Using Administrative Data in a Traditional Census

Select Topics in International Censuses[1]

Released May 2022

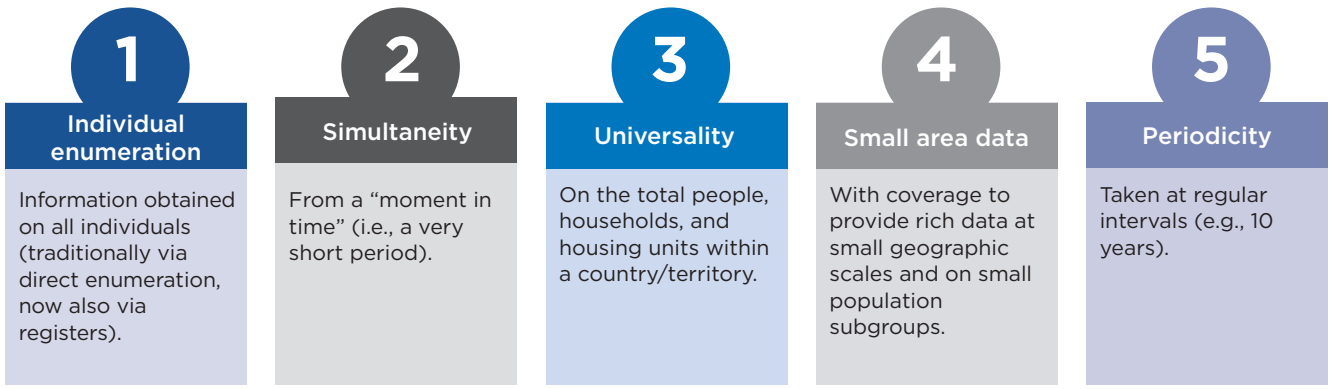
INTRODUCTION

Traditionally, population and housing censuses have been based on the essential features of individual enumeration, simultaneity, universality, small area data, and periodicity (Figure 1). Over the last few decades, however, new methodologies have expanded this definition. These new methods use administrative data to either supplant or support field enumeration for household-based population and housing censuses (UNECE, 2015). As National Statistical Offices (NSOs) consider redefining population and housing censuses in view of these methodologies, it remains imperative that they adhere to the essential features as closely as possible.

Administrative data sources (ADS) are data holdings that contain information collected primarily for administrative (not research or statistical) purposes (UNECE, 2018). These data are collected by government departments, public bodies, and other organizations for the purposes of registration, transaction, and record-keeping of individuals, businesses, and households. The use of these data in censuses varies; some countries use them to supplement a traditional census, others to conduct a combined census,

[1] This technical note is part of a series on Select Topics in International Censuses (STIC) exploring matters of interest to the international statistical community. The U.S. Census Bureau helps countries improve their national statistical systems by engaging in capacity building to enhance statistical competencies in sustainable ways. Any views expressed are those of the author(s) and not necessarily those of the Census Bureau.

Figure 1.
The Five Essential Features of a Traditional Census



Source: United Nations Economic Commission for Europe, 2015 and 2018.

U.S. Department of Commerce
U.S. CENSUS BUREAU
census.gov

United States Agency for International Development
www.usaid.gov

United Nations Population Fund
www.unfpa.org

and still others to construct a fully register-based census (Box 1). For the 2020 census round, of the 113 responding non-European/North American nations, 12.4 percent reported that they used or will use ADS in their country's census (UNSD, 2020).

This technical note focuses on the various types of ADS that may be used to supplement a traditional census, a framework for integrating ADS, and the various challenges and opportunities ADS use presents. In so doing, this note will help NSOs to incorporate administrative data with the population and housing census to produce higher quality census data.

GENERIC TYPES OF ADS

Many government agencies collect, maintain, and store various types of ADS that are generated in the normal course of their operations. These can potentially be used both in the preenumeration as well as post-enumeration phases of a census. Table 1 presents typical forms of ADS (a variety of registers). Registers are structured lists of units (e.g., individuals, businesses), with their associated attributes (e.g., gender, industry type), which are regularly updated (e.g., annually) (UNECE, 2011).

Box 1.

Three Types of Censuses

Traditional
The direct and exhaustive count of all individuals, households, and housing units, and simultaneous collection of information on their characteristics.

