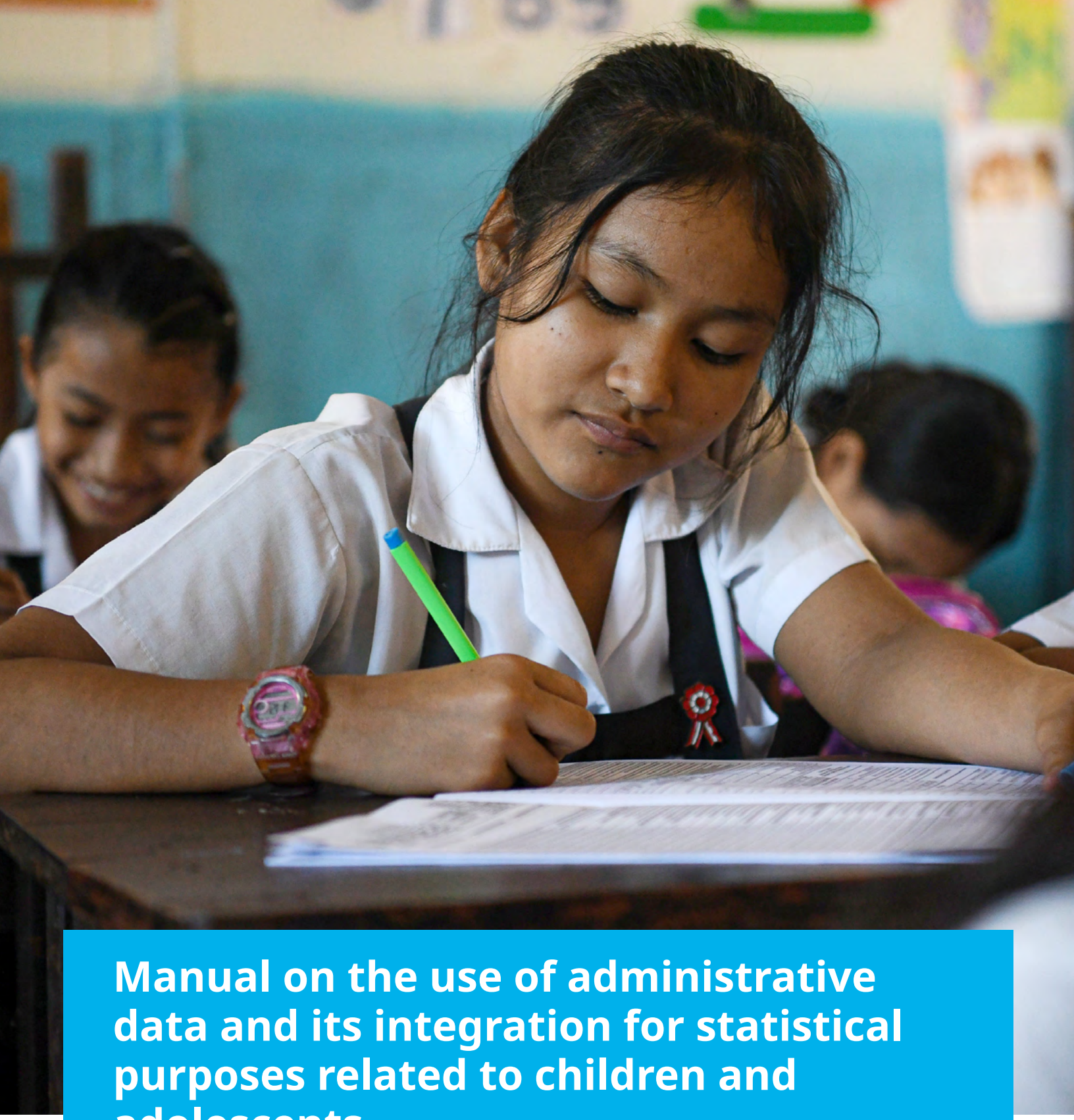


Manual on the use of administrative data and its integration for statistical purposes related to children and adolescents

United Nations
ECLAC

Statistical Conference of the Americas of ECLAC

INE Instituto Nacional de Estadísticas · Chile

URUGUAY INE Instituto Nacional de Estadística

unicef for every child

Manual on the use of administrative data and its integration for statistical purposes related to children and adolescents

ISBN: 978-92-806-5714-2

This manual was conceptualized and developed by the UNICEF Regional Office for Latin America and the Caribbean with leadership and technical guidance from the National Statistics Institute of Chile and the National Institute of Statistics of Uruguay, in their capacity as co-chairs of the Working Group on Children's Statistics of the Statistical Conference of the Americas of the Economic Commission for Latin America and the Caribbean.

The working group includes national institutions from member countries: the National Institute of Statistics and Census of Argentina; the Barbados Statistical Service; the Statistical Institute of Belize; the National Statistical Institute of Bolivia; the Brazilian Institute of Geography and Statistics; the National Institute of Statistics and Census of Costa Rica; the National Office of Statistics and Information of Cuba; the National Statistics Office of the Dominican Republic; the National Institute of Statistics of Honduras; the National Institute of Statistics and Geography of Mexico; the Statistics Department of Montserrat; the National Institute of Statistics and Censuses of Nicaragua; the National Institute of Statistics and Census of Panama; the National Institute of Statistics of Paraguay; the National Institute of Statistics and Informatics of Peru; the Department of Statistics of St. Kitts and Nevis; the General Bureau of Statistics of Suriname; the Central Statistical Office of Trinidad and Tobago; and the National Statistical Institute of Venezuela.

ECLAC: LC/CEA.13/DDR/5

Cover photo: © UNICEF/UNI415135/Romani

Design: Tinto Estudio, S.A.

UNICEF
Latin America and Caribbean Regional Office
Building 102, Alberto Tejada St., City of Knowledge
PO Box: 0843-03045
Panama, Republic of Panama
Tel: +507 301-7400
www.unicef.org/lac
uniceflac@unicef.org

Manual on the use of administrative data and its integration for statistical purposes related to children and adolescents

Contents

List of Figures

Executive Summary

This manual provides guidance on the use of administrative data for statistical purposes related to children and adolescents, with a focus on data linkage and cross-sectoral integration. Prepared by the UNICEF Regional Office for Latin America and the Caribbean with leadership and technical guidance from the National Institute of Statistics of Chile and the National Institute of Statistics of Uruguay, in their capacity as co-chairs of the Working Group on Child Statistics of the Statistical Conference of the Americas of the Economic Commission for Latin America and the Caribbean, the manual aims to strengthen the capacity of National Statistical Offices (NSOs) to produce high-quality, timely, and relevant statistics on children. Using administrative data, it supports decision-making that upholds the rights of the child as defined in the Convention on the Rights of the Child (CRC) and contributes to the achievement of the Sustainable Development Goals (SDGs) by generating critical data.

Administrative data—collected through public services such as health, education, social welfare, and civil registration—can provide cost-effective and timely insights for child-related statistics, particularly when linked across sectors. Such data are essential for fulfilling the CRC and for monitoring progress on the 2030 Agenda for Sustainable Development.

The manual is organized into three chapters, each designed to guide NSOs in the strategic and technical use of administrative data for statistics on children and adolescents:

Chapter 1 frames administrative data for child-focused statistics, highlighting its role in producing timely, disaggregated, and policy-relevant insights. It presents UNICEF's *Conceptual Framework and Guidance for Integrating a Children's Rights Perspective into Administrative Data-Based Statistics,* emphasizes the importance of cross-sectoral data linkage and integration, and introduces models for the effective use of administrative data.

Chapter 2 examines the role of administrative data in monitoring and advancing SDGs related to children and adolescents. It underscores how linked administrative data improve measurement, analysis, reporting, and the understanding of interconnections between child-related SDG indicators. Case studies from Chile, Brazil, Uruguay, and Peru illustrate best practices. Chile's study on migration statistics focuses on children and adolescents through the integration of census data and administrative records. Brazil's study examines the linkage of vital statistics with health systems to estimate under-reporting of births and deaths. Uruguay has developed the register-based statistical system SIREE which enables the analysis of linked administrative data on children and adolescents. In Peru small-area estimation techniques have been applied to create a monetary poverty map, incorporating administrative data to better target child poverty interventions.

Chapter 3 provides a practical roadmap for incorporating administrative data into statistical systems related to children and adolescents. It outlines eight key steps: (1) defining statistical needs and building the statistical business case; (2) assessing feasibility and readiness; (3)

evaluating legal and ethical alignment with statistical mandates and data protection frameworks; (4) ingesting, storing, and managing administrative data; (5) conducting linkage and preparing the linkage-execution report; (6) developing and executing the analytical strategy; (7) developing and executing the dissemination strategy; and (8) institutionalizing, promoting sustainability and scaling (see Figure 1). The chapter includes guiding questions, resources, and tools such as *A Brief Introduction to Linkage Methodologies and Software* (Appendix 1) and *A Methodological Protocol for Data Linkage* (Appendix 2). It also stresses the importance of quality assessment, sound linkage methodologies, and privacy-preserving practices.

In conclusion, this manual is a comprehensive resource for NSOs, offering practical guidance to effectively integrate administrative data into statistics focused on children and adolescents. While this manual was developed with NSOs in the Latin America and Caribbean region, its relevance is global. By emphasizing data linkage, cross-sectoral integration, and adherence to sound statistical frameworks, it strengthens the capacity of NSOs to generate high-quality, timely, and policy-relevant statistics. The inclusion of case studies and best practices demonstrates the transformative potential of administrative data in advancing child rights and monitoring progress toward the SDGs. Strengthening governance, safeguarding data privacy, and fostering inter-institutional collaboration are critical to maximizing the utility of administrative data. Ultimately, the manual supports the development of evidence-based policies that improve the well-being of children and adolescents across the region.

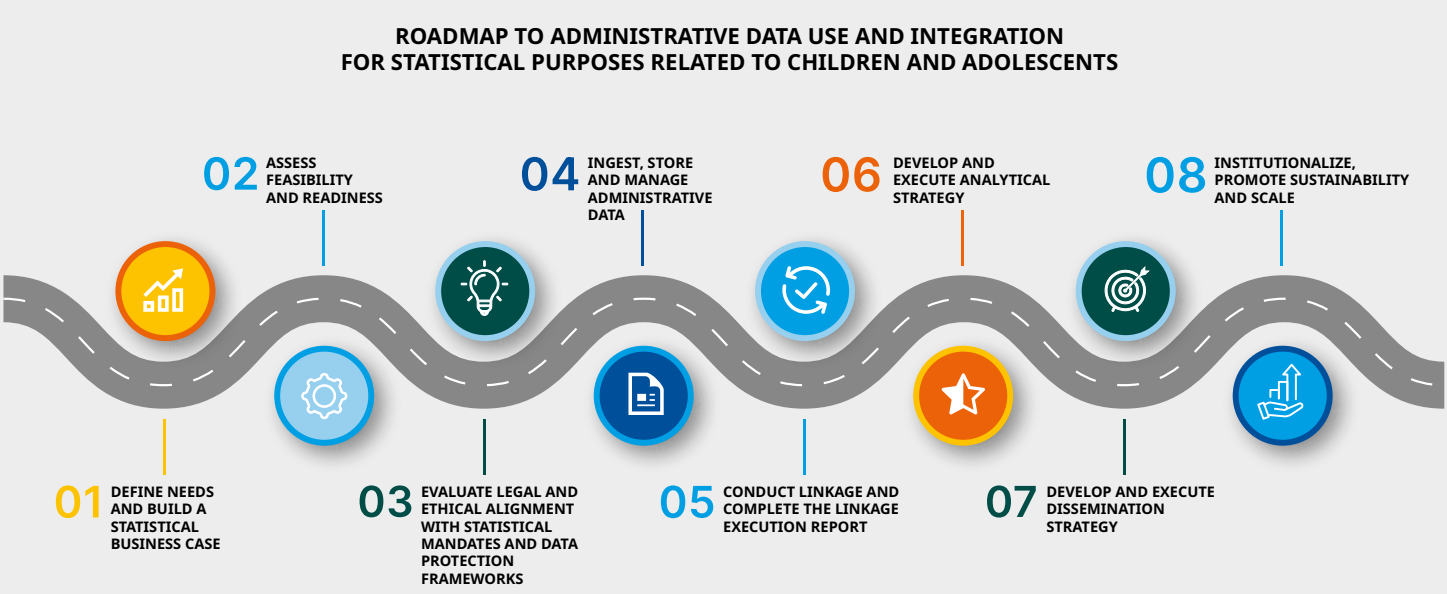


Figure 1. Roadmap to Administrative Data Use and Integration for Statistical Purposes Related to Children and Adolescents

Acknowledgements

This manual was conceptualized and developed by the UNICEF Regional Office for Latin America and the Caribbean (Liliana Carvajal Vélez and Jenneke Le Moullec) with leadership and technical guidance from the National Institute of Statistics of Chile (Aylin S. Flores Hinojosa and Julibeth Josefina Rodríguez León) and the National Institute of Statistics of Uruguay (Valentina Santo and Federico Segui), in their capacity as co-chairs of the Working Group on Children's Statistics of the Statistical Conference of the Americas of the Economic Commission for Latin America and the Caribbean.

The working group includes national institutions from member countries: the National Institute of Statistics and Census of Argentina; the Barbados Statistical Service; the Statistical Institute of Belize; the National Statistical Institute of Bolivia; the Brazilian Institute of Geography and Statistics; the National Institute of Statistics and Census of Costa Rica; the National Office of Statistics and Information of Cuba; the National Statistics Office of the Dominican Republic; the National Institute of Statistics of Honduras; the National Institute of Statistics and Geography of Mexico; the Statistics Department of Montserrat; the National Institute of Statistics and Censuses of Nicaragua; the National Institute of Statistics and Census of Panama; the National Institute of Statistics of Paraguay; the National Institute of Statistics and Informatics of Peru; the Department of Statistics of St. Kitts and Nevis; the General Bureau of Statistics of Suriname; the Central Statistical Office of Trinidad and Tobago; and the National Statistical Institute of Venezuela.

We gratefully acknowledge the contributions of all working group members to the discussions, conceptualization, and refinement of this manual.

We acknowledge the valuable support of the following experts for contributing with valuable insights:

- José Eduardo de Oliveira Trindade, Luiz Fernando Lima Costa, Ailton José Lima Martins Furtado, Paulo Cesar Dick, Brazilian Institute of Geography and Statistics
- Antônio Tadeu Ribeiro Oliveira, Brazil Ministry of Health
- Maria Fatima Ludovico de Almeida, Pontifical Catholic University of Rio de Janeiro, Brazil
- Olga Barquero Alpïzar, National Statistics Institute, Chile
- Norkis Plasencia Padrón, National Office of Statistics and Information, Cuba
- Eric Manuel Rodríguez Herrera, the National Institute of Statistics and Geography, Mexico
- Elvis Manayay Guillermo and Ana Naupari Rivas, National Institute of Statistics and Informatics, Peru
- Dinoira Moren, National Statistical Institute, Venezuela
- Xavier Mancero, Economic Commission for Latin America and the Caribbean (ECLAC)
- Claudia Cappa, Sebastian Palma, Eva Quintana, Friederike Schueuer, Enrique de la Monica, UNICEF Headquarters

We recognize the efforts from across national statistical systems in the case studies included in this manual:

- Case Study I on Migration Statistics: Department of Studies of the National Migration Service of the Government of Chile
- Case Study II on Integration of Vital Statistics Systems and Estimates of Under-reporting of Births and Deaths, with an Emphasis on Child Statistics in Brazil: Brazil Ministry of Health, Pontifical Catholic University of Rio de Janeiro Brazil, as well as the Research Department and the Coordination of Population and Social Indicators areas of (COPIS) of the Brazilian Institute of Geography and Statistics (IBGE)
- Case Study III on Integrated System of Statistical Records and Surveys: National Institute of Statistics of Uruguay
- Case Study IV on Provincial and District Monetary Wealth Map in Peru: National Institute of Statistics and Informatics, Peru

Introduction

In recent years, the importance of administrative data in producing statistics on children and adolescents has gained increasing recognition. As governments strive to design evidence-based policies that safeguard the rights and well-being of children and adolescents, administrative data has emerged as a vital resource. The ability of National Statistical Offices (NSOs) to harness and integrate these data across sectors is critical for informing public policies, monitoring progress on international goals such as the 2030 Agenda for Sustainable Development, upholding the Convention on the Rights of the Child (CRC)[1] (OHCHR, 1989), and ensuring that no child is left behind in the pursuit of equitable development.

As defined in UNICEF's Administrative Data Maturity Model (ADaMM), "administrative data is collected through the routine provision of a service such as health, education, or social welfare—most often by a government provider, but potentially also through private sector services" (UNICEF, 2022a). These structured data, in digital or physical records, are typically generated on an ongoing basis for operational, legal, or regulatory purposes rather than specifically for statistical use. Employing such data for statistical purposes offers several advantages, including cost-effectiveness, timely availability, and the potential for cross-sectoral and longitudinal analysis that can track the experiences of children and adolescents over time. However, integrating and using administrative data for statistical purposes also presents significant challenges, particularly with respect to data quality, privacy protection, and gaps in coverage and consistency.

The purpose of this manual is to provide concrete guidance on the statistical use of administrative data concerning children and adolescents, with a specific focus on linking and integrating data. Drawing on case studies from the region—including Chile, Brazil, Uruguay, and Peru—this manual highlights best practices and successful approaches to data linkage and integration. It showcases the potential of administrative data to support evidence-informed decision-making on issues related to children and adolescents, the Sustainable Development Agenda, and, ultimately, the realization of child rights. These cases offer practical insights into how countries can address and overcome challenges related to data quality, interoperability and system compatibility, as well as coverage and access[2].

1 The United Nations Convention on the Rights of the Child (UN CRC), a treaty that recognizes specific rights for children and provides international standards and benchmarks for the fulfillment of child rights (OHCHR, 1989), is the most widely ratified human rights treaty in history and has helped transform the lives of children around the world.

2 As part of the Children's Statistics Working Group Biennium 2024/25, member NSOs from across the Latin American and Caribbean region were invited to submit case study proposals for inclusion in this *Manual on the Use of Administrative Data and Its Integration for Statistical Purposes Related to Children and Adolescents.* A maximum of four case studies were selected based on (1) case study definition and alignment, (2) technical feasibility, and (3) project management and key actors. As a result, case studies from Brazil, Chile, Peru, and Uruguay are included in Chapter 2 of the manual, showcasing best practices in using administrative data for statistical purposes related to children and adolescents.

This manual also serves as a tool to advance the implementation of the recommendations from the *Administrative Data Readiness Diagnostic,* conducted by the Children's Statistics Working Group during the 2022/23 biennium (ECLAC & UNICEF, 2025) (Sierra Castillo et al., 2025). The diagnostic identified four key areas that pose challenges to the effective use of administrative data—namely, data demand, governance, data quality, and other enabling conditions—which are comprehensively addressed throughout this document.

A central theme of the manual is the recognition that the production of statistics on children and adolescents differs in important ways from the production of other types of statistics. While administrative data serve the broader statistical infrastructure of population statistics, they must also meet specific analytical requirements to capture the realities of children's and adolescents' lives and experiences. Moreover, all data concerning children and adolescents must be handled with particular care to ensure privacy and confidentiality, given the vulnerability of this population. Administrative data bring additional considerations in this regard.

In summary, this manual aims to support NSOs in developing coherent strategies for using administrative data to produce high-quality, timely, and relevant statistics on children and adolescents. By providing clear guidelines though an 8-step roadmap and practical examples, it seeks to enhance the capacity of NSOs to use and integrate administrative data for statistical purposes, thereby contributing to the evidence-based development of policies that promote the rights and well-being of children and adolescents across the region.

© UNICEF/UNI462690/Castro-Mendívil

Chapter 1:
Framing Administrative Data for Child-Focused Statistics

The production and use of administrative data for statistics on children and adolescents are crucial for informing evidence-based policy decisions. These data are also essential for monitoring child rights and tracking progress across sectors. Administrative data can provide timely, reliable, and disaggregated information on children and adolescents, complementing data gathered through surveys, which remain an important source of information. Administrative data are collected through systems such as birth registration, education, health, and social protection. These sources can be used to monitor progress on both national and international priorities. Most notably, they support tracking progress on the 2030 Agenda for Sustainable Development and the United Nations Convention on the Rights of the Child.

This chapter lays the groundwork for understanding how administrative data can strengthen statistical systems focused on children and adolescents. It introduces key concepts, highlights critical frameworks and models, and presents a child-rights perspective, providing a conceptual entry point to the applied case study discussion in Chapter 2 and to the technical guidance on administrative data–based statistics on children and adolescents in Chapter 3.

The Value of Administrative Data for Child- and Adolescent-Focused Statistics

Administrative data may be underutilized in statistical programs because of concerns about data quality or access. However, when properly managed, administrative data can provide detailed, timely, and high-quality insights into children's and adolescents' experiences across life stages and domains. Their value increases further when combined with other data sources, enhancing the depth and reliability of analysis. Figure 2 below presents common statistical uses of administrative data within national statistical systems.

