



► Room document: 20*

Making full use of administrative data

A case for administrative registers as a complementary source of labour statistics



21st International Conference of Labour Statisticians
Geneva, 11-20 October 2023

► Acknowledgements

This document was drafted by Rosina Gammarano of the Statistical Standards and Methods Unit of the ILO Department of Statistics. Valuable comments and guidance were provided by Kieran Walsh of the Statistical Standards and Methods Unit of the ILO Department of Statistics and Edgardo Greising of the Knowledge Management Solutions Unit of the ILO Department of Statistics.

► Contents

	Page
Acknowledgements	3
Contents	4
1. Introduction	7
2. Administrative registers as a complementary source of labour statistics	9
2.1. What are administrative records and registers?	9
2.2. Administrative data in the big data revolution	9
2.3. Administrative registers as a source of labour statistics	10
2.3.1. Advantages of administrative registers as a source of labour statistics	10
2.3.2. Disadvantages of administrative registers as a source of labour statistics	11
2.3.3. Overview of types of records used for producing labour statistics	12
2.4. Administrative data and survey data as complementary	15
2.4.1. Each source type has its advantages and disadvantages	15
2.4.2. No single data source can meet all needs	18
2.4.3. Labour Market Information Systems integrate all sources of labour statistics	18
3. Administrative registers' usability for labour statistics	20
3.1. Desired properties of administrative registers for statistical use	20
3.2. From administrative to statistical registers	21
3.3. Factors influencing the statistical use of administrative data	22
3.3.1. Legal framework	22
3.3.2. Institutional collaboration, communication, and exchange	22
3.3.3. Consolidated data systems	23
3.3.4. Inventory of administrative registers	24
3.3.5. Transparency and awareness	24
3.3.6. Optimization of administrative tools for statistical purposes	24
3.4. Data quality assurance	25
4. Uses of administrative registers for labour statistics	28
4.1. Administrative registers as a direct source of labour statistics	28
4.2. Administrative registers to enhance survey data quality and cost-efficiency	30
4.3. Combined use of administrative and survey data for integrated statistics	31
5. Resources and tools	35
5.1. Good Practices Database by the UN Statistics Division	35
5.2. Inventory of Resources by the Collaborative on Use of Administrative Data for Statistics	35
5.3. Tools for Quality Assessment of Administrative Registers	35
5.4. Open training resources	36
6. Concluding remarks and future work	37
7. References	38

1. Introduction

Labour statistics provide crucial information on demographic, social, and economic aspects of the world of work. They cover a wide range of topics related to employment and other forms of work, including labour force participation, labour force composition, access to employment, quality of employment, working conditions, economic performance, and productivity, among many others. Labour statistics are essential to describe the world of work and the way it operates, inform the formulation, implementation and evaluation of policies, support decision-making by labour market agents, feed research, and monitor trends. Therefore, the production of official, reliable, timely, and comprehensive labour statistics is fundamental for all countries, and typically a priority for national statistical offices.

Labour statistics can be produced from different types of sources, although not all labour statistics can be produced from all sources. The characteristics of the statistical source will determine the type of labour statistics produced, their reliability, comparability, and coverage.

The quintessential source of labour statistics is the labour force survey, a household sample survey designed specifically to collect information on the structure of the labour force, employment, and unemployment. The survey questionnaire can be designed to include specific world of work issues, as necessary for monitoring and planning purposes. The labour force survey is often one of the primary surveys conducted by national statistical offices, and at the heart of Labour Market Information Systems (LMIS).

Labour statistics (especially on employees, earnings, and hours of work) can also be derived from establishment censuses and surveys. Population censuses are also used to produce a limited range of labour statistics.

These surveys and censuses are statistical data collections purposely designed to produce statistics. However, a wealth of valuable statistics can also be derived from sources not initially created for statistical purposes, such as administrative registers and big data.

In the current context of growing data needs and demands and tightening resources for national statistical offices around the world, more and more institutions are exploring or expanding the use of administrative data. Moreover, the fast digitization trend and the increasing use of electronic methods for data entry and storage adoption are greatly enhancing the availability, timeliness, granularity, and user-friendliness of administrative registers. Indeed, making full use of administrative data as a source of labour statistics is a cost-effective means of complementing survey data. Although administrative registers could never replace surveys (and especially not labour force surveys) as a source of labour statistics, they can be very beneficial in complementarity.

When the Covid-19 pandemic hit and social distancing measures and mobility restrictions were implemented around the world to contain it, field work for labour force surveys was abruptly halted in many cases, disrupting the production of official labour statistics when they were needed the most¹. Some countries tried turning to remote data collection methods such as telephone interviewing to make up for the impossibility to conduct face-to-face interviews. Nonetheless, this shift was often not a sustainable solution given its impact on response rates and accuracy of responses, associated with the lack of a complete register of contact details for sampled households. Countries with adequate register systems already established turned to administrative data to mitigate the impact of disruptions to survey data production. Although administrative data collection may also have suffered the impact of mobility restrictions under certain circumstances, the increased use of administrative data appeared as a generally positive solution to monitor labour markets at such a trying time.

There are many advantages to making more and better use of administrative data, provided the usual caveats about it are taken into consideration. The benefits of administrative data seem to be widely recognized, as witnessed by the growing number of countries conducting register-based or combined population censuses. The first totally register-based population census was conducted by Denmark in 1981, and the second by Finland in

¹ For information from a global review of impacts on labour statistics see: <https://ilostat.ilo.org/keeping-labour-data-flowing-during-the-covid-19-pandemic/>

1990. Since then, the practice has steadily expanded to reach at least 16 countries by the 2020 round of population and housing censuses, with an additional 29 countries conducting combined censuses in that round (UNSD 2022). The Nordic countries have been pioneers in the use of administrative registers as a source of official statistics, taking full advantage of some favourable pre-existing conditions in their context including the legal base, widespread public approval, unified identification systems, longstanding tradition of institutional cooperation, and the existence of rich, comprehensive, and reliable administrative register systems. The advantages of the statistical use of administrative data are such that more and more countries are following suit, finding ways to produce quality statistics from administrative registers in less ideal contexts than the Nordic one. Indeed, adapting the process to local circumstances and constraints, countries in all regions are starting to explore the use of administrative registers as a source of official statistics, including labour statistics.

It should be noted, however, that country practices in the use of administrative data to produce official statistics seem to focus more on other statistical domains, such as population statistics, vital statistics, migration statistics, education statistics, and health statistics, among others. The production of labour statistics from administrative registers poses additional challenges, including the very basic limitation of having administrative registers with contents that could satisfy the needs for key labour market indicators. The time spanning from the establishment of an administrative register to the production of official statistics from it with data quality assurance varies from one topic area to another, and across countries. The efforts to produce official statistics from administrative registers typically start with other less challenging areas, before tackling the area of labour statistics.

Although every country's context is different (including in terms of statistical development, resources for statistical activities, availability of administrative registers, and attitudes or legal limitations regarding the use of administrative data for official statistics), there is undeniable value to knowledge and experience sharing. The learnings of countries having undergone these processes can be inspirational and beneficial to others undertaking similar paths. Exploring the approaches which proved effective in countries with different levels of income and statistical infrastructure development can be enlightening.

This room document focuses on the potential of administrative registers to supply labour statistics, providing an overview of the advantages and disadvantages of administrative data, as well as the key aspects determining the usability of administrative data as a reliable source of labour statistics. It also summarizes the main potential uses of administrative data in the field of labour statistics, citing relevant country practices across regions.

Recognizing the potential of administrative registers to supplement and improve other sources of labour statistics such as labour force surveys, the ILO invites the International Conference of Labour Statisticians to reflect on the needs for ILO support to develop register-based labour statistics in terms of capacity building and technical assistance.

2. Administrative registers as a complementary source of labour statistics

2.1. What are administrative records and registers?

An administrative record is a complete list of units in a given population, with information updated (or at least updatable) on a regular basis on identification of units, geographical localization of units, variables associated with each unit, and where applicable, the relationship between units (Wallgren, A. and B. 2007).

The term “administrative record” refers to the data contained in the record, while “administrative register” refers to the data source and “administrative registry” designates the data owner.

Administrative records are created and maintained by public or private entities for their own administrative purposes, such as keeping a record of their members, activities, staff, resources, dwellings, etc. They are not designed for statistical purposes, but they do have a significant underlying statistical potential, and can be used to produce statistics as a by-product. The statistical potential of administrative records is widely under-utilized in many countries around the world, although more and more efforts are devoted to the production of register-based statistics.

In any given country, the administrative records found are numerous and varied, and cover a wide range of topics. Some examples are population registers (which contain population data), tax authorities records (income data), credit/debit card companies’ databases (consumption data), educational system records (education data), customs records (tourism, migration, and foreign trade data), justice system records (criminal data), and medical records (health data).

When it comes to labour statistics, the main administrative records used are records of employment offices or unemployment benefits (registered unemployment data), taxation records (registered employment plus potentially data on earnings) records of workers’ organizations (unionization data, strikes data, etc.), collective agreements (negotiated wages data, agreed working conditions data, etc.) and labour inspection records (occupational accidents data, inspection visits data, etc.) (ILO 2017).

2.2. Administrative data in the big data revolution

With the advent of technology, mobile devices, ubiquitous internet access, and digital social networks, data sets are growing exponentially both in number and size, with practically every online transaction, action, and search being recorded. Big data refers to all data sets associated with the four “Vs” that represent its core dimensions: volume, variety, velocity, and veracity. These characteristics mean that big data cannot easily be analysed using traditional data processing and statistical methods (ILO 2017). It goes well beyond everything we do on the internet, involving also commercial or financial transactions done via electronic devices, information from the Global Positioning System (GPS), sensor data, and video data, to name a few. Therefore, big data encompasses a wide range of types of data, which may differ significantly in structure and characteristics. While some sources of big data are owned by government agencies, most of them belong to private agents.

Just like administrative data, big data was also not created for statistical purposes, but as a result of processes in place for administrative, electronic, commercial, surveillance, or other purposes.

Big data can supply a wealth of information on labour statistics, for instance on job searches, vacancies, and skills. What is more, the information is available in real time. Even though there are still no internationally-agreed methodologies and guidelines on deriving labour statistics from big data, it is gaining ground as a statistical source, mainly among private agents. However, the use of big data is nothing short of challenging: data curation and triage are key to identifying the portion of data that is actually useful for the purposes at hand among the enormity of data available, processes need to be designed to handle the unstructured nature of big data, the concepts and definitions underlying the data need to be clarified and established, the coverage and representativeness of the source (and inherent biases) need to be carefully outlined for the interpretability of the statistics, and data quality and accuracy need to be evaluated (ILO 2017). Given the cost of accessing big data (often from private sources), the

skills and technology needed to handling it, the resources required to ensure data protection, privacy, and quality, and the efforts involved in guaranteeing sound methodologies and linking to other sources of data, big data is a less cost-efficient source of labour statistics, at least under current circumstances.

