

Methodological document
on statistical use
of economic
administrative records



UNITED NATIONS

ECLAC



Thank you for your interest in this ECLAC publication



Please register if you would like to receive information on our editorial products and activities. When you register, you may specify your particular areas of interest and you will gain access to our products in other formats.

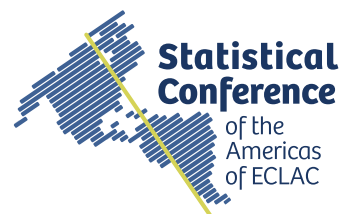


www.cep.al.org/en/publications



www.cep.al.org/apps

Methodological document
on statistical use
of economic
administrative records



This document was prepared by the working group to prepare a methodological document on conceptual aspects and statistical use of economic administrative records of the Statistical Conference of the Americas. The Group was coordinated by Colombia (National Administrative Department of Statistics (DANE)) and Mexico (National Institute of Statistics and Geography (INEGI)), with the Statistics Division of the Economic Commission for Latin America and the Caribbean (ECLAC) as technical secretariat. The following countries are members of the group: Argentina (National Institute of Statistics and Censuses (INDEC)), Brazil (Brazilian Institute of Geography and Statistics (IBGE)), Chile (National Institute of Statistics (INE)), Costa Rica (National Institute of Statistics and Censuses (INEC)), Cuba (National Office of Statistics and Information (ONEI)), Dominican Republic (National Bureau of Statistics (ONE)), Paraguay (National Institute of Statistics (INE)) and Peru (National Institute of Statistics and Informatics (INEI)).

United Nations publication
LC/CEA.11/18
Distribution: L
Copyright © United Nations, 2022
All rights reserved
Printed at United Nations, Santiago
S.21-00969

This publication should be cited as: Working group of the Statistical Conference of the Americas to prepare a methodological document on conceptual aspects and statistical use of economic administrative records, *Methodological document on statistical use of economic administrative records* (LC/CEA.11/18), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC), 2022.

Applications for authorization to reproduce this work in whole or in part should be sent to the Economic Commission for Latin America and the Caribbean (ECLAC), Documents and Publications Division, publicaciones.cepal@un.org. Member States and their governmental institutions may reproduce this work without prior authorization but are requested to mention the source and to inform ECLAC of such reproduction.

Acknowledgements

The following participated in the discussions, research and writing of this document:

- Argentina, INDEC: Javier Domench and Silvia Mario
- Brazil, IBGE: Vinicius Mendonça Fonseca, Synthia Kariny Silva de Santana, Cristiano de Almeida Martins and Raquel Rose Silva Correia
- Chile, INE: Pedro Ruz Zúñiga and Andrea Paola Marín
- Colombia, DANE: José Richard Núñez Alejo, Julieth Alejandra Solano Villa, Sayda Patricia Morera Zarate, Sheila Isabel Centeno Martínez, Mónica Lorena Ortiz Medina and Juliana Benavides Pérez
- Costa Rica, INEC: Odilia Bravo Cambronero and Rosa Cordero Peñaranda
- Cuba, ONEI: Evelyn Martínez Mendoza and Mariana Lobaina Sánchez
- Dominican Republic, ONE: Juan Arias, Rene Guzmán and Orlando Hernández
- Mexico, INEGI: Gerardo Durand Alcántara, Carmen Lizbet Corona Fuentes, Hugo Hernández Ramos, Elsa María Trueba Atienza, Juan Arturo Balderas Ávila, Adriana Herrera Canales and Adolfo Muñoz Torres
- Panama, Instituto Nacional de Estadística y Censo (INEC): Roberto Arosemena
- Paraguay, INE: Mirta Leiva and Rafael Aquino
- Peru, INEI: Marleni Orrillo Huamán and Yessica Panuera Moreno
- ECLAC: Claudio Moris

Contents

Introduction.....	9
Chapter I	
The importance of administrative records	11
Introduction	11
A. Fragmentation and integration	12
Chapter II	
Vision of a system of economic administrative records	15
Introduction	15
A. Basic system of records	15
B. Vision of statistical business register: the system of economic administrative records	16
1. Objectives of a system of economic administrative records	16
2. Objectives of a statistical business register as part of a system of economic administrative records	17
3. Components of the model of the system of economic administrative records	18
C. Statistical business register	19
1. Structured database	19
2. Unique identifier for each economic unit	19
3. Regularly updated	20
4. Territorial area	20
5. Database administered by a national statistical office	20
D. The Generic Statistical Business Process Model: framework for the system of economic administrative records	20
Chapter III	
Specification of needs	23
Introduction	23
A. Subprocesses of needs specification	23
1. Identification of needs	24
2. Consultation and confirmation of needs	24
3. Definition of objectives	25
4. Identification of concepts	25
5. Verification of data availability	26
6. Prepare the general plan and documentation	28
Chapter IV	
Design	29
Introduction	29
A. Conceptual design	29
1. Establish the objects of the register	30
2. Public-use unique identification number for observation units	31
3. Inventory of sources	34
4. Criteria for the selection of sources	35
B. Design of variable descriptions	37
1. Statistical outputs to be generated by the project	37
2. Identification and selection of variables following statistical and conceptual criteria	37
C. Design of collection	39
D. Design of processing and analysis	40
E. Design of production systems and workflows	43
1. Steps in the selection and use of sources	43
2. Steps in the design of linkages between source data and statistical register databases	44
Chapter V	
Construction	45
Introduction	45
A. Definitions at the process level	45

1. Construction of collection instruments.....	45
2. Construction or improvement of components	47
3. Construction or improvement of dissemination components.....	50
4. Organization of workflows.....	50
5. Tests of the production system	51
6. Pilot test of the statistical process	51
7. Finalization of the production system.....	51
B. System modules.....	51
1. Outreach to data providers	51
2. Collection of administrative records.....	52
3. Transformation of administrative records	52
4. Data documentation.....	53
5. Administration.....	53
Chapter VI	
Collection.....	55
Introduction	55
A. Regulatory or legal considerations	56
1. Inter-institutional agreements or memorandums of understanding	56
B. Methodological considerations	58
1. Structure and necessary information	58
2. Periodicity	58
3. Timeliness	59
4. Medium or channel of exchange	59
5. Confidentiality	59
C. Technological mechanism	60
D. Training	60
Chapter VII	
Processing	61
Introduction	61
A. Integration of data	63
B. Classification and coding.....	64
C. Revision and validation.....	65
D. Editing and imputation.....	66
E. Derivation of new variables and units.....	67
F. Finalization of data files.....	67
Chapter VIII	
Analysis	69
Introduction	69
A. Analysis of the business register.....	69
B. Subprocesses of the analysis phase.....	71
1. Preparation of draft results.....	71
2. Validation of results.....	71
3. Interpretation and explanation of results	72
4. Application of control of dissemination	72
5. Finalization of results.....	72
Chapter IX	
Dissemination	73
Introduction	73
A. Updating of output systems.....	73
B. Generation of dissemination products.....	74
C. Management of the publication of dissemination products	74
D. Promotion of dissemination products	74
E. Management of user support	74
Chapter X	
Evaluation.....	75
Introduction	75

A. Characteristics of the evaluation	75
B. Feasibility of evaluation	75
1. Institutional dimension	76
2. Specific objectives	76
3. Technical-methodological dimension	76
4. Specific objectives	76
5. Methodological approach to the evaluation	77
6. Characteristics to be considered in evaluating dimensions	78
7. Agreements or arrangements	78
8. Needs	78
9. Conceptual and analytical design	81
10. Receipt of administrative records	81
11. Integration of administrative records	81
12. Analysis and dissemination	81
13. Confidentiality	81
14. Costs and benefits	81
15. Technical assistance and registration	82
Chapter XI	83
Introduction	83
A. Towards a system of economic administrative records	83
B. Statistical process model (GSBPM)	84
C. Detection and analysis of needs	84
D. Design	85
E. Construction	85
F. Collection	86
G. Processing	86
H. Analysis	86
I. Dissemination	87
J. Evaluation	87
Bibliography	89
Annexes	91
Annex A1	91
Annex A2	100
Annex A3	101
Annex A4	104
Annex A5	109
Annex A6	121
Annex A7	123
Annex A8	124
Glossary	131
 Tables	
Table III.1 Example of an administrative register methodological fact sheet	26
Table III.2 Example of an administrative register methodological fact sheet	26
Table IV.1 Subordinate-enterprise statistical unit classification model	31
Table X.1 Institutional dimensions matrix	79
Table X.2 Matrix of technical-methodological dimensions	80
Table A1.1 Consultation matrix for identification of information needs	92
Table A1.2 Matrix of information needs and existing information	92
Table A1.3 Preliminary concept identification matrix	93
Table A1.4 Example for verifying data availability	95
Table A1.5 Example of alternatives analysis	96
Table A1.6 Example of a table identifying the stakeholder group	97
Table A1.7 Example of a constitutive act	99
Table A3.1 Administrative register inventory form	102
Table A3.2 Administrative register technical fact sheet	103
Table A4.1 Items of the channel of exchange	106