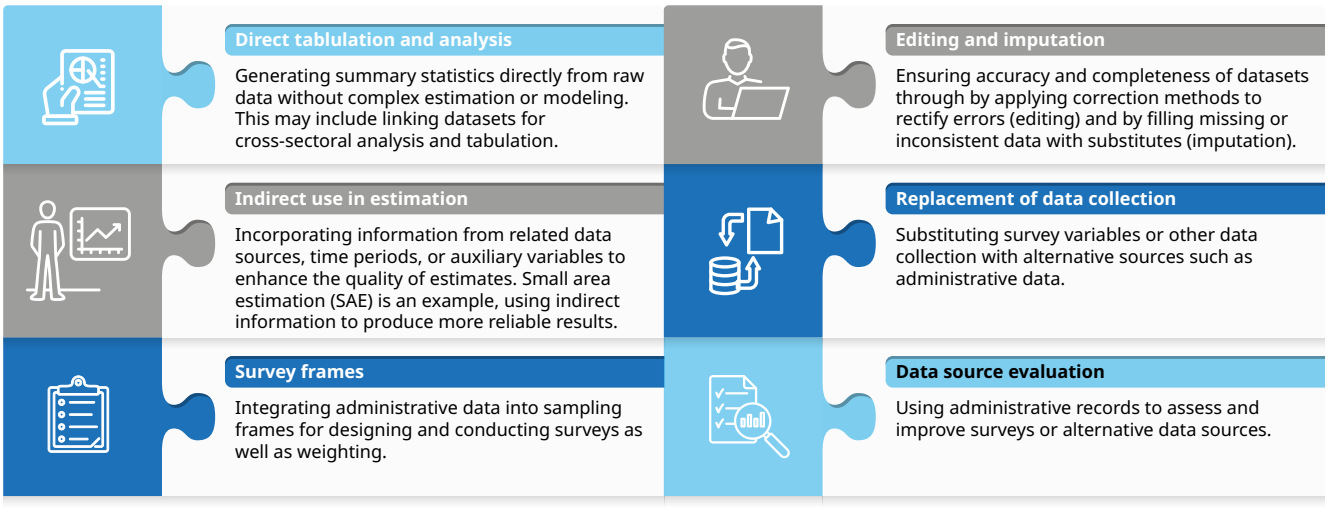


Figure 2. Statistical Uses of Administrative Data

In addition to these foundational contributions, administrative data can reveal meaningful insights about and for children and adolescents. The following examples illustrate how administrative data enable deeper analytical approaches and reveal insights that may be beyond the reach of conventional data sources:

Granular subpopulation analysis

With large coverage, administrative data enable robust disaggregation, facilitating research on small or marginalized groups that may be undersampled in surveys (e.g., Indigenous children and migrant children).

Gender analysis

Allowing disaggregation and comparison of outcomes by sex and gender across sectors, supporting evidence-based gender analysis. This enables the identification of gender-based disparities in education, health, protection, and access to social services, and helps track progress toward gender equality commitments.

Connecting context to outcomes

Integrated administrative data provide a multi-sectoral view—for example, combining education information with health services and child protection records—making it possible to explore how combinations of factors affect child and adolescent well-being.

Rare event analysis

When systems are robust, administrative data typically cover the entire population of interest (e.g., all births), allowing the identification and analysis of infrequent events. This enables thorough exploration of low-frequency but high-impact experiences relevant to children's and adolescents' well-being.

Real-world behavioural evidence

Administrative data reflect actual service use and system interactions rather than self-reported behavior, reducing recall and social-desirability bias (the tendency to give socially acceptable answers) and improving ecological validity (making the data more accurate and realistic).

Policy impact evaluation

Rich administrative datasets support quasi-experimental designs (e.g., difference-in-differences, regression discontinuity), allowing the linking of individual service histories to outcomes and enabling rigorous impact evaluations of child-related policies.

Time-sensitive policy monitoring

Because administrative data are collected on an ongoing basis, they can support near-real-time insights, such as monitoring the effects of crises (e.g., school closures during pandemics) on child and adolescent outcomes.

These analytical strengths are complemented by several practical benefits that enhance the value of administrative data use within national statistical systems:

- **Efficiency**
Once robust systems are established, the use of administrative data in statistical programs can be efficient and cost-effective, as data are routinely generated with the provision of services.

- **Consistency**
Standardized administrative systems can ensure consistent data collection over time.

- **Sustainability**
Standardized administrative systems provide reliable and comparable data collection over time, strengthening the evidence base for policymaking and monitoring.

- **Data quality**
In some contexts, administrative and survey data offer different strengths in terms of quality. For example, in health information, hospital or physician records may provide detailed clinical data, while surveys can capture patient-reported outcomes or experiences that administrative records might not reflect.

- **Reduced response burden**
Integrating administrative data into statistical production can streamline data collection by reducing the number of questions needed in surveys. This approach lessens the response burden on individuals and improves efficiency for field enumerators, while surveys continue to capture information not available through administrative sources.

- **Timeliness**
Administrative data are sometimes updated at high frequency. For instance, administrative data may be collected on an ongoing or annual basis, while a national population census may occur only every five or ten years.

Administrative Data as Complementary to Survey and Census Data

The UN Inter-Secretariat Working Group on Household Surveys (ISWGHS) identifies "enhancing the interoperability and integration of household surveys" as a technical priority area for household surveys in the next decade (Carletto et al, 2022). In this perspective administrative data is considered complementary to survey and Census data, increasing it's the cost-effectiveness, accuracy, granularity and overall value.

While administrative data offers valuable analytical potential, it is not universally suited for all purposes. Limitations in the concepts covered (e.g., opinions, attitudes, and behaviors), inconsistent definitions and standards, service-user bias (i.e., inclusion limited to individuals who interact with the administrative system), other biases, and data quality concerns (e.g., data entry errors, missing data) all indicate that, although administrative data will increasingly enhance and modernize statistical systems, surveys—including child- focused surveys—will remain essential for capturing information that administrative sources cannot provide.

Moreover, administrative data can support and enhance a population census in multiple ways: contributing to pre-census planning and frame development, replacing individual census questions, supporting post-enumeration processing and evaluation, and serving as a foundation for alternative or hybrid census models.

Integrating a Children's Rights Perspective into Administrative Data–Based Statistics

Integrating a children's and adolescents' rights perspective into administrative data–based statistics ensures that data not only monitor progress but also actively uphold children's and adolescents' dignity, inclusion, and protection. This approach is guided by UNICEF's *Conceptual Framework and Guidance for Integrating a Children's Rights Perspective into Administrative Data–Based Statistics,* which is rooted in the United Nations Convention on the Rights of the Child (CRC) (UNICEF, 2025a). The Framework defines four key dimensions: **(1)** children interact with administrative data systems differently than adults; **(2)** data governance must be fit for children; **(3)** statistical output activities must focus on the areas of greatest impact for children; and **(4)** children's data carry unique statistical challenges.

It also provides six core recommendations and eight guiding considerations to help administrative data producers and users support evidence-informed decisions for and about children, ultimately reinforcing child rights.

GUIDANCE AND RECOMMENDATIONS

For child rights, one key aim for statistical systems should be to support evidence-informed decisions for and about children. This includes:

- Establishing and maintaining high quality statistical infrastructure, methodologies, concepts and standards on child-related statistics.
- Delivering high-quality statistical outputs about children and their experiences.
- Considering the additional layers of care that are necessary when handling data for and about children.

Children interact with administrative data systems differently than adults

1. Multisectoral review of administrative data systems should be conducted for child specific statistics.

- Data should be collected to allow analysis by key sociodemographic characteristics for disaggregation.
- NSOs should be granted authority to securely share and link children's data for statistical use.
- NSOs should prioritize developing population registers that account for the unique aspects of children's data.

Data governance must be fit for children

2. NSOs should strengthen data governance for children across the administrative data value chain, from collection to dissemination of statistical outputs.

3. Data governance programmes should support the ethical and responsible use of data, and protect against harm and misuse.

- Alternatives to traditional consent and assent models, including implied consent and social licence, should considered.
- Frameworks such as necessity and proportionality should be considered.

Statistical output activities must focus on areas of greatest impact for children

4. Gaps in international standards for children's data should be addressed.

5. Focus efforts and resources on capacity building as required.

- Longitudinal and intersectoral analysis of children's data should be enabled.

Children's data carry unique statistical challenges

6. Consideration should be given to understanding data capture in children's data sources for statistical analysis.

- Children's representation in administrative data should be considered when developing statistical outputs.
- The unique challenges of children's data should be carefully considered when developing statistical outputs

Data Linkage and Integration: An Emphasis on Cross-Sectoral Statistics for and about Children and Adolescents

Although administrative data are typically sector-specific, addressing the rights and well-being of children and adolescents requires a multi-sectoral approach. Many issues affecting this population are systemic and involve multiple sectors. Understanding how these sectors interact and influence outcomes allows for more efficient and effective allocation of resources. Therefore, data linkage becomes a critical tool in this context. By combining administrative data from various sectors—such as education, health, social services, and justice—at the individual level, a comprehensive understanding of the challenges faced by children and adolescents can be achieved.

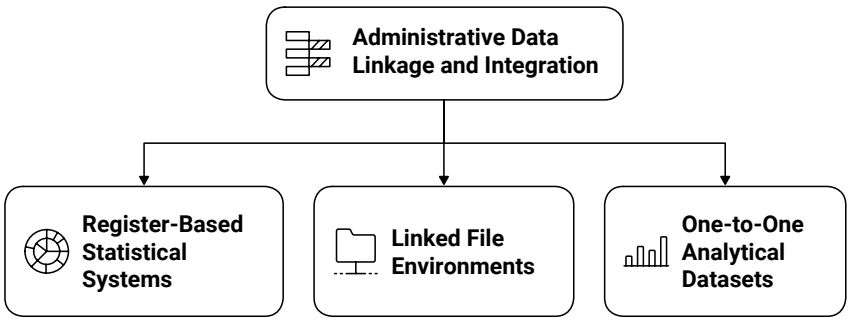
Under the United Nations Statistical Commission's Fundamental Principles of Official Statistics, NSOs are legally mandated to act impartially, uphold strict scientific standards, and resist external pressures that could lead to data misuse (UN, 2014). This positions NSOs as key actors in acquiring, managing, and linking administrative data for and about children and adolescents across sectors.

Several models exist for the linkage and integration of administrative data for use in population statistics, each offering distinct advantages and applications. These models can be broadly organized into three categories—register-based statistical systems, linked file environments, and one-to-one analytical datasets—which are described below.

Data linkage and data integration are distinct but related activities

Data linkage refers to the process of connecting or matching records about the same entity (often individuals) from different datasets based on common identifiers. Datasets linked at the entity level may then be integrated, which focuses combining data from different sources into a unified dataset. Sometimes data linkage is referred to as record linkage. *Data integration* which can extend beyond linked data, broadly refers to creating a holistic view of a dataset through merging, transforming, cleaning, and aligning data. An example of data integration is MICS Link, described later in this manual.

This *Manual on the Use of Administrative Data and its Integration for Statistical Purposes Related to Children and Adolescents* emphasizes data linkage and associated integration. It does not aim to provide an exhaustive treatment of data integration.



1. Register-Based Statistical Systems

The concept of a register-based statistical system, originating in the Nordic countries during the mid-20th century, has been articulated and formalized by Wallgren and Wallgren in *Register-Based Statistics: Registers and the National Statistical System* (Wallgren & Wallgren, 2022). Broadly defined, a register-based statistical system is an integrated network of administrative data sources, linked at the microdata level and aligned through consistent definitions and harmonized variables.

According to the Inter-American Development Bank report *Toward an Integrated Statistical System Based on Registers* by Wallgren and Wallgren, several Latin American and Caribbean countries have advanced in the development of register-based statistical systems, including Brazil, Mexico, Colombia, Costa Rica, Chile, Ecuador, and Uruguay (IADB, Wallgren & Wallgren, 2021). In the case of Uruguay, the System of Integrated Statistical Registers and Surveys (SIREE) is a flagship example of a register-based statistical system and is described in detail in Chapter 2 of this manual.

Statistical Population Registers and Register-Based Statistical Systems

Statistical population registers have served as an effective statistical data source for decades and have become central to the operations of several NSOs internationally. The term *population register* was defined in 1969 by the United Nations as "an individualized data system, that is, a mechanism of continuous recording, and/or of coordinated linkage, of selected information pertaining to each member of the resident population of a country in such a way as to provide the possibility of determining up-to-date information concerning the size and characteristics of that population at selected time intervals" (UN, 1969).

A population register, designed primarily for administrative use, enables up-to-date information about the population's size and characteristics at any given time and functions as part of an ongoing process in which notifications of specific events (such as births, deaths, or migrations)—often recorded in separate administrative systems—are linked to it in near real time. To maintain its accuracy and functionality, the population register must be legally regulated and carefully managed.

Statistical population registers differ from population registers because they are designed exclusively for statistical purposes by NSOs. A statistical population register may be derived from, or adapted from, population registers or other administrative data sources, and is processed and cleaned to meet statistical standards. Typical data sources for a statistical population register include birth, death, civil registration, and migration information.

The statistical population register can be used to generate statistics on population size, characteristics (e.g., age, sex/gender, and citizenship), and trends over time. Its focus is on ensuring that data are fit for statistical production, often through modifications that guarantee consistency, accuracy, and adherence to statistical definitions. Children are typically included in a statistical population register, as these registers aim to capture comprehensive demographic information about all residents of a country, regardless of age. As such, they include individuals from all age groups, including infants, children, and adolescents.

Within a broader register-based statistical system, the population register is considered the most important statistical register among the four key types of statistical registers: population registers, activity registers, real estate registers, and business registers. Therefore, classification as a register-based system often depends on the existence of a statistical population register.

2. Linked Analytical File Environments

A linked analytical file environment is a statistical framework that facilitates the linkage and integration of data from multiple sources into a unified analytical setting. This environment enables comprehensive statistical analysis by linking datasets across different sectors or domains, sometimes using an intermediary reference such as a statistical population register, though this is not always required.

An example of a linked analytical file environment is New Zealand's Integrated Data Infrastructure (IDI), a collection of linkable administrative data sources, census information, and surveys for analytical purposes (Milne et al., 2019).

3. One-to-one Linked Analytical Datasets

One-to-one linked analytical datasets involve identifying the same individuals across two different datasets and merging these records for analysis. Typically developed for a particular analytical purpose, these datasets do not provide the same efficiencies as register-based statistics or linked file environments when it comes to repurposing data linkages for different analytical applications. An example of a one-to-one linked analytical dataset is linking survey responses to respondents' grade information.

MICS Link: Linking and Integrating Survey and Administrative Data for Child and Adolescent Well-Being

The Multiple Indicator Cluster Surveys (MICS) is a globally recognized household survey program developed by UNICEF to collect high-quality, internationally comparable data on the well-being of children, adolescents, and women. Established in 1995, MICS has been a key resource for data on health, education, nutrition, water and sanitation, protection, and other critical indicators (Hancioglu & Arnold, 2013).

MICS Link is an innovative initiative within the Global MICS program, designed to enhance the integration of survey data with other sources, such as administrative records (UNICEF, 2025a). By linking MICS data with administrative records—including health, education, social protection, and other sectors—MICS Link seeks to:

- Enrich MICS data with auxiliary information for better contextualization.
- Alleviate respondent burden, mitigate potential social-desirability bias, and minimize recall bias.
- Promote the use of household survey data to inform sector planning and monitoring.
- Contribute to improved quality and efficiency of investments in national data ecosystems.

Models Guiding Administrative Data Use

A model provides a structured approach to using administrative data, offering guidance on acquiring, handling, processing, validating, and ultimately producing and disseminating outputs. By standardizing procedures and establishing best practices, models help NSOs improve administrative data–based statistics.

Because administrative data are often gathered from various public-sector agencies and can be highly fragmented, integrating these data requires adherence to established models to ensure high-quality statistics. The following two models provide structured approaches to address these challenges:

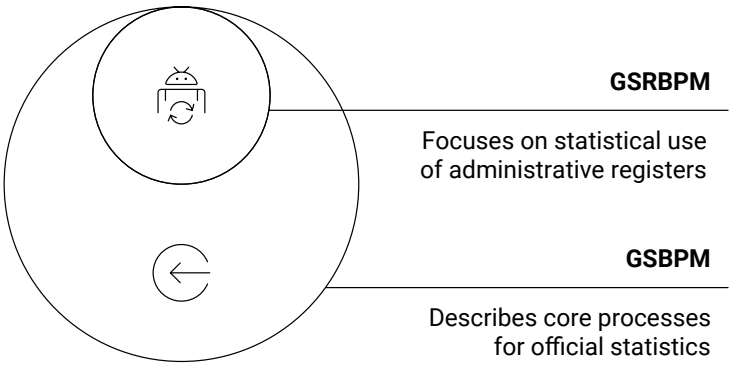
- ### Generic Statistical Business Process Model (GSBPM)

  Developed by the United Nations Economic Commission for Europe (UNECE), the GSBPM describes and defines the set of business processes needed to produce official statistics. The GSBPM should be applied and interpreted flexibly and is best viewed as a matrix through which multiple possible process paths can be followed (UNECE, 2014)

- ### Generic Statistical Register Business Process Model (GSRBPM)

  Building on the GSBPM, this model focuses on the use of administrative registers. It outlines processes for efficiently managing and linking register-based data, making it especially relevant for countries with well-developed administrative data systems (IADB & Segui Stagno, 2016). This model has been applied, for example, in Uruguay's System of Integrated Statistical Registers and Surveys (SIREE), which will be discussed further in Chapter 2.

Statistical Process Models



© UNICEF/UNI568140/Terán

Chapter 2:

Administrative Data for Children and Adolescents Related to Sustainable Development Goals

Chapter 2 focuses on the critical role of administrative data in measuring and improving progress toward reporting on the Sustainable Development Goals (SDG) indicators related to children and adolescents. In the context of global SDG commitments, achieving targets requires timely, accurate, and granular data that reflect the realities of these populations. Administrative data provide a significant opportunity to fill gaps in SDG reporting by offering detailed and timely information, thereby enhancing understanding of how various factors affect children and adolescents

A Note on Administrative Data and Gender Mainstreaming

The realities, circumstances, barriers, and needs of children and adolescents differ significantly based on gender, shaped by the social, cultural, and psychological constructs—as well as roles and expectations—associated with their sex assigned at birth within specific contexts. Administrative data offer valuable opportunities for producing statistics on gender-related issues. For example, administrative records often capture critical information on topics such as childhood pregnancy, early marriage, and gender-based violence, enabling a deeper understanding of these issues and informing targeted interventions.

However, gender information in administrative data requires careful consideration to ensure that it is accurate, ethical, and inclusive. Some administrative data systems rely on binary classifications or assumptions based on sex assigned at birth, which may fail to represent the full spectrum of gender identities. This can lead to misclassification, exclusion, or biases in analysis and reporting, potentially reinforcing stereotypes or discriminatory practices. Misinterpreting such data can exacerbate inequalities and hinder the development of inclusive policies.

Administrative Data as an Opportunity for Monitoring the Sustainable Development Goals

The *2030 Agenda for Sustainable Development,* with its premise of "leaving no one behind," includes seventeen goals, twelve of which relate directly to the rights and well-being of children and adolescents. These are: poverty, hunger, health & well-being, education, gender equality, clean water & sanitation, affordable & clean energy, decent work & economic growth, reduced inequalities, climate action, peace, justice, & strong institutions, and partnerships for the goals (UNICEF, 2018).

Within these twelve goals, there are nineteen indicators for which UNICEF serves as custodian or co-custodian. Figure 3 below outlines the twelve goals most directly concerning children and adolescents. Details on each of the twelve goals and the forty-four indicators directly related to children and adolescents can be found at https://data.unicef.org/sdgs/.

Figure 3. Child-focused Sustainable Development Goals

According to the UNICEF report Progress on Children's Well-Being: Centring Child Rights in the 2030 Agenda, which reviews progress at the midpoint of the 2030 SDG Agenda, millions of children and adolescents continue to face poverty, inequality, and social exclusion despite advancements in several areas of development (UNICEF, 2023).

Data gaps remain a complex and often under-recognized challenge affecting the rights and well-being of children and adolescents. These gaps contribute to policy blind spots, limit progress monitoring, and weaken advocacy opportunities. NSOs may face difficulties in addressing data gaps both nationally and within the global SDG indicator framework. Strengthening administrative data–based statistical systems can help address these challenges and improve SDG monitoring. Throughout Chapter 2, concrete examples from Latin American and Caribbean countries—Chile, Brazil, Uruguay, and Peru—are presented as case studies. These examples demonstrate the opportunities countries have seized in transforming administrative data into statistical information to monitor the SDGs and offer practical lessons for all countries in the region. The case studies highlight diverse contexts, illustrating how different countries have approached common issues such as integrating fragmented data sources, improving data quality, and ensuring comprehensive coverage of child and adolescent populations.

By applying a well-defined case study proposal selection process—developed by the Technical Secretariat and Co-Chairs of the Children's Statistics Working Group—this chapter ensures the inclusion of relevant and instructive examples. These real-world cases provide valuable strategies that other countries in the region may wish to explore.

Case Study I:
Migration Statistics in Chile with a Focus on Children and Adolescents

The National Institute of Statistics of Chile (INE Chile) has developed the capacity to work collaboratively with state institutions that produce administrative records suitable for use in social statistics. This capacity is reflected in a collaboration with the National Migration Service (SERMIG) to characterize the foreign population in intercensal periods since 2018. As part of this work, a joint initiative was developed with UNICEF to produce a special publication on migration statistics in Chile with a focus on children and adolescents (INE & UNICEF, 2023).

This initiative contributed to a deeper analysis of children and adolescents and provided valuable feedback on the overall process of estimating the foreign population using administrative records and data linkage methodologies.

Historically, in Chile, the Census of Population and Housing has been the primary source of information for measuring international migration. In the 2017 Census of Population and Housing, substantial increases in immigration were observed, with the migrant population reaching 4.4% of the total resident population—an all-time high not seen since 1907 (IDAC, 2023).

Following the 2017 census, there was a significant rise in residence permit applications, indicating that the census data had become outdated for estimating the total immigrant population. This emphasized the importance of administrative data in providing timely, high-quality statistical information on migratory flows and in updating census-based estimates (IDAC, 2023).

Children on the Move

Migration in search of better opportunities has been a constant feature of human history. While migration can have positive effects on children, their families, and both origin and destination communities, forced displacement has profoundly negative consequences, particularly for children.