Combined
Combine administrative data sources (ADS) with direct field enumeration.

Register-Based
Replace traditional data collection with ADS across various registers (i.e., no field enumeration).

Source: United Nations Economic Commission for Europe, 2018.

These officially collected and maintained data sources encompass a wide range of government functions and can improve an NSOs census data (UNECE, 2018 and 2021).

Table 1.

The Various Generic Types of Administrative Data Sources (ADS)

ADS Type	Description	Used to Derive
Population registers	Collected to provide a frame of people resident in a country (e.g., date and place of birth, gender, date of arrival/departure, citizenship, and marital status).	Population and population movement at national and subnational levels.
Social security registers	Collected for national contributory social insurance program administration and benefits and allowance allocation.	Sex, age, marital status, income, and disability/health status.
Tax registers	Collected for the administration and collection of income and other taxes.	Employment status, occupation, place of work, and place of usual residence.
Business registers	Collected to monitor and store information on incorporated entities.	Economic topics, especially related to industry.
Employment registers	Collected to generate employment and unemployment figures.	Economic activity, employment status, occupation, and hours/place of work.
Education registers	Collected to maintain registration and performance of students and the employment of teaching staff.	Attendance, literacy, and highest level of educational attainment.
Health registers	Collected to ensure the provision of care (i.e., insurance-based private industry or the national health system).	Health status, level of disability, and parity.
Building and dwelling registers	Collected by local land/property valuation agencies and local authorities responsible for urban planning and housing policies.	Ownership, size, and physical construction of housing units.
Address registers	Collected to generate a list of places where people might live, assign a location, and link to the population register. (Note: related to building and dwelling registers on a country-by-country basis).	Households and housing units.

Source: United Nations Economic Commission for Europe, 2018 and 2021; United Nations Statistics Division, 2022 draft.

THE ADVANTAGES OF USING ADS IN A TRADITIONAL CENSUS

Many NSOs are increasingly interested in strengthening their capacity to leverage ADS (Box 2). This decision is motivated by several key advantages gained by using ADS in a traditional census.

Lower Per Capita Costs, Greater Efficiency

The most expensive census is a traditional census, where all data about people, households, and housing units are collected through house-to-house field operations with trained enumerators. In many countries, traditional censuses with full door-to-door enumeration can run a cost of about two times the annual budget of the NSO. NSOs can reduce the time and costs of conducting censuses by taking advantage of data already available in ADS. Practice shows that by reducing the scope of field enumeration through ADS integration, NSOs can reduce costs by 22 percent compared to a traditional census. The greatest opportunity for cost reduction comes with fully register-based censuses. In the 2010 round, the median per capita cost of conducting a traditional census was $5.57 ($7.04 in 2022 inflation-adjusted dollars), while the median cost of conducting a fully register-based census was $0.24 ($0.30) (UNECE, 2014 and 2018).

Fewer Nonresponse Issues and Reduced Response Burden

When ADS provide comprehensive coverage of the entire population, these data can be used to establish and update the response frame, pre-fill questionnaires, fill in missing data due to nonresponse from a particular household, and count people who otherwise did not respond (Mule, 2021; U.S. Census Bureau, 2021). ADS (e.g., previous years' tax returns) contain information more accurate than those an NSO might produce from imputation—a statistical technique used to fill in missing data based on other available data. As ADS are integrated into censuses, respondents will need to report less information on themselves, their households, and their housing units, thus reducing the burden on respondents (UNECE, 2018).

Ensure Data Quality and Improve Coverage

ADS can be used by NSOs to corroborate respondent-provided information; these quality checks ensure that census products created from gathered data are of the highest quality. At the same time, ADS may improve coverage by updating the census address frame in countries with up-to-date postal and building registers (U.S. Census Bureau, 2020).

NECESSARY CONDITIONS FOR ADS INTEGRATION

A few important conditions must be met before integrating ADS with census planning (more detail can be found in the UN Guidelines on Registers-Based Censuses, UNSD, 2022 draft).

