FGD1, KI 6). We have a very narrow room of manoeuvre in order to produce what we like; [...] there are always some challenges no matter how advanced you are in the system (KI 2). When you share information there is so much bureaucracy (KI 1). We see a competition between different laws; so that's why we have this conflict of interests (KI 15). The other strategy is to implement fully the [...] access to information at all the ministries or departments that are lagging behind (KI 8).*

> *We see a competition between different laws; so that's why we have this conflict of interests*

c. Conflict and Cooperation

Importantly, despite the benefits of legal mandates, some respondents noted that invoking law enforcement runs counter to the interest of building long-term relationships.

*We can use sanctions [...], we, however, don't recommend using them because in order to have a healthy relationship with other institutions (FGD1, KI 6). We have this delicate balance to ask them nicely, and not punching him or her in the head with the legislation, because it doesn't foster cooperation (KI 2).*

> *We have this delicate balance to ask them nicely, and not punching him or her in the head with the legislation, because it doesn't foster cooperation*

## 4.2 DATA SYSTEMS

### Usefulness of Data for Statistical Purposes

Administrative data is sourced from various MDAs and includes data on a variety of topics from both private and public sectors that are collected over time. The unavailability of long series and sector representative data were identified as major barriers to the use of administrative data for statistical purposes. This was particularly true for Ghana and Chile.

*[...] there is a lot of activity within the private sector, and what we get from the Ministry of Health and Education, is mostly from the public side (FGD3, KI 19). [...] a lot of private hospitals [...] collect data, and these data do not feed into the Ghana health service data (FGD3, KI 20). [...] you find that the data that is available; [...] is only a short-term series [...] that limits your analysis (KI 8).*

Furthermore, inconsistencies in data management have also presented challenges to effectively re-use data.

*[...] the missing link is that whether that data we require and the data which they [ministries] are collecting [...] and that it is as per those standards which we [NSO] require [...] (KI 10). [...] our regions were coded starting from number 00 [...] the GSS is also started their coding of their regions from 01[...] (KI 16). [...] basically lots of services we do not have annoymised data. So from a lot of those records it is not possible for us to integrate information [...] (FGD1, KI 5).*

The absence of inter-sectoral guidelines leads to dissimilar data formats across MDAs. Thus, inconsistencies make it difficult to produce statistics.

In Chile [...] e*ach institution has their own protocols to share [their] own information; [...] and not all have the same standards to work with the data [...] (KI 1). [...] [Ministry] has its own standard operating procedures [...] we need to look at also developing that kind of guideline between the [Ministry] and the national statistics office [...] (KI 16).*

> *each institution has their own protocols to share [their] own information; [...] and not all have the same standards to work with the data*

### Resource Constraints to Data Sharing

Resource gaps were identified by most respondents as major barriers to data sharing. In particular, Ghana and Kenya expressed that data related activities are underfunded.

*[...] each ministry has a statistics department, and they are under resourced [...] (FGD3 KI20). [...] [ministry] activities are really seriously underfunded in our country [...] (KI 16). [...] without resources for data collection, even with the good systems, we may not have that data; [...] you find that those are funded, were able to collect as much data as possible. So resources is also key [...] (KI 8).*

> *[...] without resources for data collection, even with the good systems, we may not have that data*

Additionally, a respondent from Pakistan noted that data processes are occasionally tied to donor funded projects which are often discontinued when the project expires.

*[...] So, some of the data is sometimes linked with certain resources or projects, or donors. Once the donor goes away, the system comes back to zero level again. So, it's sometimes a donor driven system [...] (KI 10).*

Lastly, respondents from Pakistan and Ghana identified the importance of having a stable workforce and technology-enabled system to drive efficiency in administrative data sharing.

*[...] we move with persons not with the system, so much so when a person [government official] changes [transferred to another department] the whole scenario changes and you have to rebuild your old relations once again [...] (KI 10). [...] once we are able to automate the system [...] it doesn't take any human being's intervention [...] to transmit it [data] directly from the community to the centre [...] (KI 16).*

## 4.3 COMMUNICATION, COLLABORATION, AND PARTNERSHIPS

### Benefits of Communication, Collaboration, and Partnership

Many respondents emphasized that effective administrative data sharing depends on communication, collaboration, and partnership between NSOs and MDAs. Indeed, collaboration expanded access to data beyond legal mandates and improved its usefulness for statisticians.

*By sharing the data with other stakeholders, [ministries] found that [...] they could develop more data and provide more than even what is the Act of Parliament (KI 8). Authorities, they do not always remember statistics so some small amendments of registers in order to fulfill administrative needs can have huge consequences on statistics. So this dialog is a good way forward (KI 2).*

### Communication and Collaboration between NSOs and MDAs

However, some respondents explained that the degree of collaboration regarding data sharing can be uneven across sectors and disjointed across regions, inhibiting administrative data access and linking.

*[...] there is a lack of coordination among the ministries and the central government and the provincial governments. Data is available but it is available in silos, in patches [...] everything is being collected somewhere, somehow by someone, but the problem is the linking of those silo data [...] (KI 10).*

Respondents from both NSOs and MDAs emphasized that NSOs should play an active role in communication with MDAs to fill gaps in data sharing and some respondents suggested that absence of data sharing was due to lack of communication from the NSO.