The recent surge in people fleeing their communities due to conflict, violence, or poverty has elevated migration and displacement to a critical global issue, including in Latin America and the Caribbean, and this growing challenge is reflected in the Sustainable Development Goals (SDGs) (IDAC, 2023).

Unfortunately, significant gaps persist in the availability of data on migrant and displaced children and adolescents. According to the International Data Alliance for Children on the Move (IDAC), 30% of countries or territories do not produce recent age-disaggregated data on migrants. Additionally, data on Unaccompanied and Separated Children (UASC) who are refugees or asylum-seekers, collected by UNHCR, were limited at the end of 2024 to 86 out of 181 countries where UNHCR reports statistics (UNHCR, 2025).

Data on migration flows are even more limited. Many countries do not collect migration flow data at all, and age-disaggregated data on migration flows or on population movements of other mobile groups—such as internal migrants, refugees, and internally displaced persons—are even rarer (UN, 2020).

As shown in Figure 4, the migration statistics update model is based on two data components: the 2017 Census of Population and Housing and administrative records, including both regular and irregular migratory status. This general methodology was first developed for the reference year 2018 and has been improved annually. For example, since 2022, both regular and irregular migratory status (only for individuals recorded by administrative data from different sources) have been included.

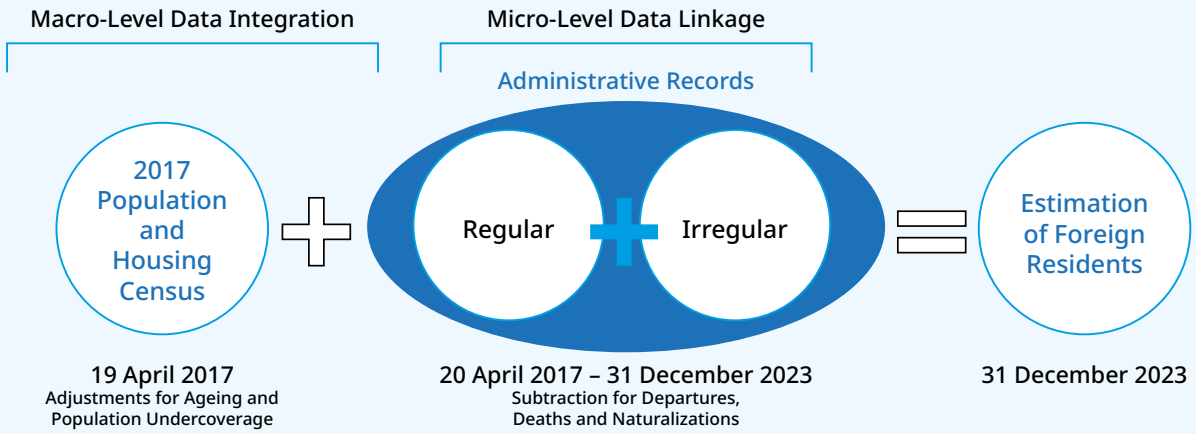


Figure 4. Components in the Methodology for Estimating Foreigners Residing in Chile (2023)[3]

The most recent publication estimating foreign residents in Chile (2023) was recently released. For the first time, it included information from the Ministry of Education on foreign students enrolled with provisional identification numbers.[4] This addition mainly affected the estimates of children over five years of age, making visible individuals who had not been accounted for in previous estimates (INE & SERMIG, 2024).

The administrative data components required strong partnerships and coordination between data-providing institutions and INE Chile, established through a collaboration agreement with the National Migration Service and other institutions. These include the National Migration Service of the Ministry of the Interior, the Ministry of Foreign Affairs, the Investigations Police, the Civil Identification and Registration Service, and the Ministry of Education. Data-sharing agreements between institutions were developed within the framework of the Chilean State's interoperability strategy (See Figure 5).

In addition, INE Chile and the National Migration Service have a bilateral cooperation agreement for the creation of the Migratory Statistics Table (MEM)[5] and for defining methodological improvements.

3 Based on Instituto Nacional de Estadísticas (INE) (Chile) & Servicio Nacional de Migraciones (SERMIG). (2024). *Documento Metodológico de la Estimación de Personas Extranjeras Residentes en Chile al 31 de Diciembre de 2023.* Retrieved from https://www.ine.gob.cl/docs/default-source/demografia-y-

4 In Chile, students without regular migratory status are enrolled in educational establishments with a unique identification number called IPE (Identificador Provisorio Escolar). Foreign students with regular migratory status are enrolled with a unique identification number called RUN (Rol Único Nacional).

5 This is an inter-institutional working group on international migration statistics that provides a framework for permanent collaboration among state institutions for the creation, processing, and analysis of international migration statistics.

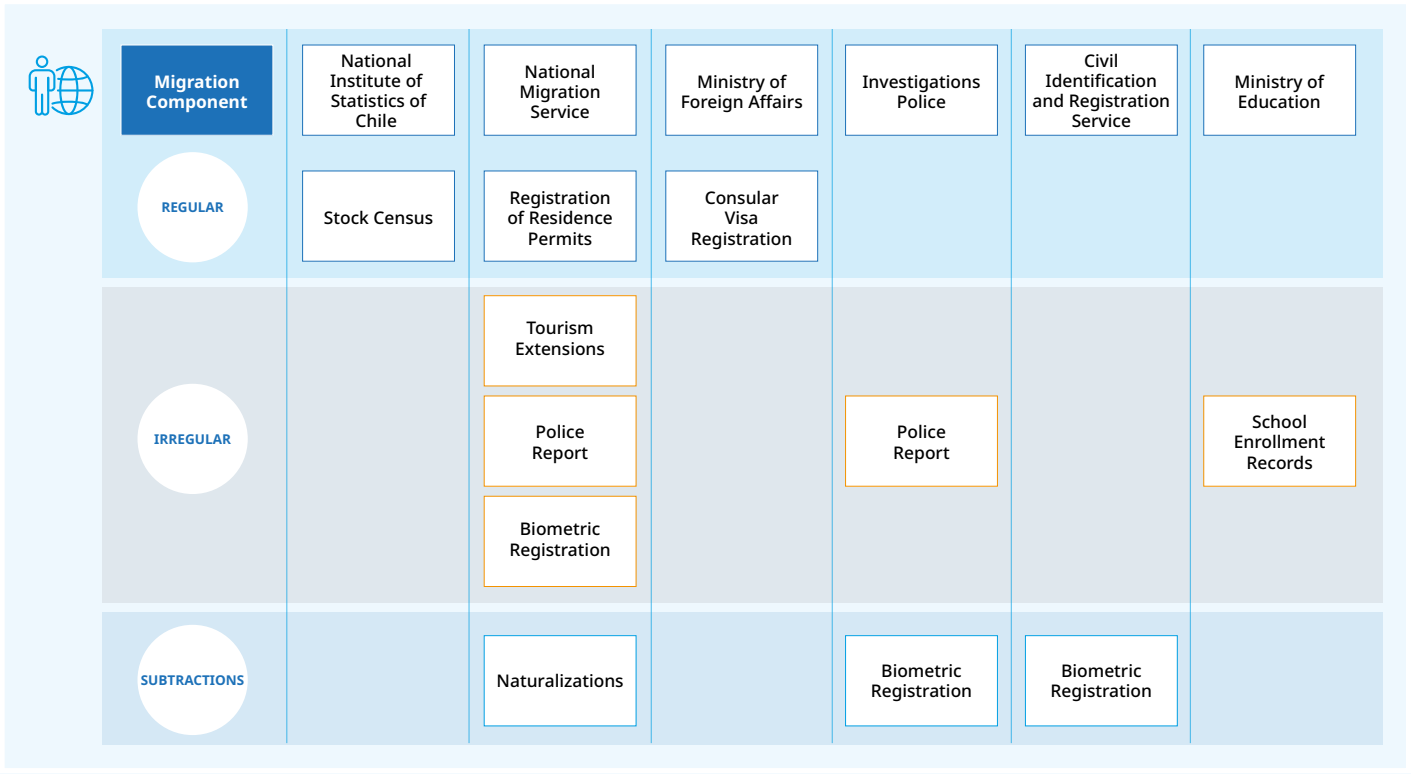


Figure 5. Data Components for Estimating Foreigners Residing in Chile

In the methodological approach for constructing the updated 2023 estimates, data integration first occurs at the micro level, linking the administrative records component where identifiers are available. Next, this component is combined with a macro-level integration using census data, applying frequencies based on variables such as age, sex, country, and region of residence, as described below:

1) **Micro integration.** The administrative records included a combination of unique and personal identifiers that enabled data linkage. Where possible, deterministic data linkage was conducted, combining records based on the national identity number. When the national identity number was not included in the administrative records, an algorithmic mixed-match approach was applied, combining multiple variables.[6]

This approach also has limitations. It is more robust for individuals with regular migratory status than for those with irregular status. For example, it is more difficult to capture all entries into the country (as only some individuals are recorded through administrative acts within the country) and to deregister individuals who leave the country through irregular crossings without those departures being captured in administrative data.[7]

2) **Macro integration.** Because the Census of Population and Housing in Chile does not include unique or personal identifiers, it could not be directly linked with administrative data sources. Therefore, the census component estimate—considered the immigration stock—is adjusted annually for census omissions and population ageing to create a new baseline on which

6 There were at least 100 deterministic combinations between variables such as passport, ID, full name, date of birth (day, month, and year), country of birth, and country of nationality. More information is available in the annex at https://www.ine.gob.cl/docs/default-source/demografia-y-migracion/metodologias/migraci%C3%B3n-internacional/metodolog%C3%ADa-epe2023.pdf?sfvrsn=6b7d9de4_15

7 More information about other integration assumptions at this level is available in Table 4 of the methodological document (in Spanish) at https://www.ine.gob.cl/docs/default-source/demografia-y-migracion/metodologias/migraci%C3%B3n-internacional/metodolog%C3%ADa-epe2023.pdf?sfvrsn=6b7d9de4_15

the balance of entries and exits tracked through administrative records is applied.

It is acknowledged that this approach has limitations, as it assumes an equal probability of omission for all individuals in the population and does not update the census at the individual level.[8]

The methodological approach also required conceptual standardization between census data and administrative records, for example in the definition of residence. A common framework was adopted, defining residents as people who have been in the country for at least six months or who intend to reside for more than six months, in alignment with the updated United Nations recommendations for measuring international migration.

The construction of the final analytical dataset enabled significant analysis of the characterization of children and adolescents over time from 2018 to 2023. This was presented in a joint publication with UNICEF in 2023. The report included detailed findings on the disaggregation of age groups, sex, and country of nationality, along with the rate of children and adolescents per 100 adults at the national and subnational levels. The full report is available online[9] (INE & UNICEF, 2023). Subsequent publications referencing the 2022 and 2023 foreign resident estimates have included indicators on children and adolescents to maintain the focus on this population.

As a result of this work, it is estimated that in 2023 a total of 302,306 foreign residents were under 20 years of age (15.8%). The population under 20 has increased since 2018, reaching its highest levels in 2023. The largest concentration within this group was in the 10–14 age segment (36.0%).

Men represented the majority of the under-20 age group in 2022 and 2023, with 52.5% and 52.3%, respectively. Women outnumbered men only in 2020, with 50.4% of the total. It was also estimated that 70,195 irregular foreign residents were under 20 years of age, representing 20.8% of the total population in irregular migratory status.

Finally, it is important to note that 77.9% of the updated migrant children and adolescent count came from the administrative data component of the update, underscoring the central role of administrative data in these estimates.

There are several key takeaways and reflections from the INE Chile case study that can serve as valuable insights for other NSOs in the region in the context of child rights statistics:

1. Administrative data sources can be a useful tool for updating estimates during intercensal periods.

2. Collaboration with data providers is key to the success of linked administrative data projects. Partners are often motivated to collaborate when the issue is of shared importance, as was the case with the observed increase in immigration to Chile.

3. Children and adolescents may represent a subpopulation in larger census or population-level administrative data statistical operations. This creates opportunities to develop subprojects focusing specifically on children and adolescents.

4. Large coverage projects, such as linking census data to administrative records, enable a high degree of disaggregation, including at the subregional level. Such analysis can produce insights that are directly actionable for policy and program development.

5. When working with special population groups such as children and adolescents, there is a risk that they may not be fully represented in official statistics. Therefore, it is crucial to have a complete mapping of the institutions holding information on these subpopulations. By incorporating educational enrollment data for foreign students registered with provisional identification numbers, Chile has addressed the underrepresentation of children and adolescents who remain in an irregular migratory situation in official statistics.

8 Other integration assumptions at this level—such as aligning country-of-birth names (in the census) with country of nationality (in the registers)—are described in Table 5 of the methodological document (in Spanish) at https://www.ine.gob.cl/docs/default-source/demografia-y-migracion/metodologias/migraci%C3%B3n-internacional/metodolog%C3%ADa-epe2023.pdf?sfvrsn=6b7d9de4_15

9 See the full report at https://www.ine.gob.cl/docs/default-source/demografia-y-migracion/publicaciones-y-anuarios/migraci%C3%B3n-internacional/estimaci%C3%B3n-poblaci%C3%B3n-extranjera-en-chile-2018/informe-estimacion-ninez-y-adolescencia-migrante.pdf?sfvrsn=8cac1976_9

Case Study II:
Integration of Vital Statistics Systems and Estimates of Under-reporting of Births and Deaths, with an Emphasis on Child Statistics in Brazil

This case study examines the Brazilian experience in the integration of vital statistics systems. The production of reliable vital statistics constitutes a fundamental pillar for social development and the protection of human rights, particularly children's rights.

This case is especially relevant to the Sustainable Development Goals (SDGs), notably Target 3.2, which aims to end preventable deaths of newborns and children under five years of age, and Target 16.9, which calls for providing legal identity for all, including birth registration.

In Brazil, two main vital statistics systems coexist: the Brazilian Institute of Geography and Statistics (IBGE) system, based on Civil Registry Statistics, and the Ministry of Health system, comprising the Live Birth Information System and the Mortality Information System.

This case study focuses on the integration of these systems—a fundamental advance in overcoming the inherent limitations of each source when used in isolation—allowing for a more complete and precise view of vital events in the country.

Birth Registration, Legal Identity, and Mortality Measurement: Advancing SDG Target 3.2 (Reducing Child Mortality) and SDG Target 16.9 (Legal Identity for All)

Birth registration is a fundamental component of legal identity and is essential for realizing children's rights and ensuring access to services such as health care, education, and social protection.

Aligned with SDG Target 16.9—which calls for providing legal identity for all, including birth registration, by 2030—the integration of administrative data systems in Brazil highlights both progress and persistent challenges.

Legal identity also plays a critical protective role: it helps prevent exploitation, such as child labor and trafficking, and strengthens governance by providing accurate demographic data for policy planning and resource allocation. Strengthening registration systems is therefore essential for inclusive development, rights-based policymaking, and reliable national statistics.

Accurately measuring mortality is equally critical for tracking progress toward SDG Target 3.2, which seeks to end preventable deaths of newborns and children under five years of age. Reliable mortality data enable governments and health systems to identify at-risk populations, design effective interventions, and allocate resources efficiently to reduce child mortality.

Furthermore, legal identity and accurate birth and death registration are interconnected: without comprehensive mortality measurement, population data remain incomplete, undermining efforts to ensure every child's right to identity and health.

© UNICEF/UNI729109/Chaves

The complementarity between the IBGE and Ministry of Health systems is essential for strengthening vital statistics. While IBGE provides a legal and administrative perspective on vital events, the Ministry of Health offers detailed insights into the health context. However, differences in timing and methodology between these systems can lead to discrepancies and gaps—particularly under-registration and underreporting—issues that are especially critical for child statistics.

In 2017, IBGE began developing national, state, and municipal-level estimates. To enhance understanding and operational effectiveness, it became necessary to produce disaggregated model estimates for additional variables, initially designated as experimental statistics. These encompassed maternal demographics and birth characteristics for live births, as well as demographics, cause of death, and geographic information for deaths.

Through rigorous testing and collaborative research with academic partners and consultants, these experimental variables achieved official statistics status with the release of 2023 data.

The core challenge identified is the incompleteness of records, reflected in under-registration in the IBGE system and underreporting in the Ministry of Health system. These limitations have serious implications for child vital statistics, hindering efforts to monitor and uphold children's fundamental rights.

To address these issues, a three-step methodology was adopted for system integration:

1. Deterministic matching of the IBGE and Ministry of Health databases, using the Declaration of Live Births and Declaration of Deaths numbers as primary linking keys. When direct matches were not possible, alternative linking keys were created using additional variables common to both systems.

2. Capture-recapture method, a statistical technique that estimates the total number of events by analyzing the overlap between incomplete sources.

3. Generalized Linear Model, applied to adjust for biases arising from source dependence and heterogeneity in capture probabilities[10] (IBGE, 2025).

10 For more detailed information, see the technical notes available on the IBGE website: (in Portuguese) https://www.ibge.gov.br/estatisticas/sociais/populacao/26176-estimativa-do-sub-registro.html?=&t=notas-tecnicas or (in English) https://www.ibge.gov.br/en/statistics/social/population/26259-under-registration-of-births-and-deaths-estimates.html?lang=en-GB&t=notas-tecnicas

The analysis revealed a high level of agreement between the systems, with overlaps exceeding 98% for live births and 96% for deaths in 2023, as shown in Figure 6 Data Integration of Live Births and Deaths Across IBGE and Ministry of Health Registries - 2023.

The integration of vital statistics systems yielded significant improvements in data completeness. Live birth registration demonstrated substantial progress, with under-registration declining from 4.21% in 2015 to 1.05% in 2023, a reduction of approximately 75%. Under-notification showed an even more pronounced improvement, decreasing from 2.01% to 0.43% over the same period.

Mortality statistics, while improving, exhibited more modest gains: under-registration fell from 4.89% to 3.55%, and under-notification decreased from 2.32% to 1.09% (IBGE, 2025).[11]

These differential improvement rates suggest that birth registration systems may be more responsive to integration efforts than death registration processes. Nonetheless, significant regional disparities persist, particularly regarding access for residents in the North and Northeast regions. These areas—often characterized by vast geographical expanses, including the Amazon biome, and enduring structural challenges—present considerable difficulties in reaching populations for vital event registration.

This under-registration in hard-to-reach populations is particularly concerning because it directly impacts the ability to formulate equitable and effective public policies focused on historically underserved municipalities and areas, thereby hindering targeted resource allocation and intervention.

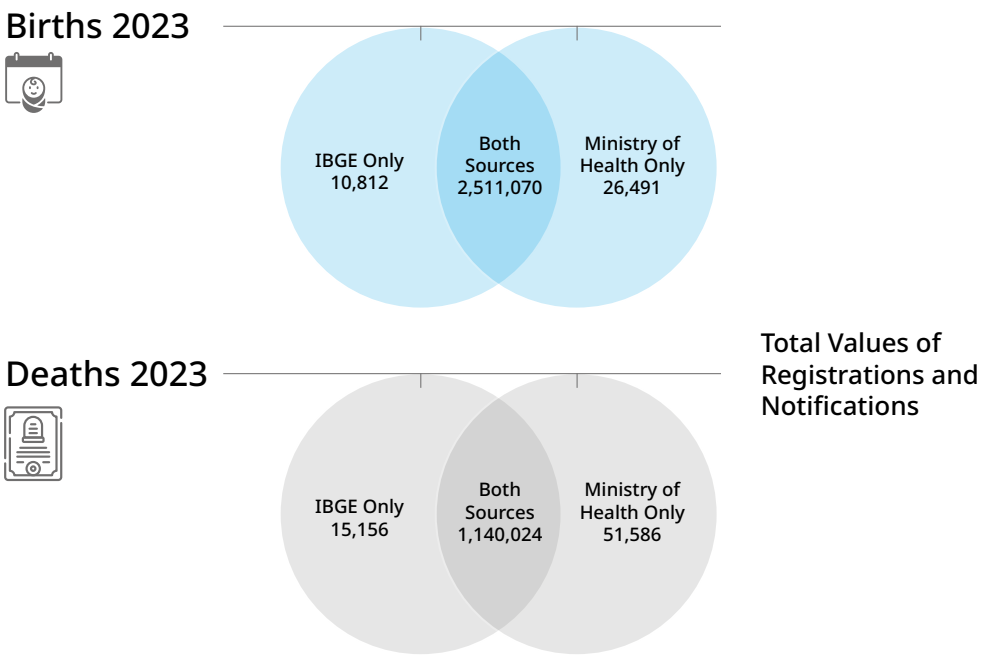


Figure 6. Data Integration of Live Births and Deaths Across IBGE and Ministry of Health Registries — 2023

11 For all results from the estimates of under-reporting and under-notification, please see the tables available at:
Portuguese: https://www.ibge.gov.br/estatisticas/sociais/populacao/26176-estimativa-do-sub-registro.html?=&t=resultados
English: https://www.ibge.gov.br/en/statistics/social/population/26259-under-registration-of-births-and-deaths-estimates.html?lang=en-GB

The cartographic analysis below depicts municipal-level under-registration patterns, differentiated by vital event type (live births in blue, deaths in green). The spatial distribution reveals pronounced regional clustering, with the North and Northeast regions demonstrating substantially higher under-registration rates than their South and Southeast counterparts (See figures 7 and 8).

This geographic inequality intensifies in peripheral municipalities characterized by lower human development indicators, highlighting the critical intersection between administrative capacity, infrastructural limitations (often exacerbated in remote areas like parts of the Amazon), and registration efficacy.

The value of precise vital statistics estimates becomes acutely apparent here, as they are fundamental for diagnosing the severity of these conditions and for informing targeted interventions. Addressing these deep-seated disparities is paramount to ensuring that statistical data truly reflect the realities of all citizens, thereby enabling more precise and just public policy.