Table A4.2	Data structure items to characterize files (tables)	108
Table A4.3	Data structure items to characterize the contents of each file.....	108
Table A5.1	Description of the software used for updating	120
Table A5.2	Determination of the business segment	120
Table A5.3	Determination of legal organization.....	121
Boxes		
Box VII.1	Key dimensions in the quality of statistics.....	62
Box A8.1	Questionnaire on agreements or arrangements.....	125
Box A8.2	Questionnaire on needs.....	125
Box A8.3	Questionnaire on conceptual and analytical design	126
Box A8.4	Questionnaire on capture of administrative records.....	126
Box A8.5	Questionnaire on integration of administrative records	127
Box A8.6	Questionnaire on confidentiality	127
Box A8.7	Questionnaire on costs and benefits.....	128
Box A8.8	Questionnaire on analysis and dissemination.....	128
Box A8.9	Questionnaire on technical assistance and registration.....	129
Diagrams		
Diagram II.1	The four basic registers proposed by Wallgren and Wallgren (2012)	16
Diagram II.2	Vision of a system of economic administrative records.....	18
Diagram II.3	Examples of unique identifiers	19
Diagram II.4	Phases and subprocesses of the Generic Statistical Business Process Model (GSBPM)	21
Diagram III.1	Subprocesses of the specification of needs phase.....	24
Diagram IV.1	Subprocesses included in the design phase	29
Diagram IV.2	Schematic diagram of the identifier database for the observation unit	33
Diagram IV.3	Schematic outline of linkage of the statistical business register with economic statistical registers	41
Diagram IV.4	Example of a model database for the system of economic administrative records	42
Diagram V.1	Subprocesses included in the construction phase	45
Diagram V.2	Different types of metadata and tools in the documentation of a register	50
Diagram VI.1	Subprocesses included in the collection phase	55
Diagram VII.1	Subprocesses included in the processing phase.....	61
Diagram VII.2	Statistical records from administrative records	63
Diagram VIII.1	Subprocesses included in the analysis phase.....	71
Diagram IX.1	Subprocesses included in the dissemination phase	73
Diagram X.1	Subprocesses included in the evaluation phase	75
Diagram XI.1	Phases and subprocesses of the Generic Statistical Business Process Model (GSBPM).....	84
Diagram A1.1	Needs specification flowchart.....	91
Diagram A1.2	Example of a schematic outline for problem or needs analysis.....	97
Diagram A1.3	Example of an outline for determining objectives.....	98
Diagram A4.1	File transfer scheme	106
Diagram A4.2	Data warehouse modelling scheme with data from the Office of the Superintendent of Pensions	107
Diagram A5.1	Process for updating the Central Directory of Enterprises and Establishments	111
Diagram A5.2	Receipt and processing of information.....	111
Diagram A5.3	Database update.....	112
Diagram A5.4	Database validation	113
Diagram A5.5	Directory georeferencing	113
Diagram A5.6	Field georeferencing process	114
Diagram A5.7	Cartographic information	115
Images		
Image A5.1	Cartographic information with zone and block codes	115
Image A5.2	Cartographic information of establishments by X and Y coordinates	116
Image A5.3	Cartographic information by census area, census tract, census block, and X and Y coordinates.....	116
Image A5.4	Cartographic information of establishments using the city cartography database of the Household Targeting System (SISFOH)	117
Image A5.5	Use of Mapcity for addresses with lot and block numbers	117
Image A5.6	Field supervision record.....	118
Image A5.7	Digitization record.....	119
Image A5.8	Status and visit result.....	119

Introduction

In recent years, several countries around the world have focused on using administrative records as a source of information for obtaining statistics. Developed countries have statistical infrastructure that contains, at a minimum, data on the identification and location of observation units (persons, dwellings, establishments, businesses, properties, and so on), which they update using administrative records. They then rely on that infrastructure to generate sampling frames for surveys. The updating is done using a unique public identifier for each observation unit adopted by government agencies, such as national statistical offices.

In Latin American and Caribbean countries, the situation is different. There are still several problems that prevent the creation of an infrastructure similar to that in developed countries, that would support the establishment of effective sampling frames to generate accurate and reliable economic statistics.

The challenges faced by the countries of Latin America and the Caribbean include the following:

- Rigid legal frameworks that prevent government agencies from sharing administrative records with national statistical offices.
- Administrative records have specific objectives that do not necessarily have a statistical purpose. They are also based on conceptual frameworks specific to the objective they pursue, and which do not always match those developed by the national statistical offices.
- Government agencies and private entities that have administrative records use classifiers that meet their own needs but differ from those used by the national statistical offices.
- National statistical offices construct the sampling frames for the statistical design of their surveys in isolation.
- Although there are various international recommendations on the statistical use of administrative records, there are no established methodologies or statistical tools or software to enable the countries of the region to build infrastructure on the basis of administrative records for potential statistical use.

Accordingly, this methodological document is intended to furnish the countries of the region with a set of guidelines and recommendations for producing innovative and timely economic statistics, and to provide greater benefits by making statistical use of administrative records. References to administrative records in this methodological document are understood as alluding to administrative records on economic issues.

The four fundamental inputs for this document are:

- (i) The book *“Register-based Statistics: Administrative Data for Statistical Purposes”* by Anders Wallgren and Brit Wallgren.

This volume contains an exhaustive presentation of the administrative records that feed the four basic registers —population register, business register, real estate register and activity register— which, in turn, support each country’s national statistical system (NSS). This document only addresses the business register.

- (ii) The Economic Commission for Europe (UNECE) Guidelines on Statistical Business Registers.

As will be seen in Chapter II, the statistical business register is the backbone of economic statistics generation processes. This document is aligned with these international recommendations, since the statistical business register offers enormous advantages in relation to economic statistics production.

- (iii) The UNECE “Generic Statistical Business Process Model (GSBPM) (Version 1.0, September 2016)”

This model is the framework around which the guidelines and recommendations for a better statistical use of economic administrative records are developed.

- (iv) The experience of the countries in the region.

One of the most important inputs for this document is the experience of the representatives of national statistical offices in the countries that make up Latin America and Caribbean region. The methodology incorporates good practices adopted by these countries and shows how they have solved problems that are common across the region.

Lastly, this methodological document consists of 11 chapters. The first chapter discusses the importance of administrative records. Chapter II provides a vision for a system of economic administrative records. Chapters III to X develop recommendations and guidelines for each of the eight phases of the Model: specification of needs, design, construction, collection, processing, analysis, dissemination and evaluation. Lastly, chapter XI contains conclusions and recommendations.

Chapter I

The importance of administrative records

Introduction

The information collected in administrative records available in public agencies and private organizations can be an input for producing basic statistics to supplement or serve as an alternative to censuses and sample surveys.¹

These agencies are regarded as the sources of statistical information, reporting sources, records offices or administrative sources. However, to use them for statistical purposes, their operational characteristics, conceptual definitions, event coverage, and territorial and temporal breakdown must all be carefully evaluated. Likewise, it is also pertinent to review the procedures, the technologies used to collect information and the evaluation of the quality of the data reported for producing statistics, in accordance with the applicable administrative regulations.

Thus, for example, the substantive functions of a country's tax offices are to receive taxes from individuals and businesses liable to tax and to enforce tax laws. Once evaluated, this information can be integrated into a statistical process enabling its use in the study of the structure, characteristics, behaviour and trends of economic considerations, such as maintaining a list of businesses that pay taxes, knowing the births and deaths of businesses, the amount of taxes paid and the income of businesses, variations over time, and so on.

In that context, the interest in making statistical use of administrative records stems mainly from their qualities, as follows:

- Coverage and continuity of data collection on events that may be of general interest.
- Coverage and geographic disaggregation of the records, particularly when they are national in scope.
- Feasible characterization of the phenomenon given the variables captured, in accordance with the thematic coverage that is confined to the legal jurisdiction of the organizations or agencies that compile the register.
- Possibilities of broadening the data collection's conceptual coverage.
- Marginal costs relative to current statistical use.

In spite of those advantages, the following types of problems are relatively common in the statistical use of administrative records:

- Complexity and diversity of substantive functions covered by the recording institution, which do not necessarily have to do with statistical objectives.
- Insufficient budgetary support for statistical functions.
- Limited technical capacity for maintaining the continuity of statistical functions in line with information quality requirements.

¹ Most administrative records come from public agencies. They are the result of the regulatory activity they carry out. However, in some cases, information available from private organizations can also be used for statistical purposes.

- Lack of interest in the register's use for statistical purpose on the part of the authorities in charge.
- Collection difficulties inherent to the units being recorded and to the circumstances in which the data are collected.
- Poor operational control and technical limitations of the personnel involved in the statistical process.
- Lack of documentation on the compilation of records and the collection methodology used.
- Modification or alteration of the structure of records by omitting or adding variables.

Administrative records constitute an input in statistical production with diverse applications in the NSS, such as:

- Statistics generation. Given the structure of the public sector at its different levels and the information available to the different agencies, the needs for the generation, integration and dissemination of official statistical information of interest to the NSS can be classified into thematic fields and associated with different sectors of public activity in order to organize statistics by thematic sectors or regions.
- Preparation of directories or lists to facilitate other statistical production methods. These data can be used to build lists and directories of organizations engaged particular fields as sampling frames or as a reference for use in combination with censuses.

Thus, the use of administrative records to produce official statistics is important because of the range of different areas in which the public sector is involved. Under a conventional thematic scheme, the information can be used to integrate sectoral systems of official statistics produced by the agency in charge of statistics in the country or through the joint participation of the official statistics agency and other public agencies according to agreed guidelines on coordination and rules in order to facilitate that participation.

Administrative records are fundamental for understanding gender gaps in areas such as demographics, education, health, labour, gender-based violence, political participation and decision-making, access to property, migration, the impact of climate phenomena and so on. The use of administrative records is important for gender analysis inasmuch as they yield data that reflect the status, needs, opportunities and specific contributions of women and girls to society.