*We [at the MDA] don't share the data with [the NSO]. I would say as far as I know we haven't received a specific request from our NSO (KI 15). [...] the NSO is not as proactive as it should be (KI 10). My first suggestion would be [for NSOs] to establish a dialogue (KI 2).*

Several respondents indicated that formal but non-legally binding agreements, such as memorandums of understanding (MOUs), between NSOs and MDAs can facilitate effective communication, build long-term relationships, and standardize administrative data sharing practices.

*[...] a memorandum of understanding on data exchange, which is not a binding document, the main purpose is to put a face on the person sitting both in statistical institution and in authority and have some general agreement, how and what frequency the data is transmitted, and exchanged. And, and in my view, this is a useful document – even not being legally binding – because it adds a human perspective (KI 2). Signature of memorandum of understanding is really important as normally the interaction we have with other institutions is not just one time, but it has to be recorded over time [...] the document has a signature, so it is like making more official (FGD1, KI 6).*

### International Partnerships

Respondents from Ghana and Pakistan mentioned that international development organizations have advocated for registries as a source of administrative data and have improved the quality of administrative data systems.

*[...] UNICEF, Plan International, who are the child centred organizations, are doing a lot of advocacy work on the benefits of birth registration [and] most of the programs are funded by our development partners, UNICEF has funded most of the programs [...] the World Bank also has come to the aid of some of the programs [...] (KI 16). Over time the quality [of data] has been improved because definitely you are engaged with different organizations [...] IMF and the World Bank is there so we are improving our quality. If you go 5 or 10 years back, the data quality may not be even [and] the dissemination on the website was not there [...] (FGD2, KI 12).*

> *Data is available but it is available in silos, in patches [...] everything is being collected somewhere, somehow by someone, but the problem is the linking of those silo data*

However, international partners can also have unintended consequences on administrative data systems. A respondent from Ghana noted that new data systems introduced by development partners to monitor vertical programs can undermine, rather than improve or supplement, existing data systems.

*You [development partners] set up your parallel system for compiling data for you to monitor whether the resources you are giving to that ministry is making progress or not [...] when you set up a parallel system, you are indirectly killing the existing system (FGD3, KI 20).*

*You [development partners] set up your parallel system for compiling data for you to monitor whether the resources you are giving to that ministry is making progress or not [...] when you set up a parallel system, you are indirectly killing the existing system*

## 4.4 AWARENESS AND ATTITUDES

### Public Attitudes and Awareness

Respondents underscored various rationales for differing levels of interest in sharing of administrative data among MDAs and NSOs. Notions of trust and privacy were commonly recognized as significant determinants of sharing data.

*In Denmark, this trust is high. [...] Let me generalize everyone believes that the government will take proper protection of privacy data (KI 2). I would say a one of the gap is that fear, that fear of misuse of data [...] (KI 8).*

Further, respondents highlighted that enhanced visibility of data allowed for more positive attitudes towards data sharing and cooperation. Particularly, this was a common assertion concerning macro-economic data.

*[...] Since they [Ministry of Finance] want the public to participate in a budget making process, they they reveal the data as much as possible for the public to get the information and to discuss that information and give feedback (KI 8). This is how you represent yourself to the world. So, therefore, you have to improve the quality (FGD2, KI 12).*

Respondents from Pakistan and Kenya emphasized the role of social media in increasing citizens' knowledge and participation. Public awareness thus was recognized as an impetus to increase transparency of data systems.

*The moment you release a number, it goes viral. It goes on social media. [...] They [public] start asking questions and when you start asking questions then you start asking for the reasons and for the solutions also and then that gives you a motivation to make your system more clear, open and refined (KI 10). They [NSO] [...] can share their data to all the citizens that can be done is okay, then you can make use of the social media [...] everybody will be able to get that information as and when it is required (KI 8).*

### Attitudes of Authorities and Stakeholders

In some cases, respondents also indicated that progress in improving data sharing was linked to attitudes of the MDA leaders.

*[...] But you find that the ministries are at different levels [...] such that you find even this sharing of data is at different stages. Those that were sharing it, the data, they increased or improved data sharing. Those other ministers that were a bit closed up, you find even what they share is not enough for, for good analysis (KI 8) But there are departments which move around a certain person. An innovative person, a dynamic person comes and it makes certain things so efficient and certainly he leaves or moves from the scene, a new person comes with a different mentality mentality, a different methodology and the system starts decreasing and decaying again (KI 10).*

> *Those other ministers that were a bit closed up, you find even what they share is not enough for, for good analysis*

### Preferences of Statisticians

Preferences for data collection were often based on variables relevant to a particular MDA sector or NSO division. However, respondents illustrated varying prioritization of administrative data in their approach to gathering data on topics of interest.