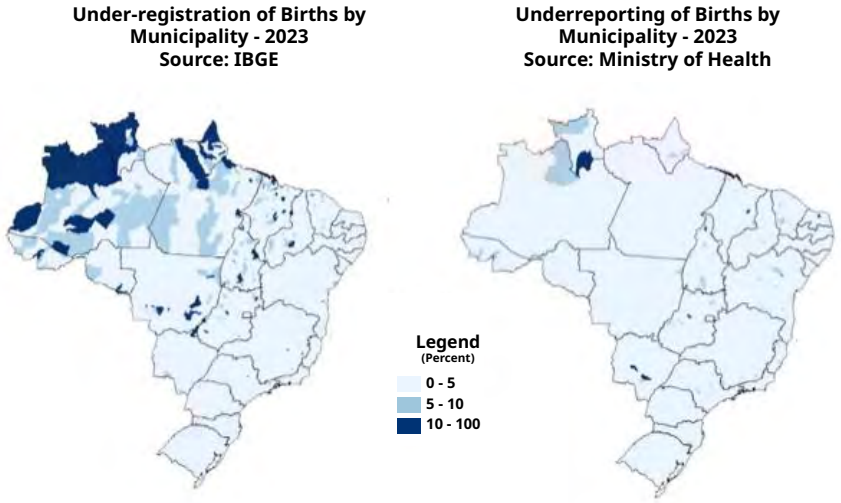


Figure 7. Percentage of Under-registration and Underreporting of Births by Municipality in Brazil — 2023. As reported by IBGE and the Ministry of Health

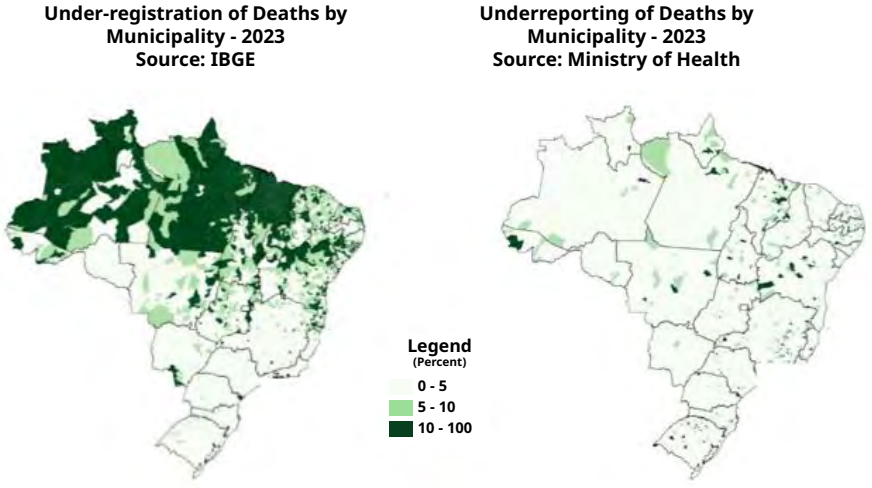


Figure 8. Percentage of Under-registration and Underreporting of Deaths by Municipality in Brazil — 2023. As reported by IBGE and Ministry of Health in 2023

By aligning statistical efforts with rights-based approaches, countries can ensure that no child is left invisible—in data or in policy.

For child mortality, the under-registration rate for deaths among children aged 0 to 1 year declined from 15.3% in 2015 to 11.4% in 2023—indicating steady progress, although considerable challenges remain. The analysis also revealed an inverse correlation between age range and under-registration rates: the broader the age range, the lower the percentage of under-registration.

Additionally, the registration of births to adolescent mothers remains a major concern, with consistently higher under-registration rates compared to other maternal age groups. This finding is particularly troubling given its implications for access to essential services and the protection of fundamental rights for both adolescent mothers and their children.
In this context, this case study addressed two Sustainable Development Goals (SDGs):

SDG 3 – Good Health and Well-Being (Target 3.2)

This target aims to end preventable deaths of newborns and children under five years of age. Achieving it requires accurate, timely, and disaggregated data on births and child mortality. Such data are essential for tracking progress, identifying regional and social disparities, and informing public health strategies. Strengthened administrative records make it possible to monitor trends more precisely and to address the specific needs of vulnerable populations, directly contributing to improved child survival outcomes.

SDG 16 – Peace, Justice and Strong Institutions (Target 16.9)

This target focuses on ensuring legal identity for all, including universal birth registration. Civil registration is a fundamental right and a gateway to public services such as healthcare, education, and social protection. It is also vital for social inclusion and the full exercise of citizenship.

The integration of administrative data systems enhances the ability to monitor this target by providing more complete and accurate statistics. This, in turn, enables UNICEF and public policymakers to detect gaps, design evidence-based interventions, and track the effectiveness of actions taken—ultimately strengthening the protection of children's rights and ensuring that no child is left behind.

The Brazilian case illustrates that the effective integration of vital statistics systems depends not only on technical capacity and adequate infrastructure but also on a sustained commitment to upholding children's rights and on a nuanced understanding of the specificities of child-related data.

The observed improvements in data coverage and quality—particularly in historically vulnerable regions—highlight the transformative potential of integrated statistical systems in advancing child and adolescent well-being and safeguarding fundamental rights.

This model of system integration provides valuable lessons for other countries in the region, demonstrating how strengthening civil registration and vital statistics systems can directly support the achievement of the SDGs and the full realization of children's rights. By aligning statistical efforts with rights-based approaches, countries can ensure that no child is left invisible—in data or in policy.

The IBGE Brazil case study offers valuable lessons and practical recommendations that can guide other NSOs in the region. Key takeaways include:

1. Establish robust legal frameworks that enable secure, transparent, and regulated data sharing with the NSO as an essential foundation for integrated statistical systems.

2. Invest in technological infrastructure and continuous technical training to ensure the proper maintenance, updating, and scalability of data integration platforms.

3. Strengthen coordination among sectors to improve consistency, completeness, and reliability of vital records and administrative data.

4. Encourage international cooperation to facilitate the exchange of experiences, methodologies, and technologies that can accelerate progress in administrative data integration.

5. Adopt an experimental approach when publishing new linked administrative statistics, allowing NSOs to test innovative methods, explore new data sources, and apply emerging technologies without the constraints of fully standardized frameworks.

6. Maximize the reuse of existing data sources across statistical programs to enhance efficiency, reduce costs, and expand analytical capacity.

Hard-to-Reach versus Hard-to-Count Populations, Data Linkage, and Capture-Recapture Methods

In administrative data–based statistics, certain population groups may be hard to reach or under-represented in statistical estimates due to a lack of capture in administrative data systems. According to the UNECE Task Force on Hard-to-Reach Groups in Administrative Data Sources, some groups tend to be underrepresented for a variety of reasons, including:

- They are hard to identify.
- They choose not to be identified.
- They are systematically excluded (e.g., people living in institutions).
- They are physically hard to reach (e.g., those living in remote areas or without a fixed address).
- They are difficult to record even once identified (e.g., people living with dementia, people who do not speak the national language, and young children).

(UNECE, 2023).

Data linkage and capture-recapture methods combine data from multiple sources to estimate the size and characteristics of a population that may be underrepresented in a single data source.

Capture-recapture methods—originally developed and extensively used in ecology to estimate wildlife populations—involve linking individuals across multiple datasets and "capturing" individuals in one dataset, "recapturing" them in another, and using statistical models to estimate the total population, including those not observed in any dataset. This approach is often referred to as Multiple Systems Estimation.

Case Study III:
National Institute of Statistics Uruguay's Integrated System of Statistical Records and Surveys

In Uruguay, the System of Integrated Statistical Registers and Surveys (SIREE) represents a comprehensive and dynamic approach to statistical infrastructure. It provides a holistic framework with interconnected elements and features a data warehouse as its most visible component.

The SIREE is underpinned by several key elements, including master data and metadata management, the Generic Statistical Register Business Process Model (GSRBPM) (adapted from the Generic Statistical Business Process Model, or GSBPM), and a conceptual and methodological framework.

The system is based on three fundamental pillars—people, processes, and technology—in alignment with INE Uruguay's strategy. The SIREE's approach is similar to the register-based statistical model common in the Nordic countries (IADB & Segui Stagno, 2016). See Figure 9 below for a representation of SIREE's components (INE, 2021).

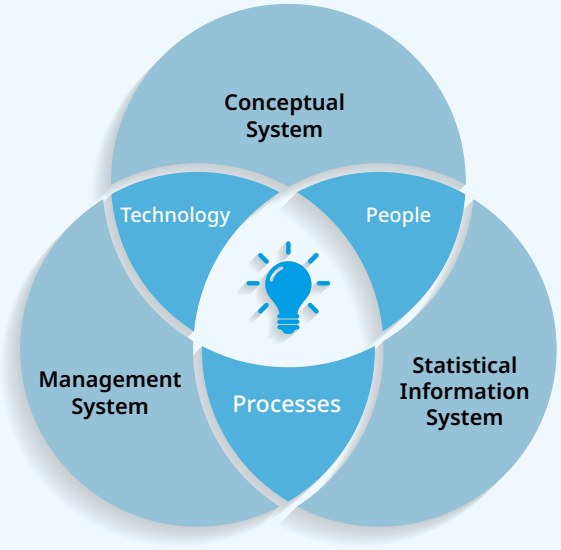


Figure 9. SIREE System Components

Register-based statistics and SDG reporting

Statistical population registers, as described in Chapter 1, are derived from administrative population registers or other administrative data sources that are processed and cleaned to meet statistical standards. They differ from population registers because they are designed exclusively for statistical purposes by NSOs.

A statistical population register—combined with a statistical business register, statistical building register, and statistical activity register—forms a register-based infrastructure that maximizes the use of actionable information from data already collected. This approach improves the timeliness of official statistics and reduces response burden.

A register-based system can be leveraged in a combined traditional–administrative data population census or even in replacing traditional census data collection altogether.

Key benefits of the register-based approach include: lower costs, increased data quality, opportunities for greater disaggregation, including for children and adolescents, higher data frequency, and reduced response burden. These benefits extend directly to SDG indicator reporting.

As a system of integrated statistical registers, SIREE links and integrates a set of administrative, survey, and census data sources.

See Figure 10 for the SIREE data logic model and Figure 11 for details on the population register, including the incorporation of children and adolescent administrative data sources.

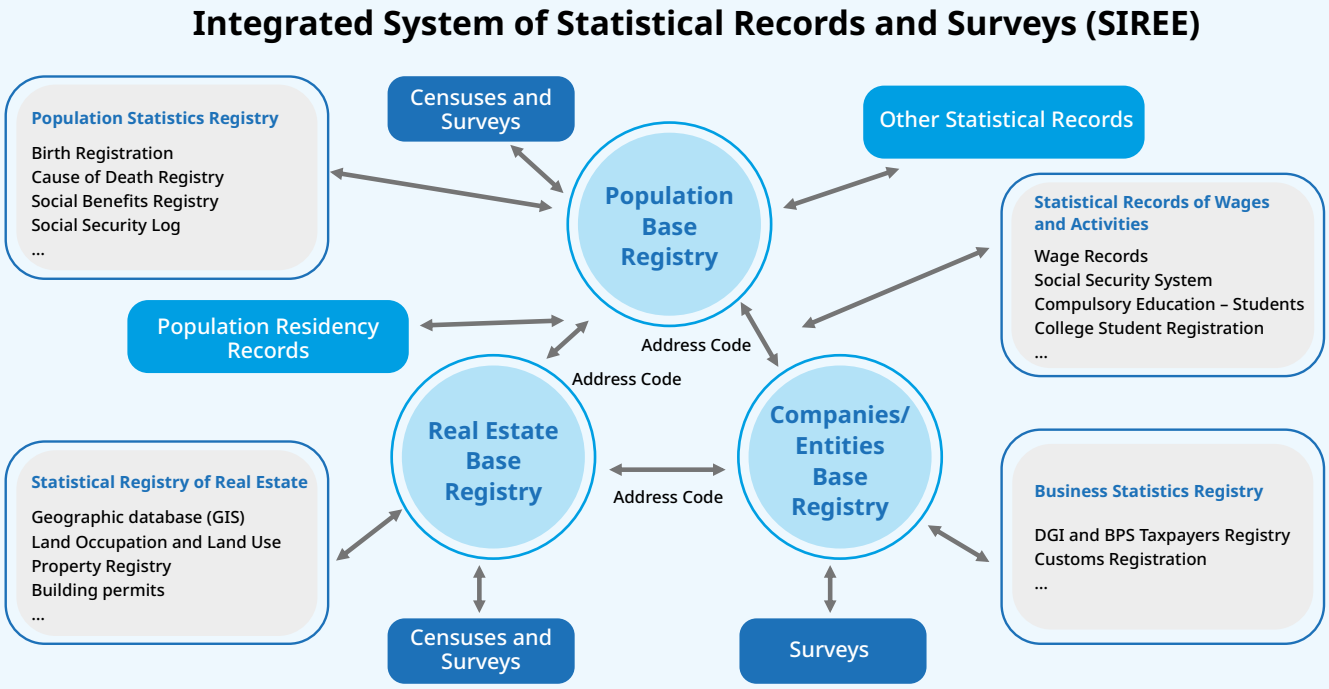


Figure 10. SIREE Data Logic Model

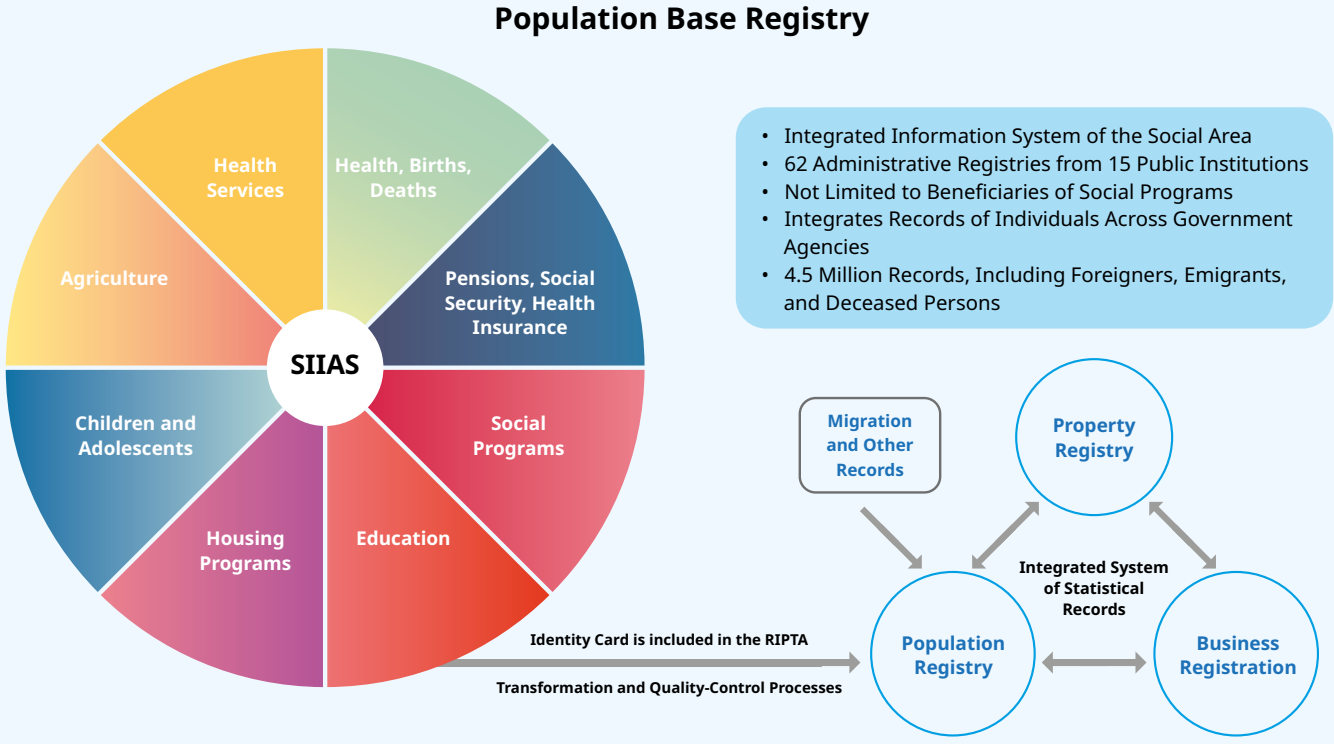


Figure 11. SIREE Statistical Population Register Components

Guided by a strong data governance and stewardship program, SIREE provides secure access to data users both within INE Uruguay and across institutions through a microdata library.

This approach securely unlocks data to be accessed for a range of research needs. For example, in 2023, INE Uruguay successfully completed the first administrative-data and direct-collection census in the LAC region using SIREE (INE, 2024).

In this process, children under 14 years of age who were not successfully enumerated were added to the census household in cases where mothers were enumerated. This was accomplished by using the child's birth certificate to obtain the unique identifiers of both the child and the mother, allowing the linkage of the mother's census questionnaire.

There are several key takeaways and reflections from the SIREE case study that can serve as helpful insights for other NSOs in the region in the context of child rights statistics:

1. Adopting a register-based system approach enables NSOs to establish durable statistical infrastructure serving a diverse range of data needs.

2. A mature register-based system can be used to reinforce surveys or census statistics, as demonstrated in Uruguay, which strengthened its census statistics in 2024.

3. By including administrative data sources on children and adolescents in the statistical population register, coverage of this subpopulation is improved.

4. Secure inter-institutional microdata sharing unlocks data for a wide range of research needs, enhancing their utility.

Data Analysis on Child Development in Uruguay

With the support of UNICEF, Uruguay conducted an analysis of early childhood development data, focusing on information from the Ministry of Public Health under the performance goals of the National Integrated Health System.

Presented in the report *Data Analysis of Child Development in Uruguay,* this work examines prenatal care, newborn health, and child development outcomes for 2018–2022, based on administrative data, with the aim of strengthening preventive measures and addressing disparities (Ministerio de Salud Pública (Uruguay) & UNICEF, 2023).

This initiative is an excellent example of administrative data linkage and analysis within the larger national statistical system. A comprehensive, harmonized, and longitudinal linked administrative database was constructed for analysis, based on the child and mother's identity cards, and included both cross-sectional and longitudinal measures that were used for indicator development.

Case Study IV:
National Institute of Statistics and Informatics of Peru – Provincial and District Monetary Wealth Map

In Peru, the National Institute of Statistics and Informatics (INEI) undertook a monetary poverty mapping initiative at the provincial and district levels under the supervision of the National Poverty Advisory Committee, with engagement and advisory support from development banks and other institutions. The objective of this work was to illustrate the geographic distribution of monetary poverty in smaller areas (provinces, districts, and sub-districts) to identify and prioritize the poorest regions of the country.

The approach involved integrating a large set of data sources and employing small area estimation (SAE) techniques. This tool is critical for improving public policy decisions, enhancing social welfare, and defining priorities for public resource allocation—particularly because more aggregated data do not reflect the heterogeneity that exists within provinces or districts (INEI, 2020).

The report employs a micro-level estimation methodology proposed by the World Bank, which extends small area estimation techniques.

Primary data sources included a combination of the 2017 and 2018 national household surveys, the 2017 Population Census, and 20 administrative databases. Among the 22 total databases used is the conditional cash transfer program Juntos, which provides monthly payments to households meeting specific criteria, such as having at least one pregnant woman or at least one child under 17 living in the household.

According to the results, 24 provinces have a poverty rate of 48.8% or higher. Fifty provinces fall within a poverty range of 38.3% to 48.7%, while 45 provinces range from 27.3% to 38.2%. Thirty-nine provinces have a poverty rate between 16.0% and 27.2%, and 38 provinces report less than 16.0% poverty. See Figure 12 for monetary poverty estimates by province.

Child poverty and SDGs

Sustainable Development Goal (SDG) 1 aims to end poverty in all its forms across the globe. Children and adolescents experience poverty differently from adults due to their developmental stage and associated vulnerabilities, which can have lifelong consequences.

Child poverty entails the lack of public and private material resources—such as good health, clean water, education, and shelter—needed to realize child rights. Unfortunately, these differences are often not fully recognized, and child poverty can be camouflaged when derived from other measures of poverty.

For example, disaggregating household poverty by age to measure childhood poverty may be insufficient because a household could be considered non-poor based on adult indicators, even though children in that household suffer deprivations.

Despite these important differences, Peru's Provincial and District Monetary Poverty Map remains highly useful in advancing child rights and SDGs related to children. It provides granular baseline geographic information on monetary poverty that can be combined with other data sources to target areas where child poverty is likely to be concentrated, especially when overlaid with data on households with children.

Small Area Estimation and Poverty Mapping

Small area estimation (SAE) is a collection of statistical techniques used to produce reliable estimates for subgroups or geographic regions where direct survey data are insufficient due to small sample sizes. This is achieved by complementing survey data with auxiliary information, often drawn from administrative and census records.

In small area estimation, "borrowed strength" is taken from these auxiliary variables, typically through modelling.

Poverty mapping applies SAE techniques to estimate poverty indicators for small geographic regions. By reaching the most granular levels of geographic disaggregation, policies and programs can be better informed and ultimately improve outcomes for those living in poverty.

More information on using SAE for SDGs can be found in the United Nations Statistics Division *Toolkit on Using Small Area Estimation for SDGs* (UN, 2023).