Legal and Policy Base

NSOs must work within their governments to ensure that legal provisions prescribing their access to—and protection of—administrative data are established. Ideally, the government should enact additional provisions for the NSO to have authority on the creation, revision (or deletion), and sharing of the data (and metadata) to be used in statistical registers. (UNECE, 2011)

Organizational Framework and Cooperation

The NSO should also work with data-holding bodies across the private and public sectors to develop a framework detailing the following:

- Key players in the data transfer.
- Standards for metadata, frequency of data sharing, data and metadata quality requirements, and confidentiality rules.
- Payment (preferably free).
- Period of agreement.
- Contingencies for changes.
- Procedures for resolving disputes.

This framework is no substitute for a transparent and open working relationship with other data holding bodies that must exist to ensure data holders provide good quality data on time (UNECE, 2011 and 2018).

Stakeholder and Public Approval

Opinions toward the sharing of data vary within and between countries (UNECE, 2011). For example, some people may not trust that their confidentiality can be maintained as their data is transferred amongst various data-holding bodies, and outrightly reject ADS use. To avoid confusion, distrust, and gain public approval, NSOs must ensure transparency, meaning that the public is informed of how their data will be used (e.g., publish clear limits and rules for use), and that data users are communicated with regarding any changes in access to specific data types (UNECE, 2018).

Unified Identification System and Reliable Statistical Register System

ADS integration (Figure 2) requires time for linking individuals across registers and censuses. In the ideal scenario, NSOs can link individuals through established unique IDs that map individuals and households across disparate sources, greatly facilitating register-based censuses (UNECE, 2018). These registers, of course, must be comprehensive and reliable (i.e., contain accurate, relevant, and timely data) for NSOs to integrate them into their census structure, and must typically undergo a rigorous transformation process prior to integration.

METHODS AND PROCESSING FOR PRODUCING AN INTEGRATED TRADITIONAL CENSUS

As many countries transition from traditional censuses to combined and register-based censuses, a common framework (Figure 3) may be useful for streamlining the integration process (UNECE, 2018). This section briefly describes the five key processes identified in Figure 3.

Identify and Collect Data Sources

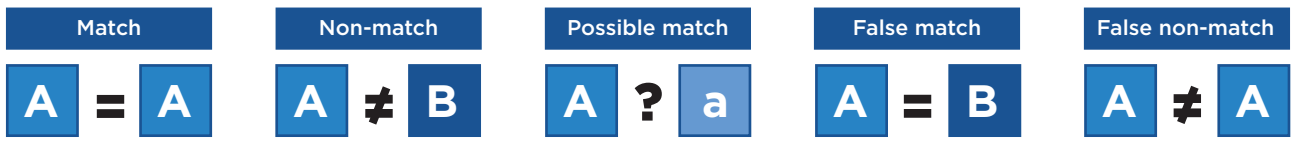
Potential, promising, and relevant data must be identified through meetings between the NSO and administrative data suppliers. NSOs should be a driving force in this process, establishing themselves as a data and communications hub in this evolving network of connections. These connections are key as initial data transfer (i.e., collection) and update mechanisms must be established prior to any data integration.

Record Linkage

Record linkage is one of the most important processes for using ADS in a census. In the absence of unique IDs (i.e., common identifiers) across ADS, NSOs can use common identifying variables (i.e., matching keys) such as name, sex, and date of birth to link across sources (UNECE, 2011). Keys should be chosen based on their uniqueness, or "distinguishing power." For example, "Robert Smith, b. 1962" has a much lower distinguishing power than "Robert Herbert Smith, b. 13 June 1962, Beaumont, TX." These matching routines, which express possible record matches probabilistically, are likely to result in fewer false matches and false nonmatches, and reduce the need for further human investigation of automated matching processes to assess pair accuracy (Figure 2).