2.3. Administrative registers as a source of labour statistics

2.3.1. Advantages of administrative registers as a source of labour statistics

There are several advantages to the production and use of official statistics, and particularly labour statistics, from administrative registers, as summarized in the bulleted list below. In general terms, administrative registers are a cost-effective, rather inexpensive source of labour statistics with a potentially exhaustive coverage of all the units in their reference universe. However, these advantages need to be interpreted and valued taking into consideration the limitations of administrative data as well.

- **Ubiquitous source of statistics**

By their very nature and their existence in response to administrative needs of public and private entities, administrative records are virtually everywhere. Administrative registers cannot replace survey data collections such as labour force surveys given the targeted way these are designed to collect information on the specific labour market issues of interest and following agreed methodologies and quality standards. Yet, in the absence of a regular or recent survey, administrative registers can provide basic labour market indicators to monitor trends and patterns.

For instance, in countries where the most recent labour force survey or similar household survey dates back too many years to be relevant, statistics on registered unemployment can be derived from social insurance or employment office records and statistics on employment can be derived from social insurance, taxation, labour inspection and/or business records. Although these do not relate to the internationally-agreed concepts of unemployment and employment derived from a labour force survey (and are therefore not comparable) they still provide valuable information for policymakers in times of need.

Similarly, when labour force survey field work was abruptly halted due to the mobility restrictions implemented to contain the Covid-19 pandemic, many countries turned to administrative data to provide the much-needed labour market information to policymakers (at least to the extent possible).

- **Cost-effective source of statistics**

Given that administrative records are created and maintained by the corresponding agency for its own administrative purposes, the information is readily available, and no further compilation effort is needed, which implies no additional burden for (would be) respondents.

There will be costs associated with the process of turning administrative records into statistical registers and linking various records where needed, but these are minimal compared to the typical costs of organizing a full data collection. The more systematized and integrated the administrative register system (via, for instance, the use of unique unit identifiers and harmonized concepts, definitions, and classifications), the less the cost of building a register-based statistical system.

- **Exhaustive coverage within the scope**

Administrative records, by their very definition, have typically a complete (exhaustive) coverage of all units in its scope. The extent to which this is true will depend on the registration procedures, rules, regulations, and incentives. For instance, if registration is mandatory for all individuals by a certain date every year and there are penalties for failing to register (credibly enforced), the record will most likely tend to be exhaustive.

The exhaustive coverage implies that there are no sampling errors (such as the ones associated with sample surveys), allowing to produce reliable statistics on small population groups or geographic areas.

Where problems of under-coverage or under-reporting are known or suspected, measures can be taken to correct them or account for them.

- **Data accuracy**

Provided there are sufficient mechanisms in place to ensure sound data entry in the administrative process (and correction of potential errors), administrative records should contain highly accurate data on the items covered. Contrary to proxy interviews often used in household sample surveys, administrative records contain information usually collected directly from the unit concerned, or an entity having accurate information on the units concerned based on its mandate.

Also, individual respondents tend to fill in more accurately administrative forms, considering them for official purposes.

- **Topic coverage**

There are many topics for which the labour force survey remains the preferred (if not only) source, but administrative registers appear as a fundamental source for topics which are not easily covered by traditional statistical sources. There may be many reasons for the labour force survey not to be a solid source of data on some topics, including scepticism or sensitivity of respondents to answer truthfully to some questions, usual mis-recollection, or sampling concerns.

Indeed, administrative registers are typically the preferred source for statistics on occupational safety and health (including on occupational accidents and injuries), industrial relations (including collective bargaining coverage, collective bargaining outcomes, and strikes and lockouts), and social protection coverage (including on pension coverage and level, maternity insurance, and registered unemployment).

It should be noted, however, that from their very nature administrative registers are also a flawed (and sometimes even impossible) source of data on other topics, such as the informal economy, unemployment (as per the internationally-agreed definition), and other forms of labour underutilization which rely on criteria or coverage that administrative registers are unlikely to be able to apply.

- **Consistency and stability**

Rules and regulations underlying administrative registers are typically stable over time, which implies that registration methodology is also rather consistent and stable. Nonetheless, for this to be true, the administrative registers has to have reached a sufficient level of maturity with serious registration procedures gauging for data quality in place.

- **Timeliness**

For administrative registers with robust registration and maintenance procedures, the timeliness of data production can be much quicker than for surveys. Indeed, for some digitalized records the information is updated in real time. In some cases, the updates involve only the units which have recorded changes, making the process much simpler and cost-effective.

Given that agencies keeping administrative registers often apply the same methodology over time (even though this methodology is most likely not in line with international labour statistics standards), robust time series can be available for useful longitudinal studies.

2.3.2. Disadvantages of administrative registers as a source of labour statistics

Despite the many advantages of using administrative registers as a source of labour statistics, there are also some limitations to keep in mind, as summarized below. These limitations can be overcome with appropriate statistical processes to ensure data quality, caution in data analysis and interpretation, transparency in disclosing data caveats, and especially, using administrative data in combination with survey data.

- **Methodology determined by administrative processes**

Since administrative records were not created for statistical purposes, their methodology will most likely not follow international (or even national) statistical standards. Definitions, concepts, and classifications are determined by the administrative processes and purposes, rather than by statistical standards. Similarly, the units used, the universe, and the topics covered, also depend on the objective and mandate of the agency keeping the record, rather than on statistical needs.

In due time, where integrated register-based statistical systems are developed with strong institutional cooperation at their core, national statistical offices could have some influence on the methodology and contents of records.

- **Contents determined by administrative processes**

Likewise, the contents of administrative records (in terms of topics, units, and variables covered) are determined by the administrative purposes for which they were created. Therefore, the labour market indicators that could potentially be produced depend on data contained in records by default, and do not necessarily respond to statistical needs. Indeed, administrative records often have limited information on individual or household characteristics, whereas surveys typically collect a rather complete socio-demographic profile of respondents, allowing to produce labour statistics with interesting disaggregations.

It should also be noted that administrative records contain no information on those not participating in the record, whereas a survey captures in a consistent and integrated manner all individuals. For instance, while a labour force survey would enable the consistent analysis of people employed, unemployed and outside the labour force, an administrative record would typically provide data on those registered as employees only, or on those receiving unemployment benefits only, hindering the comparability of indicators to those that would be produced by a survey. This may be mitigated to some extent by linking or combining different administrative sources, but it would be extremely unlikely to achieve the same level of coverage of topics that can be achieved through a survey.

Additionally, administrative records often lack references to the territory, a crucial feature to allow for the reliable production of official statistics. At best, administrative records include an address associated with each unit, which will have to be geolocated using Geographic Information System (GIS) technology for statistical production. Contrarily, surveys always locate the statistical unit in a certain reference area, part of a complete division of the national territory made for census purposes and not necessarily coincident with the administrative division of the territory.

Similarly, the reference period for administrative data may not be as straightforward as in surveys, where questions refer to a precise moment in time.

- **Data accuracy, completeness, and timeliness determined by administrative processes**

As previously mentioned, data from administrative registers can be highly accurate, timely, and comprehensive, but this all depends on the rules and regulations guiding the registration and administrative process.

Where there are no strict or enforced mechanisms to ensure proper and complete registration, and where there are no pre-established regular updates, issues of data accuracy, representativeness, and timeliness may arise. Therefore, it is crucial to have information on the main characteristics of an administrative register, including registration and update methodology, in order to use it as a source of labour statistics.

It should be mentioned that quality assurance procedures in place for administrative registers may focus on variables of interest for administrative purposes with less attention paid to variables mainly serving statistical purposes. Therefore, running checks with validation rules focusing on the intended statistical use of the data is advisable.

- **Coverage limitations**

Although administrative registers (with solid registration procedures) tend to have an exhaustive coverage of units within their universe, the universe may be limited, or at least less extended than desired for labour statistics purposes. For instance, some high-quality, exhaustive registers may cover only one country region, and information from all regions will need to be integrated together to have nationally representative statistics. Also, employment statistics from social insurance records will only cover formal employees or self-employed workers registered with the social insurance, employment statistics from labour inspection records will only cover workplaces liable to labour inspection, and so on. In the case of administrative records on establishments an important issue is the process by which new establishments are entered onto the record and old/closed ones are removed – a limitation which may also apply to records covering individuals. If these are subject to a time lag it may mean that the coverage at any given point of time is not up to date.

2.3.3. Overview of types of records used for producing labour statistics

Administrative data are ubiquitous. In any given country, administrative records found are numerous and varied. In the field of labour statistics, administrative records are sometimes used to provide data on employees, registered

unemployment, wages, working time, occupational accidents and injuries, labour inspection, strikes, unionization, and collective bargaining coverage, among others. Administrative data can also be used in combination with surveys or other sources to provide information on some key variables that would otherwise be collected through survey questionnaires, for instance sociodemographic characteristics such as educational level or migration status. The list below gives an overview of some of the main types of administrative records used for labour statistics.

- **Population registers**

These registers result from the continuous recording and/or coordinated linkage of selected information pertaining to each member of the resident population of a country to provide up-to-date information concerning the size and characteristics of that population at selected time intervals (UN 1969). They usually cover the whole country and provide reliable information for government administrative purposes such as budgeting, planning, and taxation.

Population registers are far from a novelty. In fact, their central role for official decision-making has led governments to develop population registers many centuries ago. The first known registers date back to the early seventh century in China, Korea and Japan. However, the centralization of population registers is more recent, emerging in the second half of the twentieth century. Iceland was the first country to create a national population register in 1952-1953. During the 1960's the remaining Nordic countries (Denmark, Finland, Norway, and Sweden) all established their own national population registers with unique personal identification numbers. Since then, more and more countries around the world are creating or improving national population registers, taking advantage of technological advances.

Population registers have been used for producing official statistics for decades in many countries.

In labour statistics, they may serve as sample frame (alone or in combination with other sources), they may provide some very basic labour-related characteristics, or key socio-demographic characteristics to be linked to the labour-related data kept in other registries.

- **Business registers**

These registers are structured databases of economic units in a territory, typically a country. National business registers are often maintained by a public entity in accordance with national legislation. The business register includes current information on the number and type of registered businesses, usually with data on key aspects such as their legal form, headquarters, capital, and legal representatives. If there is more than one kind of economic unit covered, the links between units are also included.

Business registers differ from one another in terms of their content, updating frequency, legal checks, legal value of the registered information, and the ability to access documents. They can provide a sampling frame for the selection of establishments, for example, for employment-based establishment surveys (ILO 2013), however, registers will not cover informal establishments making their coverage low in many countries.