Therefore, statistics on a range of issues can be compiled from different types of administrative records as a by-product of administrative processes. Other examples include health statistics, which are compiled from hospital records; job statistics, which are obtained from employment offices; demographic statistics, which are put together from civil registration systems, and education statistics, which are obtained from enrolment reports from the ministry of education.

In addition to government agencies, it is very common for private organizations to have information on companies that carry out some type of economic activity. For example, there may be a chamber of commerce that acts as an umbrella for all the companies that engage in some type of commercial activity and, therefore, have information such as directories, contact information, telephone numbers, e-mails and web pages, among other data. This makes them a source for NSS updates.

A. Fragmentation and integration

One of the advantages for countries that have developed their statistical system by integrating different administrative records has to do with the existence of a unique public identifier for observation units, whatever they may be.

That identifier-based integration stems from the fact that the different offices of each administration —be they public or private, national or subnational— disaggregate observation units according to whatever it is they are keeping records on.

In the case of economic units, there are administrative records on taxes in national and subnational tax agencies; administrative labour records in agencies involved in labour-related matters; real estate or cadastral administrative records; administrative records on electricity, gas, water, telephone and Internet consumption, among other things; all of them in public or private organizations at the national and subnational levels.

In a traditional statistical operation such as an economic census, each of these topics of administrative records is a dimension or variable of the observation unit surveyed on a structured survey form.

The systematization of fragmented administrative records consists of integrating them on the basis of that unique public identifier.

However, that fragmentation or dispersion occurs both horizontally and vertically. Horizontal, when in the same government jurisdiction (e.g., the national administration) different agencies keep records on tax, labour, cadastral and other matters. Vertical, when records are kept in different jurisdictions or levels of government. Thus, there are national and subnational taxes, each with their respective fiscal agencies.

These characteristics of administrative records should be considered at the time of integration since, depending on whether the countries are federal or unitary, the subnational states also have records that should be considered for integration.

Chapter II

Vision of a system of economic administrative records

Introduction

An outline of the role of administrative records and their relationship to the statistical business register is provided in order to facilitate a conceptual understanding for increasing the statistical use of administrative records, which are of fundamental importance for the following:

- Building and updating the statistical business register
- Production of economic statistics through:
 - Traditional exploitation of administrative records.
 - Production of statistics by linking administrative records with the statistical business register.
 - Integration of a system of economic administrative records from the statistical business register and economic administrative records, with the aim of identifying interactions for addressing the 2030 Agenda for Sustainable Development.

A. Basic system of records

Wallgren and Wallgren (2012) point to the need to build a system of administrative records that ensures that all data can be efficiently integrated, and their statistical use enhanced. To that end, they identify four basic registers:

- (i) Population register.
- (ii) Activity register.
- (iii) Real estate register.
- (iv) Business register.

The links between them form the backbone of the system of records.

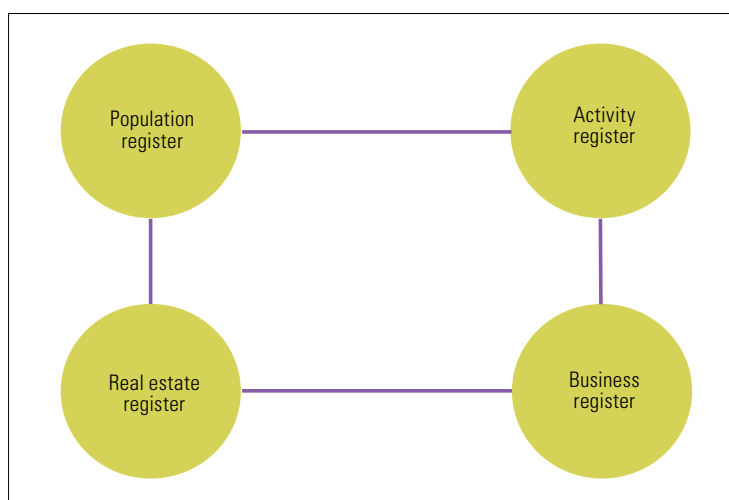
According to them, each basic register should have the following characteristics:

- Have clearly defined types of statistical units.
- Define sets of statistical units or populations of standardized statistical units.
- Contain links from other basic registers.
- Contain links to other records related to the same type of object.

They also argue that the information supporting the basic registers should be of high quality and well documented (see diagram II.1).

Diagram II.1

The four basic registers proposed by Wallgren and Wallgren (2012)



Source: A. Wallgren and B. Wallgren, *Estadísticas basadas en registros: aprovechamiento estadístico de datos administrativos*, Aguascalientes, National Institute of Statistics and Geography (INEGI), 2012.

The purpose of this document is to develop guidelines and recommendations to enable Latin American and Caribbean countries to build or consolidate their basic business register.

B. Vision of statistical business register: the system of economic administrative records

It is suggested that national statistical offices have a system of economic administrative records containing the statistical business register, with basic information on the country's economic units that is updated periodically, reflects the reality of the object of interest, allows the sampling frames needed for economic surveys to be extracted and produces indicators on business demographics and longitudinal studies on economic units.

Administrative records can yield information on jobs, social security, taxes, labour inspection, working hours and businesses registered by size and industry. Statistical products can be produced; for example, tax records can provide an indicator of the amount of tax paid monthly by all the businesses in a country, which contributes to analysis of the country's economic evolution.

Analysing that information from a gender perspective is extremely relevant for highlighting the obstacles faced by women in different dimensions of their economic autonomy. Hence the importance of making every effort to increase the number of variables available in this area.

This source of information can also serve to compile statistics for smaller geographic areas and population groups, and even to provide longitudinal information for long reference periods, if the units receive unique identifiers in the records. This is relevant for the study of women's labour market histories and the performance of women-led businesses.

1. Objectives of a system of economic administrative records

- Generate a higher quality of statistical information on an ongoing basis and reduce overburdening of reporting entities.
- Link administrative records with statistical records in each country through a flexible, efficient records system for the production of combined economic statistics.

- Establish a common framework in the Latin American and Caribbean for the creation of statistical business registers used for statistical purposes.
- Create a homogeneous system that facilitates the linking of economic information from all the countries and makes it possible to leverage economic statistical studies in the region.
- Use administrative data for the detection, updating and classification of statistical units and the formation of sampling frames.
- Complement statistical information based on data imputation and validation of information, without the need to expend resources on surveys and forms that make statistical processes more expensive.

2. Objectives of a statistical business register as part of a system of economic administrative records

- Detect and build the statistical units of a country's production system and link them with the rest of the records of the system of economic administrative records.
- Have a tool for the preparation and coordination of economic surveys and the extrapolation of the results of obtained from those surveys.
- Be catalogued as a source of information for the study and analysis of the business population and its demographic data.
- Enable longitudinal analysis of economic units over time.
- To have a tool, structured by statistical units, for the transfer of administrative data to statistical records, especially at a geographic or other level of disaggregation.
- Serve as a tool for integration and dissemination of the economic activity of a country and that of the Latin American and Caribbean region.
- Be used as a dissemination unit for a central reference directory that can provide lists of companies with certain data, in accordance with national restrictions.

For the information contained in the statistical business register to be effective, it must be updated periodically from administrative registers that refer to the same type of economic unit (group of companies, enterprise, or establishment) and that are harmonized conceptually and informatically. In other words, in addition to concepts and IT issues, administrative records should use the same catalogues for categorical variables.

Economic censuses and surveys generate their own statistical products (tabulations, query systems, graphs, etc.). That said, in countries where economic censuses are carried out, they can be linked through the statistical registration of businesses with administrative records to increase the number of variables on the same economic unit, which increases the statistical potential of the information. This linkage is easily made by means of the unique identifier that each business entity has in a system of economic administrative records.

The updating of the statistical business register will be synchronized in a time-based manner, prioritizing sources (census, surveys or administrative records). Since the statistical business register contains current and historical information on each economic unit, longitudinal studies can be carried out based on it.

The role of the national statistical office is mainly one of permanent coordination with government entities that have administrative records in order to ensure that those records meet the technical criteria for their statistical use, which should be duly recorded in the corresponding metadata.

In that context, each economic unit needs a unique identifier that must be used by all governmental or private entities with administrative records to refer to the same economic unit. Consequently, an identifier is assigned for each new business entity, and shared with the various governmental and private entities that keep or use administrative records.