*[...] our first thought is not administrative data, it's like, well, we don't have the data, so we need to make a survey (FGD1, KI 4). [...] in 2015, we did an assessment, and we realized that look, administrative data will need to form the larger basis of computing the SDG indicators [...] over the years, the focus has been more on the survey and census programs than administrative data, except for economic statistics [...] (FGD3, KI 20).*

# 5. Discussion

In-depth interviews with specialists and experts from five different countries – Chile, Denmark, Ghana, Kenya, and Pakistan – provided key insights into the existing challenges to effective re-use of administrative data. A combination of factors, grouped across four broad themes – namely, legislation, data systems, attitudes, and partnerships – were identified as key determinants of data sharing. These factors, by virtue of their interconnectedness, cannot be addressed in isolation if efficient and effective data sharing and use is to be achieved.

According to the findings, the value of using administrative data to inform policy decisions is well-understood by MDAs. However, in practice, the dissemination of administrative records is met with legal and ethical challenges that are inherently complex. While it is widely agreed that legal mandates promote greater access to data, there is a high degree of reluctance towards implementing legal enforcement when NSOs are faced with poor cooperation. Instead, countries such as Chile, Denmark, and Ghana prefer to rely on interpersonal relationships to maintain engagements.

However, relying on relationships alone can lead to inconsistent data-sharing, especially in the event of staff turnover. Consequently, time and effort are required to rebuild relationships, during which there is a halt in data sharing between the NSO and MDA. For instance, in some cases, MDAs collect required variables based on informal requests, however, changes in staff leads to discontinuity of that data. Respondents from many countries, including Chile, Denmark, Kenya, and Pakistan, noted that formal but non-legally binding agreements, such as MOUs, are valuable for balancing friendly interagency relationships with stability of data sharing flows. Overall, a combination of cooperative relationships, formal partnerships, and legislative support is needed to foster sustainable and effective administrative data sharing.

Additionally, respondents from Chile, Ghana, and Pakistan noted that technical barriers also limit the usefulness of administrative data for statistics. Specifically, lack of common identifiers across MDAs can prevent data linking from different sources. In some cases, due to lack of standardization, extensive reformatting was necessary to make administrative data suitable for statistical purposes, disincentivizing data sharing. Moreover, in some cases, the usefulness of administrative data for research, particularly longitudinal studies, is further restricted by lack of historical data. Additionally, respondents from Ghana indicated that administrative data is under-representative of the general population because it is mostly derived from public sources. These technical limitations undermine the potential benefits of using administrative data to

produce national statistics. Despite these challenges, many respondents emphasized progress in overcoming technical barriers through capacity-building efforts supported by NSOs.

Resource availability was identified as a catalyst for data-related activities. For example, MDAs that were adequately funded amassed a greater amount of data in Kenya. However, project specific donor financing provided only short-lived data improvements in Pakistan. While the NSO in Pakistan emphasized the importance of a stable data system, a respondent from Kenya suggested that resource deficiency prevented efficient systems from functioning at adequate capacity. Additionally, while development partners have been instrumental in advocating for increased investment in data-related activities, evidence from Ghana suggested that donor-funded parallel systems undermine existing data structures.

This combination of persisting resource shortages and piecemeal donor involvement make it difficult to overcome the technical barriers described above, compounding the obstacles to administrative data re-use in LMICs. Ensuring that administrative data systems are consistently supported by adequate resources is a complex problem without a single, straightforward solution. However, one facet of the solution is to raise awareness of the value of administrative data for informing national statistics and improving government efficiency in the long run, particularly among government leaders and others with the power to influence resource allocation. Thus, the findings underscored that improvements in data will require national government commitment in addition to donor investments in the development of sustainable data systems.

In addition to potentially mitigating resource limitations, increased public awareness of the benefits of re-using administrative data is crucial to building political will for increasing data access and enhancing data quality. Evidence from Pakistan suggested data pertaining to economic indicators benefited from increased public feedback. This particularly motivated the NSO to increase data transparency and collaboration for improving data quality. Similarly, respondents from Kenya and Pakistan emphasized the critical role social media can play in increasing public acknowledgment of data. Additionally, a respondent from Denmark noted that high levels of public trust in the government facilitates the country's extensive administrative data sharing. Thus, by maintaining data visibility and public trust, NSOs can foster positive attitudes toward data sharing and facilitate its re-use to inform evidence-based policy.

In identifying four interconnected themes across the five countries, the findings offer direction for enhancing the re-use of administrative data. Given the diversity of challenges presented and complex relationships among the themes, a coordinated, multi-sectoral approach is required to

effectively improve data sharing landscapes. This includes directed efforts towards increasing collaboration between NSOs and MDAs as well as engagement of international partners, government officials, researchers, and the public. Overall, data system strengthening coupled with legislative support and increased awareness of the benefits of data sharing are paramount to bringing long-term solutions. Further research is encouraged to investigate these findings in greater depth and to tailor solutions to local contexts.

# 6. Limitations

The findings and discussion of this report should be considered in light of some limitations. Firstly, the respondents indicated many sources of barriers to administrative data sharing beyond the direct jurisdiction of MDAs and NSOs. Since interviews were limited to representatives from MDAs and NSOs, this report includes limited discussion of the roles of other relevant actors. Future research could further investigate the perspectives of other actors by expanding the scope of the sample.

A second limitation of this report is that the research process was conducted in English, which is not the official language of some of the countries included in the study. Consequently, non-English documentation, reports, and publications were not included as secondary sources for background research. Since respondents were recruited in English, language barriers may have discouraged non-English speakers from participating, potentially restricting the sample. Interviews were also conducted in English, which may have increased the risk of miscommunication. The interview questionnaires were provided in advance and a translator was included in one of the interviews to mitigate these limitations.