- **Insurance records**

These are records of social insurance institutions and private insurance companies, generally including information on persons insured, contributions made, services incurred, and benefits granted. These records contain key information for the insurance institution itself, policymakers, and labour market agents, on contributors, affiliates, and people receiving benefits. Their coverage is by definition limited to the population insured or covered by the corresponding institution, and this depends highly on the reach of the insurance system. Therefore, the statistics derived from insurance records refer to insured or compensated events, not to all events (for instance, statistics on occupational injuries from insurance records refer to compensated injuries only).

A key common use of social insurance records for policymakers and analysts is the production of statistics on registered unemployment (people receiving unemployment benefits).

It is advisable to interpret insurance records statistics alongside statistics from other sources whenever possible, to complement the analysis (for example, analysing statistics on registered unemployment alongside statistics on unemployment based on a labour force survey applying international statistical standards, and analysing statistics on compensated injuries with injuries notified by employers to the Labour Ministry or injuries recorded by the Labour Inspection). However, for many of the reasons already described comparability cannot necessarily be expected between sources so the emphasis of

analysis should not be direct comparison, instead focussing on supplementing the information available from other sources in a transparent way recognising the difference between the sources.

- **Employment office records**

These are records from government offices (sometimes called public employment services) which, through the declaration of registered employers, employees, and unemployed persons, collect and provide information on job vacancies and jobseekers.

- **Taxation records**

Taxation or tax records are records kept by the authority competent in matters of tax administration and/or collection, including tax returns, income declarations, and any other item relevant under the tax code or legislation. Taxation records of establishments or individuals can provide valuable information on employees and earnings, with varying levels of detail depending on the degree of compliance and registration. Taxation records are often considered a reliable source of data on taxable earnings of formally employed workers. In some cases, they may cover self-employed workers (either within the same record as employees or separately). The comprehensiveness and accuracy of the data will be closely linked to the level of tax compliance.

Difficulties in data access may hinder or prevent the statistical use of taxation records. Tax authorities may be reluctant to share their records covered by tax secrecy rules or legislation. Here too, legislation entitling the national statistical office (or other competent authority) to access administrative data kept by official entities including taxation records for specific statistical purposes and respecting strict data privacy and confidentiality rules will enhance the potential for statistical production.

- **Collective bargaining agreements**

According to the ILO Collective Agreements Recommendation (recommendation n°91, adopted in 1951) collective agreements are all agreements in writing regarding working conditions and terms of employment concluded between an employer, a group of employers or one or more employers' organizations, on the one hand, and one or more representative workers' organizations, or, in the absence of such organizations, the representatives of the workers duly elected and authorized by them in accordance with national laws and regulations, on the other.

Institutions such as Labour Ministries may keep administrative records on collective bargaining agreements with valuable information on the coverage, content, date of signature, and duration of the agreements.

It should be noted that collective bargaining rarely covers self-employed workers, and some sectors of economic activity or groups of workers may be excluded from the right to collective bargaining.

- **Labour inspection records**

Labour inspection records keep a tally of the activities and resources of the competent authority in the area of labour inspection. Labour inspection activities recorded include, notably, labour inspection visits, that is, workplace inspections carried out by labour inspectors who are public officials responsible for securing the enforcement of labour legislation, working conditions regulations, and the protection of workers.

It should be borne in mind that some types of workplaces and workers are harder to cover by labour inspection activities, and the labour inspection may not reach isolated geographic areas.

Also, since most of the data is collected by the inspectors when carrying out their duties (mainly visits), the data quality highly depends on the methods and tools used for field data collection.

- **Records of workers' organizations**

These are records kept by workers' organizations (including unions and union confederations) on their members, staff, and activities.

When using statistics derived from these records, the legal and social context should be kept in mind, since in some cases workers' organizations cover only employees (excluding the self-employed), while in other cases they not only cover the self-employed but also some persons not in employment (unemployed, retired, students). Also, some groups of workers may be excluded from the right to unionize.

The accuracy of the data in records of workers' organizations also depends on the procedures in place, as well as on the motivations and incentives to keep an updated tally.

- **Records of employers' organizations**

These are records kept by employers' organizations on their members, staff, and activities. The records may cover different types and sizes of employer entities, and their contents and completeness will depend on the conditions and procedures for registration.

2.4. Administrative data and survey data as complementary

2.4.1. Each source type has its advantages and disadvantages

There are different types of possible sources of labour statistics. The labour force survey is perhaps the most emblematic source of labour statistics, specifically designed for that purpose and with the ability to cover the labour market issues of interest to policy and decision-makers, and with the possibility to design the sample to achieve complete coverage of the population. Other types of household surveys, such as household income and expenditure surveys, household budget surveys, or living standards measurement surveys, can also serve as sources of statistics on some labour-related topics (although in less detail than labour force surveys and often with smaller sample sizes). Population censuses also provide data on some basic labour market issues. Establishment surveys and censuses are useful sources of information on employees, working time, and earnings, among other labour-related topics pertaining to establishments. In some countries the national accounts are used as a source of labour statistics, often drawing across sources (including the labour force survey) to generate estimates, for example of labour input to enable calculation of labour productivity. Big data can be used as well to produce some labour statistics, although the complexities associated with the nature of big data are such that official production of labour statistics from big data remains scarce and exploratory. And of course, the object of this document, administrative records also have a great statistical potential for monitoring the labour market.

Each of these potential sources of labour statistics has its advantages, disadvantages, and specificities. These need to be considered when interpreting the resulting statistics, and users must be informed of them through clear and extensive descriptive metadata accompanying the data. The absence of such information may lead to misinterpretations of the information.. The table below summarizes the key advantages and disadvantages of the main sources of labour statistics.

► Summary of advantages and disadvantages of main sources of labour statistics

Main sources of labour statistics	Advantages	Drawbacks
Labour force surveys and other types of household surveys	<ul style="list-style-type: none"> • Collection specifically designed for purposes at hand, so coverage of agreed topics of interest and use of agreed methodological standards and definitions. • Detailed questions for measurement of specific topics and probing for precision. • Ability to design samples to cover the entire population of interest (can be subject to limitations in practice such as difficult to reach areas or groups etc). • Consistent framework for the study of employment across sectors and statuses, unemployment, persons outside the labour force, and even potentially participation in other forms of work than employment. • Less costly than population censuses, so can be conducted more frequently, allowing for the study of short-term trends. • Designed over a household/dwellings frame which ensures the reference to the territory. 	<ul style="list-style-type: none"> • Sampling may prevent the production of reliable estimates for small population groups or certain geographic areas. • Sampling errors. • Measurement errors - data quality depends on accuracy of responses related to various methodologies such as questionnaire design, field implementation practices such as approach to proxy interviewing, non-response etc. • The lack of a regular program of labour force surveys may lead to sparse survey rounds with methodological differences across rounds.
Population and housing censuses	<ul style="list-style-type: none"> • Exhaustive, universal coverage. • Reliable statistics on small population groups and all geographical areas. • Ability to apply statistical concepts subject to space limitations. • Can serve as sample frame for surveys. 	<ul style="list-style-type: none"> • Very costly in its traditional form with full field enumeration (less so if combined or registered-based methods are used). • Not frequent (usually every 10 years in its traditional form with full field enumeration). • Provides basic information only, no detailed information on specific topics (such as labour). • Data quality depends on accuracy of respondents. • Measurement errors - data quality depends on accuracy of responses related to various methodologies such as questionnaire design, field implementation practices such as approach to proxy interviewing etc.

Main sources of labour statistics	Advantages	Drawbacks
Establishment censuses and surveys	<ul style="list-style-type: none"> • Comprehensive coverage of large businesses and some smaller business depending on registers and sample design. • Payroll records provide consistent and reliable data on some topics (for example on earnings and working time). • Surveying establishments is less expensive and more cost-effective than surveying households. 	<ul style="list-style-type: none"> • Typically poor (or no) coverage of small and unregistered economic units. • Difficult identification of informal economic units, where covered. • High non-response rates (where participation is not mandatory). • Sampling may prevent the production of reliable estimates for small population groups or certain geographic areas. • Possible topics covered limited to information available to employers or establishments. • Coverage limited to those employed in economic units within the scope of the survey, often excluding the self-employed. Thus, does not allow for the consistent study of employment, unemployment, and persons outside the labour force.
Administrative registers	<ul style="list-style-type: none"> • Cost-effective supply of data. • Data readily available, no compilation efforts required, only processing, and linking where applicable. • Usually exhaustive coverage of the universe. • Total count allows for maximum detail and estimates for small population groups and geographic areas. • Ubiquitous, particularly useful when there is no regular household or establishments survey. 	<ul style="list-style-type: none"> • Not created for statistical purposes, so methodology often not in line with statistical standards. • Topics covered, units used, and variables included depend on administrative processes, not on statistical needs. • Sometimes limited scope (small universe). • Data quality, consistency, completeness, and timeliness depend on the rules, regulations and incentives governing the record (for instance, whether registration is mandatory or not).
Big data	<ul style="list-style-type: none"> • Wealth of real-time information available. • Can allow for increased granularity and timeliness of labour market information. • Cost-effective source if data is publicly owned or publicly accessible (although this is rarely the case), and once up-front necessary costs have been incurred. 	<ul style="list-style-type: none"> • No statistical standards or guidelines yet. • Depending on the data owner, can be difficult and/or expensive to access some types of big data. • High up-front cost in skills and technology needed to process big data. • Unstructured nature of big data require time-consuming processes to derive statistics. • Extensive data triage and curation processes required. • Limited application to date in practice for the generation of official statistics
National accounts	<ul style="list-style-type: none"> • Coherent, consistent, and integrated set of macroeconomic accounts, allowing for the study of items within the framework of the economy. • Provides consistent information on output and compensation of employees. • Based on internationally-agreed concepts, definitions, classifications, and accounting rules. 	<ul style="list-style-type: none"> • Designed for purposes of economic analysis, less relevant for labour market analysis. • Employment in national accounts requires the integration of data from many sources, and the results are not comparable with employment statistics from other sources.

By recognizing the relative strengths, limitations, and complementarities of the various types of sources available, the spectrum of labour statistics becomes wider and more robust. The limitations of each type of source are mitigated by the combined use of different sources.

There are numerous ways in which different source types can complement and enrich each other. For instance, population census data can be used for benchmarking and to develop sample frames for surveys, just as population or business registers. Also, survey data can provide between-census estimates and be used to monitor short-term trends. Furthermore, administrative data can complement survey and census data, and at the same time, survey data can be used to offset potential problems of under-registration in administrative registers.

However, for this to be possible (that is, to have an integrated statistical system of labour market information reconciling the various source types available), it is necessary to ensure the coherence between sources in terms of the methodological guidelines used (concepts, definitions, classifications, reference periods, etc.). Ideally, this coherence would be obtained by all sources adhering to the relevant international standards on labour statistics.