3. Components of the model of the system of economic administrative records

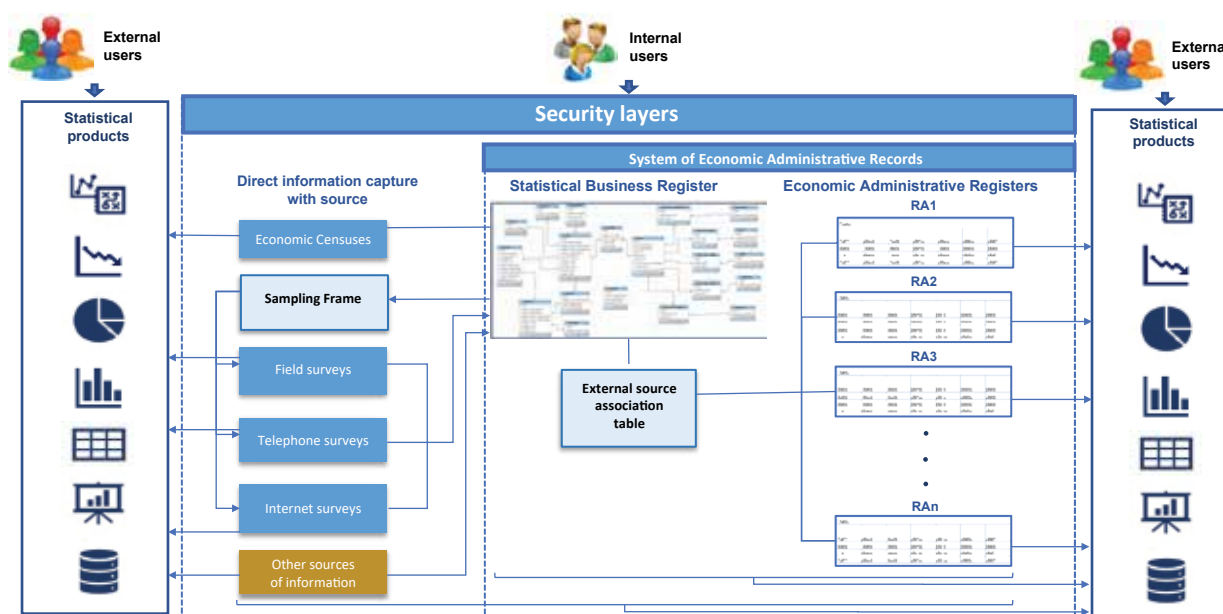
The model of the system of economic administrative records comprises the following elements:

- (1) System of economic administrative records
 - (a) Statistical business register.
 - (b) Economic administrative records.
 - (c) Relationship within and between components.
- (2) Databases built through direct collection of information.
 - (a) Economic censuses (if conducted in the country).
 - (b) Sampling frames.
 - (c) Field surveys.
 - (d) Telephone surveys.
 - (e) Internet surveys.
 - (f) Other information sources.

In addition, relationships between different components are established. For example, administrative records are linked to the statistical business register by means of a unique identifier for each economic unit. The administrative records are linked to each other via the external source association table. Diagram II.2 shows the data model of the system of economic administrative records, which includes the administrative records and the statistical business register.

Diagram II.2

Vision of a system of economic administrative records



Source: Economic Commission for Latin America and the Caribbean (ECLAC), on the basis of National Institute of Statistics and Geography (INEGI).

C. Statistical business register

In the UNECE (2015) document, the statistical business register is a structured database of economic units in a territorial area, managed by a national statistical office. This database is updated regularly and used for statistical purposes.

The following is a definition of some concepts necessary for a better understanding of what a statistical business register is.

1. Structured database

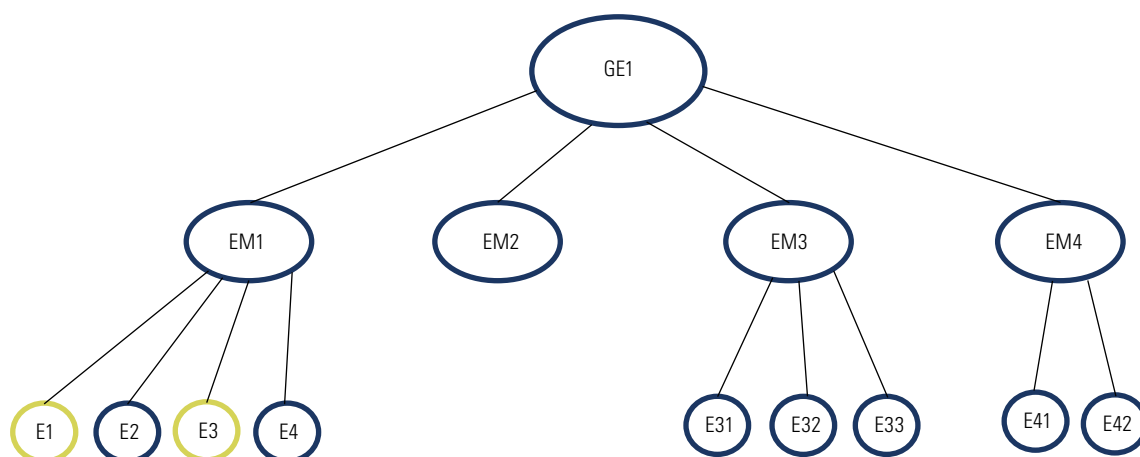
It is a database in which the characteristics of each economic unit are stored, as well as the relationships that exist between the different types of business entities. For example, establishments belonging to the same company, or companies that are part of a corporate group.

2. Unique identifier for each economic unit

It is very important that each economic unit has a unique identifier within the statistical business register, and that such an identifier is used for all types of economic units. This makes it possible to refer to a specific economic unit, so identifiers should not be repeated. The unique identifier assigned to a group of companies refers to all the companies belonging to that group, and the unique identifier of a company encompasses all the establishments that make up that company.

Diagram II.3 shows an example of a group of companies whose unique identifier is GE1, which is made up of companies with the unique identifiers EM1, EM2, EM3 and EM4. Company EM1 company is made up of the establishments whose unique identifiers are E1, E2, E3 and E4; company EM3 consists of establishments E31, E32 and E33; while establishments E41 and E42 make up company EM4.

Diagram II.3
Examples of unique identifiers



Source: National Institute of Statistics and Geography (INEGI) of Mexico.

It is important to note that the unique identifiers should not be reused for any other economic unit, as they will refer to the economic unit even after its life cycle has ended.

Chapter IV details the methodology for the use of a unique identifier.

3. Regularly updated

A register that is not updated regularly quickly loses its value, since the economic units are continuously changing in terms of both their economic variables and their structure. For example, new economic units may be created, some economic units may cease to operate permanently, mergers or splits may occur, as may changes in location or economic activity, and even their sales value or the number of persons employed may vary.

The frequency at which the statistical business register is updated depends on the availability of data, i.e. administrative records.

4. Territorial area

A statistical business register is formed to cover all economic units located in a specific territorial area, which usually corresponds to the territory of a country. Generally speaking, an economic unit affiliated with a company located in the country but situated in another country will not be found in the statistical business register.

On the other hand, all the information of the economic units of the statistical business register should be contained in the same database. In other words, the information should not be separated by region, since that would defeat the purpose of the register.

5. Database administered by a national statistical office

Administrative records are used by administrative authorities for specific purposes related to their function, so the coverage, updating methods and, in general, everything related to the administrative register is designed to address administrative matters. For this reason, the national statistical office should work with entities that provide administrative records to ensure conceptual and informatic validation in order to facilitate the creation and updating of the statistical business register, which must necessarily be managed by the national statistical office.

One of the purposes of the statistical business register is to provide the sampling framework for economic surveys, so it is essential that the register be based on generally accepted statistical concepts, preferably those contained in international recommendations.

D. The Generic Statistical Business Process Model: framework for the system of economic administrative records

In order to ensure that national statistical offices are provided with sufficient elements to have a system of economic administrative records, the following chapters develop guidelines that consider the context and level of statistical progress by countries in Latin America and the Caribbean.

The Generic Statistical Business Process Model (GSBPM) proposed by the UNECE (2016) is considered for the development of the guidelines. It describes and defines the phases and subprocesses necessary to generate official statistics.

The GSBPM provides a standard framework and harmonized terminology to help statistical organizations modernize statistical production processes and share methods and components across different statistical programmes. It can also be used to integrate standards in data and metadata, as a model for documenting processes, harmonizing statistical computing infrastructure and providing a frame of reference for quality process improvements and evaluations.

The GSBPM should be applied and interpreted in a flexible manner. It is not a rigid framework in which all steps must be followed in a strict order, but rather identifies possible steps within the statistical process and the interdependencies between them.

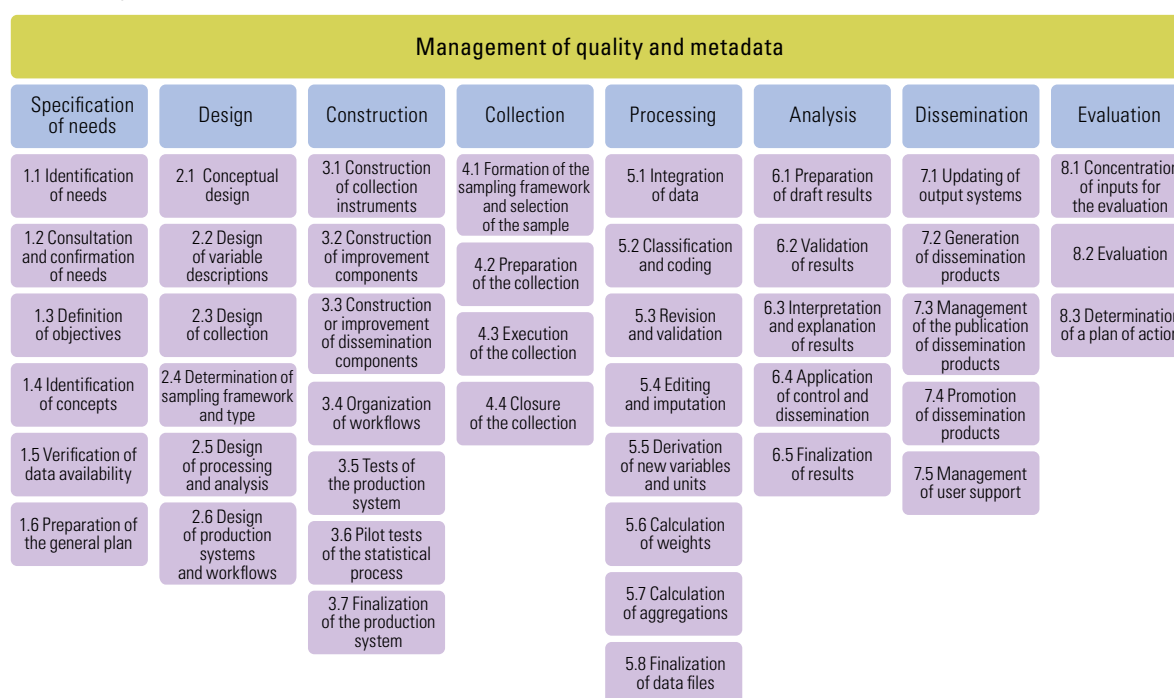
Although the GSBPM will be arrayed in a logical sequence of steps in most statistical processes, components of the model may appear in a different order, depending on different circumstances. In addition, some subprocesses will be reviewed several times, forming iterative cycles, especially in the processing and analysis phases.