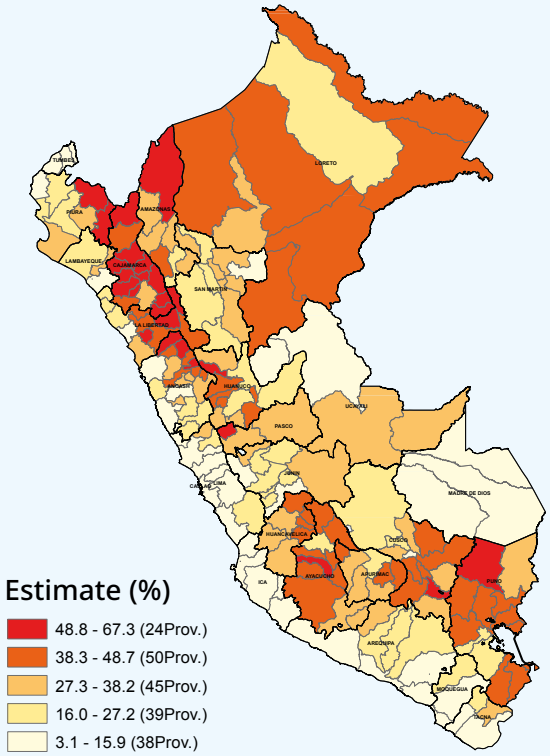


Figure 12. Peru – Monetary Poverty by Province, 2018

There are several key takeaways and reflections from the Peru case study that can serve as helpful insights for other NSOs in the region in the context of child rights statistics:

1. Apply advanced statistical techniques—such as small area estimation—combining survey and administrative data to improve estimates. These approaches can be applied to both geographic and non-geographic estimation across a range of topics beyond poverty.
2. Ensure strong statistical legislation to support the sharing of administrative data for large-scale administrative data projects.
3. Leverage children and youth registers, such as household conditional cash transfer programs involving youth, as inputs for large-scale modelling projects.

Chapter 3:

Roadmap to Administrative Data Use and Its Integration for Statistical Purposes Related to Children and Adolescents

Administrative data hold enormous potential to enhance the quality, timeliness, and relevance of official statistics—particularly those related to children's and adolescents' rights, well-being, and development.

When effectively governed and methodologically integrated, administrative sources can offer a cost-effective and sustainable means of producing disaggregated data that supports national planning, international reporting, and evidence-informed decision-making to uphold the rights and well-being of children.

The *Roadmap to Administrative Data Use and Its Integration for Statistical Purposes Related to Children and Adolescents* outlines the core steps that National Statistical Offices (NSOs) can follow to responsibly access, integrate, and use administrative data for the production of official statistics about children and adolescents.

The pathway is designed to be both practical and adaptable. It reflects international best practices, incorporates a child rights perspective, and emphasizes the importance of planning, data governance, quality assurance, and stakeholder collaboration throughout the process.

The eight steps presented in this roadmap include:

1. Defining the statistical need and building a strong statistical business case (Step 1).
2. Assessing feasibility and readiness (Step 3).
3. Evaluating legal and ethical alignment with statistical mandates and data protection frameworks (Step 3).
4. Ingesting, storing, and managing administrative data (Step 4).
5. Conducting data linkage activities and completing linkage execution report (Step 5).
6. Developing and executing analytical strategy (Step 6).
7. Developing and executing dissemination strategy (Step 7).
8. Institutionalizing, promoting sustainability, and scaling successful practices (Step 8).

This roadmap is intended as a living guide. It can be used to support administrative data initiatives at any stage and serves as a shared reference for modernizing data systems and promoting children's rights and well-being through improved administrative data–based statistics.

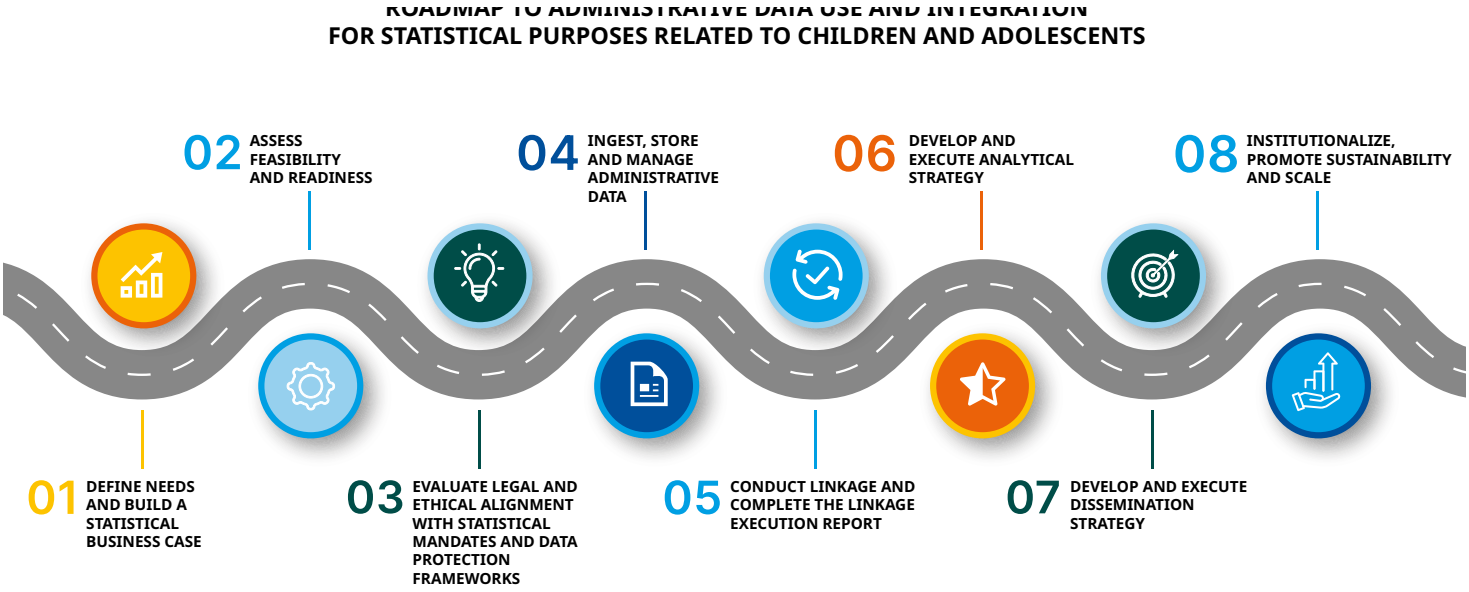


Figure 13. Pathways to Administrative Data-based Statistics on Children and Adolescents

STEP 1:
Define Needs and Build Statistical Business Case

PURPOSE

This section establishes the foundation for incorporating administrative data into child- and adolescent-focused official statistics. It emphasizes the importance of identifying statistical needs, addressing data gaps, and aligning outputs with national and international priorities, including the UN Convention on the Rights of the Child (CRC) and the Sustainable Development Goals (SDGs).

By clearly defining objectives and building a robust business case, National Statistical Offices (NSOs) ensure that administrative data are used strategically and responsibly to support evidence-informed decisions for and about children and adolescents.

1.1 Identify Statistical Need and Project Scope

In official statistics, defining data needs is the first and critical step when incorporating new data sources into statistical programs. This process typically involves, (1) identifying the demand for statistical products and the specific information needs; (2) setting high-level objectives for the intended statistical outputs; (3) determining the relevant concepts and variables requiring data; (4) assessing whether existing data collections or methodologies can satisfy these needs; and (5) preparing a business case to secure approval for producing the statistics.

Key questions to consider when determining whether administrative data related to children fulfill the statistical need and project scope:

1. Is there an information gap, and if so, how can it be described?
2. What data sources may exist to fill this gap?
3. Which other government departments (ministries or agencies) manage the relevant data?
4. Are new administrative data sources required, or are there other data sources and methods to fill this information gap?
5. What is the minimum data required to fill this gap?
6. Does this information need cross more than one sector or life domain? If so, is data linkage required to fulfill this information need?

Resources

- **United Nations Economic Commission for Europe's *Generic Statistical Business Process Model (GSBPM)***
  Developed by the United Nations Economic Commission for Europe (UNECE), the GSBPM describes and defines the set of business processes needed to produce official statistics. The GSBPM should be applied and interpreted flexibly and is best viewed as a matrix through which multiple process paths can be followed (UNECE, 2014).

- **United Nations Statistical Institute for Asia and the Pacific's Using Administrative Data to Produce Official Statistics training**
  This self-directed training addresses the increasing reliance on administrative data in national statistical systems and equips participants with techniques to integrate administrative records into official statistics effectively.

  Key topics include defining administrative data, approaches to mapping data sources, metadata, data sharing, and data linkage. Designed for government statisticians and data analysts, the training incorporates case studies and practical applications to enhance statistical production.

  By the end of the course, participants will be better prepared to utilize administrative data for evidence-based policymaking (UN SIAP, 2025).

1.2 Align with National and International Children and Adolescent Statistical Priorities

The United Nations Convention on the Rights of the Child (UN CRC) is a treaty that recognizes specific rights for children and provides international standards and benchmarks for the fulfillment of child rights (OHCHR, 1989). It is the most widely ratified human rights treaty in history and has helped transform the lives of children and adolescents around the world. NSOs play an important role in supporting the monitoring of children's rights through the provision of official statistics for evidence-informed decision-making for and about children. While there is no minimum core indicator set (or core administrative indicator set) that NSOs are required to produce under the CRC, State Parties are required to report regularly on the implementation of child rights with the support of reliable disaggregated data.

Another key source to guide alignment with national and international children's statistical priorities is the Sustainable Development Goals (SDGs). Of the seventeen SDGs, twelve relate to the rights and well-being of children and adolescents: These are: poverty, hunger, health & well-being, education, gender equality, clean water & sanitation, affordable & clean energy, decent work & economic growth, reduced inequalities, climate action, peace, justice, & strong institutions, and partnerships for the goals (UNICEF, 2018).

Within these twelve goals, there are nineteen indicators for which UNICEF is custodian or co-custodian. Details on each of these twelve goals and the 44 indicators directly related to children and adolescents can be found at: https://data.unicef.org/sdgs/.

In the broader context of child rights monitoring, no official minimum core indicator set has been established. Nevertheless, a set of commonly recognized domains and indicators provides a basis for countries to monitor the realization of children's rights and their well-being. Domains relevant to administrative data, along with key disaggregation variables, are presented in Figure 14 and Figure 15.

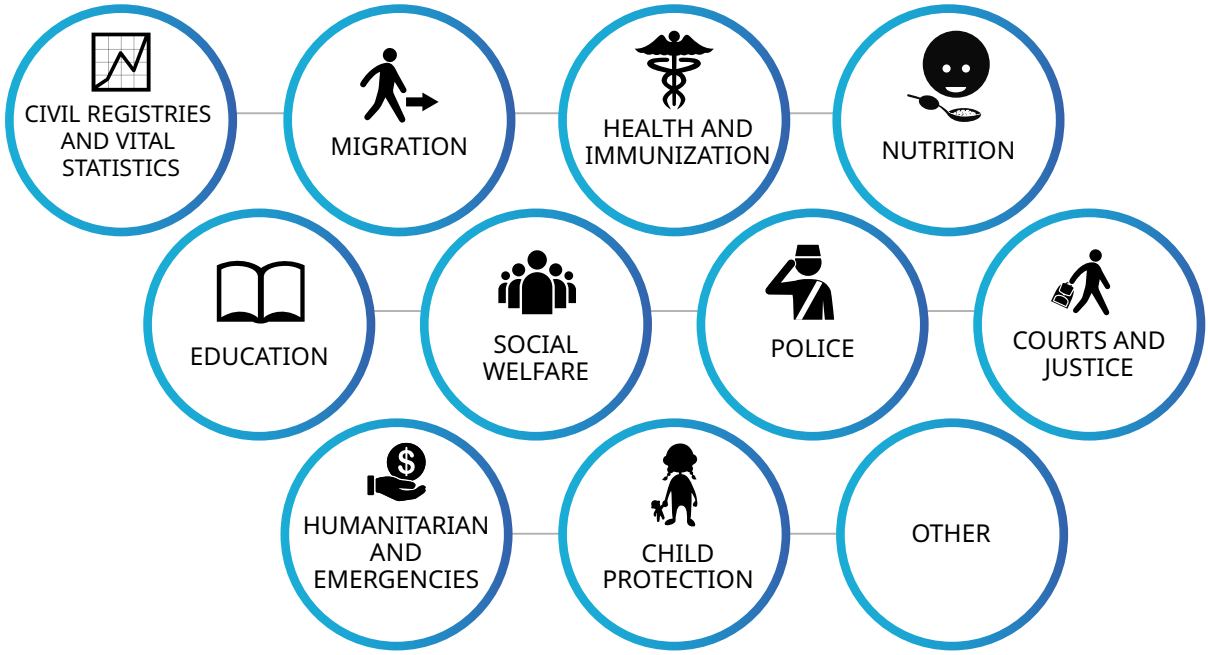


Figure 14. Administrative Data Sectors Commonly Considered Relevant in Informing Decisions for and about Children and Adolescents

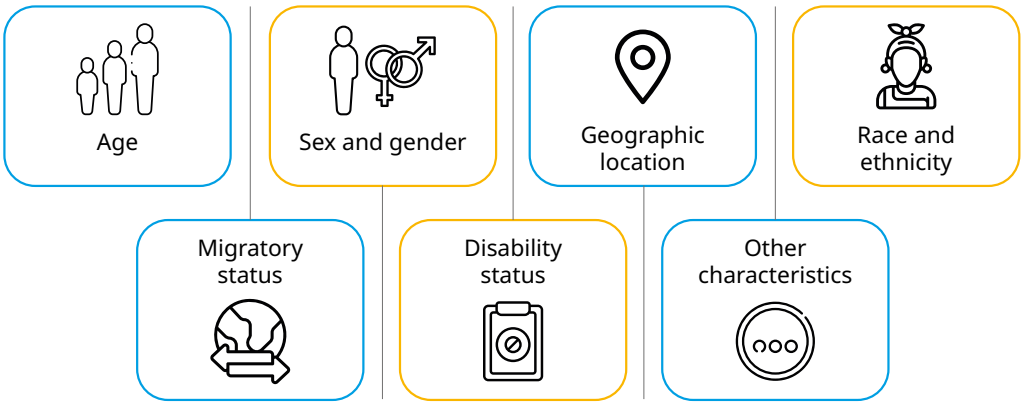


Figure 15. Key Characteristics for Disaggregation

UNICEF's *Conceptual Framework and Guidance for Integrating a Children's Rights Perspective into Administrative Data-Based Statistics* (UNICEF, 2025a) provides an additional reference for aligning with national and international priorities in children's statistics.

It asserts that one key aim for NSOs should be to support evidence-informed decisions for and about children and includes the following four key dimensions: (1) children interact with administrative data systems differently than adults; (2) data governance must be fit for children; (3) statistical output activities must focus on the areas of greatest impact for children.; and (4) children's data carry unique statistical challenges.

The framework provides a conceptual guide for NSOs and offers structured and strategic considerations and recommendations to help administrative data producers and users support evidence-informed decisions and ultimately strengthen child rights.

Key questions to consider when determining whether there is a need to produce administrative data–based statistics aligned with national or international statistical priorities for and about children and adolescents:

1. Is a child rights perspective mainstreamed throughout the statistical system?

Resources

- United Nations *Convention on the Rights of the Child*
  Adopted by the United Nations in 1989, the CRC sets out the civil, political, economic, social, and cultural rights of children (OHCHR, 1989).

- UNICEF's *Using Data to Achieve the Sustainable Development Goals (SDGs) for Children*
  This data hub provides information and resources on the 19 global SDG indicators for which UNICEF serves as custodian or co-custodian (UNICEF, 2025b).

- UNICEF's *Conceptual Framework and Guidance for Integrating a Children's Rights Perspective into Administrative Data-Based Statistics*
  This conceptual framework guides NSOs and line ministries—particularly in the Latin America and Caribbean region—by offering strategic considerations and recommendations to integrate a children's rights perspective into administrative data-based statistics and support evidence-informed decision-making (UNICEF, 2025a).

- UNICEF Eastern and Southern Africa Regional Office & PARIS21 *Toolkit for Mainstreaming Child Rights into Strategic Planning for Official Statistics*
  This toolkit provides guidelines for integrating child rights into national statistical systems. It highlights the need for child-specific data and indicators in policy development and decision-making, aligning national systems with international frameworks such as the CRC, and stresses ethical data collection practices (UNICEF & PARIS21).


© UNICEF/UNI568126/Terán

2. Are data gaps related to the CRC being addressed?

3. Are there gaps or improvements to be made on child-related SDG indicators that can be addressed with administrative data?

4. What policy-relevant or developmental gap does the proposed project aim to fill?

5. How will the planned statistical output improve decision-making?

1.3 Develop Business Case

Once the questions in Sections 1.1 and 1.2 have been addressed, a business case for leveraging administrative data—whether by using existing data sources within the NSO or by acquiring new sources—can be formally articulated.

This provides the foundation to move forward with Step 2: Assessing Feasibility and Readiness.

Consideration for Gender Mainstreaming

Beyond national and international children's statistical priorities, it may be important to consider additional priority areas such as gender mainstreaming. This refers to the systematic integration of gender equality perspectives into institutional practices, including the production of official statistics. The following references may be helpful in this regard:

- UNICEF's *Gender Statistics and Administrative Data Systems: An Annotated Resource Guide*
  This resource guide outlines how administrative data systems can be used to strengthen gender statistics, especially for children and marginalized groups. It organizes resources into thematic areas such as frameworks, capacity building, and country case studies to help address gender data gaps and improve national statistical systems (UNICEF, 2020).

- UNICEF's *Strengthening Administrative Data Systems to Close Gender Data Gaps*
  UNICEF leverages its Administrative Data Maturity Model (ADaMM) to provide practical guidance to national statistical offices and line ministries on enhancing administrative data systems—particularly those in health, education, and civil registration—to improve the availability, granularity, timeliness, and alignment of gender data with international standards.

  Developed with the Inter-Agency and Expert Group on Gender Statistics, this guidance highlights both challenges and opportunities in making gender-responsive administrative data systems more effective (UNICEF, 2022b).

STEP 2:
Assess Feasibility and Readiness

PURPOSE

This section focuses on evaluating the feasibility of using administrative data for statistical purposes. It guides NSOs in assessing data quality and availability, technical infrastructure, and human resource capacity required to manage and use administrative data.

Special attention is given to the unique challenges of data linkage, including privacy, data structure, and the discriminatory power of identifiers. These assessments are essential for determining whether administrative data can be transformed into robust, high-quality official statistics.

2.1 Assess Administrative Data Quality

A significant challenge in using administrative data for statistical purposes is that the data were typically not originally collected for statistical use. NSOs do not control the collection and processing of these data; these tasks are managed by the original data producers.

Similarly, the units and variables within administrative data are defined by administrative needs, which may not align directly with the requirements of the NSO. Therefore, it is crucial to assess the quality of a potential administrative dataset early in the process to ensure it is fit for purpose.

Figure 16 presents the twelve dimensions of administrative data quality. These dimensions draw on the United Nations National Quality Assurance Framework (UN NQAF), which provides a structured approach to ensuring the quality of official statistics (UN, 2019).

This framework has been expanded by the United Nations Collaborative on Administrative Data (CAD) through its *Toolkit for Quality Assessment of Administrative Data for Official Statistics,* which provides practical guidance and templates (UN Women, 2022).

Note: When data linkage is involved, the availability of unique identifiers is especially critical.

In addition, UNICEF's Administrative Data Maturity Model (ADaMM) provides a complementary framework for evaluating the readiness and quality of administrative data systems for statistical use.

The model outlines multiple dimensions that help NSOs assess not only the technical characteristics of administrative datasets but also the institutional and systemic conditions that influence their usability.

As a child-focused evaluation tool, ADaMM is particularly relevant when working with data related to children and adolescents.

Key questions to consider when assessing the quality of administrative data:

1. Is the administrative data under consideration of sufficient quality?
2. Is it possible to obtain information from the administrative data owner regarding the quality dimensions?
3. How mature are the administrative systems being evaluated?