2.4.2. No single data source can meet all needs

Labour statistics are produced to cater for policymakers, researchers, journalists, and labour market agents in general. Policymakers need labour market information to inform policy formulation, implementation, and evaluation, on many different labour-related issues. Labour market agents, including workers and employers, need quality and timely data to support their decisions related to the labour market.

However, no single data source can meet all needs, in large part given the wide range of topics and issues covered by labour statistics, including labour force participation, employment, unemployment, time-related underemployment, inactivity, informality, job search, working time, earnings, economic units and output, labour productivity, vacancies, collective bargaining, unionization, industrial disputes, occupational accidents, occupational diseases, social protection coverage and levels, skills mismatch, labour migration, child labour, unpaid domestic and care work, volunteer work, unpaid trainee work, and own-use goods production, to name a few. Different source types are complementary. All the main sources mentioned should contribute to an overall integrated system of national labour statistics, such as a Labour Market Information System (LMIS).

2.4.3. Labour Market Information Systems integrate all sources of labour statistics

A Labour Market Information System (LMIS) is a network of institutions, persons and information that have mutually recognized roles, agreements, and functions with respect to the production, storage, dissemination and use of labour market related information and results to maximize the potential for relevant and applicable policy and program formulation and implementation (ILO 2023).

The main purpose of a LMIS is the production of information and analysis for policymakers and other labour market agents and stakeholders, in a centralized and consolidated manner. The key functions of a LMIS include facilitating labour market analysis, providing the basis for policy formulation, implementation, and evaluation, and promoting institutional coordination across all relevant users and producers of labour market information and analysis, such as national statistical agencies, labour ministries, social insurance institutions, research agencies, and workers' and employers' organizations.

LMIS have four main components, namely: data collection and compilation, a data repository, analytical capacity and tools, and institutional arrangements and networks. These four components are strongly interdependent, that is, the efficiency, performance, and sustainability of LMIS depend on the solidity of the four components.

Effective labour market information systems pool all major data sources, facilitating their joint analysis. Ideally they must provide a complete picture of the labour market situation and its context, which requires bringing together information from a variety of sources covering numerous labor market-related topics. The data collection and compilation of a LMIS typically encompasses labour force surveys, other types of household surveys, population censuses, establishment surveys and censuses, a broad spectrum of administrative registers, and other data sources. As previously mentioned, each type of source has its advantages and disadvantages and provides

information on specific topics of interest, so an effective LMIS builds on all of them to combine their strengths and mitigate their weaknesses.

Since LMIS centralize the production, storage, analysis, and dissemination of labour market information by a wide range of institutions and actors, institutional arrangements and networks are essential to ensure their efficiency and sustainability. Institutional arrangements will enable, formalize, and regulate the use and dissemination of labour market information by providing timely access to data and metadata, facilitating analysis sharing, and promoting coordination of methodologies and decision-making. This is all the more important for the use of labour statistics from administrative registers. Since registers are typically owned by diverse entities, and since the official statistical mandate and expertise typically rests with the national statistical office, collaboration and cooperation is key to ensure the production of reliable, timely, and comparable labour statistics from administrative registers.

3. Administrative registers' usability for labour statistics

3.1. Desired properties of administrative registers for statistical use

Not any administrative register can become a reliable source of statistics, and even less so of labour statistics. For this to be possible, the administrative register needs to fulfill some basic requirements in terms of its contents and methodology. The bulleted list below summarizes the main conditions to be met by administrative registers to ensure the reliable production of official statistics (UNECE 2007).

- **Relevance of contents**

Since the register was created for administrative purposes, it is not guaranteed that its contents will be fully relevant to meet official statistical needs, particularly in the field of labour statistics. The very first condition for the statistical production process to be worthwhile is for the contents of the register to be relevant. That is, the register needs to cover enough data items (in terms of units, variables, and scope) with the potential to satisfy the end users' needs for labour statistics.

- **Data quality and accuracy**

Naturally, for the statistics derived to be deemed reliable, the underlying administrative register needs to have sufficient levels of data quality and accuracy. Quality requirements and quality assurance checks in the administrative registration process ensuring precision and completeness of the information recorded are crucial for statistical production. For the producer of statistics to assess whether data quality and accuracy levels are acceptable, proper documentation on methods and protocols needs to accompany the administrative register.

- **Units and universe**

For statistical use, units need to be well defined. The typical units used in administrative processes which could serve to support statistics production are persons, dwellings, and economic units.

Unit identifiers play a key role here, both in the maintenance of administrative registers and in the process of deriving statistics from them. Unit identifiers are especially important for linking sources (whether it is linking various types of administrative records together or linking administrative data to survey data). The use of unique unit identifiers in a country greatly eases these processes, but in their absence, probabilistic or non-probabilistic linking methods can be used.

It should be noted that units used for administrative purposes do not always coincide with the statistical units needed to produce key labour market indicators. For instance, dwellings do not necessarily correspond to households (basic statistical unit of household surveys such as labour force surveys). Therefore, some additional unit conversion techniques may be necessary, highlighting again the importance of unit identifiers.

Also, ideally, all units within the administrative register's scope (universe) should be covered. That is, registers should be comprehensive. Incomplete coverage needs to be assessed to determine the extent and type of biases that would be transferred to the statistics produced. In cases of known under-coverage, it is necessary to evaluate whether it is systematic or not. If the under-coverage is not systematic, it would logically not result in biases. However, under-coverage most often is systematic, and in such cases biases would arise (for example, if migrants or workers in informal employment are excluded from a given register).

- **Time references**

The time dimension plays a key role in statistics, which is why time references are of special importance when producing register-based statistics. In administrative registers, the date of an event often differs from the date of recording the information. Therefore, ideally, any data item in the administrative register should be accompanied by two dates: the date of the event and the registration date. The most crucial time reference for statistical use is the date of the event, but registration dates help assess accuracy and

timeliness. For some registers, it may not be possible or practicable to register events in real time, hence the importance of registration dates.

The timeliness and punctuality of administrative registers are key determinants of data quality. Incomplete, infrequent, or unpredictable updates will void the statistical usability of an administrative register.

- **Methodological stability**

In order to facilitate the statistical use of administrative registers, the methodology underlying the registration process and the record maintenance should be sufficiently stable and consistent over time (and ideally, across agencies). Where concepts and methods are applied over the longer term in the administrative register, comparable figures can be produced over time, allowing for the study of trends.

Legislative, regulatory, or procedural changes will result in changes to the data content of administrative registers. Sometimes it will be possible to adjust for these changes, other times breaks in series or even series discontinuation will be inevitable. In any case, documentation is key to assess the methodological robustness of the administrative register, covering all necessary metadata.

The metadata and documentation on administrative registers should include information on the owner and keeper of the record, the structure and content of the data, the questionnaires instructions used for data collection, any changes to processes and procedures. With a view to building an integrated national statistical system consolidating several administrative data sources as well as surveys, censuses, and other sources (such as a Labour Market Information System -LMIS-), documentation and metadata on all sources should follow common standards, methods, and rules. For convenience and transparency, they should also be stored in a common repository and referenced to the data whenever possible, easily accessible to anyone involved in data collection or production activities.

- **Cooperation between record keepers and statistics producers**

Naturally, the producers of labour statistics will not have the same level of control or influence over the contents and methodologies of administrative data collections that they do for survey data collections. Indeed, it is not guaranteed that registers will cover the relevant units, variables, and data items in general for statistical needs. Also, owners and keepers of administrative registers may benefit from maximizing the statistical potential of registries, by using the results to improve their processes and inform their decision-making.

The close cooperation between record keepers and statistics producers is essential to the statistical use of administrative registers, which in turn is for the common (statistical) good. Eventually, and without forgetting the core administrative purposes of the records, statisticians and record keepers may agree on methodology improvements to optimize data quality (and perhaps even changes to data content).

A national statistics act governs in many countries the sound cooperation between the competent statistical authority and owners of relevant administrative records.

3.2. From administrative to statistical registers

Given that the records were created for administrative purposes and not statistical ones, they have to undergo certain statistical processes for the administrative data to become reliable statistics, including the study of definitions, concepts and classifications used, editing the data to correct any inconsistencies found, coding variables and creating new ones if need be, and deciding on the imputation method to apply for the treatment of missing data and units. Data processing to meet statistical needs covers data cleaning, editing, validating, and quality assurance.

Although administrative records may have protocols in place to ensure data quality, these may not follow the same criteria than those established by the competent statistical authority for official statistics, which is why statistical processing is required.

3.3. Factors influencing the statistical use of administrative data

3.3.1. Legal framework

The [Fundamental Principles of Official Statistics](#) adopted by the United Nations Economic and Social Council in 2013 stress that, in order to be effective, the fundamental values and principles that govern statistical work have to be guaranteed by legal and institutional frameworks and respected at all political levels and by all stakeholders in national statistical systems.

Numerous countries across regions have a statistical act or law providing the basis for the production and dissemination of official statistics. The UN Statistics Division keeps a repository of [Legislation Country Practices](#) sharing materials on legislation governing official statistics around the world. This repository covers 195 countries and territories at the time of writing, conveying the commonality of national statistical acts.

Legislation such as a national statistical act provides a key foundation for the use of administrative records for statistical purposes. Indeed, legislation stipulating the statistical use of administrative records provides a formal framework for the efforts and can promote the public's understanding and approval. Legislation can greatly facilitate the statistical use of administrative records by giving the legal mandate to access and treat administrative data to the national authority.

Legislation can emphasize the need for cost-efficiency in the production of official statistics, including avoiding duplication of data collection efforts (for instance, facilitating the use of existing administrative data by the national statistical authority to avoid inclusion of the same data items in an additional data collection exercise).

Legislation establishes official statistics as a common public good, and promotes the use of all available sources, including administrative registers, for this purpose. Legislation can explicitly give the national statistical authority the right to access relevant administrative registers with the level of completeness and detail required, setting appropriate data confidentiality and protection mechanisms. The national statistical authority would be granted access to administrative data for statistical and research purposes only.

In some cases, legislation can also stipulate the use of unique identifiers, ensuring their widespread use by all relevant agencies.

In addition to national legislation governing official statistics, there may be other laws and regulations guiding the statistical use of administrative registers, including the mandate, mission, data confidentiality and data protection rules of the agency owning and keeping each administrative register. Legal protections may hinder data sharing and linking, and legal interpretation of these rules may be necessary to reach an agreement on the responsible use of administrative data by the national statistical authority.

3.3.2. Institutional collaboration, communication, and exchange

Institutional collaboration and open communication are fundamental for the development and good functioning of a national statistical system taking advantage of existing administrative data sources. Close collaboration and communication channels are needed between the owners and keepers of administrative records and the national statistical authority in charge of official statistics production to ensure administrative data sharing is done in a timely, complete, and reliable manner.