The GSBPM should, therefore, be seen more as a matrix in which there are many possible paths. In this way, the GSBPM is intended to be generic enough to be broadly applicable and to promote a standardized view of the statistical process, without becoming too restrictive or abstract and theoretical.

As diagram II.4 shows, the GSBPM has eight phases: needs specification, design, construction, collection, processing, analysis, dissemination and evaluation. The GSBPM has been adapted by some national statistical offices to their own needs in terms of the organization of the statistical output they are in charge of.

Diagram II.4

Phases and subprocesses of the Generic Statistical Business Process Model (GSBPM)



Source: Economic Commission for Europe (UNECE), "Generic Statistical Business Process Model GSBPM (Version 1.0 in Spanish, September 2016)", Geneva, 2016.

Each phase is divided into subprocesses. There are 44 subprocesses in total. Not all are necessarily applicable to these guidelines because some have to do with very specific aspects that fall within the realm of other traditional methods for preparing statistics, such as sample surveys.

The following eight chapters contain recommendations that national statistical offices may consider for a system of economic administrative records. Each phase includes the following.

- Phase 1: Specification of needs (Chapter III)
 - Identification of needs
 - Consultation and confirmation of needs
 - Definition of objectives
 - Identification of concepts

- Verification of data availability
 - Preparation of the general plan
- Phase 2: Design (Chapter IV)
 - Conceptual design
 - Design of variable descriptions
 - Design of collection
 - Design of processing and analysis
 - Design of production systems and workflows
- Phase 3: Construction (Chapter V)
 - Construction of collection instruments
 - Construction or improvement of components
 - Construction or improvement of dissemination components
 - Organization of workflows
 - Tests of the production system
 - Pilot test of the statistical process
 - Finalization of the production system
- Phase 4: Collection (Chapter VI)
 - Preparation of the collection
 - Execution of the collection
 - Finalise collection
- Phase 5: Processing (Chapter VII)
 - Integration of data
 - Classification and coding
 - Revision and validation
 - Editing and imputation
 - Derivation of new variables and units
 - Calculation of aggregations
 - Finalization of data files
- Phase 6: Analysis (Chapter VIII)
 - Preparation of draft results
 - Validation of results
 - Interpretation and explanation of results
 - Application of control of dissemination
 - Finalise outputs
- Phase 7: Dissemination (Chapter IX)
 - Updating of output systems
 - Generation of dissemination products
 - Management of the publication of dissemination products
 - Promotion of dissemination products
 - Management of user support
- Phase 8: Evaluation (Chapter X)
 - Gathering of inputs for evaluation
 - Evaluation
 - Determination of a plan of action

As can be seen, this document does not consider the subprocesses relating to sampling, since they do not apply to the system of economic administrative records.

Chapter III

Specification of needs

Introduction

As a starting point, an analysis of the information needs of users, both internal and external, should be conducted to put together a strategy to meet the information demand for creating a system of economic administrative records, the main component of which is the statistical business register. This structure must also be flexible so that, over time, it can adjust to evolving needs and priorities in a changing environment.

The needs specification phase consists of having an overview of the information needs that a system of economic administrative records should meet in each country in Latin America and the Caribbean. That would make it possible to improve the production of economic statistics in order to meet the current and growing demand from users. National statistical offices would be given the possibility to join current and future information generation arrangements, both national and international.

The scope of the needs specification phase will reveal why the decision is made to create a system of economic administrative records. It will also make it possible to determine the national and international background that led to the detection of the need to create it, the information requirements that currently exist in terms of business registers and the methodological dimensions that should be taken into account in the system's design.

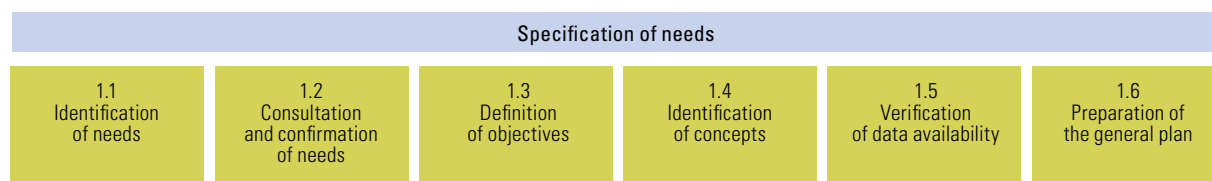
In this phase, it is important to analyse, in particular, the data needs for the development of policies on gender equality and women's empowerment and to identify a framework of monitoring indicators with clear data sources. Likewise, it is important to consult not only with ministries for different sectors, but also with mechanisms for the advancement of women and feminist organizations. That will serve to identify gaps between the data required and the data available, as well as measures to fill those gaps.

A. Subprocesses of needs specification

Needs specification is divided into the six subprocesses shown in diagram III.1, which make it possible to: (i) identify the need for statistics; (ii) confirm the statistical needs of the interested parties in greater detail; (iii) establish high-level objectives for the statistical results; (iv) identify the relevant concepts and variables for which the information is required; (v) check the extent to which current data sources can meet those needs, and (vi) prepare the necessary justification with a view to obtaining approval to produce statistics.

Diagram III.1

Subprocesses of the specification of needs phase



Source: Economic Commission for Europe (UNECE), “Generic Statistical Business Process Model GSBPM (Version 1.0 in Spanish, September 2016)”, Geneva, 2016.

1. Identification of needs

This subprocess consists of conducting an analysis of information needs. They may be the result of a need at the country or institutional level, or an express request from an international organization or governmental body. The purpose is to obtain information that will allow public and private policy decisions to be made.

This identification of needs must be based on a country overview that allows the creation of a single system of economic administrative records based on a statistical business register and administrative records to be maintained by each country’s national statistical office. Its objective would be to avoid duplication of efforts, lower costs, and supply information in a timely and effective way. The system of economic administrative records would be used to obtain a statistical framework for economic surveys or studies, precluding additional inquiries to businesses, since any such queries could be disposed of using administrative records already existing in the country.

For this reason, the information needs of all users (external and internal) should be included in order to avoid duplication of efforts and so that the national statistical office does not make repeat requests for business administrative records to a given institution. In this way, the institution will be able to work in a coordinated manner and with the same outlook, with clarity about the criteria for homogenizing information and classifications, and so on.

The system must also offer international comparability. To that end, this needs identification process should include research on the production of similar statistics or practices implemented by the national statistical office and national and international organizations. This information could serve as reference in the proposal for generating the required statistics (for example, a bibliographic review of international organizations on the minimum information requirements that a statistical business register based on administrative records must meet).

2. Consultation and confirmation of needs

It is vital to know the information needs of the various potential users identified in advance, in order to have a road map of the most important aspects to consider in the design phase of the system of economic administrative records. To the extent possible, the disaggregation of variables by sex should be included as part of the needs, which will allow internal and external needs to be met.

A common mistake is to consider only external needs (either at the country or international level). The internal needs of the national statistical office are as important as the external ones, since there may be several processes or departments that are users or providers of the system of economic administrative records.

By performing the consultation internally (in the first instance), duplication of efforts and costs, among other considerations, can be avoided. In addition, that will allow for better coordination and thus strengthen the use of administrative records.²

With respect to external consultation, that refers to institutions that make up the NSS, which in some countries may be identified by law. For example, with respect to Costa Rica, in the recently enacted Law No. 9694 (of 30 May 2019), Article 24 of Section V, “User access to information and consultation of their needs,” provides: “NSS institutions shall clearly identify the main users of the statistics under their responsibility and consult them regularly on the relevance of the results published, the accuracy and quality observed, new information needs, and the level of access to statistical information, in addition to considering their opinions in the improvements of the statistical service” (Legislative Assembly of Costa Rica, 2019).

In addition, Article 35 states: “INEC shall consult users, by the means it deems pertinent, on current statistics and new information needs. To develop new official statistics, the feasibility and cost of meeting such requirements will be analysed, and their preparation will be subject to additional funding being obtained for INEC or to contributions secured from interested parties, and those statistics must be incorporated into the National Statistical Plan” (Legislative Assembly of Costa Rica, 2019).

Information needs should be updated with the information users. For example, it is useful to detect those that could have been captured in the past but were not, or those that at some point were captured but were not consulted. It is also necessary to analyse the requirements of users and the feasibility of meeting them, confirm those that can be met and by what date, prioritize those requirements and those that cannot be met, and document the justification why (an example how to document is presented below, under point 5 of annex A1).