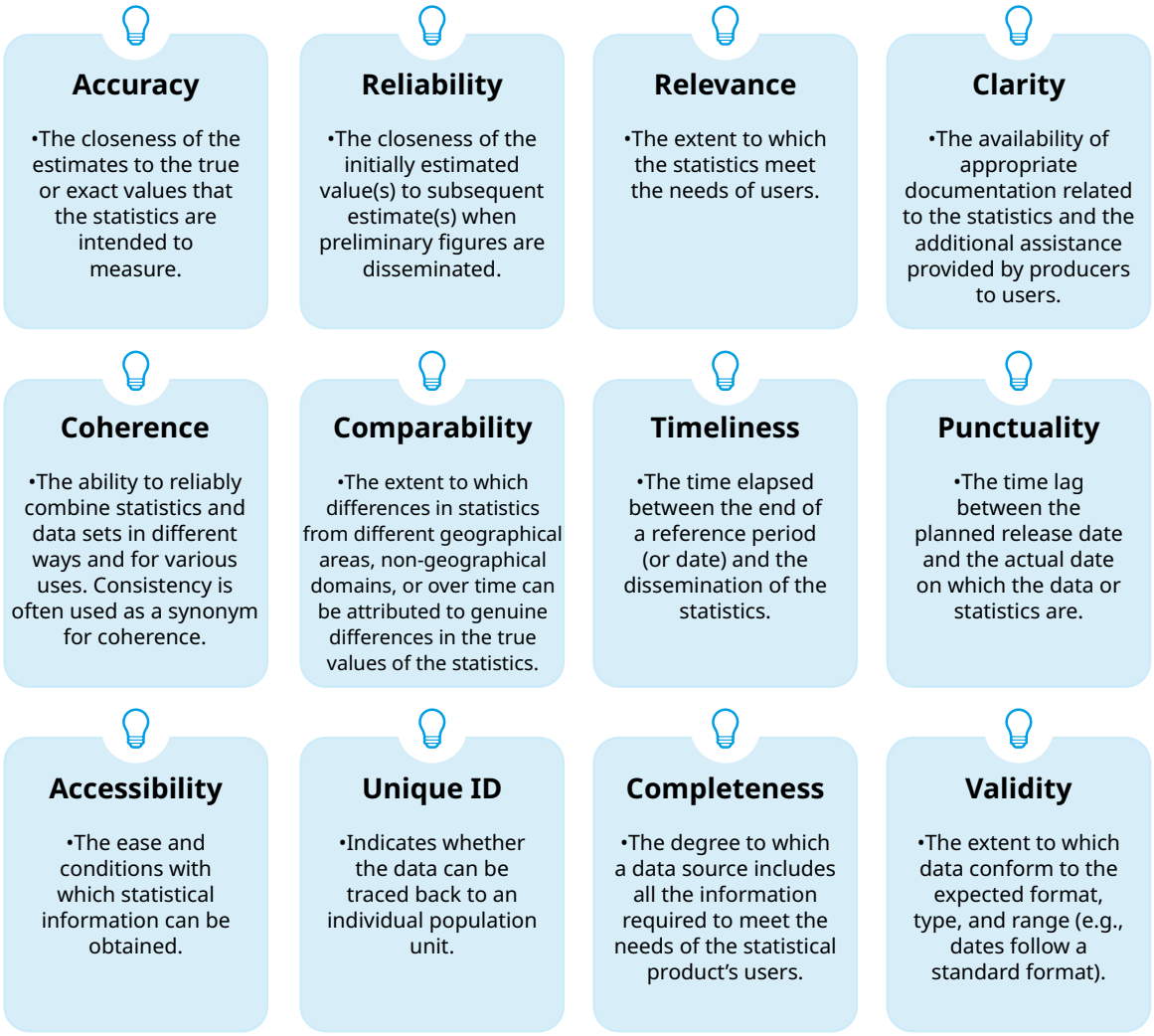


Figure 16. Quality Dimensions for Administrative Data – Toolkit for the Quality Assessment of Administrative Data for Official Statistics

Resources

- United Nations Department of Economic and Social Affairs – *United Nations National Quality Assurance Frameworks Manual for Official Statistics, Module for Quality Assurance when using Administrative and Other Data Sources to produce Official Statistics*
  This module serves as a comprehensive guide for statistical agencies seeking to enhance the quality of statistics derived from administrative and other data sources, ensuring they meet the highest standards of reliability and user trust (UN, 2019).

- United Nations Collaborative on Administrative Data – *Toolkit for Quality Assessment of Administrative Data for Official Statistics*
  This toolkit expands on the above module and provides practical guidance and templates for assessing administrative data quality (UN Women, 2022).

- UNICEF's Administrative Data Maturity Model (ADaMM)
  ADaMM is a framework designed to evaluate and improve the use of administrative data within national statistical systems, particularly in relation to children and adolescents. It helps countries assess their current data management capabilities and supports the development of more effective, integrated, and ethical use of administrative data for policymaking and reporting on child and adolescent-related outcomes. It also promotes continuous improvement in data practices to protect children's rights (UNICEF, 2022a).

- United Nations Economic Commission for Europe – ***Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices***
  This UNECE handbook provides a comprehensive guide for integrating administrative and secondary data into official statistics. It emphasizes the importance of establishing legal frameworks, ensuring data quality, and fostering collaboration between data providers and statistical agencies. It also highlights the benefits of using such data sources—including cost-effectiveness and improved timeliness—while addressing challenges like data confidentiality and standardization (UNECE, 2011).

2.2 Evaluate Technical Readiness

A foundational element of any administrative data statistical project is the technical environment. The environment must be: robust enough to handle potentially large volumes of data; sophisticated enough to support linkage algorithms when necessary; and secure enough to protect sensitive information.

Key components of the technical environment for administrative data–based statistical production include:

- **Secure data transfer and ingestion.** Administrative data may originate from external providers; therefore, the technical environment must support secure transmission protocols and verification mechanisms to ensure data integrity and confidentiality during transfer.
- **Processing power.** Administrative data and data linkage can involve manipulating large datasets, requiring substantial computational resources. In some cases, high-performance servers or cloud-based platforms may be necessary.
- **Data storage.** Large administrative datasets require scalable and secure storage systems.
- **Software and linkage algorithms.** The technical environment should support the required software, including tools for data linkage techniques and other advanced methodologies.
- **Data processing and standardization.** The system should include software to handle missing data, inconsistent formats, typographical errors, and duplicate records.
- **Security.** The technical environment must implement strong data security protocols to protect sensitive information.

Key questions to consider when evaluating technical readiness:

1. Can the computing infrastructure handle encrypted data transfers, secure storage, and high-performance processing?
2. Are there budget considerations regarding data storage, processing, or required tools?

2.3 Evaluate Human Resource Readiness

Having skilled personnel is essential for successful administrative data projects, as these demand expertise in: data management and software tools, linkage methodologies (if applicable), and legislative and technical contexts.

Employees must be proficient in the methodologies required to manage computing environments, ensure secure data storage, and implement privacy protections. Additionally, they should also have a strong grasp of data governance and legal requirements to ensure compliance. Effective collaboration among technical teams, data custodians, and decision-makers is key to delivering accurate and secure results.

Key questions to consider when assessing human resource readiness:

1. Do employees need training in producing official statistics using administrative data?
2. Are there other skill gaps in supporting areas such as legal or information technology?

2.4 Assess Linkage Feasibility

Interacting with identifiers is inherent to data linkage. Privacy in data linkage is fundamentally different from other statistical operations, where it is typically defined as protecting the identity of individuals in a dataset.

In data linkage, the ultimate objective is to identify unique individuals in one administrative dataset with accuracy such that these individuals can be identified in another dataset to create linked records at the individual level for the production of statistics.

The feasibility of a linkage project depends largely on the quantity and quality of the identifying information available in the datasets being linked.

Before initiating a linkage project—which may involve resource-intensive activities such as developing data-sharing agreements and data acquisition procedures—it is important to assess whether high-quality linkage is possible given the available identifiers and their associated discriminatory power.

Key questions to consider when evaluating linkage feasibility:

1. What identifiers are available in the administrative datasets?
2. Can deterministic, probabilistic, rules-based, or hybrid record linkage approaches be used?
3. Has a methodological protocol for data linkage—as discussed in Step 5: Conduct Linkage Activities and Appendix 2 of this roadmap—been completed?

Resources

- **Appendix 1: A Brief Introduction to Linkage Methodologies and Software**
  Provides a concise overview of data linkage methodologies and commonly used software tools relevant for official statistics. It supports NSOs in identifying suitable methods and technologies to execute high-quality linkage projects related to children and adolescents.
- **Appendix 2: A Methodological Protocol for Data Linkage**
  Offers a standardized methodological protocol template for data linkage. It outlines key elements needed to design, document, and validate a linkage project, including: project objectives, datasets and metadata, legal and ethical considerations, linkage methodology, data cleaning, software, validation, risk management, and dissemination policies. This ensuring transparency, quality, and accountability in producing linked datasets for statistical use.
- **Data4Now's *Training in Record Linkage – National Administrative Department of Statistics (DANE) Colombia***
  Delivered under the Data4Now initiative, this training focused on strengthening the capacity of Colombia's NSO to apply record linkage techniques using administrative data. It covered methodological principles, practical applications, and tools for linking datasets in support of official statistics (UN, 2022a).

STEP 3:
Evaluate Legal and Ethical Alignment with Statistical Mandates and Data Protection Frameworks

PURPOSE

This section outlines the legal and governance foundations required for integrating administrative data into official statistics for and about children and adolescents.

It emphasizes the importance of a coherent legal framework that combines statistical, privacy, and sectoral laws; and the development of formal and informal engagement strategies with data providers—often other government departments.

In parallel, it highlights the need for strong data governance practices that promote accountability, ethical use, and the protection of children's sensitive information. Together, these elements ensure that administrative data initiatives are lawful, ethical, and trusted, and that they uphold child rights.

3.1 Determine Legal Framework and Data Provider Engagement Strategy

While administrative data can significantly improve children's and adolescents' lives by informing better decisions, they also carry the responsibility to prevent harm.

Strong legislative, data governance, and ethical environments are essential for NSOs in producing administrative data–based statistics because they ensure that these initiatives are trustworthy, legally compliant, and socially responsible.

A comprehensive statistical legal framework for administrative statistics typically includes the following components:

- **National statistical legislation.** In each country, national statistical legislation should include provisions regarding access to and sharing of administrative data, considering both aggregate and microdata. In Latin America and the Caribbean, the *Code of Good Practice in Statistics for Latin America and the Caribbean* was revised in 2023 to include a new principle on the use of secondary data sources (ECLAC, 2024). Furthermore, the Generic Law on Official Statistics establishes the mandate for access to administrative records (ECLAC, 2020).

- **Other laws and their interactions with national statistical legislation.** Because administrative data are primarily collected by other government agencies, they are often governed by sector-specific laws. Such data may be subject to personal data protection acts, privacy regulations, and transparency laws.

  Together with national statistical legislation, these laws form the statistical legal framework for administrative statistics. Ensuring coherence between national statistical legislation and sectoral laws that govern administrative data and its access is critical.

Once the legal framework has been established and it is determined that data sharing with the NSO is legally permitted, a formal data-sharing mechanism can be developed.

Memorandums of Understanding (MOUs) are an effective way to formalize data sharing. They clarify key aspects such as: data use and purpose, roles and responsibilities, aggregate or microdata exchange, periodicity and duration of the agreement, confidentiality and

security requirements, data transfer protocols, data quality standards, and data documentation and metadata provisions.
Beyond formal agreements, non-formalized engagement with data providers is equally important. Maintaining cooperation, demonstrating transparency and trust, and showing the value of administrative data-based statistics for informing policy and program decisions—and ultimately for upholding the rights and well-being of children and adolescents—are essential practices.

These relational strategies foster long-term collaboration, reduce resistance to data sharing, and build institutional alignment beyond legal mandates.

By consistently communicating the ethical use of data, the benefits of improved evidence, and the shared accountability for child and adolescent outcomes. NSOs can sustain resilient, adaptive, and responsive partnerships within evolving data landscapes.

Key questions to consider when determining the legal framework and data provider engagement strategy:

1. What legislation needs to be considered?
2. Have legal experts been consulted?
3. What informal engagement strategies with data providers will be implemented?
4. If data linkage is involved, can identifiers be shared?

Resources

- Economic Commission for Latin America and the Caribbean – ***Code of Good Practice in Statistics for Latin America and the Caribbean***
  The 2023 edition of the Code of Good Practice, developed by ECLAC, provides updated regional guidelines to strengthen the production and use of official statistics. It emphasizes institutional coordination, data quality, transparency, and the strategic use of administrative records. Notably, the revised version introduces a new principle promoting the use of secondary data sources, reinforcing their role in improving efficiency, coverage, and timeliness in statistical systems (ECLAC, 2024).

- Economic Commission for Latin America and the Caribbean – ***Generic Law on Official Statistics***
  This *Generic Law* offers a regional model for Latin American countries seeking to establish or reformulate the legal foundation of their National Statistical Systems (NSS) and the production of official statistics. It specifically establishes a mandate for access to administrative records (ECLAC, 2020).

- United Nations Statistics Division – ***Guide and Template for a Memorandum of Understanding on Data Sharing Between National Statistical Offices and Owners of Administrative Data.***
  Developed by the United Nations Statistics Division, this guide provides practical steps for NSOs to formalize data-sharing arrangements with government agencies through Memorandums of Understanding (MOUs) (UN, 2022b).

3.2 Confirm Data Governance Approach

Strong data governance underpins administrative data–based statistical activities for and about children and adolescents.

It provides clear policies, standards, and protocols for accessing, sharing, and using data, which help mitigate risks such as breaches, misuse, or inaccuracies, ensure alignment with national laws and global standards, foster transparency and accountability in partnerships between data providers and NSOs.

Because administrative data often involve sensitive topics—such as health, income, or education—it is equally important to foster an ethical environment and adhere to a formal ethical framework. Given the particular vulnerability of children and adolescents, who may not be developmentally mature enough to understand how to protect their data, additional layers of care are required when handling children's administrative data.

As noted in Section 1.2 – Align with National and International Children Statistical Priorities, UNICEF's *Conceptual Framework and Guidance for Integrating a Children's Rights Perspective into Administrative Data-Based Statistics* provides structured and strategic recommendations for understanding the role of administrative data producers and users in supporting evidence-informed decisions for and about children and, ultimately, in upholding child rights.

Within this framework, data governance fit for children is one of the four key dimensions, offering concrete recommendations for NSOs (UNICEF, 2025a).

Key questions to consider when determining the data governance approach:

1. What access controls will be implemented?
2. Are there documented protocols for data access, retention, and disposal?
3. How will ethical oversight be ensured?
4. Has a data custodian of the administrative dataset been identified?
5. How will data be anonymized or pseudonymized, especially when linking identifiable information?
6. Has a data custodian of the linked dataset been identified?
   (Note: a linked dataset is considered a new data asset and should be governed accordingly.)
7. How does the data governance approach help to uphold data provider and public trust?

Resources

- UNICEF – *Conceptual Framework and Guidance for Integrating a Children's Rights Perspective into Administrative Data Based Statistics*

  This framework provides structured and strategic recommendations for NSOs and line ministries to support evidence-informed decisions for and about children, ensuring the protection of child rights. Within the framework, data governance that is fit for children is highlighted as one of the four key dimensions, with practical guidance to design governance approaches tailored to children's specific vulnerabilities and rights (UNICEF, 2025a).

STEP 4:
Ingest, Store and Manage Administrative Data

PURPOSE

Once Steps 1 through 3 are in place, administrative data can be securely transferred and stored within the NSO. This step ensures that data integrity is maintained, processing environments meet statistical and privacy standard; and administrative data are properly staged for further transformation, potential linkage, and analysis.

4.1 Secure Data Transfer and Ingestion

Because administrative data are often sourced from other government departments, secure and reliable transfer and ingestion protocols are critical.

Recommended practices include:

- Using secure file transfer methods that protect the data and require login credentials.
- Establishing secure receiving areas where incoming data can be safely checked and validated.
- Keeping detailed records of the transfer, including confirmation that files were received intact and complete.
- Clearly documenting all transfer and security steps in the data sharing agreement or memorandum of understanding (MOU).

4.2 Establish Data Storage, Management, and Access Protocols in Line with the Data Governance Approach

Robust data storage, management, and access protocols are essential in administrative data–based statistical projects. As part of a broader data governance program, these protocols help ensure that data are protected, ethically handled, and maintained for long-term usability, trust, and legal compliance. Key best practices include:

- **Data minimization.** Collect, share, and use only the data strictly necessary to reduce privacy risks.
- **Anonymization.** Remove or modify personally identifying information to prevent re-identification.
- **Access control.** Restrict data access based on roles and necessity.
- **Encryption and secure storage.** Protect data while in transit and in storage.
- **Active transparency.** Clearly communicate why data are collected, how they are managed, and how they are used.
- **Data custodianship.** Assign clear responsibility for managing, protecting, and monitoring the data lifecycle to ensure accountability and compliance.

  Note: Linked datasets should be treated as new data assets and placed under explicit custodianship.
- **Retention policies.** Establish and enforce time-bound retention schedules based on purpose and sensitivity, ensuring secure deletion or archiving when data are no longer required for approved statistical use.
- **Data separation.** Store identifying and analytical data separately to enhance confidentiality.

The separation principle—keeping personally identifying information separate from analytical datasets—strengthens privacy protection and reduces the risk of unauthorized access. Figure 17 illustrates how the separation principle can be implemented in a data linkage environment.

It is important to note that the separation principle is relevant to the storage of all administrative data, not only that in a linkage environment.

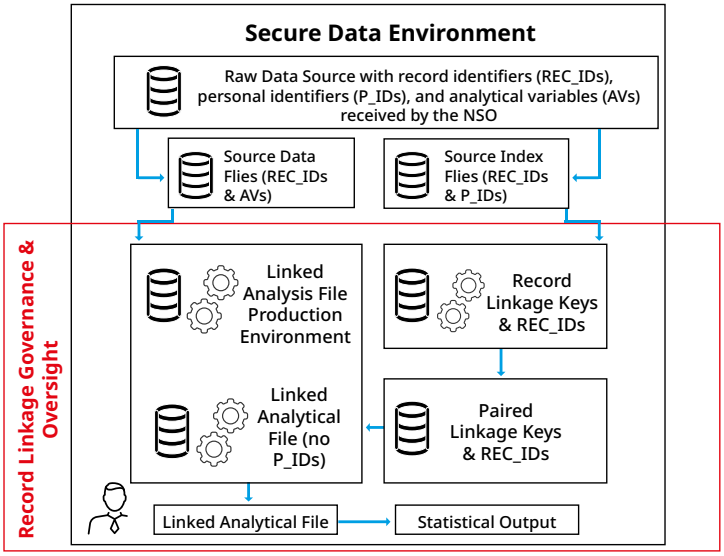


Figure 17 Example Implementation of the Separation Principle

In addition, the *Five Safes* framework is an internationally recognized operational model within broader data governance systems. It provides a structured approach to balance data utility and privacy (Desai, Ritchie & Welpton, 2016). The Five Safes has already been adopted in many NSOs within the Latin American and Caribbean region and is a useful model to implement as part of a data governance program.

1. **Safe people:** Ensure that only qualified and authorized individuals, who understand their ethical and legal obligations, are granted access.

2. **Safe projects:** Confirm that data are used exclusively for approved research projects serving the public interest.

3. **Safe settings:** Provide controlled, secure environments where data access is monitored and restricted to prevent misuse or breaches.

4. **Safe data:** Guarantee that data are de-identified or controlled to reduce the risk of individual identification.

5. **Safe outputs:** Require review of research results prior to publication to avoid disclosing sensitive or identifiable information.

Key questions to consider when designing data ingestion and management approaches:

1. How will data be securely transferred to the NSO?

2. Where will data be stored, and is that environment secure?

3. Who is responsible for managing and protecting the data?

4. Who will be allowed to access the data, and under what conditions?

5. Will personally identifiable information be kept separate from analytical datasets?

6. What measures will be taken to protect individual privacy?

Resources

- **Tanvi Desai, Felix Ritchie, and Richard Welpton – *Five Safes***

  The *Five Safes* framework provides a structured approach to managing data access for research while balancing data utility and confidentiality. It addresses five dimensions: Safe people, Safe projects, Safe settings, Safe data, and Safe outputs.

  Widely adopted by national statistical offices and research institutions, the framework supports secure access to sensitive administrative and linked data while maintaining public trust (Desai, Ritchie & Welpton, 2016).

STEP 5:
Conduct Linkage and Complete Linkage Execution Report

PURPOSE

This step outlines how to execute the record linkage process and document its results. Its goals are to ensure that: the linkage follows the defined methodology, the process is transparent and reproducible, and the resulting data meet quality standards for analysis, dissemination, and decision-making.

Preparing a linkage execution report strengthens accountability, supports independent review, and confirms whether the linked dataset is fit for its intended statistical purpose.

5.1 Develop Methodological Protocol for Linkage

As referenced in Section 2.4: Assess Linkage Feasibility, it is critical to plan and define the linkage methodology in advance. A methodological protocol helps guarantee that linkage is feasible, accurate, and of high quality. This protocol should include at least the following components (see Appendix 2 for a template):

- **Data source assessment.** Evaluate linkage feasibility in two dimensions: (1) Identifier variables: verify that each dataset contains sufficiently strong identifiers to allow accurate linkage. Even with advanced methods, if identifiers are incomplete or weak, linkage is not possible. (2) Analytical variables: assess whether the combined variables will provide the analytical utility needed (e.g., quality of sources, reference periods, and relevance of included variables).
- **Preprocessing and standardization of linkage variables.** Preprocessing and standardization of identifiers for record linkage involves cleaning and harmonizing key variables—such as names, dates of birth, and addresses—to reduce inconsistencies. Techniques such as formatting text (e.g., lowercase conversion), applying phonetic algorithms, and correcting common errors improve match quality. These steps ensure comparability across datasets and lay the foundation for accurate linkage.
- **Linkage strategy.** A linkage strategy outlines how records from different datasets will be matched. It defines the matching variables, blocking rules to reduce comparisons, and thresholds for accepting, rejecting, or manually reviewing matches. A clear strategy ensures the linkage is efficient, transparent, and minimizes errors like false matches or missed links.
- **Linkage quality validation.** The validation process assesses the accuracy and reliability of the record linkage by measuring how well-matched records reflect true links. Describing this process in a feasibility protocol ensures that the linkage method is evidence-based, minimizes errors, and supports confidence in the resulting dataset.

Key questions for developing a linkage methodological protocol:

1. Have the datasets been thoroughly assessed for identifier and analytical suitability?
2. Is the proposed linkage strategy methodologically sound and clearly documented?
3. Will the resulting linked dataset meet the information needs defined at the outset?

Resources

- **Appendix 1: A Brief Introduction to Linkage Methodologies and Software**
- **Appendix 2: A Methodological Protocol for Data Linkage report template**

5.2 Conduct Linkage and Prepare Linkage Execution Report

After finalizing the methodological protocol, execute the linkage strictly following the approved steps and specifications. This ensures that the process remains methodologically sound and reproducible.

Upon completion, it is best practice to prepare a Linkage Execution Report that includes: a detailed description of how the linkage was carried out, any deviations from the approved protocol and their justifications, key linkage statistics (e.g., match rates, error estimates), and important decisions made during execution.

This documentation provides transparency, supports quality assurance, and facilitates future replication or auditing. It also serves as essential evidence to confirm that the linked dataset is fit for purpose in official statistical production.


© UNICEF/UNI594351/Dejongh

STEP 6:
Develop and Execute Analytical Strategy

PURPOSE

Developing an analytical strategy is a critical step in transforming administrative data into meaningful, high-quality official statistics. It provides a structured roadmap that defines what will be measured, how data will be prepared and analyzed, and how the results will be interpreted and used.