This also applies to Labour Market Information Systems (LMIS), which require institutional arrangements and strong coordination among various actors to ensure timely access to data from a wide range of sources, including administrative registers. The actors involved and the nature of the institutional arrangements depend on the country context, but close and sustained collaboration between key labour market stakeholders and statistical stakeholders is always needed to ensure the performance and sustainability of a LMIS. Typically, a comprehensive LMIS will involve the National Statistical Office, the Ministry of Labour, as well as other relevant ministries and other government entities, employment services, labor inspection services, workers' organizations, and employers' organizations, among others.

Institutional collaboration and data exchange can be guaranteed or enhanced through formal agreements involving all relevant partners. Indeed, uneven collaboration across agencies or areas will result in data gaps or

differences in data quality and timeliness across topics. The lack of adequate coordination mechanisms and institutional arrangements can lead to difficulties in accessing data from relevant administrative registers by the national statistical authority. Coordination mechanisms should be formalized in institutional agreements that establish the main characteristics of the collaboration and the responsibilities of each party, to minimize the risk of coordination problems. Formal agreements also promote the sustainability of the collaboration, ensuring the continuity of the statistical production based on administrative registers.

In this respect, data use agreements, data sharing agreements, or memoranda of understanding may be established to agree on what data will be shared, for what purpose, by when, for how long, under which access restrictions, etc. In this sense, it is crucial to consider the legal framework governing the use of the data.

The United States Federal Committee on Statistical Methodology, dedicated to improving the quality of federal statistics, lead a project to examine successful data sharing experiences and draw lessons to promote such agreements for the statistical use of administrative data. The core challenges identified for the establishment of data sharing agreements include financial challenges, legal challenges, technical challenges, the management of internal processes, and the management of interagency relations. The elements of success identified include strong high-level commitment in the involved agencies to the process of data sharing for statistical use of administrative records, having narrow but flexible goals for the agreement, assigning sufficient resources and infrastructure to the efforts, and finding mutual interest of the agencies involved. In other words, to be successful, data sharing arrangements should benefit all agencies involved, and these should devote reasonable resources to implement them.

The lack of sustained institutional collaboration and communication can lead to parallel data systems being developed by different agencies, hindering the reliability of each system. A high-level explicit commitment to create a unique integrated register-based statistical system ensures the centralization of efforts as well as their sustainability. The lack of high-level commitment means that any institutional, managerial or leadership change could hinder or even stop the process. This commitment must be formalized to ensure its continuity even after institutional or leadership changes.

A good example of this was seen in recent years in New Zealand, where the Central Government sent a clear message on the need to minimize the response burden of businesses in data collections. In 2003, a Parliament government compliance cost committee required Statistics New Zealand to assess the compliance costs of all its information requests and consider the use of alternative existing data sources. This request pushed Statistics New Zealand to enhance the use of administrative data for the production of official statistics, with a clear high-level mandate. The strategic outcomes of Statistics New Zealand include maintaining the cooperation of data providers with more effective use of existing data sources, particularly administrative data.

3.3.3. Consolidated data systems

Having a consolidated data system greatly eases and enhances the centralized and integrated use of administrative records for official statistics. Consolidated data systems refer to the consolidation and harmonization of everything related data production and dissemination, including data attributes, variables, classifications, and formats. Indeed, documentation and metadata are essential components of data systems. Consolidated data systems also imply pre-established management processes which ensure respect of data privacy, confidentiality, and quality criteria.

Ideally, these core harmonized guidelines (concepts, definitions, variables, classifications, etc.) should be established at the beginning of the system design process, to avoid costly or laborious changes later on, once the system is already in operation. A strong data governance structure is needed to ensure the implementation of these agreed harmonized guidelines.

Integrated register systems link together various administrative registers to create a more comprehensive, complete, and relevant database. Harmonization of methodologies and procedures across agencies and within agencies facilitates the data linking process. A high-level commitment to create the centralized data system, an official recognition of the leading agency and other agencies involved in the process, and the assignment of sufficient resources are all central elements in the process.

One major factor simplifying the production of statistics from administrative registers is the use of unified identification systems across records, agencies, and sources, especially for the linkage and integration of various

administrative registers and other sources. Nordic countries have a widespread use of unified identification systems, including unified personal identity codes. This was the cornerstone for their early development of register-based statistical systems.

In the absence of unified identification systems, different linking methods can be used to integrate administrative registers with other registers or other sources, including deterministic and probabilistic methods. In a probabilistic method (the most rigorous), in order to determine whether two units from different registers correspond to the same one, several variables are used (typically first name, last name, and date of birth for persons, among others).

3.3.4. Inventory of administrative registers

Naturally, one key factor determining the potential for production of official and labour statistics from administrative registers is the existence of a robust set of relevant administrative records, covering topics of interest with adequate levels of timeliness and completeness.

In this sense, it is useful to count on an inventory of administrative registers listing and characterizing the registers available, to assess their potential statistical use. This inventory is a crucial input for the implementation of an integrated system of statistical registers.

To build this inventory, it is necessary to identify all relevant available administrative registers and do a quick assessment of their potential statistical value based on some basic minimum criteria. Brief identification cards can be created for each relevant administrative register, covering basic information on availability of the data, the design of the record, the metadata, and quality assurance (Segui 2016).

The United Nations Economic Commission for Latin America and the Caribbean developed (under the auspice of the Knowledge Transfer Network of the Statistical Conference of the Americas) a brief questionnaire to be sent out to all relevant agencies to collect information on administrative registers kept by the responding agency and assess their potential for statistical production. This questionnaire was tested by many countries in the Latin American region, typically administered by the national statistical office. The country experiences are described in one of the open resources presented in section 5.4.

3.3.5. Transparency and awareness

The general attitude by government and the public towards the statistical use of administrative data is a key factor underlying the feasibility and success of the endeavor. Therefore, increasing public and government awareness of the benefits of statistical use of administrative data is essential, highlighting the gains in terms of cost-efficiency and reduced response burden. The public should be reassured that data is used for specific and reasonable purposes, in line with national legislation, as well as data confidentiality and data protection regulations. The statistics produced should be made available to the public, alongside relevant metadata and documentation. Standards and methods should be open and transparent.

Trust is also central to data sharing across agencies. Openness and transparency of the processes, data visibility, close collaboration, and favourable leadership all contribute to building positive attitudes towards data sharing. In fact, trust in government and a general sense of transparency is often quoted as one of the pre-conditions that in Nordic countries greatly eased the use of administrative data for statistical purposes.

3.3.6. Optimization of administrative tools for statistical purposes

Administrative records are numerous and varied, but in all cases the data collection procedure will greatly determine data quality, accuracy, and completeness.

The actual tool used for data collection for administrative records is typically a user form or questionnaire, to be filled in by a variety of respondents (individuals, insured persons, victims of accidents, institutions, associations, employers, establishments, etc.) and submitted electronically, via mail, or in person, depending on the register's guidelines.

Some examples of data collection activities in administrative records (useful to labour statistics) are union membership forms filled in by individuals to join a union, information recorded by labour inspectors during inspection visits in their forms, enrolment forms for the educational system, and claims for unemployment benefits.

The use of modern, electronic equipment for data collection and registration can always be beneficial. It allows to systematize data collection procedures. Data quality can be greatly improved by minimizing as much as possible the manual entry of data, and hence reducing the risk of typos. Moreover, having an electronic system considerably accelerates the process of producing statistics, since the information could be transferred to the competent persons or agencies much faster. Having all the data gathered in electronic format also facilitates running consistency checks, spotting errors, treating, using, and interpreting the data.

3.4. Data quality assurance

Data quality is a major concern for official statistics production, including from administrative registers.

Data accuracy, completeness, reliability, relevance, and timeliness, all contribute to data quality. The data quality assurance process seeks to guarantee that the statistics disseminated respond to minimum quality requirements through data cleaning, editing, and validation.

Many countries around the world have their own data quality assurance frameworks or guidelines for official statistics. What is more, some countries have even adopted data quality assurance frameworks specifically for administrative data. For example, in the United Kingdom, the Administrative Data Quality Framework (ADQF) was developed by the Methodological Research Hub and the Data Quality Hub, both based in the Methods and Quality directorate of the Office for National Statistics. In New Zealand, Statistics New Zealand developed the *Guide to reporting on administrative data quality*, which includes, a framework for understanding strengths and limitations of administrative datasets. In Canada, a framework for the evaluation of administrative data was conceived by Statistics Canada covering data relevance, accuracy, timeliness, accessibility, interpretability, and coherence.

Some quality assurance frameworks for administrative data have also been developed by regional or international agencies, such as the *Toolkit for Quality Assessment of Administrative Data for Official Statistics* by UN Women and the UN Statistics Division, created as an instrument for countries that do not have their own quality assessment systems (yet) to be able to assess administrative data sources.

At the international level, the UN National Quality Assurance Framework for Official Statistics adopted in 2019 by the Statistical Commission is the core instrument to assess data quality in different country contexts. It covers all sources of official statistics, including administrative data. It is aimed at assuring the quality of official statistics throughout the entire national statistical system.

Additionally, the Generic Statistical Business Process Model (GSBPM) provides a framework listing all stages followed by national statistical offices in the process of producing statistics, in a standardized manner. The GSBPM is intended to apply to all statistical production activities (either by national or international statistics producers) and all data sources. However, to increase its usefulness for administrative data sources, some agencies have undertaken an adaptation of the GSBPM for administrative registers, including the Uruguayan National Institute of Statistics.

► United Nations recommendations on statistical quality assurance

The following recommendations are based on the 10 Fundamental Principles of Official Statistics², and are meant to guide and support national statistical offices and other agencies involved in national statistical systems in assuring the availability and quality of official statistics.