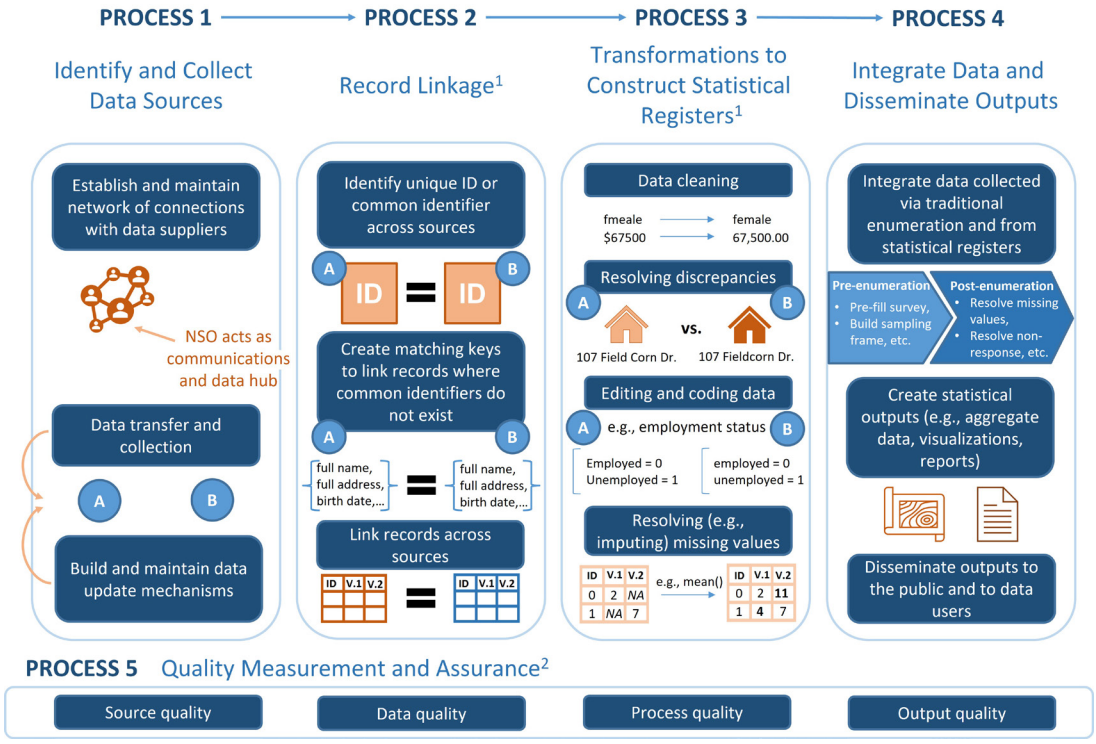
There are five main potential outcomes from probabilistic record pair matching: (1) Match: a pair representing the same entity in reality, (2) Nonmatch: a pair representing two different entities in reality, (3) Possible match: a pair without enough information to determine whether it is a match or nonmatch, (4) False match: a pair wrongly designated as a match (i.e., false positive), and (5) False nonmatch: a pair matching in reality, but not matched in the mapping process (i.e., false negative) (UNECE, 2011).

Figure 2.
The Main Potential Outcomes from Matching Pairs of Records



Source: United Nations Economic Commission for Europe, 2011.

Figure 3.
Common Framework for Integrating Administrative Data Sources in Census Outputs



[1] More detail on processes 2 and 3 can be found in United Nations Statistics Division (UNSD), 2022 draft.
[2] More information on process 5 can be found in United Nations Economic Commission for Europe, 2021; UNSD, 2022 draft.
Source: United Nations Economic Commission for Europe, 2018 and 2021.