Additionally, the process of recruiting respondents and conducting interviews was administered online via email and Zoom. Consequently, potential respondents without reliable, high-speed internet access may have been unable to respond or discouraged from participating, which could have further limited the response rate and restricted the sample. This limitation was mitigated by allowing a written response in one case. Moreover, the online format of the interviews may have limited the flow and depth of the conversations, especially for the focus groups. If feasible, future research could benefit from conducting in-person interviews in the official language of the country.

# 7. Recommendations and Conclusion

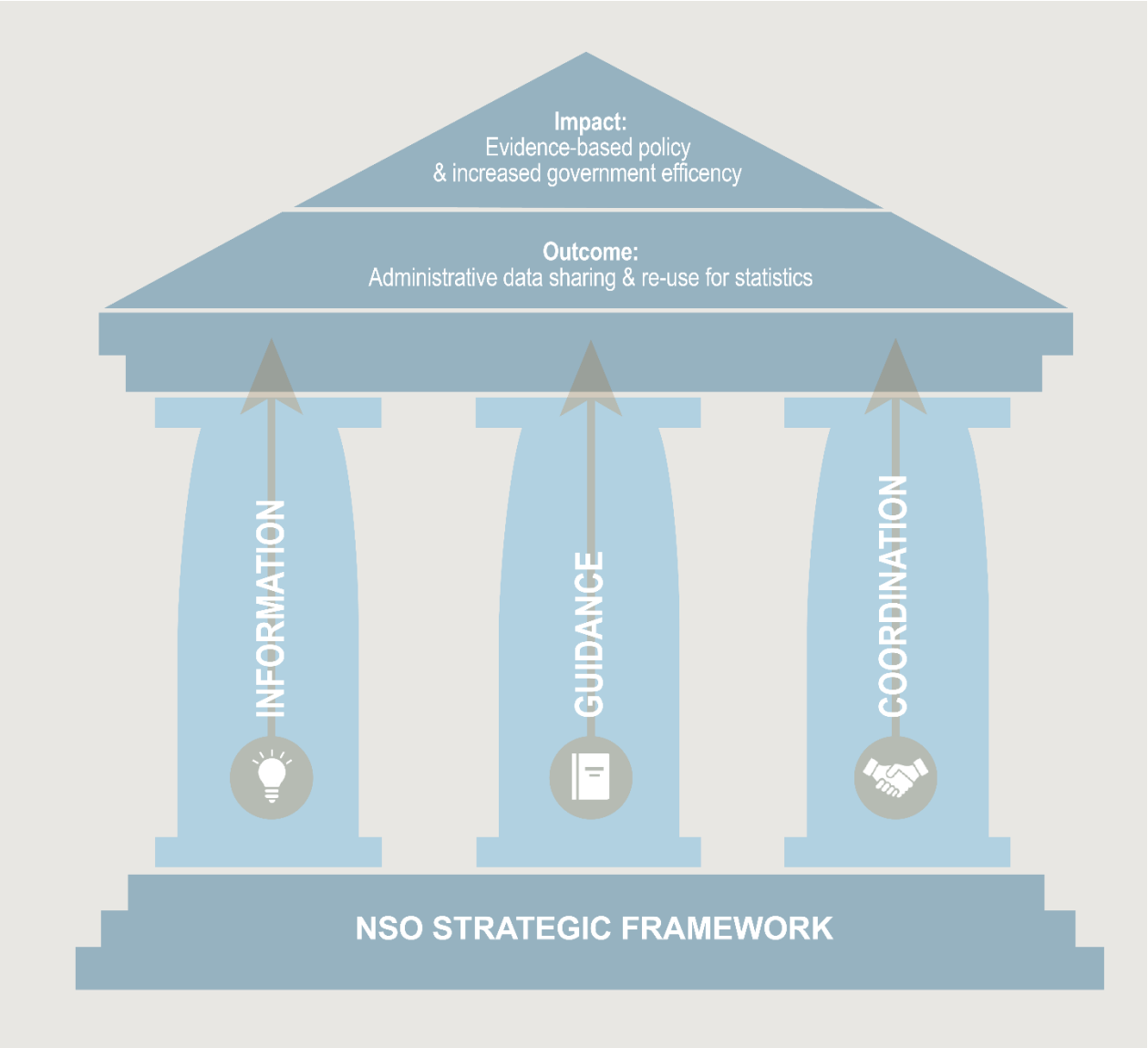
Increasing administrative data sharing and re-use requires coordinated actions and aligned objectives of diverse actors participating in the various levels of data flows, from its collection to its application. Given their centralized authority over national statistics production and understanding of country contexts, NSOs play a particularly crucial role in initiating and unifying changes in data practices at the national level. Thus, to cultivate an interest in administrative data sharing and increase awareness of the value of administrative data for decision making, we urge NSOs to produce strategic frameworks outlining their approach to administrative data collaboration and long-term objectives for enhancing such multilateral and multisectoral relationships. We propose a strategic plan framework comprising three pillars: distribution of critical information, development of explicit guidance, and facilitation of organized coordination.