- **Recommendation 1.** Legal and institutional frameworks should guarantee the fundamental values and principles of official statistics, to be respected by all stakeholders in national statistical systems.
- **Recommendation 2.** National statistical legislation and other legislation mandating the production of statistics for official use should include the requirements of quality assurance.
- **Recommendation 3.** Countries should establish a national quality assurance framework for official statistics and all members of the national statistical system should continually assess, improve, and report on the quality of official statistics.
- **Recommendation 4.** The national quality assurance framework for official statistics should be in line with or consider the United Nations National Quality Assurance Framework or similar frameworks.
- **Recommendation 5.** The national quality assurance framework should be implemented at the national statistical office and throughout the entire national statistical system, and where applicable, to data and statistics produced outside of the national statistical system.
- **Recommendation 6.** Statistics at all levels, including the local level, should be planned, designed, developed, produced and disseminated on an impartial basis, and according to strictly professional considerations.
- **Recommendation 7.** Statistics and data should be presented in a way that facilitates their correct interpretation, which implies that appropriate metadata, such as that relating to data sources, methods and procedures used, needs to be made available in conjunction with the released data or statistics.
- **Recommendation 8.** All members of the national statistical system should comment on erroneous interpretation and misuse of official statistics in their respective subject-matter domain(s) of statistical production and dissemination, as required.
- **Recommendation 9.** Countries should include in their statistical laws: (a) the mandate of producers of official statistics to collect needed information to compile statistics directly from respondents if it is not already available in the national statistical system and cannot be obtained from existing data; and (b) the entitlement to select data sources based on professional considerations, including “new” sources such as big data.
- **Recommendation 10.** Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, should be kept strictly confidential and used exclusively for statistical purposes.
- **Recommendation 11.** The laws, regulations, and measures under which the national statistical systems operate should be made public.
- **Recommendation 12.** The work of the national statistical system should be properly coordinated to achieve consistency and efficiency. Therefore, the statistical law of the country should: (a) clearly identify the roles and responsibilities of the individual members within the national statistical system; and (b) establish a body that is responsible for managing and coordinating system-wide activities, including promotion of the national quality assurance framework.
- **Recommendation 13.** The national statistical office and all other members of the national statistical system, including members at the local level, should use international statistical concepts, classifications and methods, thereby ensuring the consistency of official statistics and the efficiency of statistical systems at all levels.
- **Recommendation 14.** Countries should participate in bilateral and multilateral cooperation in statistics to improve official statistics in all countries.

Source – United Nations. Department of Economic and Social Affairs. 2019. *United Nations National Quality Assurance Frameworks Manual for Official Statistics - Including recommendations, the framework and implementation guidance*. ST/ESA/STAT/SER.M/100. New York: UN.

► Principles of the United Nations National Quality Assurance Framework

The UN National Quality Assurance Framework includes 19 principles of data quality, organized into four categories, as presented below.

- **Level A. Managing the statistical system**
 - **Principle 1.** Coordinating the national statistical system
 - **Principle 2.** Managing relationships with data users, data providers and other stakeholders
 - **Principle 3.** Managing statistical standards
- **Level B. Managing the institutional environment**
 - **Principle 4.** Assuring professional independence
 - **Principle 5.** Assuring impartiality and objectivity
 - **Principle 6.** Assuring transparency
 - **Principle 7.** Assuring statistical confidentiality and data security
 - **Principle 8.** Assuring commitment to quality
 - **Principle 9.** Assuring adequacy of resources
- **Level C. Managing statistical processes**
 - **Principle 10.** Assuring methodological soundness
 - **Principle 11.** Assuring cost-effectiveness
 - **Principle 12.** Assuring appropriate statistical procedures
 - **Principle 13.** Managing the respondent burden
- **Level D. Managing statistical outputs**
 - **Principle 14.** Assuring relevance
 - **Principle 15.** Assuring accuracy and reliability
 - **Principle 16.** Assuring timeliness and punctuality
 - **Principle 17.** Assuring accessibility and clarity
 - **Principle 18.** Assuring coherence and comparability
 - **Principle 19.** Managing metadata

Source – United Nations. Department of Economic and Social Affairs. 2019. United Nations National Quality Assurance Frameworks Manual for Official Statistics - Including recommendations, the framework and implementation guidance. ST/ESA/STAT/SER.M/100. New York: UN.

4. Uses of administrative registers for labour statistics

4.1. Administrative registers as a direct source of labour statistics

Perhaps the most obvious and straightforward use of administrative data in the field of labour statistics is as a direct source. That is, using administrative registers as a direct stand-alone source of labour statistics in the absence of other sources or for indicators which cannot be produced from other sources.

Indeed, administrative registers are particularly useful in contexts where there are no regular household or establishment surveys in place, to serve as an alternative source of labour statistics. In these cases, administrative registers help bridge data gaps and provide some of the much-needed labour statistics to monitor labour markets. Namibia's experience (described in the box below) follows this idea to take advantage of administrative data in the absence of a recent labour force survey.

► Namibia – Administrative data in the absence of a recent labour force survey

In Namibia, the labour force survey is recognized as a central source of labour statistics, key to support policy formulation and labour market analysis. Indeed, there was an intention to conduct a national labour force survey on an annual basis, but financial and resource constraints hindered these plans. The survey was carried out every year from 2012 to 2014, and every two years from 2014 to 2018, when the last round took place. There has not been a labour force survey since 2018.

Recognizing that the irregular frequency of the national labour force survey is not enough to meet the needs for labour statistics of policy and decision-makers, the Namibia Statistics Agency is exploring the use of administrative data to support labour market monitoring. In collaboration with the Ministry of Labour, the Social Security Commission, and other relevant agencies, the Namibia Statistics Agency envisages the integration of different administrative registers to feed a central system of regular and timely labour statistics in the absence of a recent labour force survey, in the context of the establishment of a LMIS.

Of special interest to this project are the records kept by the Ministry of Labour in relation to the Namibia Integrated Employment Information System (including information on vacancies, employers, employment by occupation, and jobseekers), the Occupational Safety and Health Inspections (including information on inspections conducted by industry, occupational accidents, and injuries), and the Employment Equity Commission (including information on affirmative action).

Source – Presentation by Namibia Statistics Agency for the [Expert clinic on the Production of labour statistics using administrative data](#), organized by the Collaborative on Use of Administrative Data for Statistics in May 2021.

When the Covid-19 pandemic hit and mobility restrictions were enforced around the world to try to contain it, field data collections for labour force surveys based on face-to-face interviews were abruptly halted, affecting the regular production of labour indicators. Some countries could turn to remote modes of data collection (such as computer-assisted telephone interviews) but this was not an option everywhere, since it required the existence of reliable contacts databased with telephone information for sampled households. Other countries, such as the United Arab Emirates and Malaysia, resorted to administrative data to produce basic labour market indicators given the impediments to the labour force survey.

It should also be noted that for some specific labour-related topics, administrative registers are the preferred source, or even the only possible source. There are some topics for which labour force surveys could never provide data (or at least not reliably), given their nature and statistical unit.

In fact, countries have committed to reporting on the Sustainable Development Goals Global Indicator Framework with a view to monitoring progress towards the achievement of the goals. Within this framework, SDG indicator 8.8.1 relates to the rate of fatal and non-fatal occupational injuries per 100,000 workers. Data for this indicator would typically come from administrative records such as labour inspection records, insurance records, or records kept by the labour ministry. Likewise, the preferred source for SDG indicator 1.3.1 on the proportion of the population covered by social protection systems are social insurance records.

In Peru, the register that employers are mandated to use to inform the competent authority on its employees, retired employees, contractors, and trainees serves as a source of valuable statistics on several labour-related topics hardly covered by the labour force survey. This experience is described in more detail in the box below.

► **Peru – Electronic forms submitted by employers supply data on many labour-related topics**

In Peru, employers are required to fill in and submit the *Planillas Electrónicas* (electronic forms) on a monthly basis. These are documents maintained by each employer, recording information on employees, pensioners, service providers, trainees, etc., in order to give the workers access to the services of the Social Security system. These forms are submitted (via an electronic system) every month to the *Superintendencia Nacional de Aduanas y de Administración Tributaria* (tax agency of Peru), responsible for the *Planillas Electrónicas* jointly with the Ministry of Labour and Employment Promotion. They include information on each worker's nationality, educational level, occupation, economic activity, pension system, type of labour contract, income, deductions applied by the employer, union membership and coverage by collective agreements, among others.

Based on this data, it is possible to produce reliable statistics on many labour-related topics, including union membership and collective bargaining coverage, with useful disaggregations.

Source – [Labour indicators at district level](#), by Labour and Employment Promotion Ministry of Peru, retrieved on September 12, 2023.

Administrative registers can also be used to produce labour statistics to be disseminated and interpreted alongside key labour market indicators derived from the labour force survey. That is, the analysis of some labour market aspects can be enriched by the joint use and interpretation of indicators from different sources.

For example, it is widely recognized that official unemployment statistics refer to the internationally-agreed definition of persons not in employment, available for employment, and actively seeking employment. For this indicator, the preferred source is a labour force survey. However, for policymakers, it may be useful to also consider the number of unemployment claimants (or registered unemployment). That is, persons receiving unemployment benefits according to social insurance records or employment office records. The case of Netherlands provides an example of the joint use of these indicators (described in the box below). Sufficient metadata should accompany the data to make the users aware of the differences between indicators when disseminated jointly.

► **Netherlands – Joint dissemination of unemployment and registered unemployment data**

The Central Bureau of Statistics (CBS) of Netherlands publishes both the official unemployment estimate (from the labour force survey) and data on those receiving unemployment benefits (from the administrative register) in the same monthly release, highlighting the comparison of the two indicators' trends.

Although the indicator given prominence is the official unemployment estimate from the labour force survey, its analysis alongside recipients of unemployment benefits adds value to the monitoring of labour market issues and the understanding of the conditions of the unemployed.

It is interesting to note that the two data series generally follow similar trends, but in 2022 the official unemployment estimate increased while the unemployment benefits series decreased.

Source – [Unemployment release](#) by the CBS of Netherlands, retrieved on September 12, 2023.

At the height of the Covid-19 pandemic, recognizing the increased need for relevant labour statistics allowing policymakers to react to an unprecedented labour market shock, some countries decided to disseminate additional indicators derived from administrative registers to supplement the labour force survey indicators. For instance, in Australia, although the labour force survey data collection was not too affected by mobility restrictions since it was already done mostly remotely, the Australian Bureau of Statistics started releasing administrative data on paid jobs and wages on a fortnightly basis. Similarly, Denmark increased its use of administrative data for labour statistics, disseminating daily data on newly registered unemployed (although disruptions to the labour force survey had not been major).

It should be noted that, even when used as direct source of labour statistics, administrative registers should be consistent with other registers and sources (where applicable) in terms of concepts, definitions, classifications, reference periods and methodologies.

4.2. Administrative registers to enhance survey data quality and cost-efficiency

In addition to their use as a direct independent source of labour statistics, administrative registers can also be used in combination with other sources, such as censuses and surveys. Even without going as far as the full combination of administrative and survey data with data integration and linkages (presented in section 4.3), administrative data can be used to increase surveys' data quality and cost-efficiency. In the field of labour statistics, this use of administrative data typically relates to supporting and improving the labour force survey.

Administrative records can be very useful in sampling and stratification processes. Indeed, a comprehensive, timely, official population register can serve as a sampling frame for the labour force survey. If the population register is used in combination with other registers containing information on additional individual characteristics, samples may be drawn for desired sub-groups.

When the Covid-19 pandemic hit and mobility restrictions impeded face-to-face data collection for labour force surveys, countries such as Chile and Ecuador explored the use of administrative data to identify contact information of selected households to reach them by telephone. The experience of Ecuador is further described in the box below.