Transformations to Construct Statistical Registers

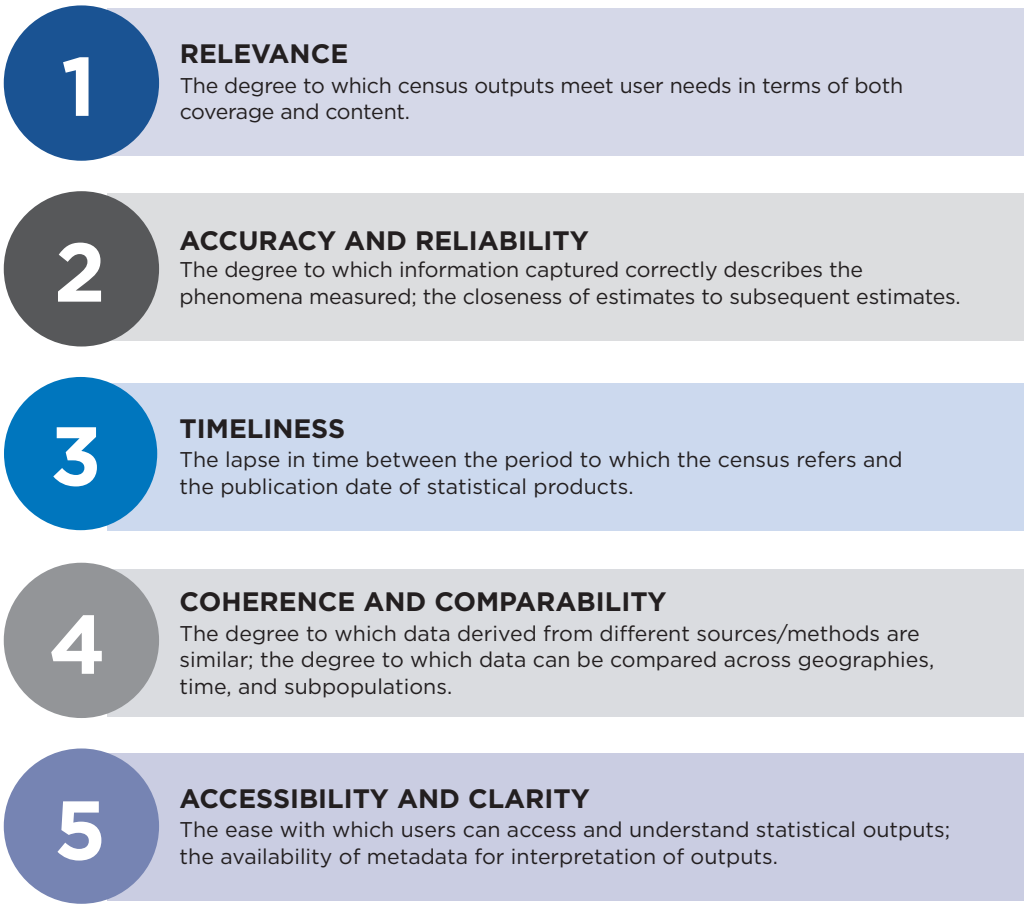
Data transformation occurs after linkage has been established, whether to account for under or overcoverage, to resolve discrepancies between sources, or to impute missing data. These processes take administrative data and transform them through statistical algorithms and data cleaning into statistical registers. Transformation processes include:

- Data cleaning.
- Resolving discrepancies (e.g., ensuring addresses match across sources).
- Editing and coding data (e.g., align codes across sources).
- Conversion of administrative units (e.g., tax and legal) to statistical units (e.g., people and households).
- Resolving (e.g., imputing) missing values.
- Selecting records that meet the population group of interest. These transformed and linked statistical registers are post-processed databases that can be used to produce census-type outputs. Common linked statistical registers include population registers, address/dwelling registers, business registers, and activity registers. The entities included in registers are individuals, addresses, and businesses.

Integrate Data and Disseminate Outputs

These statistical registers can then be used, in combination with data collected through traditional enumeration, to produce a range of standard census-type statistics with associated measures of output quality (Figure 4). Successful integration of statistical registers will allow for NSOs to produce and disseminate more frequent and timely census-type statistics and products for data users and the public.

Figure 4.
Output Quality Dimensions



Source: United Nations Economic Commission for Europe, 2021.

Quality Measurement and Assurance

Throughout each of the processes in this framework, NSOs must conduct quality checks and perform quality assurance. Broadly, NSOs can breakdown quality measurement into input quality (i.e., quality of the sources and input data), process quality (i.e., quality changes as each process is conducted), and output quality (i.e., the quality of the resulting statistics disseminated to the public and other data users) (Figure 4). More information can be found in UNECE, 2021 and UNSD, 2022 draft.

THE CHALLENGES OF ADS INTEGRATION

Though there are many benefits of utilizing ADS, there are also many challenges. It is important for NSOs to consider the challenges associated with integrating ADS in their censuses prior to implementation.

Limited Control Over Administrative Data Collection

Some of the greatest challenges to integrating ADS in a traditional census structure are due to NSOs' limited control over administrative data collection. NSOs rely on public authorities, which collect data for specific administrative purposes that are likely to differ from the NSOs needs. This may lead to differences in concepts and definitions, units of analysis, and reference periods (UNECE, 2011 and 2018). These differences must be accounted for by NSOs through, for instance, the creation of classification systems to map definitions between ADS and censuses, or the conversion of administrative units (e.g., tax, legal) to statistical units (e.g., people, households), which may be costly. These conversion and linking processes, which enable NSOs to use ADS for purposes other than those which the data were originally intended, can raise privacy, security, and legal concerns (UNECE, 2018).