I. INFORMATION
- Utilize social media and other interactive platforms to increase public awareness and citizen-led accountability for administrative data re-use.
- Organise briefings with government leaders to encourage investment in sector-specific data collection and management.
- Engage in political discourse to frame administrative data as an asset for evidence-based policymaking.
- Facilitate and advocate for research using administrative data to demonstrate its applications.

II. GUIDANCE
- Lobby for legislative reforms to ensure sufficient resources for administrative data collection and access.
- Develop and disseminate standards of practice on data privacy and cybersecurity.
- Establish common identifiers to facilitate data linking and support standardized data formatting.

III. COORDINATION
- Establish a taskforce for multisectoral collaboration with defined terms of reference (ToR) including:
    - Identifying region-specific examples of data systems successes to promote progress.
    - Sustaining dialogue between stakeholders to exchange strategies and tools for data collection and sharing.
    - Building accountability structures to enforce data sharing and adherence to standards of practice.
- Produce guidelines for international donors collecting data to align with existing data systems and contribute to data sharing flows.

*Figure 9: Framework for NSO strategic plans*




While progress towards leveraging administrative data for statistics requires proactive leadership by NSOs, other actors are also urged to implement changes to support effective administrative data collection, sharing, and re-use. Based on the findings of this report, we recommend that MDAs allocate adequate resources for administrative data quality assurance and designate focal persons to maintain collaborative relationships with NSOs. Additionally, government leaders, legislators, and policy makers should consult NSOs in the development of budgets, laws, and policies that impact administrative data collection, management, and access. Finally, we encourage development partners and international organisations, including the United Nations Statistics Division, to avoid creating short-term, parallel data governance and instead channel resources towards strengthening the existing administrative data systems and supporting NSOs in their strategic plans to enhance administrative data re-use for statistics.

# Bibliography

About Us | *Pakistan Bureau of Statistics* (no date). Available at: https://www.pbs.gov.pk/content/about-us (Accessed: 12 March 2021).

Allard, S. W. et al. (2018) 'State Agencies' Use of Administrative Data for Improved Practice: Needs, Challenges, and Opportunities', *Public Administration Review*, 78(2), pp. 240–250. doi: https://doi.org/10.1111/puar.12883.

Bharadwaj, P., Løken, K. V. and Neilson, C. (2013) 'Early life health interventions and academic achievement', *American economic review*, 103(5), pp. 1862–1891.

Boden, P. and Rees, P. (2010) 'Using administrative data to improve the estimation of immigration to local areas in England', *Journal of the Royal Statistical Society. Series A* (Statistics in Society), 173(4), pp. 707–731.

Boyatzis, R. E. (1998) Transforming Qualitative Information: Thematic Analysis and Code Development. *SAGE*.

Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', *Qualitative Research in Psychology*, 3(2), pp. 77–101. doi: 10.1191/1478088706qp063oa.

Burnard, P. et al. (2008) 'Analysing and presenting qualitative data', *British Dental Journal*, 204(8), pp. 429–432. doi: 10.1038/sj.bdj.2008.292.

Card, D. E. et al. (2010) 'Expanding Access to Administrative Data for Research in the United States', *SSRN Electronic Journal*. doi: 10.2139/ssrn.1888586.

Connelly, R. *et al.* (2016) 'The role of administrative data in the big data revolution in social science research', *Social Science Research*, 59, pp. 1–12. doi: 10.1016/j.ssresearch.2016.04.015.

*Denmark Statistics | About us* (no date). Available at: https://www.dst.dk/en/OmDS (Accessed: 26 March 2021).

Dinkelman, T. and Martínez A., C. (2014) 'INVESTING IN SCHOOLING IN CHILE: THE ROLE OF INFORMATION ABOUT FINANCIAL AID FOR HIGHER EDUCATION', *The Review of Economics and Statistics*, 96(2), pp. 244–257.

Enayati, H. and von Schrader, S. (2016) 'Using Administrative Data', in Bruyère, S. M. (ed.) Disability and Employer Practices. 1st edn. Cornell University Press (Research across the Disciplines), pp. 79–

96. Available at: https://www.jstor.org/stable/10.7591/j.ctt18kr56q.9 (Accessed: 12 March 2021).

Figlio, D., Karbownik, K. and Salvanes, K. (2017) 'The Promise of Administrative Data in Education Research', Education Finance and Policy, 12(2), pp. 129–136. doi: 10.1162/EDFP_a_00229.

Gavrielov-Yusim, N. and Friger, M. (2014) 'Use of administrative medical databases in population-based research', Journal of Epidemiology and Community Health (1979-), 68(3), pp. 283–287.

Ghana Statistical Services. (2020). *About Us*. https://statsghana.gov.gh/aboutgss.php?category=MzQzMDg4NDY0Ljg5NDU=/webstats/r410p12p51

Goroff, D., Polonetsky, J. and Tene, O. (2018) 'Privacy Protective Research: Facilitating Ethically Responsible Access to Administrative Data', The ANNALS of the American Academy of Political and Social Science, 675(1), pp. 46–66. doi: 10.1177/0002716217742605.