► Ecuador – Improving the labour force survey data with administrative data

The National Institute of Statistics of Ecuador (INEC, for its acronym in Spanish) sought to improve the quality of contact and identification data of respondents to the National Survey of Employment, Unemployment and Underemployment (ENEMDU, for its acronym in Spanish) with administrative data kept by the Civil Registry.

By comparing the identification data of ENEMDU respondents to the data contained in the Civil Registry administrative record, the INEC was able to confirm and/or recover the identity of respondents and integrate the survey information with other administrative sources for a richer analysis. This allowed the generation of variables not included in the ENEMDU and improvements to data editing and imputation processes for outliers or missing data.

The matching of these two data sources was possible for 76 per cent of those interviewed in the ENEMDU-COVID19 (a special round of the ENEMDU to measure the socioeconomic impact of the Covid-19 pandemic). Therefore, it was possible to integrate data of 76 per cent of surveyed individuals with administrative records, to add variables or complete missing information and to improve the labour market analysis with data on social security, health or other administrative records and the impact of the pandemic. The variables used to confirm individuals' identity were their names, surnames, age, and sex, in addition to their identification number.

Naturally, this data integration was done for purely statistical purposes and respecting data confidentiality.

Source – Case study of Ecuador in the [Inventory](#) of the Collaborative on the Use of Administrative Data for Statistics.

After survey data collection, administrative records can be used to verify and complement the information, enhancing data quality and widening the breadth of variables covered.

Comparing the survey data collected to the administrative data in the register will allow for the correction of some errors arising from the data entry. Similarly, administrative data can be used to compensate for non-response or missing information. For instance, in Australia, administrative data was used to improve occupancy determination and imputation for the 2021 population census. For non-responding households, administrative data was used to determine whether the corresponding dwelling was occupied, and if it was, to impute data from similar households.

Nevertheless, it should be mentioned that it is not recommended to impute information on data used to identify the individuals' labour status as it is considered necessary to base this on information provided in response to a survey applying the latest statistical standards. Indeed, European Union regulations prohibit the use of external data sources for this purpose. Therefore, the use of administrative data to impute information for a labour force survey should be limited to additional information where there is a match in the concepts being applied, for

example, where registers exist on education level, migration status, nationality, etc. matching the concepts that would be applied in a survey.

Administrative data can also inform post-stratification techniques such as calibration. In calibration, data on population distribution from the administrative register is used to estimate new weights with a view to reducing sampling error.

Furthermore, sample surveys will usually not provide reliable estimates for small areas. Data from administrative registers can be used to increase the granularity of survey data and enable the production of estimates for small areas of interest.

Some administrative records are updated in real time, providing more timely data than labour force surveys. In these cases, the timeliness of administrative data can be used to the benefit of labour force survey data, gauging for patterns and trends. The case of unemployment statistics in Ireland presented in the box below relates to taking advantage of the timeliness of administrative data.

► Ireland - Unemployment estimate updated with the registered unemployment trend

The Central Statistics Office (CSO) of Ireland regularly disseminates both unemployment statistics from the labour force survey, considered the official source, and data on those claiming unemployment benefits from the administrative source known as the Live Register, based on data access agreements with the Department of Social Protection.

Each month, the CSO releases a monthly unemployment estimate series using the most recent labour force survey quarterly estimate as a benchmark. That estimate is then updated using the trend in the Live Register series of unemployment benefit claimants. That is, for example, if the Live Register estimate increases by 1 per cent, this percentage increase is applied to the latest labour force survey benchmark for unemployment.

When the new labour force survey results become available, the corresponding monthly series is revised.

Here, both sources of information are used for the value of their joint analysis, but also, their integration allows to benefit from the timeliness of the administrative data and the reliability of the official labour force survey unemployment estimate.

The decision to use the Live Register for short-term forecasting of unemployment data (subject to revision by the following labour force survey results) was informed by studies showing that the short-term trends of the two series generally coincided. However, since it may not always be the case, revisions may be needed.

Source - [Unemployment release](#) by the CSO of Ireland, retrieved on September 12, 2023.

4.3. Combined use of administrative and survey data for integrated statistics

Data from one or more administrative registers can be linked to data from surveys such as the labour force survey to produce an integrated, consolidated data system. This data linkage and integration allows to take full advantage of the strengths of each data source while mitigating their individual limitations.

Indeed, the accuracy and timeliness of administrative data adds to the targeted nature of survey data in terms of the topics and variables covered. Administrative data can supplement survey data, providing additional variables not covered by the survey data collection. What is more, this data integration allows to have shorter labour force survey questionnaires since it is no longer necessary to collect information via the survey on the items covered by administrative records, making the labour force survey more cost-effective. Also, using administrative records for certain variables may enhance data quality since respondents may not have as accurate information as the records, or sensitivity or recall issues may affect their answers.

In Finland, for example, the labour force survey data collection focuses on the most essential labour-related questions such as labour force participation and working time, while demographic and educational information is pulled from administrative registers.

In integrated statistical systems, the contents of each data source can be limited but their integration maximizes their potential, broadening the scope in terms of topics, variables, and units covered. The boxes included in the following pages present some examples of data integration for labour statistics in New Zealand, the United States, Uruguay, and Colombia.

However, it should be noted that data linkage and integration require that the various datasets concerned have some overlapping information on which to base the matching procedure. The ideal situation here would be that all data sources use the same set of unique identifiers, but in their absence, combinations of available information can be used (such as name, sex, date of birth, address, etc.). While linking at the micro-level is the most obvious (or straightforward) way to do it, there is also the possibility to do the integration at the aggregate level - for instance, linking one source with income information to another with labour force information using common disaggregations to present a wider set of data, or to model or impute some values from one to the other.

The national statistical office is usually the best-placed agency to lead these efforts, not only given its statistical expertise and mandate but also based on legislation which often grants it access to non-anonymized administrative data (and other sources) for statistical purposes and under non-disclosure rules.

► **New Zealand – The Linked Employer-Employee Database (LEED) and the Integrated Data Infrastructure**

A number of recent initiatives in the use of administrative data for the production of official statistics in New Zealand have involved the integration of administrative data (with other administrative data or survey data) to fill a specific information need or increase the explanatory power of an administrative collection. One example of this is the Linked Employer-Employee Database (LEED), which results from the linking of taxation system data with business data to measure labour market dynamics and provide insight into the operation of the labour market. The initial success of the LEED has led to its gradual expansion to include data on benefit type for individuals receiving an income-replacing main benefit, and potentially tertiary students' enrolments and completions data.

Official quarterly statistics produced from the LEED measure labour market dynamics, and their level of detail helps explain employment changes seen in other official labour market statistics. Indeed, the LEED improves the knowledge base for policy formulation in employment, business performance, productivity, firm creation and growth, employee turnover, and many other areas.

In a broader way, Statistics New Zealand uses data linking to maximize the use of the data it manages. For this purpose, it created the Integrated Data Infrastructure (IDI), which is a large research database. It holds de-identified microdata about people and households, covering a wide range of topics such as population, education, income, work, benefits, social services, housing, migration, justice, and health. The information comes from government agencies, Statistics New Zealand surveys, and non-government organisations (NGOs). The data is linked together, or integrated, to form the IDI.

Naturally, the IDI respects data privacy and confidentiality.

Source – [Official website](#) of Stats NZ Tauranga Aotearoa, retrieved on September 12, 2023.

► United States - The Longitudinal Employer-Household Dynamics (LEHD) program

The Longitudinal Employer-Household Dynamics (LEHD) program of the United States Census Bureau combines federal, state and Census Bureau data on employers and employees under the Local Employment Dynamics (LED) Partnership, with a view to providing policy and decisionmakers with key timely information.

The LED Partnership was created in 1999 as a voluntary federal-state partnership to merge data from workers and from employers to produce the Quarterly Workforce Indicators (QWI). Under the LED Partnership, states share Unemployment Insurance earnings data and the Quarterly Census of Employment and Wages data with the Census Bureau. The LEHD program combines these administrative data, additional administrative data and data from censuses and surveys, to create statistics on employment, earnings, and job flows with geographical and industrial granularity.

LEHD data products include online applications, public-use data, and restricted-use microdata, which are all available online for public use, and of great value to researchers and labour market analysts.

Source – United States Census Bureau, [Longitudinal Employer-Household Dynamics](#), retrieved September 12, 2023.

► Uruguay – Building an integrated register statistical system with a register-based wage survey

The Uruguayan National Statistical Institute set out to build a register-based statistical system based on the Nordic model, with three core base registers: the population register, the business register, and the real estate register. Unique unit identifiers allow to link base registers.

As part of this endeavour, the National Statistical Institute built a wages and activities register, fed with data on public sector employees from the Ministry of Finance and on private sector employment from the Ministry of Labour and the Social Security Authority, through formal data exchange agreements. Once this register reached a satisfactory level of maturity with assurance of data quality, it was the basis for revising and improving the wages and salaries survey. Indeed, it is estimated that records in the wages and activities register cover around 87 per cent of public sector workers and 99 per cent of private sector workers. As of March 2022, the wages and salaries survey used to produce the wage index has become a register-based survey.

In a broader way, the National Statistical Institute combines survey data (including labour force survey data) with administrative data to increase the quality, cost-efficiency, and topic coverage of labour market statistics. Linking the population, business, and activities registers, it is able to produce, for instance, valuable statistics on labour relations and labour market transitions. Combining labour force survey data and administrative data by the Ministry of Labour, it is able to increase the quality of formal employment statistics by sex, age and country region, since labour force survey data is thought to over-estimate formal employment. Combining labour force survey data with various administrative records also allows to analyse the targeting and impact of specific cash transfer programs.

It should be noted that there is in Uruguay an Integrated Information System on Social Areas, kept by the Ministry of Social Development integrating data from over 60 administrative registers from over 15 agencies. The National Statistical Institute uses this system to feed the base population register.

Source – Presentation by expert of Uruguay's National Statistical Institute for the [Expert clinic on the Production of labour statistics using administrative data](#), organized by the Collaborative on Use of Administrative Data for Statistics in May 2021, and [official website of the National Statistical Institute](#), retrieved on September 12, 2023.

► Colombia – Integrating social security records and the labour force survey to meet new data needs

The Great Integrated Household Survey (GEIH, for its acronym in Spanish) is at the core of the national statistical system in Colombia, resulting from the integration of three previous household surveys: the Continuous Household Survey, the National Income and Expenditure Survey, and the Life Quality Survey. The motivation behind this integration was to increase the cost-efficiency of the statistical collection while increasing data quality and keeping a vast coverage of topics. The Labour Market Continuous Household Survey became the central component of the GEIH. The GEIH data collection has been conducted continuously since 2006.

In 2019, the National Administrative Department of Statistics (DANE, for its acronym in Spanish) started a thorough revision of the GEIH, to update it in line with latest international labour statistics standards and emerging statistical needs.