It is important to keep in mind that this consultation process (internal or external) may require more than one contact by the national statistical office with the identified stakeholders (information users or providers), so the initial approach with each stakeholder is key to the process.

3. Definition of objectives

Once the information needs of the various stakeholders have been consulted and confirmed, this subprocess must convert the expected results into specific objectives by which to meet the needs identified in the consultation.

The aim is to set an objective that will allow the creation of a system of economic administrative records, so that the advantages gained from their use can be maximized. Where appropriate, these can be a complement to other sources of information, so that more statistical results that are better in terms of quality, timeliness and efficiency can be produced.

4. Identification of concepts

In the case of economic administrative records that are of interest and for which sufficient and quality information is available, this consists of an initial identification of the concepts to be used as part of the subject matter of the system of economic administrative records. Those concepts will be important inputs for standardizing business definitions for statistical purposes, among others.

In the case of the statistical business register, it is important to have concepts that refer to aspects such as identification, location, size and economic activity. In addition, business demographics, exports and imports, and environmental issues, among others, can be included. To cover these topics, it is necessary to have various

² Ideally, the use, access, safekeeping and security, and so on, of administrative records should be regulated and standardized in the statistical office, so as to ensure efficient use and a commitment to the information providers.

administrative records and to identify the thematic concepts. Thus, for example, for work with the administrative register of the tax collection office, some of the thematic concepts could be income, taxes, tax regime, natural person (individual) or legal person (entity), among others. Identifying these concepts will be very useful for determining the variables that can be used from the administrative record.

5. Verification of data availability

Research should be done to determine which existing administrative records could be key inputs for obtaining the information required for generating economic statistics.

That research could include making an inventory of existing economic administrative records, as described in section 4.1.3. Costa Rica's NSS, among others, has an inventory of statistical operations that could serve as an example. Tables III.1 and III.2 show two examples of methodological fact sheets for the administrative register.

Table III.1

Example of an administrative register methodological fact sheet

Costa Rica: administrative register methodological fact sheet	
Name of administrative register	
Name of institution responsible	
Name of the administrative unit that prepares the administrative record	
Overarching objective of the administrative register	
Sector or topic	
Subsector or subtopic	
Geographic coverage	
Geographic disaggregation	
Statistical units	
Main variables	
Periodicity of data capture	
Periodicity of dissemination	
Dissemination methods	
Legal framework	
Name of the person in charge of the administrative register	
Position	
Email	
Telephone no.	
Website	

Source: National Institute of Statistics and Censuses (INEC) of Costa Rica, *Informe de resultados del Inventario Nacional de Operaciones Estadísticas*, San José, 2018.

Table III.2

Example of an administrative register methodological fact sheet

Mexico: outline of topics, categories and variables to be collected in the inventory of administrative registers	
Administrative registers identified	
Identification code	
Official name	
Other names	
Date of inclusion in the inventory of administrative records	
Type of record	
Complementary records	
Operating status	
Administrative registers with information stored in electronic media	
Database format	

Table III.2 (concluded)

Mexico: outline of topics, categories and variables to be collected in the inventory of administrative registers	
Administrative registers with online inquiry access	
URL address	
Administrative registers used to generate statistics	
Name of the institution generating the statistics	
Type of users of statistical information	
Name of related statistical projects	
Means of presentation of statistics	
Physical storage medium	
Administrative registers linked to statistical projects	
Status of inclusion in the statistical business register	
Administrative registers that disclose statistics on the Internet	
URL address	
Administrative registers with other participating institutions	
Type of participation	
Administrative registers operational	
Name of institution responsible	
Institution initials	
Sphere of government	
Administrative unit responsible	
Official in charge	
Position of the person in charge	
Official address	
Telephone no.	
Email	
Other participating institutions	
Name of the normative basis of operation	
Date of last publication in the Official Gazette	
Name of related procedures or services	
Main topic, subtopic, section	
Objective	
Subjects of the register	
Variables characterizing subjects of the register	
Coverage of subjects of the register	
Geographic coverage	
Geographic disaggregation	
Availability of information	
Deadline	
Physical recording medium	
Name of the registration form or file	
Frequency of uptake	
Physical storage medium	
Entity where data are processed	
Names of available concept papers	
Names of available glossaries	
Names of available manuals	
Names of available instruction books	
Type of users of the register's information	
Means of dissemination or consultation of the register's information	
Use of information to generate statistics	

Source: National Institute of Statistics and Geography (INEGI) of Mexico.

Once the available economic administrative records have been identified, an analysis should be done of the legal framework of each source, access and use of the information, quality, periodicity and other characteristics. The aim is to identify if they are the most apt, or whether to prioritize them over other sources. It is therefore recommended, based on the above-mentioned inputs, to prepare an inventory of economic administrative records and suggest the variables it should contain.

Work should also be done on formal arrangements or documents, such as agreements or letters of understanding (in the event that there is no law that makes release of the records mandatory) supporting the release of the data (including forms, delivery instruments, periodicity and so on).

6. Prepare the general plan and documentation

The documentation of a process is important, since it provides evidence of the actions carried out during this phase, so that it can be used for approval by the corresponding authorities and, therefore, for the implementation of the process or statistical operation.

This document should include a description of the current status of the system of economic administrative records (if it already exists), with information on how current statistics are produced. Any areas of opportunity that need to be addressed should be highlighted. A proposed solution should be included that details how the process will be developed to produce the new statistics, or statistics that already exist but require improvements or adjustments and, finally, provides an evaluation of the costs and benefits.

Chapter IV

Design

Introduction

The purpose of this phase is to design the statistical information products that will meet the structured information needs, as determined from the documentary components collected in the previous phase. In other words, the aim is to define the outputs, concepts, methodologies, data collection instruments, protocols and channels of exchange, as well as the general strategies for the following phases of the GSBPM model (see diagram IV.1).

Diagram IV.1

Subprocesses included in the design phase



Source: Economic Commission for Europe (UNECE), "Generic Statistical Business Process Model GSBPM (Version 1.0 in Spanish, September 2016)", Geneva, 2016.

Note: Subprocess 2.4 "Determination of sampling frame and type", which does not apply to administrative records, is omitted, since it is oriented towards considerations covered by other traditional methods for compiling statistics, such as sample surveys.

A. Conceptual design

In the design phase, the owners of the system of economic administrative records will define the outputs, concepts, methodologies, collection instruments, protocols and channels of exchange. They will also define the overall strategies for the construction, collection, processing, analysis and dissemination phases, the method of execution and other aspects considered relevant within the information production process.

It should not be pretended from the outset that the system of economic administrative records will provide all the information required for the set of terms and variables necessary for the statistical analysis of the economy. It is best to consider achievable goals in the short term —on the basis of information from accessible administrative and statistical records— that meet the expected quality requirements, in order then to devise an improvement plan based on objectives and goals to be met in the long term.

The requirements for a statistical business register may actually be different from those of an administrative business register. Even if the statistical business register is fed by sources of information that constitute administrative registers, the concepts, characteristics, methodologies, database maintenance and other elements, of the statistical business register need to be based on statistical concepts, preferably on international recommendations.

An important feature of the system of economic administrative records is that it is only created to be used for statistical purposes. It cannot serve other purposes, especially administrative ones. This is due to the need to protect the confidentiality of individual information collected by the information systems. It must be ensured that data on individuals or businesses collected by the information systems for statistical compilation are kept strictly confidential and used exclusively for statistical purposes.

1. Establish the objects of the register

The main theoretical framework on which to base the design for the statistical observation units of the system of economic administrative records and the statistical business register as the core of the system is the System of National Accounts (SNA) 2008. References to terms and definitions on which the statistical register must be consolidated can be found in it.

Both the basic observation units of the system of economic administrative records and those of the statistical business register, as part of that system, have the same classification. They can be:

- Institutional units, which are the economic units that can engage in the full range of transactions, own assets and incur liabilities on their own behalf.
- Institutional units that are resident in the economy are grouped together into five mutually exclusive sectors composed of the following types of units:
 - (i) Non-financial corporations, which are institutional units principally engaged in the production of market goods and non-financial services.
 - (ii) Financial corporations, which are institutional units principally engaged in the provision of financial services, including financial intermediation.
 - (iii) General government, which, social security funds included, consists of institutional units that, in addition to fulfilling their political responsibilities and their role of economic regulation, produce services (and possibly goods) for individual or collective consumption mainly on a non-market basis and redistribute income and wealth.
 - (iv) Non-profit institutions serving households, which are legal entities that are principally engaged in the production of non-market services for households or the community at large and whose main resources are voluntary contributions.
 - (v) Households, which are institutional units consisting of one individual or a group of individuals. All physical persons in the economy must belong to one and only one household. The principal functions of households are to supply labour, to undertake final consumption and, as entrepreneurs, to produce market goods and non-financial (and possibly financial) services. The entrepreneurial activities of a household consist of unincorporated enterprises that remain within the household except under certain conditions.

The system of economic administrative records will only take into account as observation units for basic statistics from administrative records those that produce goods or services, which are:

- non-financial corporations;
- financial corporations;
- general government, and
- non-profit institutions that produce goods and services.

Administrative units classified as non-profit institutions that do not have any type of production, as well as households, are not part of the system of economic administrative records.

Institutional units that engage in economic activities involving the production of goods and services are commonly classified as enterprises. Enterprises may be linked, forming business groups, and a company may in turn be divided into other units subordinate to them. These can be constituted as statistical units within the statistical business register. Different classifications of statistical units referring to companies can be observed, depending on the form of division:

- (i) Institutional units.
- (ii) Business groups.
- (iii) Enterprises.
- (iv) Establishments (local units by activity).
- (v) Local units.
- (vi) Units by type of activity.
- (vii) Other (specify).