Hand, D. J. (2018) 'Statistical challenges of administrative and transaction data', Journal of the Royal Statistical Society: Series A (Statistics in Society), 181(3), pp. 555–605. doi: 10.1111/rssa.12315.

Harron, K. et al. (2017) 'Challenges in administrative data linkage for research', Big Data & Society, 4(2), p. 2053951717745678. doi: 10.1177/2053951717745678.

Harron, K., Goldstein, H. and Dibben, C. (2015) Methodological Developments in Data Linkage. John Wiley & Sons.

*History of KNBS - Kenya National Bureau of Statistics* (no date). Available at: https://www.knbs.or.ke/?page_id=86 (Accessed: 26 March 2021).

Holman, C. *et al.* (2008) 'A Decade of Data Linkage in Western Australia: Strategic Design, Applications and Benefits of the WA Data Linkage System', *Australian health review : a publication of the Australian Hospital Association*, 32, pp. 766–77. doi: 10.1071/AH080766.

International Monetary Fund (2007) 'Chile: Report on the Observance of Standards and Codes: Data Module, Response by the Authorities, and Detailed Assessment Using the Data Quality Assessment Framework (DQAF)', *IMF Staff Country Reports*, 07(318), p. 1. doi: 10.5089/9781451951523.002.

Jutte, D. P., Roos, L. L. and Brownell, M. D. (2011) 'Administrative Record Linkage as a Tool for Public Health Research', *Annual Review of Public Health*, 32(1), pp. 91–108. doi: 10.1146/annurev-publhealth-031210-100700.

Khalid, A. M., Sharma, S. and Dubey, A. K. (2020) 'Data Gap Analysis, Indicator Selection and Index Development: A Case for Developing Economies', Social Indicators Research, 148(3), pp. 893–960. doi: 10.1007/s11205-019-02225-6.

Kumar, M. et al. (2018) 'Research gaps in routine health information system design barriers to data quality and use in low- and middle-income countries: A literature review', The International Journal of Health Planning and Management, 33(1), pp. e1–e9. doi: https://doi.org/10.1002/hpm.2447.

Lavrakas, P. (2008) Encyclopedia of Survey Research Methods. 2455 Teller Road, Thousand Oaks California 91320 United States of America: Sage Publications, Inc. doi: 10.4135/9781412963947.

Mazzali, C. and Duca, P. (2015) 'Use of administrative data in healthcare research', *Internal and Emergency Medicine*, 10(4), pp. 517–524. doi: 10.1007/s11739-015-1213-9.

Ministry of Foreign Affairs of Denmark (2019). *Statistics*. Denmark in Ghana. Retrieved March 25, 2021, from https://ghana.um.dk/en/sector-cooperation/statistics/

Nacional, B. del C. (2008) *Biblioteca del Congreso Nacional | Ley Chile*, *www.bcn.cl/leychile*. Available at: https://www.bcn.cl/leychile (Accessed: 11 March 2021).

Nagler, P. (2013) 'How unemployment insurance savings accounts affect employment duration: evidence from Chile', *IZA Journal of Labor & Development; Heidelberg*, 2(1), pp. 1–25. doi: http://dx.doi.org.gate3.library.lse.ac.uk/10.1186/2193-9020-2-9.

Nowell, L. S. et al. (2017) 'Thematic Analysis: Striving to Meet the Trustworthiness Criteria', International Journal of Qualitative Methods, 16(1), p. 1609406917733847. doi: 10.1177/1609406917733847.

*Pakistan | Data* (2019b). Available at: https://data.worldbank.org/country/pakistan (Accessed: 12 March 2021).

Palinkas, L. A. et al. (2015) 'Purposeful sampling for qualitative data collection and analysis in mixed method implementation research', Administration and policy in mental health, 42(5), pp. 533–544. doi: 10.1007/s10488-013-0528-y.

Petrila, J. (2018) 'Turning the Law into a Tool Rather than a Barrier to the Use of Administrative Data for Evidence-Based Policy', The ANNALS of the American Academy of Political and Social Science, 675(1), pp. 67–82. doi: 10.1177/0002716217741088.

*Population, total - Ghana | Data* (no date). Available at: https://data.worldbank.org/indicator/SP.POP.TOTL?locations=GH (Accessed: 25 March 2021).

Robinson, O. C. (2014) 'Sampling in Interview-Based Qualitative Research: A Theoretical and Practical Guide', Qualitative Research in Psychology, 11(1), pp. 25–41. doi: 10.1080/14780887.2013.801543.

Rubin, H.J. and Rubin, I.S. (2004). Qualitative Interviewing: The Art of Hearing Data. 2nd ed. Sage Publications, Thousand Oaks, CA.

Sanhueza, G. E. (no date) 'Exploring correlates of Prison Violence in Chilean Prisons: Examining nationwide, administrative data', p. 123.

Saunders, M. N., Lewis, P., Thornhill, A. (2016). Research Methods for Business Students. 7th Edition [VitalSource Bookshelf version]. Retrieved from vbk://9781292016641

Schnell, R. (2014) 'An efficient privacy-preserving record linkage technique for administrative data and censuses', Statistical Journal of the IAOS, 30(3), pp. 263–270. doi: 10.3233/SJI-140833.