When the Covid-19 pandemic hit, the GEIH data collection mode had to be abruptly changed from face-to-face interviews to computer-assisted telephone interviews. Adapting to this new mode of data collection under strenuous circumstances meant reducing the household survey questionnaire, which affected the calculation of the informality rate, a key labour market indicator in Colombia.

Therefore, the DANE set out to estimate informal employment using a predictive model linking administrative data from the social security record to the survey data.

In addition, the DANE explored the use of administrative data on pensions and government allowances for imputation of household income sources, which was the basis for the calculation of poverty and inequality rates for 2020.

The DANE has also created the RELAB, a statistical register of labour relations with information on dependent and independent labour relations, paid or unpaid leave, retirement, and income with disaggregations such as by economic sector, age group, and country region. The data results from the integration of records kept by four different entities, with the Ministry of Health as the main holder of the data.

Source – Presentation by expert of Colombia's National Administrative Department of Statistics (DANE) for the [Expert clinic on the Production of labour statistics using administrative data](#), organized by the Collaborative on Use of Administrative Data for Statistics in May 2021, and case study of RELAB in the [Inventory](#) of the Collaborative on the Use of Administrative Data for Statistics.

5. Resources and tools

5.1. Good Practices Database by the UN Statistics Division

The UN Statistics Division maintains since 2000 a [Good Practices Database](#) with material from countries around the world on their official statistics practices. The information is organized according to the [Fundamental Principles of Official Statistics](#), that is, relevance, impartiality and equal access, professionalism, accountability, prevention of misuse, cost-effectiveness, confidentiality, legislation, national coordination, international coordination, and international statistical cooperation. The database includes an additional category for reference materials.

The database features examples of policies and practices around the implementation of these principles in the national statistical system. Many of these examples pertain to the statistical use of administrative data, and the governance of this exercise. Administrative data use for official statistics is particularly showcased in the database category for cost-effectiveness, which covers minimizing reporting burden, form design, and access to administrative records.

5.2. Inventory of Resources by the Collaborative on Use of Administrative Data for Statistics

Recognizing that administrative registers are gaining ground as a source of official statistics and that in connection with this, there is a need for capacity building, the United Nations Statistics Division and the Global Partnership for Sustainable Development Data convened the Collaborative on Administrative Data. This multi-stakeholder collaborative brings together countries, regional and international agencies to share resources, tools, best practices, and experiences. This rich exchange is expected to strengthen the capacity of countries to use administrative records for statistical purposes and raise awareness about the benefits of using administrative registers as a complementary source of official statistics.

The work of the Collaborative is organized into three task teams: Task team 1 on institutional framework, coordination and partnerships, Task team 2 on data management and standardization, and task team 3 on technical interoperability and linking.

The Collaborative set up a rich and useful [inventory](#) to make available materials on recommendations and practical examples of the statistical use of administrative data.

5.3. Tools for Quality Assessment of Administrative Registers

Data quality assurance is a key concern for official statistics, and this is particularly true for official statistics derived from administrative registers given their initial design for administrative purposes and not statistical ones. We mentioned in section 3.4 the existence of frameworks for data quality assurance both at the national and international levels. In order to implement such frameworks in the field of administrative data, some practical tools have been created. Some useful examples are listed below.

- The [Questionnaire to Evaluate the Quality of Administrative Records](#) (CECRA, for its acronym in Spanish) is an instrument to capture information on quality criteria for different stages of the development process of administrative records. It was recommended by the Working Group on Administrative Records in the framework of the Statistical Conference of the Americas. Its development was coordinated by the United Nations Economic Commission for Latin America and the Caribbean.
- The [Self-Assessment of the Quality of Administrative Records](#) (SAQUAR) is an adaptation of the CECRA to capture information on the characteristics and functioning of administrative records in order to assess their potential as a source of data for statistical purposes, in particular, to generate useful indicators for the Labour Market Information System (LMIS).

- The [***Toolkit for Quality Assessment of Administrative Data for Official Statistics***](#) by UN Women and the UN Statistics Division was developed as an instrument for countries that do not have their own quality assessment systems (yet) to be able to assess administrative data sources.

5.4. Open training resources

Given the increasing demand for technical assistance, cooperation, and capacity building in the statistical use of administrative registers, more and more workshops and trainings are being organized around the world, usually with an emphasis on sharing country experiences. The materials and lectures of some of these are openly available and remain as valuable resources. Some key capacity building activities are referenced below.

- Regional seminar on [***Enhancing the use of administrative records for statistical purposes for monitoring the 2030 Agenda***](#), organized by the United Nations Economic Commission for Latin America and the Caribbean in October 2017.
- Expert clinic on the [***Production of labour statistics using administrative data***](#), organized by the Collaborative on Use of Administrative Data for Statistics in May 2021.
- United Nations Expert Group Meeting on the [***Draft handbook on register-based population and housing censuses***](#), held in December 2021.
- Webinar on [***Experience sharing on the modernization and use of administrative data and records for statistical purposes***](#), organized by the United Nations Economic Commission for Africa in June 2022.
- Online course on the [***Statistical use of administrative records***](#), organized by the Inter-American Development Bank in its online academy in August-September 2023.

6. Concluding remarks and future work

In our current world of rapidly changing labour markets and increasing data needs for evidence-based employment policy making, taking advantage of the full statistical potential of readily available data sources makes absolute sense, especially considering common resources constraints for statistical data collections.

Administrative records have a great potential to provide labour statistics on various topics, in the absence of or in complementarity with labour force surveys. Indeed, administrative registers can be a highly accurate, timely, and cost-effective source of statistics on some specific labour-related topics. However, they cannot replace labour force surveys given their targeted design to measure labour market issues of interest, although they can add value to labour market monitoring and analysis. The combined use of administrative and survey data allows to benefit from the advantages of each data source while mitigating their individual limitations, as long as it is done in a transparent manner emphasising the supplementary nature of the sources.

Administrative records are a key pillar of Labour Market Information Systems, whose performance depend on their coverage of all relevant labour data sources in the country. In fact, there are some indicators relevant to LMIS which can only be derived from administrative data.

Many countries around the world are already producing some sets of official labour statistics based on administrative data, and many others are exploring how to do so in the near future. Countries with different levels of statistical development and infrastructure have undertaken the statistical use of administrative data, highlighting the value of these efforts. Country practices and experiences range from the simple use of some administrative data as a direct source of basic labour indicators in the absence of any other source, to the creation of an integrated register-based statistical system bringing together hundreds of administrative records from across agencies.

However, country efforts to produce official statistics from administrative registers have tended to focus on other areas than labour statistics, probably due to the specific challenges of monitoring the world of work. Although some countries are developing interesting experiences to produce register-based labour statistics, the ILO recognizes the need for more widespread guidance and capacity building.

Taking all of this into consideration, the ILO invites participants to the ICLS to indicate their interest in administrative data use and whether they believe some work to develop further guidance and capacity building would be of value. They are also invited to share with the ILO any country experience they deem relevant in this regard, and to provide guidance to the Office on potential priority future activities in this area.

7. References

- Connelly R., Playford, C., Gayle, V., Dibben, C. 2016. *The role of administrative data in the big data revolution in social science research*, Social Science Research, Volume 59, 2016, Pages 1-12.
- DAAS, Peter and VAN NEDERPELT, Peter. 2012. *49 Factors that Influence the Quality of Secondary Data Sources*. Statistics Netherlands.
- EUROSTAT. 2010. *Business Registers – Recommendations Manual*.
- ILO. 2013. *Decent work indicators - Guidelines for producers and users of statistical and legal framework indicators*. ILO Manual, second edition. Geneva: ILO.
- ILO. 2016. *Guide on the Harmonization of Labour Inspection Statistics*. Geneva: ILO.
- ILO. 2017. *Quick Guide on Sources and Uses of Labour Statistics*. Geneva: ILO.
- ILO. 2020. *COVID-19 impact on labour market statistics - COVID-19 impact on the collection of labour market statistics*. ILO Technical Note. 11 May 2020. Geneva: ILO.
- ILO. 2023. *Labour Market Information Systems (LMIS)*. ILOSTAT. Retrieved September 6, 2023, from <https://ilostat.ilo.org/resources/labour-market-information-systems/>
- NORDBOTTEN, Svein. 2010. *The Use of Administrative Data in Official Statistics – Past, Present, and Future – With Special Reference to the Nordic Countries*. Official Statistics – Methodology and Applications in Honour of Daniel Thorburn, pp. 205-225.
- O'Hara, Amy. 2020. *Model Data Use Agreements: A Practical Guide*. In: Cole, Shawn, Iqbal Dhaliwal, Anja Sautmann, and Lars Vilhuber (eds.), *Handbook on Using Administrative Data for Research and Evidence-based Policy*. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab.
- PEMBER, Robert. 1998. *Compilation and presentation of labour statistics based on administrative records*. Bangkok: ILO.
- Segui, Federico. 2016. *Guide for management of administrative registers inventories in national statistical systems*. Statistics Finland.
- Statistics Finland. 2004. *Use of Registers and Administrative Data Sources for Statistical Purposes – Best Practices of Statistics Finland*.
- Thomsen, I., Kleive Holmøy, A. 1998. *Combining Data from Surveys and Administrative Record Systems. The Norwegian Experience*. International Statistical Review, August 1998, Vol. 66, No. 2 (Aug., 1998), pp. 201-221.
- UNECE. 2007. *Register-based statistics in the Nordic countries - Review of best practices with focus on population and social statistics*. New York and Geneva: UNECE.
- UNECE. 2015. *Guidelines on Statistical Business Registers*.
- United Nations Department of Economic and Social Affairs, and London School of Economics and Political Science. 2022. *Re-using Administrative Data for Statistics: Case Studies from Five Countries*.
- United Nations Department of Economic and Social Affairs. 2019. *United Nations National Quality Assurance Frameworks Manual for Official Statistics - Including recommendations, the framework and implementation guidance*. ST/ESA/STAT/SER.M/100. New York: UN.
- United Nations Economic and Social Council. 2013. *Fundamental Principles of Official Statistics*. Resolution adopted by the Economic and Social Council on 24 July 2013.
- United Nations Statistics Division. 2022. *Handbook on Registers-Based Population and Housing Censuses*. Background document for the Fifty-third session of the Statistical Commission (1 – 4 March 2022). Second draft – subject to final substantive and copy editing.
- United Nations. 1969. *Methodology and evaluation of population registers and similar systems*. Studies in methods. Series F, n°15. New York: UN.

United Nations. 1969. *Methodology and Evaluation of Population Registers and Similar Systems*.

WALLGREN, Anders and Britt. 2007. *Register-based Statistics – Administrative data for statistical purposes*. Wiley series in survey methodology. Wiley.

Contact details**International Labour Organization**

Route des Morillons 4
CH-1211 Geneva 22
Switzerland

T: +41 22 799 6831
E: statistics@ilo.org