This section describes the core elements of an analytical plan: confirming data suitability, defining statistical outputs, selecting appropriate analytical methods, and ensuring quality assurance.

Where data linkage is involved, additional measures are highlighted to address integration quality and cross-source consistency.

6.1 Develop Analytical Plan

An analytical plan is a detailed, step-by-step guide for producing official statistics from administrative data. It ensures that the analysis is meaningful, transparent, and reproducible. Key components include:

- **Purpose of analysis.** Clearly state the statistical or policy objectives—such as monitoring service access, informing program delivery, or generating SDG- or CRC-related indicators. Define the target population and unit of analysis (e.g., children, households, schools) and ensure alignment with the information need and business case defined in Step 1.

- **Confirm data suitability.** Drawing on earlier assessments, confirm and document that the dataset is current, complete, and relevant. Check that essential variables (dates, service interactions with administrative data systems, demographic and geographic details) are present and of sufficient quality, and that population coverage includes all necessary subgroups and represents the population of interest.

- **Specify the statistical outputs.** Define and document intended outputs—such as indicators, rates, proportions, averages, or model-based estimates—and clarify calculation methods and reference standards. Identify required disaggregations (age, sex/gender, disability status, region) and confirm that necessary variables are available.

- **Prepare the dataset for analysis.** The dataset must be reviewed and prepared to ensure that variables are accurate and usable. This includes removing duplicate records, checking for inconsistencies, and correcting or flagging obvious data entry errors. Missing values must be handled according to a documented strategy, such as imputation or exclusion. Variables should be formatted consistently (e.g., dates in standard format, classification codes standardized). Where needed, derived variables should be created using reproducible code or procedures.

- **Additional preparation for linked datasets.** For analyses based on linked datasets, additional planning steps are necessary. For example, the plan must document variable origins and address potential conceptual or temporal inconsistencies (e.g., school-year vs. calendar-year reporting).

- **Select analytical methods.** Appropriate analytical techniques must be chosen based on the data and objectives, and

well-documented. Methods must be suitable for the structure and quality of the administrative data and must be documented clearly. As the range of administrative data applications is broad, including survey frame construction, imputation, indicator development, or other analytical files, so are the analytical methods.

- **Quality assurance procedures.** Processes must be established to review and verify the analysis before outputs are published. These include peer review, code checking, result verification, and comparison with historical trends or external benchmarks. All analytical steps must be documented clearly, including assumptions, limitations, and treatment of outliers or anomalies. A metadata file or technical note should accompany the results to ensure methodological transparency.

Guiding questions for developing the analytical plan

1. Have all details of the analytical process been fully specified and documented?

2. Is the analytical plan methodologically sound and aligned with the project's objectives?

Resources

- **Because analytical plans for administrative data–based statistics can vary widely, there is no single standard template; plans should be tailored to the context and objectives of each project.**

6.2 Execute Analytical Plan and Complete Analysis Summary Report

After finalizing the analytical plan, carry out the statistical analysis strictly following the defined objectives, data-preparation steps, and methodological specifications. Executing the analysis in line with the approved plan ensures outputs are rigorous, transparent, and fit for purpose.

Upon completion, prepare an Analytical Summary Report documenting: the analytical process and any deviations from the plan, data quality issues encountered, methods applied, and key notes for interpreting results.

This report strengthens the credibility and reproducibility of the findings, supports quality assurance and peer review, and allows for future replication or updates. It also provides essential context for assessing whether the resulting statistics are fit-for-purpose and suitable for dissemination.


© UNICEF/UN0778873/Paz Soldán

STEP 7:
Develop and Execute Dissemination Strategy

PURPOSE

Dissemination is not simply the release of numbers—it is a core step in building trust, fostering evidence-based decision-making about children, and ensuring transparency in the use of administrative data related to children and adolescents.

Disseminating the results of administrative data–based statistics ensures that the knowledge generated is accessible, useful, and actionable for a wide range of stakeholders, including policymakers, service providers, civil society, the general public, and children and adolescents themselves.

Effective dissemination means presenting results in a manner that is transparent, respectful, inclusive, and protective of children's dignity and privacy. A thoughtful dissemination strategy guarantees that statistics are not only technically sound but also socially meaningful, ethically communicated, and applied to advance child rights and well-being.

7.1 Develop Dissemination Strategy

A dissemination strategy should outline:

- **Child-rights lens.** Identify target audiences, the intended use of statistical outputs, and consider child-friendly dissemination formats to promote inclusion and accessibility.
- **Output formats and dissemination channels.** Specify the types of outputs—such as analytical reports, dashboards, or disaggregated tables—and establish the appropriate dissemination channels for each.
- **Disclosure controls.** Apply strict safeguards to prevent identification of individuals, especially when dealing with small or sensitive subpopulations.

  Note: linked datasets require their own dedicated disclosure control protocols.
- **Metadata and explanatory notes.** Provide comprehensive supporting information to enable correct interpretation and appropriate use of the data.

7.2 Execute Dissemination Strategy

During implementation:

- **Review final outputs.** Ensure privacy, accuracy, and clarity through rigorous internal checks and quality assurance.
- **Ensure accessible dissemination channels.** Select platforms and formats that are relevant and easy to use for audiences focused on children and adolescents.
- **Promote effectively.** Consider outreach activities—such as social media campaigns, webinars, or stakeholder presentations—to maximize visibility and impact.
- **Engage stakeholders.** Encourage responsible use of the data and invite feedback to strengthen future dissemination efforts.

  Note: Not all administrative data–based statistical activities require a full dissemination strategy. For example, creating a survey frame using administrative data may not require public dissemination.

Guiding questions

1. Have all aspects of the dissemination strategy been fully considered and documented?

2. Does the dissemination plan explicitly incorporate a child rights perspective?

Resources

- **Because dissemination strategies for administrative data–based statistics vary widely depending on context, no universal template exists. Strategies should be tailored to the specific project, data type, and intended audience.**

STEP 8:
Institutionalize, Promote Sustainability and Scale

PURPOSE

Institutionalizing the use of administrative data for official statistics—particularly for and about children and adolescents—means moving beyond one-time initiatives toward long-term, system-level integration.

This requires embedding practices into legal, organizational, and technical structures; ensuring sustainable capacity, financing, and partnerships; and expanding successful methods to new domains.

A child-rights perspective reinforces the need for continuity, accountability, and responsiveness to evolving data needs. The steps below outline how to support the long-term use and impact of administrative data–based statistics in this context.

8.1 Institutionalize Practices

Key pathways to embed administrative data use within national statistical systems include:

- **Embedding practices in policies and procedures.** Formalize data-sharing agreements, linkage protocols, analytical frameworks, and dissemination strategies to create institutional continuity.

- **Integrating into national statistical plans.** Include administrative data use and specific objectives related to children's statistics in national strategies for the development of statistics (NSDS), SDG reporting frameworks, and other multi-year plans.

- **Assigning institutional roles.** Clearly define the ongoing responsibilities of NSOs, line ministries, data providers, child-focused agencies, and oversight bodies in managing, updating, and using administrative data.

- **Establishing long-term governance.** Create or strengthen inter-institutional coordination mechanisms—such as child statistics working groups or data governance boards—to ensure oversight, continuity, and child-sensitive guidance.

8.2 Promote Sustainability

To maintain momentum and secure long-term gains:

- **Invest in capacity and infrastructure.** Provide continuous training for technical staff in data management, linkage, analysis, and ethics; plan for staff succession; and maintain secure, scalable systems capable of handling administrative and linked data.

- **Secure sustainable financing.** Allocate and protect funding for administrative data systems and child-related statistical capacity.

- **Foster interoperability.** Strengthen the connections between data systems by promoting common understanding, compatible structures, and shared practices that enable smoother information exchange, integration, and collaboration between the NSO and data providers.

- **Develop reusable tools and templates.** Standardize protocols, metadata models, and analytical plans to reduce duplication, improve efficiency, and ensure consistent data quality over time.

- **Embed a culture of data use.** Foster continuous use of administrative data for monitoring, evaluation, and decision-making at all levels of government and within child- and adolescent-serving institutions.

8.3 Scale Successful Approaches

Scaling involves expanding successful practices in a systematic manner:

- **Replicate proven models.** Apply successful linkage, analysis, and governance approaches to new sectors such as health, education, or child protection.

- **Explore applications in statistical registers.** Use administrative data to build and maintain statistical registers that can serve as foundational sources for surveys, censuses, and longitudinal analyses. Registers can improve efficiency, reduce duplication, and support longitudinal tracking of child and adolescent outcomes when developed with strong legal, methodological, and privacy frameworks.

- **Document and share lessons.** Disseminate tools, case studies, and evaluations to help other countries adopt and adapt best practices.

- **Advance regional collaboration.** Engage in regional platforms to align standards, share innovations, and strengthen child rights–focused data systems across countries.

Guiding questions for institutionalization, sustainability, and scaling

1. Are key practices documented and embedded into official procedures?
2. Are systems and human resources in place for long-term implementation?
3. Have successful pilots or models been adapted for broader application?
4. Is there ongoing engagement with child rights advocates to guide and sustain these efforts?

Resources

- Children's Statistics Working Group – *Diagnosis and Recommendations to Integrate Administrative Records Related to Children Readiness Diagnosis*
  Evaluates the readiness of 21 Latin American and Caribbean NSOs to link and integrate administrative records related to children and adolescents for statistical purposes. Provides a maturity typology of four system types and 11 actionable recommendations across governance, data quality, stakeholder engagement, and IT infrastructure. Emphasizes legal reform, trust-building, and institutional coordination to advance evidence-based policy, SDG monitoring, and child rights (ECLAC & UNICEF, 2025).

- The International Development Bank (IDB) & Federico Segui Stagno – *Generic Statistical Register Business Process Model (GSRBPM)*
  Builds on the GSBPM by focusing on the use of administrative registers. Outlines processes for managing and linking register-based data efficiently, making it highly relevant for countries with or transitioning to large-scale administrative data systems (ECLAC & UNICEF, 2025).

Conclusion

This manual offers a comprehensive and practical roadmap for National Statistical Offices (NSOs) to fully harness the potential of administrative data in producing high-quality statistics about children and adolescents.

By integrating a child-rights perspective, robust technical methodologies, and strong governance safeguards, the manual demonstrates how administrative data can strengthen statistical systems, close critical data gaps, and drive evidence-informed policymaking for children and adolescents.

Through the inclusion of real-world case studies, practical tools, and structured guidance, it supports countries in institutionalizing successful practices, promoting sustainability, and scaling innovations that protect privacy, enhance interoperability, and ensure that the rights and well-being of every child and adolescent are fully visible in official statistics.


© UNICEF/UN0843619/Izquierdo

Appendix 1:

A Brief Introduction to Linkage Methodologies and Software

This appendix provides a concise overview of data linkage methodologies and commonly used software tools relevant for data-linkage projects in official statistics. It aims to support NSOs in identifying appropriate methods and technologies for executing high-quality linkage projects related to children and adolescents.

Overview of Linkage Methodologies

Linkage methodologies vary according to the structure, quality, and availability of identifiers in the data sources. Broadly, they fall into three main categories: deterministic, probabilistic, and rules-based or hybrid. A brief description of each—together with suitable use cases and key limitations—is provided below. Further details are presented in the section Conducting the Record Linkage.

Linkage methodology	Description	Appropriate use cases	Limitations
Deterministic Linkage	A method that uses exact matching on identifiers such as national identification numbers to connect records across datasets. Example: linking hospitalization records and child-protection records using a child's national ID number.	• Unique, high-quality identifiers exist. • A low tolerance for false matches is required. • A fast, straightforward method is preferred.	• Cannot accommodate minor discrepancies or typographical errors.
Probabilistic Linkage	A method that matches records across datasets by estimating the likelihood that different identifiers refer to the same entity, allowing for minor variations or missing data. Example: linking a child's hospitalization records to social-service records using a combination of name, sex, birth year, and address.	• Unique identifiers are missing or unreliable. • Indirect identifiers are inconsistent. • Maximizing the number of linked records is important, up to a defined quality threshold.	• Methodologically complex and potentially computationally demanding. • Requires explicit setting of match-quality thresholds.
Rules-based and Hybrid Linkage	Combinations deterministic and/or probabilistic methods can be applied. For example, in deterministic linkage, a rules-based mixed-match approach can be implemented where different combinations of variables are used to create unique identifiers for matching. For example, FirstName+LastName+DateofBirth, LastName+Address. Additionally, hybrid linkage approaches may include a combination of deterministic and probabilistic methodologies.	• Both high-quality and lower-quality unique or indirect identifiers are present. • A staged strategy is needed to maximize linkage rates while maintaining precision.	• Methodologically complex and may be less transparent to non-specialists.

Assessing Linkage Feasibility and Discriminatory Power

The feasibility of record linkage depends largely on the quality and discriminatory power of identifiers—variables that distinguish individuals across datasets. Typical identifiers include:

1) Unique (direct) identifiers (e.g., citizen registration number, birth registration number)
2) Indirect identifiers (e.g., name, date of birth, sex, address)

A citizen registration number may often be considered an ideal identifier in the data linking process, since it is a unique number; however, a citizen registration number may not be available for the entire target population; for example, some countries, children are not assigned a citizen registration number.

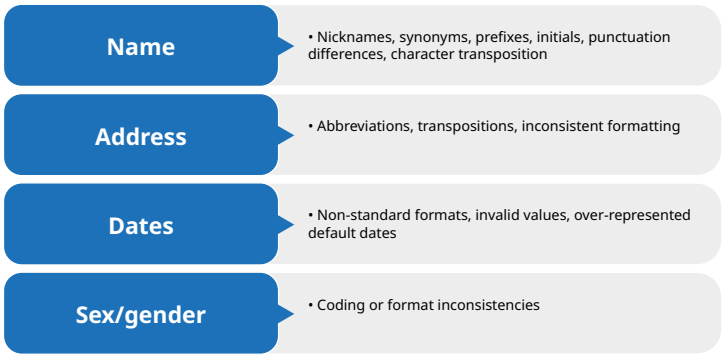
The quality of an identifying variable may include its completeness and accuracy. If the existence and the quality of the identifying variables are not sufficient, then linkage is not possible. The discriminatory power of an identifying variable refers to its ability to differentiate between different entities (and in this case, individuals). To take a simple example, *Month of Birth* which only has 12 possible values has less discriminatory power than a full address, which has many more possible values.

Before initiating a linkage project, which may involve resource-intensive activities such as developing data sharing agreements and data acquisition procedures, it is important to assess whether a linkage is possible given the quality and completeness of identifiers and their associated discriminatory power. This discriminatory power cannot be evaluated solely by the presence of identifiers or the amount of missingness for these identifiers; however, it can be quantified using Information Theory and the concept of Shannon's entropy, which measures the unpredictability associated with a random variable. Other approaches, such as that developed by Cooke et al., (Cook, Olson & Dean, 2001) compute match weights for each identifier and sum them across all variables to determine if the total exceeds a threshold corresponding to the desired match confidence.

Data Cleaning and Standardization

Because administrative data are collected primarily for administrative purposes and then repurposed for statistical use, these data may not be collected with the same level of precision as a survey or census conducted by an NSO. Therefore, personal identifiers collected in administrative systems may be subject to typographical and data entry errors.

Below are common examples of variations found in personal identifiers used for data linkage:



Standard data-processing techniques in statistical software—such as case normalization, trimming, or pattern matching—facilitate cleaning and standardization.

Additional string-matching methods (e.g., phonetic encoding algorithms) can help reconcile spelling variations and improve match quality before record linkage.

Conducting the Linkage

Deterministic Linkage

Deterministic record linkage applies explicit matching rules to link records across different datasets, assuming that unique identifiers exist and can be used definitively to link records.

Example: Two administrative sources are considered:

Dataset A - Hospitalization data				
Health registration number	**National Identification Number**	**First name**	**Last name**	**Birth date**
2002	00001	Ana	Ruiz	28/01/2012
2003	00002	Jorge	Lopez	15/03/2020
2004	00003	Maria	Hernandez	17/11/2018

Dataset B - Child protection data				
Child protection registration number	**National Identification Number**	**First name**	**Last name**	**Birth date**
3005	00001	Ane	Ruiz	28/01/2012
3006	00002	Jorge	Lope	15/03/2020
3007	00004	Julio	Torres	11/04/2009

A simple deterministic rule might be:

> IF National Identification Number in hospitalization data = National Identification Number in child protection data
>
> THEN records are a match
>
> ELSE records are not a match

Applying this rule produces a set of matches and non-matches based on exact agreement of the identifier.

National Identification Number	Health registration number	Child protection registration number	Outcome
00001	2002	3005	Match
00002	2003	3006	Match
00003	2004	-	Nonmatch
00004	-	3007	Nonmatch

Probabilistic Linkage

Probabilistic record linkage is an evidence-based approach to record linkage that compares indirect identifiers for possible pairs across two data sets and estimates the likelihood that the two records are referring to the same entity. To demonstrate this, consider the same hospitalization and child-protection datasets from above, but assume that the National Identification Number does not exist. Using the indirect identifiers of First Name, Last Name, and Birth Date for the first record in the hospitalization data (Health registration number 2002) and (Child protection number 3005) in probabilistic record linkage may give an overall probability that the records are a match.

Hospitalization data				Child protection data		
First name	Last name	Birth date	Match probability	First name	Last name	Birth date
Ana	Ruiz	28/01/2012	90%	Ane	Ruiz	28/01/2012

The overall match probability combines the prior probability of a match with agreement weights for each field. Evidence thresholds can then be applied to classify pairs as matches, non-matches, or potential matches requiring clerical review.

Rules-based and Hybrid Linkage

Administrative datasets often lack perfect identifiers or contain inconsistencies. Rules-based or hybrid linkage strategies address these challenges by combining deterministic and probabilistic techniques.

> The Fellegi–Sunter method is a probabilistic record linkage approach used to identify matching records across datasets when a unique identifier is unavailable. It compares pairs of records based on common fields (e.g., name, date of birth), calculates the probability of agreement assuming both match and non-match scenarios, and assigns weights accordingly. These weights are summed into a total match score, and records are classified as matches, non-matches, or potential matches based on predefined thresholds. The method accounts for uncertainty and data quality issues, making it widely used in linking administrative and survey data (William E. Winkler, 1999).

Rules-based Linkage (Mix-Match Strategy)

Rules-based linkage expands deterministic logic by defining superkeys—combinations of indirect identifiers used sequentially for matching. Often referred to as mix-match, multiple variable combinations are employed to create matching rules across records, which are tested in a pre-determined order.

Consider again the hospitalization and child-protection datasets, but without national identification numbers and with address information:

Dataset A - Hospitalization data				
Health registration number	**First name**	**Last name**	**Birth date**	**Address**
2002	Ana	Ruiz	28/01/2012	Calle 45 No. 123, Bogotá
2003	Jorge	Lopez	15/03/2020	Av. Central 10, Medellín
2004	Maria	Hernandez	17/11/2018	Calle 12 No. 5-20, Cali

Dataset B - Child protection data				
Child protection registration number	**First name**	**Last name**	**Birth date**	**Address**
3005	Ane	Ruiz	28/01/2012	Calle 45 No. 123, Bogotá
3006	Jorge	Lope	15/03/2020	Avenida Central 10, Medellín
3007	Julio	Torres	11/04/2009	Calle 12 No. 5-20, Cali

Rules could be applied as follows:

Rule 1: First name + Last name + Birth date

Apply the rule:

IF First name + Last name + Birth date match exactly
THEN records are a match

Results:

3005 vs. 2002: Ana ≠ Ane ---> No match
3006 vs. 2003: Lopez ≠ Lope ---> No match
No records match using Rule 1

Rule 2: Last name + Address

Apply the rule:

IF Last name + Address match exactly
THEN records are a match

Results:

3005 vs. 2002: Last name = Ruiz (exact), address = Calle 45 No. 123, Bogotá ---> Exact match
3006 vs. 2003: Lopez ≠ Lope ---> No match
1 record matches using Rule 2

This staged approach illustrates how sequential rules can capture different aspects of potential linkage.

Hybrid Linkage

Hybrid linkage strategies also address limitations such as missing unique identifiers or inconsistencies in how data are recorded by integrating deterministic and probabilistic methods. Typically, a hybrid approach involves:

1. Applying deterministic rules to identify high-confidence, exact matches.
2. Using probabilistic methods to evaluate remaining record pairs, assigning weights based on the likelihood that they refer to the same entity.
3. Setting thresholds to automatically accept or reject matches, with optional manual review for borderline cases.

Hybrid linkage offers a practical solution for real-world administrative data, where identifiers may be inconsistent, incomplete, or recorded inconsistently. It combines the precision of deterministic logic with the flexibility and tolerance of probabilistic scoring, making it an ideal option for complex linkage environments.

Additional methods and innovations

Beyond these core approaches, other techniques can enhance or complement linkage:

- Expectation–Maximization (EM) algorithms improve probabilistic estimates when data are missing or incomplete.
- Bayesian linkage methods incorporate prior knowledge and uncertainty directly into the matching process.
- Privacy-preserving record linkage (PPRL) enables cross-dataset matching while maintaining confidentiality through encryption or secure multi-party computation.
- Machine-learning-based approaches offer adaptive, scalable solutions for large or complex linkage tasks, increasing automation and match accuracy.

Assessing Linkage Errors and Validation

In data linkage, Type I and Type II errors describe the incorrect classification of record pairs as matches or non-matches:

- Type I (false positive) – two records are incorrectly linked. The algorithm mistakenly identifies them as a match when they are, in fact, from different entities. This error results in records being falsely connected, leading to inaccurate conclusions about linked datasets.
- Type II (false negative) – two records that truly represent the same entity are not linked. The algorithm incorrectly identifies them as non-matches, causing the true match to be missed.