Sexton, A. et al. (2017) 'A balance of trust in the use of government administrative data', Archival Science, 17(4), pp. 305–330. doi: 10.1007/s10502-017-9281-4.

The World Bank Group (2018) *Chile | Data*. Available at: https://data.worldbank.org/country/chile (Accessed: 18 March 2021).

The World Bank Group (2019a). *Denmark | Data*. Available at: https://data.worldbank.org/country/denmark?view=chart (Accessed: 10 March 2021).

The World Bank. (2019c). *Ghana | Data*. https://data.worldbank.org/country/GH

The World Bank (2019d). *Kenya | Data*. Available at: https://data.worldbank.org/country/KE (Accessed: 10 March 2021).

Woollard, M. 2014. Administrative data: problems and benefits. A perspective from the United Kingdom. In: Duşa, A. et al. (eds). *Facing the future: European research infrastructures for the humanities and social sciences.* [Online]. Berlin: Scivero, pp. 49-60. [Accessed 25 March 2021]. Available from: https://www.allea.org/wp-content/uploads/2015/09/2014_06_04-FACING_THE_FUTURE.pdf

# Appendices

## Appendix one: Terms of Reference

| Project Proposal | |
|---|---|
| Organisation and Department | UN Department for Economic and Social Affairs, Statistics Division Collaborative on administrative data |
| Project Working Title | Improving the efficiency of government through re-use of administrative data sources for statistical purposes |
| Background: | The UN Statistics Division (UNSD) has a core mandate in supporting National Statistical Offices (NSOs) in their work to provide timely and reliable data and statistics for informed decision making. It has the secretariat role for the annual Statistical Commission where Chief Statisticians from NSOs meet to discuss and agree on recommendations and practical implementation of statistics production and publication. Many countries are in a situation where there is a strong need for capacity development support in the area of statistics and UNSD is therefore supporting NSOs and their partners in countries through various projects and activities.<br><br>The Collaborative on administrative data was established in May 2020 and has since then worked with a selected number of countries and international agencies to help re-use of data collected for administrative purposes (f.ex. health records, birth and death records, custom records or company data) for official statistics production. Key challenges are linked to trust and collaboration among agencies, agreeing on common standards and processes to ensure sufficient data quality. |

| | |
|---|---|
| Question: | What can be done to improve the interest and understanding among Government entities in Chile, Denmark, Ghana, Kenya, and Pakistan that there is a benefit from working together to share and improve quality of data collected? And how to create the same interest among policy and decision makers to push for these developments to increase the availability of evidence for decision making? |
| Objective: | The main objective with this assessment is to understand more in detail what are the main challenges to sharing data between health, education and economy sectors within Chile, Denmark, Ghana, Kenya, and Pakistan? Additionally, provide recommendations on how to bring about a change? If successful, the work will in the longer run ideally lead to more timely and disaggregated data for the Sustainable Development Goals (SDGs) as well as other national statistics. |
| Methodology: | The students will carry out desk research, focus group discussions and interview a number of statisticians and policy makers to better understand the challenges and good experiences that countries are facing. Based on this they will present the finding and propose new approaches to improve collaboration to increase the use of administrative data for statistical purposes. The proposals should take into account the low budgets and national governance contexts.<br><br>The students will be working with a team of experts at UNSD and the wider Collaborative on administrative data. They will also be able to share their experiences and results through online workshops and meetings. |

## Appendix two: Interview Questions

Questions for NSO:

1. Please describe your role, particularly how you use administrative data.
2. Is sharing administrative data with the NSO mandated for any ministries?
    a. How frequently is data shared from the ministries?
3. What are the main sources of administrative data for the NSO?
4. Does the NSO have a preference for data based on:
    a. Government sectors
    b. Indicators or types of data
    c. Are resources and efforts allocated in accordance with preference?
5. Are standards of practice or data sharing guidelines readily available to ministries?
   If yes:
        i. What are some of the main guidelines?
        ii. To what extent are the guidelines followed?
        iii. Is there consensus over the requirements from the NSO?
   If no:
        iv. Would data sharing guidelines be useful to ministries?
6. To your knowledge, please comment on administrative data sharing trends over time. If changes have or have not occurred:
    a. Why?
7. What is the attitude toward administrative data sharing among ministries?
8. Are there any limitations or gaps in access to administrative data for the NSO?
9. What barriers/limitations are there to using administrative data for statistical or research purposes?
10. What strategies are most helpful in facilitating the use and reuse of administrative data for statistical purposes?

Questions for MDAs:

1. Please describe your role, particularly how you use administrative data.
2. What types of administrative data are collected by your ministry?
    a. How is the data collected, stored, and shared?
    b. Have there been any notable changes in the types of administrative data collected by your ministry, or in the way administrative data is collected, stored, and shared over the past several years?
3. Is any administrative data collected by your ministry routinely shared with the NSO?
    i. If yes:
        a. Are there standards of practice or guidelines for data sharing?

b. How frequently is data shared?

ii. If no, why do you think this is the case?