Box 2.

Administrative Data Use Across the Collaborative

In 2020, the United Nations Statistics Division and the Global Partnership for Sustainable Development Data convened a collaborative of stakeholders with the express aim of strengthening the capacity of NSOs to leverage ADS for statistical purposes. This group, called the Collaborative on the Use of Administrative Data for Statistics (hereafter, *Collaborative*), acts as a platform for sharing resources, tools, best practices, and experiences across diverse country, regional, and international contexts.

Countries within the Collaborative have a broad range of experiences with ADS integration. Some countries, like Ghana, are in the initial phases of integration, which involve garnering political support and championing ADS use. In the Ghanaian case, officials have held events and conferences where smaller pilot projects, which illustrated the data transfer process from ministries to the Ghana Statistical Service, demonstrated ADS feasibility at larger scales. Other countries, like Ecuador, have been working with ADS for several years and have existing protocols and methodologies for their integration. In the Ecuadorian case, officials have been able to improve the quality of identification data of the ENEMDU (National Survey of Employment, Unemployment and Underemployment) COVID-19 Survey interviewees by linking these data with the "Cedulados" database from the Civil Registry of Ecuador.

Through the Collaborative, developed and developing countries come together in response to the urgent need for more timely and disaggregated statistical data to address the 2030 SDGs.

Source: United Nations Statistics Division and the Global Partnership for Sustainable Development Data, 2021.

Inconsistencies Between Sources and Missing Data

Challenges also arise when multiple sources of administrative data appear to contradict one another and when relevant data are missing entirely. In the first case, NSOs must make decisions regarding which ADS to utilize (i.e., most likely to be up to date and reliable), and which contradicting sources to ignore, or develop processes for harmonizing administrative data across various sources. In the latter case, NSOs must implement processes for either collecting additional data or for data imputation.

Resistance to Change and Maintaining Knowledge

Beyond public and stakeholder approval, staff statisticians may not trust the quality of data they do not collect themselves—or may focus on the negative aspects of ADS integration over the benefits (UNECE, 2011). This resistance to change within the NSO can lead to limited ADS use, undermining innovative ADS integration. Additionally, staff turnover at NSOs may result in valuable knowledge loss between statistic-producing years (UNECE, 2018).

CONCLUSION

Administrative data sources are of critical and increasing importance in improving the efficiency and quality of traditional censuses. When considering the integration of ADS, NSOs should work hand in hand within their government and with potential data partners to ensure the lawful and timely provision of quality data. In this technical note, some of the most important activities that should be managed and evaluated during ADS integration are discussed. The ultimate goals of ADS integration are to improve the data quality, timeliness, and efficiency of traditional censuses.

REFERENCES

Mule, Thomas, "Administrative Records and the 2020 Census," U.S. Census Bureau, Decennial Statistical Studies Division, April 2021.

United Nations Economic Commission for Europe, "Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices," New York and Geneva, 2011.

_____, "Measuring population and housing: Practices in UNECE countries in the 2010 round of censuses," New York and Geneva, 2014.

_____, "Guidelines on the Use of Registers and Administrative Data for Population and Housing Censuses," New York and Geneva, 2018.

_____, "Guidelines for Assessing the Quality of Administrative Sources for US in Censuses," Geneva, 2021.

_____, "Recommendations for the 2020 Censuses of Population and Housing," New York and Geneva, 2015.

United Nations Statistics Division, "Handbook on Registers-Based Population and Housing Censuses," second draft, March 2022.

_____, "Report on the results of the UNSD survey on 2020 round population and housing censuses," March 2020.

United Nations Statistics Division and the Global Partnership for Sustainable Development Data, "Collaborative on the Use of Administrative Data for Statistics," 2021.

U.S. Census Bureau, "Intended Administrative Data Use in the 2020 Census," 2020.

U.S. Census Bureau, "Administrative Data Used in the 2020 Census," 2021.

The Select Topics in International Censuses (STIC) series is published by International Programs in the U.S. Census Bureau's Population Division. The United States Agency for International Development sponsors production of the STIC series, as well as the bilateral support to statistical organizations that inform authors' expertise. The United Nations Population Fund collaborates on content and dissemination, ensuring that the STIC series reaches a wider audience.