Non-profit institutions that produce goods or services or other units not constituted as enterprises can be found in item vii (other).

The model in table IV.1 can be used for the classification of subordinate-enterprise statistical units.

Table IV.1

Subordinate-enterprise statistical unit classification model

		Divided by locality	
		No	Yes
Divided by activity	No	Enterprise	Units per location
	Yes	Units by activity	Establishments (units by type of activity and by locality)

Source: European Commission and others, *System of National Accounts 2008*, New York, 2016; United Nations, *International Standard Industrial Classification of All Economic Activities: Revision 4* (ST/ESA/STAT/SER.M/4/Rev.4), New York, 2009.

2. Public-use unique identification number for observation units

The observation unit identification number is used to establish a link with the totality of the information reported by all sources of information of the system of economic administrative records, whether administrative or statistical. This makes it possible for individual economic analyses to be performed and to ensure the transfer of information from administrative records to statistical records. For example, structural analyses of economic activity can be done by linking statistics and data from administrative records on production, customs, taxes, finance, labour, wages and others.

The statistical business register contains all the units that perform economic activities related to a unique identification number, also called a key number.

The design and use of the unique identification number should be defined separately in each country, always according to the premise that it should be a unique code for each observation unit of the statistical business register, even when the unit physically disappears. It is recommended that it be an alphanumeric code, with one or two digits that identify the type of unit registered according to one of its most important classifications. The classification criteria can be selected from among the most significant ones, such as type of activity, sector of ownership, main product, and so on, with four or five digits then added that respond to a consecutive number, although the structure is not relevant.

The statistical business register must be capable of containing attributes that identify the observation unit, particularly those aspects of its characterization that are not subject to frequent changes. It may have components that identify it according to its fundamental characteristics, such as those listed below:

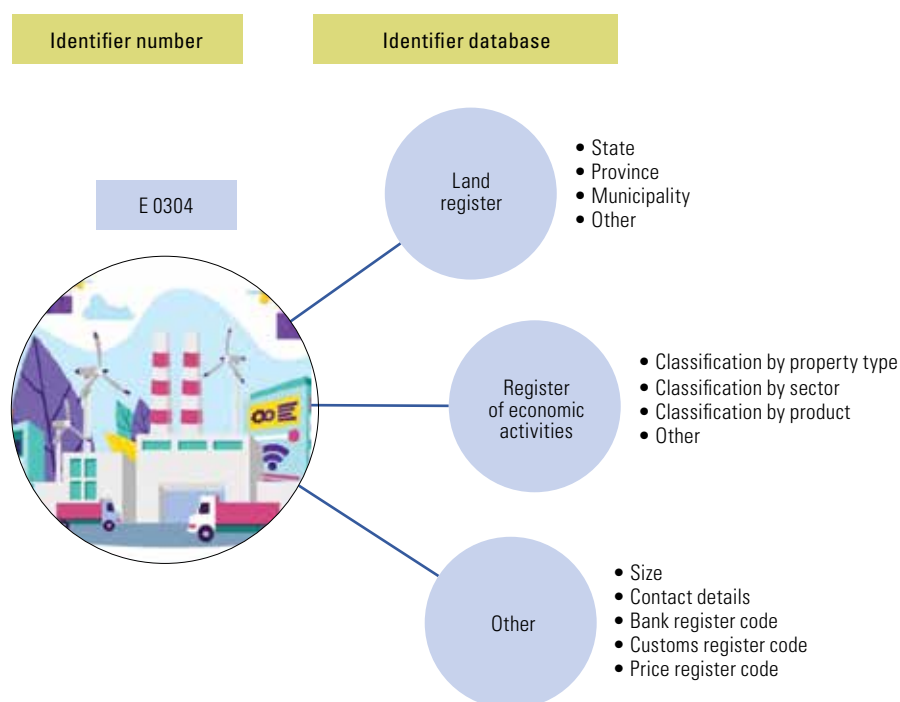
- Trade name: the name by which the unit is known in its economic relations.
- Residence address: the residence of each institutional unit is the economic territory with which it has the strongest connection, in other words, its centre of predominant economic interest. Each observation unit is a resident of a single economic territory. Each enterprise belonging to a business group is a resident of the economy in which it is located, not of the economy where the head office of the corporate group is located.
- The residence address should independently provide data for each territorial division, according to the legal and geographical framework established by the country and, if possible, the georeferencing of each unit. In this way, analyses can be obtained at different territorial levels.
- Contact details: including telephone, e-mail, fax and postal address.
- Institutional sector, classified according to SNA 2008.
- Classification of the business sector according to ownership (state, private, mixed, or also, domestic, foreign, mixed).
- Type of production (market producers, own account producers or non-market producers) according to SNA 2008.
- Core products (goods or services) (minimum of 1) (Central Product Classification (CPC) version 2).
- Classification of the economic activities it performs.
- Categorization according to size (micro, small, medium, large).
 - Date of emergence or creation of the unit
 - Start date of operations
 - Date of definitive cessation of activities
 - Operational status (e.g., high, low, halted, bankrupt) that describes the level of functionality at the current time
 - Unit to which it belongs and subordinate units (linked with the corresponding unique identifier number, so that they can be linked in the database)
 - Category by contribution to GDP
 - Unique identifier number used for the incorporation of the enterprise
 - Unique identifier numbers for the unit issued during the enterprise's lifetime due to modifications in the original registration, from which the enterprise can be considered a continuation of other previously registered enterprises.
 - Number of employees, by gender and salary category.

When each observation unit is opened in the statistical business register, in addition to the information as to the unique identification number and the name of the unit, the identification key will provide access to the necessary and available information. Diagram VI.2 shows an example of this characterization.

These attributes may be part of the statistical business record itself or may be contained in another database. In other words, the statistical business register may contain only a list of statistical units with their identification number. The variables associated with the register are documented in another database and can be accessed in reports in response to inquiries made using their identifier number. Those identifying variables are useful for statistical analysis based on the combination of the economic information received that is associated with the identifier number of the basic unit registered.

Diagram IV.2

Schematic diagram of the identifier database for the observation unit

**Source:** National Office of Statistics and Information (ONEI) of Cuba.

In future developments of the system of economic administrative records, following the creation of statistical business registries in all countries in the region, the database identifying each country's observation units could be useful to consider establishing a unique identifier number design at the regional level, using the identifying variables as design sources for the digits in the unique identifier. This would be useful to obtain more information on the units registered by their identifier number, without having to access the unit identification data for each country through its identifying record. However, it has the disadvantage that some of these variables may change during the active life of the unit, which would complicate serial statistical analyses.

The unique identification code should be assigned the moment the entity is created, which varies depending on the law in each country. Some countries use the tax identification code, or one is assigned by the national statistical office once incorporated into the statistical business register. Correlations between different identification codes can also be used, for example, between the statistical business register code and the tax identification number.

The legal recognition of this unique identification code or the synergy between the national statistical offices and the different institutions where institutional units, such as tax offices, banks and customs offices, must register, guarantees the widespread use of this identification code and contributes to its application and generalization within the country's system of economic administrative records.

The key or unique identifier number associated with each observation unit, in addition to having the function of linking the associated information, contributes to economic statistical analysis with other administrative records, whether or not they are economic objectives. It can also be used by these administrative registers to link the information corresponding to their functions with the basic economic unit, thus creating a two-way flow of information.

3. Inventory of sources

Generally speaking, administrative records are not created for statistical purposes. Hence the importance of evaluating their potential as a possible source of information to generate statistics within the system of economic administrative records or to serve as complementary or contrasting information in other statistical operations of the system of economic administrative records.

Each country should carry out a study to select the appropriate sources of information for obtaining the statistical information required by the system of economic administrative records. The selection approach, in its first phase, is to determine the value of administrative sources for providing information for developing the system of economic administrative records, taking into account the needs presented. In other words, it is a matter of determining how effectively the data can be used. Other considerations that can be defined prior to the quality analysis of the administrative source include:

- (i) Legal and institutional framework: the administrative register is protected by mandatory laws and resolutions governing its use. It has procedures that define the processes for updating and maintaining data, as well as other descriptive ones. These procedures are periodically evaluated with a view to improving the quality of the administrative record.
- (ii) Geographic coverage: a decision is made as to what level of territorial subdivision (established in the country by law) units will be geographically grouped at (municipality, county, province, state, etc.) and it is verified that the administrative source can meet the desired coverage.
- (iii) Business coverage: it would be best if the selected administrative sources can cover all levels and classifications of the country's production units, based on the classifications regulated by the national statistical institutes, as part of their function: international and national classifications, type of activity, sectors, size and other economic and territorial classifications. It would be acceptable, however, to choose those that administrative records can cover if their variables meet the quality requirements established for statistical data (relevance, accuracy, timeliness and punctuality, accessibility and clarity, comparability, coherence and completeness) (see section IV.B on the design of variable descriptions).
- (iv) Range: administrative sources can be chosen that provide total information on the population of observation units collected in the statistical business register, or partial information as long as they provide a representative percentage of the total number of observation units that can be assumed to be valid in the country or that can be completed with information from other registers.
- (v) Referential integrity: a unique identification number for each unit, which can be validated with one from the statistical business register. Standardized classifiers are used that can be validated with those used internationally. The databases are documented by standardized metadata.
- (vi) Computer technologies: a review is performed of the administrative records that provide or exchange information with the system of economic administrative records to determine which ones operate with computer technologies that are accepted as valid in the country in terms of ways of accessing and updating the data in the administrative record and that also have the capacity to link with the computer technology with which the system of economic administrative records operates.
- (vii) Use of administrative records for analysis, delivery and dissemination of information: assess whether the functions of administrative records include data analysis, delivery and dissemination, as well as the frequency and quality of those processes.