4. Are you aware of any existing or previous partnerships or collaborations between your ministry and the NSO?
5. What is the attitude toward sharing administrative data with the NSO within your ministry?
6. In your opinion, are there any benefits to sharing administrative data with the NSO?
7. What do you see as the major challenges to increasing administrative data sharing with the NSO?
8. Have there been any efforts to increase administrative data sharing between your ministry and the NSO?
    a. If so, what were they and to what extent were they successful or unsuccessful?
9. What do you think would be the most effective way to improve administrative data sharing between your ministry and the NSO?

## Appendix three: Information Sheet

Research Topic:
Re-using Administrative Data for Statistics: Case Studies from Five Countries

Name of researchers: [names removed]
Health and International Development, LSE

Information for participants
Thank you for considering participating in this study which will take place during the months of January 2020 and March 2021. This information sheet outlines the purpose of the study and provides a description of your involvement and rights as a participant, if you agree to take part.

What is the research about?
The main objective of this research is to investigate the main challenges to sharing data among the health, education, and economy sectors, and the national statistics office in the following countries: Chile, Denmark, Ghana, Kenya, and Pakistan. Additionally, this research seeks to provide recommendations for addressing these challenges. Findings from this work will in the long run lead to more timely and disaggregated data for the Sustainable Development Goals (SDGs) as well as other national statistics. Focus group discussions and interviews with statisticians, policy makers, and administrators will be conducted to better understand the challenges and good experiences that countries are facing.

Do I have to take part?
Participation in this research is entirely voluntary. If you decide to take part, we will ask you to sign a consent form which you can sign and return in advance of the interview/focus group meeting.

What will my involvement be?
You will be asked to take part in an interview/focus group discussion about your experience/knowledge of administrative data collection, flow, submission, and sharing in your country. This interview will take approximately 40 minutes.

How do I withdraw from the study?

You can withdraw from the study at any point until the 31st of March 2021, without having to give a reason. If any questions during the interview/focus group discussion make you feel uncomfortable, you do not have to answer them. Withdrawing from the study will have no effect on you. If you withdraw from the study, we will not retain the information you have given thus far, unless you are happy for us to do so.

What will my information be used for?
We will use the collected information for a research project to better understand more in detail what the main challenges and barriers are to data sharing between health, education, economy sectors, and the respective national statistics office in the following countries - Chile, Denmark, Ghana, Kenya, and Pakistan?

Will my taking part and my data be kept confidential? Will it be anonymised?
The records from this study will be held internally and kept strictly confidential. Only the team members and an academic supervisor at the London School of Economics will have access to the interview files and recordings. Your data will be anonymised – neither your name nor any identifiable information will be included in reports or publications resulting from the study without your consent. [1] All digital files, transcripts and summaries will be given codes and stored separately from any names or other direct identification of participants. Any physical copies of research information will be kept in locked files at all times.

Limits to confidentiality: confidentiality will be maintained as far as it is possible, unless you tell us something which implies that you or someone you mention might be in significant danger of harm and unable to act for themselves; in this case, we may have to inform the relevant agencies of this, but we would discuss this with you first.

Who has reviewed this study?
This study has undergone ethics review in accordance with the LSE Research Ethics Policy and Procedure.[2]

Data Protection Privacy Notice
The LSE Research Privacy Policy can be found at:
https://info.lse.ac.uk/staff/divisions/Secretarys-Division/Assets/Documents/Information-Records-Manage ment/Privacy-Notice-for-Research-v1.1.pdf

The legal basis used to process your personal data will be Students "Legitimate interests. The legal basis used to process special category personal data (e.g., data that reveals racial or ethnic

origin, political opinions, religious or philosophical beliefs, trade union membership, health, sex life or sexual orientation, genetic or biometric data) will be for scientific and historical research or statistical purposes. To request a copy of the data held about you please contact: glpd.info.rights@lse.ac.uk

What if I have a question or complaint?
If you have any questions regarding this study please contact any of the following: [names and emails removed]

If you have any concerns or complaints regarding the conduct of this research, please contact the LSE Research Governance Manager via research.ethics@lse.ac.uk .

If you are happy to take part in this study, please sign the consent sheet attached.

# Appendix four: Consent form

Research Topic: Re-use of Administrative Data for Statistics: Case Studies from Five Countries

Researcher: [names removed]

PARTICIPATION IN THIS RESEARCH STUDY IS VOLUNTARY

| | |
|---|---|
| I have read and understood the study information dated [__/__/____], or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction. | YES / NO |
| I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and that I can withdraw from the study at any time up until March 31st 2021, without having to give a reason. | YES / NO |
| I agree to the [interview/focus group] being audio recorded | YES / NO |
| [For focused group discussion only!] I agree to maintain the confidentiality of the focus group discussions. | YES/NO |
| I understand that the information I provide will be used for a consultancy project with the UN Department for Economic and Social Affairs, Statistics Division Collaborative on administrative data and that the information will be anonymised. | YES / NO |
| I agree that my (anonymised) information can be quoted in research outputs. | YES / NO |
| I understand that any personal information that can identify me – such as my name and contact information – will be kept confidential and not shared with anyone beyond the study team. | YES / NO |
| I give permission for the (anonymised) information I provide to be deposited in a data archive so that it may be used for future research. | YES / NO |

Please retain a copy of this consent form.

Participant name:

Signature: _____ Date _____

Interviewer name:

Signature: _____ Date _____