Both error types compromise the accuracy and reliability of the linked dataset. Managing this trade-off is essential when designing linkage algorithms.

A standard confusion matrix can be used to evaluate performance:

		Algorithm Output	
True Match Status		Match	Nonmatch
	Match	A	B
	Nonmatch	C	D

"A" refers to all true matches that are accurately identified as matches.
"B" includes all true matches that are mistakenly classified as nonmatches.
"C" consists of all true nonmatches that are incorrectly identified as matches.
"D" represents all true nonmatches that are correctly classified as nonmatches.

From this matrix, key quality metrics are derived:

Sensitivity = A / (A + B).
Measures the ability to correctly identify true matches.

Specificity = D / (C + D).
Measures the ability to correctly identify true non-matches.

Positive Predictive Value (PPV) = A / (A + C).
Indicates the proportion of record pairs classified as matches that are actually true matches.

Negative Predictive Value (NPV) = D / (B + D).
Indicates the proportion of record pairs classified as nonmatches that are truly nonmatching.

Linkage Software Tools

Multiple open-source tools support deterministic and probabilistic record linkage. The table below summarizes widely used options and key references.

Tool	Summary
fastLink	fastLink is an open-source probabilistic record linkage library in R available in CRAN. It implements a Fellegi-Sunter probabilistic record linkage model following the methods described in Enamorado, Fifield, and Imai "Using a Probabilistic Model to Assist Merging of Large-scale Administrative Records" <doi:10.1017/S0003055418000783> and is available at <https://imai.fas.harvard.edu/research/linkage.html> (Enamorado, Fifield & Imai, 2023). The library is available at https://doi.org/10.32614/CRAN.package.fastLink
RELAIS	RELAIS (Record Linkage At Istat) is a record linkage toolkit developed by the Italian National Institute of Statistics, implementing the deterministic and probabilistic (Fellegi-Sunter) record linkage model. RELIAS has been implemented in Java and R and has a MYSQL database architecture. The toolkit is available at: https://www.istat.it/en/classifications-and-tools/methods-and-software-of-the-statistical-process/process-phase/data-integration/relais/ (ISTAT, 2024).
Splink	Splink is an open-source probabilistic record linkage library in Python developed by the UK Ministry of Justice based on the Fellegi-Sunter probabilistic record linkage model available at: https://moj-analytical-services.github.io/splink/ (Linacre et al., 2022). Further information can be found in the related tutorial: https://www.robinlinacre.com/intro_to_probabilistic_linkage/ (Linacre, 2023).
Note: When applying phonetic encoding (e.g., Soundex, Metaphone), ensure the algorithm or package is appropriate for the language and regional naming conventions of the data.	

Appendix 2:

A Methodological Protocol for Data Linkage

1. Project Overview

Project Title: _____

Contact Person: _____

Date of Preparation: _____

2. Objectives and Use

Statistical Objective:

_____

Target Population:

_____

Expected Outputs:

_____

Intended Use (check all that apply):

❍ Direct tabulations and analysis ❍ Editing and imputation ❍ Indirect use in imputation ❍ Replacement of data collection ❍ Survey frames ❍ Data source evaluation ❍ Other: _____

3. Source Datasets and Metadata

Dataset Name	Data provider	Years Covered	Population Represented	Purpose of Collection	Identifiers Available

Attach detailed metadata files, record layouts, and data dictionaries as annexes.

4. Legal and Ethical Considerations

Legal Basis (Statistical Act, Sector-specific law):

_____

Memorandum of Understanding:

_____

Data Governance and Management:

_____

5. Linkage Design

5.1 Linkage Methodology:

❍ Deterministic ❍ Probabilistic ❍ Rule-based ❍ Hybrid

5.2 Identifiers Used for Matching:

Field Name	Identifier Type (Direct/Indirect)	Preprocessing Applied	Discriminatory Power

5.3 Matching Strategy:

_____

6. Data Cleaning and Standardization

Planned Cleaning Activities (check all that apply):

❍ Remove duplicate records

❍ Handle missing values: _____

❍ Normalize formats (names, dates, addresses)

❍ Code harmonization across datasets

Phonetic Encoding Methods Used: _____

7. Software and Tools

Tool Name	Type / Language	Purpose	Configuration Notes

8. Linkage Quality and Validation Plan

Validation Strategy:

_____

Accuracy Metrics:

Sensitivity =

Specificity =

Positive predictive Value =

Negative predictive value =

9. Output Description

Name of Final Linked Dataset: _____

File Format: ❍ CSV ❍ Parquet ❍ SQL ❍ Other: _____

Variable List or Schema (attach as annex):

Weighting or Statistical Adjustment Applied: ❍ Yes ❍ No

If yes, describe method: _____

10. Risk Assessment

Identified Risk	Proposed Mitigation Strategy

11. Dissemination and Retention Policy

Planned Release Type:

❍ Public Use File ❍ Internal Analytical File ❍ Restricted Research Access

Retention Schedule:

- Raw source data: _____ months

- Final linked dataset: _____ months

Data Custodian:

12. Sign-Off and Approval

Name	Title	Institution	Date	Signature

Glossary

Term	Definition
Administrative Data	As defined in UNICEF's Administrative Data Maturity Model (ADaMM), administrative data are collected through the routine provision of a service—such as health, education, or social welfare—most often by a government provider, but potentially also through private sector services. These data are essential for improving service provision, identifying and acting on development goals and targets, and reporting against international commitments such as the Sustainable Development Goals (UNICEF, 2022a).
Administrative Data Maturity Model (ADaMM)	A model developed by UNICEF to assess and strengthen the readiness and effectiveness of administrative data systems (UNICEF, 2022a).
Capture-Recapture Method	A statistical technique used to estimate population size by linking multiple datasets and identifying overlapping records.
Child Rights Perspective	An approach that integrates the principles of the United Nations Convention on the Rights of the Child into statistical data collection, analysis, and dissemination.
Conceptual Framework and Guidance for Integrating a Children's Rights Perspective into Administrative Data-Based Statistics	A UNICEF-developed framework providing six core recommendations and eight guiding considerations to help administrative data producers and users support evidence-informed decisions for and about children, ultimately strengthening the realization of child rights.
Convention on the Rights of the Child (CRC)	The United Nations Convention on the Rights of the Child (UN CRC) is the most widely ratified human rights treaty in history. Adopted in 1989 (OHCHR, 1989), it recognizes specific rights for children and provides international standards and benchmarks for fulfilling those rights, helping transform the lives of children worldwide.
Cross-Sectoral Statistics	The practice of combining data from multiple sectors—such as health, education, and social welfare—to produce comprehensive child-related statistics.
Data Governance	A framework of policies, standards, and procedures that ensures data are managed ethically, securely, and in compliance with legal and institutional requirements.
Data Integration	The process of bringing together data from multiple sources to create a holistic dataset, including merging, transforming, cleaning, and aligning information—beyond record linkage alone.
Data Linkage	The process of combining records from different datasets based on common identifiers at the record level to enhance data analysis and statistical accuracy.

Term	Definition
Deterministic Linkage	A record linkage method that uses exact matching on identifiers—such as national identification numbers—to connect records across datasets. A mix-and-match approach in this context might involve using different combinations of identifiers (for example, concatenating name and birth date information).
Discriminatory Power	Discriminatory power refers to how well a variable can distinguish between different individuals in a dataset during the process of record linkage.
Five Safes Framework	A risk-based approach to secure data access, encompassing safe people, safe projects, safe settings, safe data, and safe outputs (Desai, Ritchie & Welpton, 2016).
Generic Statistical Business Process Model (GSBPM)	A standard framework developed by the United Nations Economic Commission for Europe (UNECE) to describe and manage statistical production processes (UNECE, 2014).
Generic Statistical Register Business Process Model (GSRBPM)	An extension of the GSBPM that focuses specifically on the use of administrative registers in official statistics (IADB & Segui Stagno, 2016).
Hard-to-Reach Populations	Groups that may be underrepresented in statistical estimates due to barriers in administrative data capture, such as migrants, displaced persons, or children in vulnerable situations.
Linked Analytical File Environment	A statistical framework that supports the integration of data from multiple sources into a unified analytical dataset.
National Statistical System	The national statistical system (NSS) is the group of organizations within a country, including the national statistical office (NSO) that jointly collect, process, and disseminate official statistics on behalf of the national government. This plays a key role with administrative data.
Probabilistic Linkage	A statistical technique that matches records across datasets by estimating the likelihood that different identifiers refer to the same entity, allowing for minor variations or missing data.
Register-Based Statistical System	A statistical framework that uses administrative registers as primary data sources for producing official statistics.
Rules-based and Hybrid Linkage	Variants of deterministic and probabilistic record linkage approaches. For example, in deterministic linkage, a mix-match approach can be implemented where different combinations of variables are used to create unique identifiers for matching—for example, FirstName+LastName+DateofBirth, LastName+Address. Additionally, hybrid linkage approaches may include a combination of deterministic and probabilistic methodologies.
Separation Principle	A data security approach that requires personally identifiable information to be stored separately from analytical datasets to strengthen privacy protection.
Sustainable Development Goals (SDGs)	A set of global objectives established by the United Nations in 2015 to promote social, economic, and environmental sustainability by 2030.

References

Brazilian Institute of Geography and Statistics (IBGE). *System of Vital Statistics,* IBGE, 2025, Rio de Janeiro, <www.ibge.gov.br/en/statistics/social/population/26259-under-registration-of-births-and-deaths-estimates.html?lang=en-GB>.

Carletto, C., H. Chen, and F. Perucci, 'Positioning Household Surveys for the Next Decade', Statistical Journal of the IAOS, 2022, pp. 923-946.
<https://doi.org/10.3233/SJI-220042>.

Cook, L.J., L.M. Olson and J.M. Dean, 'Probabilistic record linkage: Relationships between file sizes, identifiers, and match weights', *Methods of Information in Medicine,* vol. 40, July 2001, pp. 196–203, PMID 11501632.

Desai, T., F. Ritchie and R. Welpton. *Five Safes: Designing Data Access for Research,* University of the West of England, 2016, Bristol, <https://doi.org/10.13140/RG.2.1.3661.1604>.

Economic Commission for Latin America and the Caribbean (ECLAC) and United Nations Children's Fund (UNICEF), *Diagnosis and Recommendations to Integrate Administrative Records Related to Children: Use of Administrative Records for Better Statistics on Children and Adolescents,* ECLAC and UNICEF, 2025, Santiago de Chile, <www.cepal.org/en/publications/81697-diagnosis-and-recommendations-integrate-administrative-records-related-children>.

Economic Commission for Latin America and the Caribbean (ECLAC), *Generic Law on Official Statistics for Latin America.* ECLAC, 2020, Santiago, <https://repositorio.cepal.org/server/api/core/bitstreams/7c04bd8d-fa22-4b02-a50e-11cca1347aa4/content>.

———, *Code of Good Practice in Statistics for Latin America and the Caribbean 2023,* ECLAC, 2024, Santiago, <www.cepal.org/en/publications/80478-code-good-practice-statistics-latin-america-and-caribbean-2023>.

Enamorado, T., B. Fifield and K. Imai, *fastLink: Fast Probabilistic Record Linkage (version 0.6.1) [software].* CRAN, 2023, Vienna, <https://doi.org/10.32614/CRAN.package.fastLink>.

Hancioglu, A. and F. Arnold. 'Measuring Coverage in MNCH: Tracking Progress in Health for Women and Children Using DHS and MICS Household Surveys', *PLOS Medicine,* vol. 10, no. 5, 7 May 2013, p. e1001391, <https://doi.org/10.1371/journal.pmed.1001391>.

Instituto Nacional de Estadística (INE) (Uruguay), *Marco conceptual y metodológico del Sistema Integrado de Registros Estadísticos y Encuestas – SIREE,* INE, 2021, Montevideo, <https://www5.ine.gub.uy/documents/Informes/Marco%20conceptual%20y%20metodol%C3%B3gico%20del%20SIREE-DW-INE.pdf>.

———, *Metodología del Censo Combinado 2023,* INE, 2024, Montevideo, <https://www5.ine.gub.uy/documents/CENSO%202023/Metodolog%C3%ADa%20del%20Censo%20Combinado%202023.pdf>.

Instituto Nacional de Estadística e Informática (INEI) (Peru), *Mapa de pobreza monetaria provincial y distrital 2018,* INEI, 2020, Lima, <www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1718/Libro.pdf>.

Instituto Nacional de Estadísticas (INE) (Chile) and Servicio Nacional de Migraciones (SERMIG), 2024. *Documento metodológico de la estimación de personas extranjeras residentes en Chile al 31 de diciembre de 2023,* INE and SERMIG, 2024, Santiago, <https://serviciomigraciones.cl/wp-content/uploads/estudios/Estimaciones/Estimacion-extranjeros-2023.zip>.

Instituto Nacional de Estadísticas (INE) (Chile) and United Nations Children's Fund (UNICEF). *Niñez y adolescencia migrante en Chile: Estimación de población,* INE and UNICEF, 2023, Santiago, <www.unicef.org/chile/media/8721/file/Informe%20migrantes%20INE%20V15.pdf>.

Inter-American Development Bank (IADB) and F. Segui Stagno, *Project on Population and Real Estate Statistics Based on the Use of Official Administrative Registers in the Andean Community,* IADB, 2016, Washington, DC.

———, A. Wallgren and B. Wallgren, *Toward an Integrated Statistical System Based on Registers,* IADB, 2021, New York, <https://publications.iadb.org/en/publications/english/viewer/Toward-an-Integrated-Statistical-System-Based-on-Registers.pdf>.

International Data Alliance for Children on the Move (IDAC). *Manual on Child-Specific Data Capacity Strengthening on Children on the Move,* IDAC, 2023, Piraeus, <https://elstat-outsourcers.statistics.gr/UNICEF_MANUAL_2023%20booklet%20HYPERLINKS%20(1).pdf>.

Istituto Nazionale di Statistica (ISTAT) 'RELAIS (REcord Linkage At Istat)', ISTAT, 2024, Roma, <www.istat.it/en/classifications-and-tools/methods-and-software-of-the-statistical-process/process-phase/data-integration/relais/>.

Linacre, R., et al. 'Splink: Free software for probabilistic record linkage at scale', International Journal of Population Data Science, vol. 7, no. 3, p. 23, 2022, <https://moj-analytical-services.github.io/splink/>.

Linacre, R. *An Interactive Introduction to the Fellegi-Sunter Model for Data Linkage/Deduplication* [tutorial], Robin Linacre, 2023, London, <www.robinlinacre.com/intro_to_probabilistic_linkage/>.

Milne, B., et al., 'Data Resource Profile: The New Zealand Integrated Data Infrastructure (IDI)', *International Journal of Epidemiology,* vol. 48, no. 3, June 2016, pp. 677–677e. <https://doi.org/10.1093/ije/dyz014>.

Ministerio de Salud Pública (Uruguay) and United Nations Children's Fund (UNICEF). *Análisis de datos de desarrollo infantil en Uruguay.* El aporte de las metas asistenciales del SNIS. Ministerio de Salud Pública and UNICEF, 2023, Montevideo, <www.gub.uy/ministerio-salud-publica/comunicacion/publicaciones/analisis-datos-desarrollo-infantil-uruguay-aporte-metas-asistenciales>.

Sierra Castillo, J.D., et al., 'Children Administrative Data Linkage and Integration: Readiness Diagnosis for National Statistical Offices in Countries in Latin America and the Caribbean', *Statistical Journal of the IAOS,* vol. 41, no. 1, 1 March 2025, pp. 28–37. <https://doi.org/10.1177/18747655251315875>.

United Nations (UN). Department of Economic and Social Affairs, *Methodology and Evaluation of Population Registers and Similar Systems.* UN, 1969, New York, <https://unstats.un.org/unsd/publication/SeriesF/Seriesf_15e.pdf>.

———. Department of Economic and Social Affairs, Economic and Social Development Home, *Fundamental Principles of Official Statistics,* UN, 2014, New York, <https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>.

———. Department of Economic and Social Affairs, Statistics Division, *United Nations National Quality Assurance Frameworks Manual for Official Statistics,* UN, 2019, New York, <https://unstats.un.org/UNSDWebsite/nqaf/Attachment%20B-UNNAQF%20Manual%20in%20PDF.pdf>.

———. Department of Economic and Social Affairs, Population Division, *International Migration 2020. Highlights,* UN, 2020, New York, <www.un.org/en/desa/international-migration-2020-highlights>.

———. Department of Economic and Social Affairs, Statistics, Data4Now: *Training in Record Linkage for National Statistical Office Colombia (DANE),* UN, 2022a, Bogotá, <https://unstats.un.org/UNSDWebsite/capacity-development/data-for-now/events-details/D4N-record-linkage>. P.61

———. Statistics Division, Memorandum of Understanding. *A guide and template for improving coordination and cooperation within the national statistical system and increasing access to administrative data for statistics production* [draft], UN, 2022b, New York, <https://unstats.un.org/capacity-development/admin-data/docs/mou-guide-and-template>.pdf. P.64

———. Statistics Division, *Toolkit on Using Small Area Estimation for SDGs.* UN, 2023, New York, <www.unsdglearn.org/resources/toolkit-on-using-small-area-estimation-for-sdgs/>.

United Nations Children's Fund (UNICEF), *Progress for Every Child in the SDG Era.* UNICEF, 2018, New York, <https://data.unicef.org/wp-content/uploads/2018/03/Progress_for_Every_Child_V4.pdf>.

———, *Gender Statistics and Administrative Data Systems. An Annotated Resource Guide* UNICEF, 2020, New York, <https://data.unicef.org/wp-content/uploads/2020/07/Admin-and-gender-data-annotated-resource-guide.pdf>.

United Nations Children's Fund (UNICEF), *Progress on Children's Well-Being: Centering Child Rights in the 2030 Agenda,* UNICEF, 2023, New York, <https://data.unicef.org/resources/sdg-report-2023/>.

———, *The Administrative Data Maturity Model (ADaMM),* UNICEF, 2022a, New York, <https://data.unicef.org/resources/the-administrative-data-maturity-model-adamm/>.

———, *Strengthening Administrative Data Systems to Close Gender Data Gaps,* UNICEF, 2022b, New York, <https://data.unicef.org/resources/strengthening-administrative-data-systems-to-close-gender-data-gaps/>.

———, *Conceptual Framework and Guidance for Integrating a Children's Perspective into Administrative Data-Based Statistics*, UNICEF, 2025a, New York, <https://www.unicef.org/lac/en/reports/conceptual-framework-guidance-integrating-childrens-rights-perspective-data-based-statistics>.

———, *Using Data to Achieve the Sustainable Development Goals (SDGs) for Children,* UNICEF, 2025b, New York, <https://data.unicef.org/sdgs/>.

United Nations Children's Fund (UNICEF). Eastern and Southern Africa Regional Office and The Partnership in Statistics for Development in the 21st Century (PARIS21), *Toolkit for Mainstreaming Child Rights into Strategic Planning for Official Statistics,* UNICEF and PARIS21, Nairobi, 2024, <www.unicef.org/esa/documents/toolkit-mainstreaming>.

United Nations Economic Commission for Europe (UNECE), 2011. *Using Administrative and Secondary Sources for Official Statistics. A Handbook of Principles and Practices,* UNECE, 2011, New York and Geneva, <https://unece.org/DAM/stats/publications/Using_Administrative_Sources_Final_for_web.pdf>.

———, *Generic Statistical Business Process Model. High-level Workshop on Modernization of Official Statistics* [workshop], UNECE, 2014, Nizhni Novgorod, <https://unece.org/fileadmin/DAM/stats/documents/fund.principles/2014/1-Generic_Statistical_Business_Process_Model_EN.pdf>.

———, *Terms of Reference for the Task Force on Hard-to-Reach Groups in Administrative Sources,* UNECE, 2023, Cardiff, <https://unece.org/sites/default/files/2023-11/07_TF_Hard%20to%20reach%20groups%20in%20admin%20sources_ToR_approved.pdf>.

United Nations High Commissioner for Refugees (UNHCR), *Global Trends. Forced Displacement in 2024.* UNHCR, 2025, Copenhagen, <www.unhcr.org/global-trends-report-2024>.

United Nations Statistical Institute for Asia and the Pacific (UN SIAP), *Using Administrative Data to Produce Official Statistics* [online course], UN SIAP, 2025, Chiba, <https://siap-elearning.org/enrol/index.php?id=260>.

United Nations Women (UN Women), *Guidance and Toolkit for Quality Assessment of Administrative Data for Official Statistics. Eastern and Southern Africa. August 2022,* UN Women, 2022, New York, <https://unstats.un.org/capacity-development/admin-data-toolkit/>.

United Nations. Office of the High Commissioner for Human Rights (OHCHR), *Convention on the Rights of the Child,* OHCHR, 1989, Geneva, <www.ohchr.org/en/instruments-mechanisms/international-human-rights-instruments/convention-rights-child>.

Wallgren, A. and B. Wallgren. *Register-Based Statistics: Registers and the National Statistical System.* John Wiley & Sons, New Jersey, 2022, <https://doi.org/10.1002/9781119632672>.

Winkler, W.E., *The State of Record Linkage and Current Research Problems,* Bureau of the Census, 1999, Washington, DC, <https://www.census.gov/content/dam/Census/library/working-papers/1999/adrm/rr99-04.pdf>.

UNICEF
Latin America and Caribbean Regional Office
Building 102, Alberto Tejada St., City of Knowledge
PO Box: 0843-03045
Panama, Republic of Panama
Tel: +507 301-7400
www.unicef.org/lac
uniceflac@unicef.org

United Nations
ECLAC

Statistical Conference of the Americas of ECLAC

Instituto Nacional de Estadísticas · Chile

Uruguay INE Instituto Nacional de Estadística

unicef for every child