Initially, the coverage, ranges and technologies provided by the different sources consulted, which have been recognized as valid due to their compliance with data quality requirements, can be accepted, as will be seen below. Subsequently, by means of development agreements or by modifying laws limiting use, the limits of acceptability

and accessibility can be deepened. National statistical offices can work with owners of administrative records by carrying out methodological and convincing preparation, so that the characteristics of the administrative record are adjusted to the needs of its statistical use.

An example of a questionnaire for an inventory of administrative records for statistical purposes applied in Cuba can be found in annex A3.

In the case of Peru, a computerized system has been developed for the inventory of administrative records, which includes questionnaires that determine the characteristics of providers and the content of administrative records. The Inventory System for Administrative Records (SIRA) has provided information on municipalities and government entities.

In parallel to the selected administrative sources, other sources that would provide statistical reference information can be accepted. They can be used to ascertain that the data are as reliable as possible and meet requirements. Institutional agreements governing the data and frequency of input may provide the legal framework for the use of these sources. Examples of these are:

- The data provided by ministries that regulate the economic activities of the units of the statistical business register, which are different from or may even be linked to those contained in the administrative records, as well as meeting the same previously agreed quality requirements.
- Statistical registers that are already defined for economic and demographic statistical studies by other statistical institutions operating in the country, and that can provide the identification data and analysis variables of the observation units registered in the statistical business register, according to the different classifications for their reports. It should be verified that they have the capacity to be linked to the administrative records selected as sources.
- The records of national statistical offices that respond to requests for information envisaged in international agreements or commitments (e.g., information needed to fulfil the Sustainable Development Goals).
- The records of the national statistical offices that contain the information provided by government departments of the territorial divisions under a mandatory legal framework, with the statistical data required for the economic statistical databases, from the basic territorial unit to the national level.
- Surveys or direct reports designed to provide economic information on the websites of national statistical offices with regulated access.
- The data that national statistical offices in each country obtain from their territorial units, through traditional channels of communication with the national level.
- Economic data from the management of the units of the statistical business register in localities based on previously agreed mandatory surveys.

4. Criteria for the selection of sources

In order for administrative records serving as the sources of the system of economic administrative records to be selected, they must meet certain quality attributes. Based on the principles of statistical information quality, those requirements can be summarized as follows:

- Content: the register must include the data that allow the statistical operations necessary for its use to be prepared and carried out.
- Accuracy (or validity): assurance that the data they provide are not erroneous.
- Age: the data contributed must reference a recent update.
- Consistency: the definitions of units and variables should not change over time.

- Clarity and transparency: the definitions of units, variables and other data should be easy for users to understand.
- Availability and accessibility: record data should be easy to obtain.

Before opting to use an administrative register, it is necessary to guarantee some basic institutional and operational premises for those records (administrative records and ECLAC statistical and prospective studies):

- (i) Regular access to the administrative register recognized in some type of law or institutional agreement that addresses aspects such as the frequency and time periods in which data must be available.

The fundamental limitation that may be encountered in the first selection is the legal framework guaranteeing the accessibility of administrative records for statistical purposes.

The laws in force may prevent their use by other institutions or the exchange of information for purposes other than those for which they were conceived. For that reason, it is also necessary to review and study the laws that protect administrative records and, if necessary, propose legislative changes to make the use of those with value as a source in the creation of the statistical register more flexible. In any case, it must be ensured that the rules on confidentiality are observed.

- (ii) The important variables for the directories must meet a minimum level of quality, in addition to including a variable indicative of size (turnover, number of employees or others) for sampling purposes, for which it is necessary to know what procedures are used in the administrative data collection and cleaning processes.
- (iii) Estimated degree of coverage, error associated with each of the main variables and type of treatment adopted to identify units created and dissolved.
- (iv) Procedures used to adjust administrative data to the requirements of the statistical directory in terms of statistical definition of basic information units, especially in enterprises with more than one establishment that need to align the concept of institutional unit with that of statistical unit.
- (v) Ideally, statistical and administrative bodies should adopt a uniform activity classification system.
- (vi) Identification of a link between the units in the administrative register and the corresponding units in the statistical directory; for example, the administrative tax identification code.
- (vii) Joint effort between administrative and statistical bodies in technical discussions for introducing enhancements to data collection and processing processes.

In general, the quality of the administrative register and the databases that are to be used as sources for the system of economic administrative records must be adequate in order to meet requirements with set indicators, such as:

- Percentage of controlled units in relation to the total population.
- Percentage of units with empty data fields.
- Percentage of units registered more than once in the register or database.
- Percentage of units with false fields in relation to what is established by the procedures and metadata of the system of economic administrative records.
- Percentage of fields with last verification more than two years old.

As for the level of quality, the standards for each indicator may be as follows: less than 5% is considered good, less than 10% and more than 5% is considered average, and more than 10% is considered poor.

Once the sources that meet the basic requirements have been selected, there may be more than one source that both meets these requirements and is capable of providing the same information. In selecting between them, the one that best meets the requirements that were taken into account when evaluating the selection should be chosen.

At all points during the study of the administrative register, it should be specified how it meets the requirements related to the names and specifications regulated at the country level, or at the international level, as the case may be, as well as its advantages and drawbacks, with a view to making more exhaustive and definitive assessments in the future.

B. Design of variable descriptions

1. Statistical outputs to be generated by the project

The statistical outputs obtained from the system of economic administrative records should correspond to the information needs that resulted in the analysis in the previous chapter (on needs specification). Therefore, the statistical variables included in the catalogue must contain at least the data necessary to obtain them.

A study is made of the administrative and statistical sources capable of providing the necessary variables with the required quality (see section IV.A.4 on criteria for the selection of sources). These variables are recorded in the metadata of the system of economic administrative records and are the ones that must be captured for statistical studies until a review of the entire methodology used is scheduled, based on a new analysis of information needs that requires alterations to the design of the system of economic administrative records.

With a certain frequency specified in the regulations, a review and assessment of new needs is carried out, based on which the resulting variables and indicators can be changed.

First, the variables with which to calculate these economic indicators for qualitative and quantitative analysis must be defined, and the sources and statistical method to be used to obtain the variables must be selected. That will provide the premises to design the operation of the system of economic administrative records in terms of which databases to use and how they will be linked, in terms of data collection from the selected administrative and statistical sources, in order to provide all the relevant information for obtaining the statistical output.

2. Identification and selection of variables following statistical and conceptual criteria

The data must be processed to generate the variables of interest to the particular project or programme. The variables are chosen and designed based on two premises:

- (i) Those necessary to meet the needs established in the previous chapter.
- (ii) Those with sufficient data for the analyses that can be obtained from the selected sources.

When setting up the system of economic administrative records, it may be necessary to train a registrar or team of registrars capable of harmonizing the data reported from the sources containing the information required for the statistical register. It should be noted that sources may exceed the information that needs to be obtained or be limited in that regard. All the characteristics of the variables to be obtained from the metadata of the system of economic administrative records must be defined in advance.

The national statistical offices should have a rule that the metadata catalogues of the system of economic administrative records must contain and describe the variables contained in the register and the statistical use of those variables.

That catalogue's design must be governed by a legally endorsed and regulated mandatory document. The metadata must have standards that make it possible to determine that they meet the rigorous quality standards expected of the information provided by the register. Those standards should be referenced in this document.

The metadata catalogue of the system of economic administrative records should contain the following information:

- Identification of the administrative register
- Responsible entity
- Contact information
- Objectives
- Legal framework
- Observation unit
- Geographic coverage
- Basic concepts
- Use of nomenclatures and classifications
- Database variables
- Data collection methodology
- Instructions, manuals or guides for the register's development
- Register control and quality control
- Storage
- Use of the register
- Accessibility

Defining the following characteristics in the metadata of the system of economic administrative records would provide a guide to the requirements for the review of variables:

- Technique to be used for data transfer: data can be entered manually from a printed survey or updated from an automatic survey designed to capture information from the web or by transfer from a previously created database.
- Uptake unit of measure: harmonized units of measure should be selected for the required analyses. Based on the above, there are two variants:
 - (i) the units of measure captured from the source data is the same as that used for statistical analysis, including in the case of data capture from transfer software, and
 - (ii) the other variant would be to programme their capture along with the choice of the units of the report data, since the capture software performs the conversions to the appropriate units of measure, so that the result of the analysis is homogeneous. Establishing a single unit of measurement for data capture is recommended as the most practical approach, taking into account the possibility that the registrar or reporter may not be aware of the appropriate units of measure, which could lead to confusion and errors.
- Number of decimals allowed: the number of decimals allowed in each variable must be defined.
- Negative numbers: it is necessary to define whether the variable accepts negative data or zero values ("0"), and what steps to follow in case they are not accepted.
- Coverage of the expected variable by each source: the limitations of data use should be defined in terms of the coincidence or difference that may arise when it is necessary to use more than one source to obtain a variable. For each source, the data that are or are not desired must also be specified.
- Benchmark values: the reporting entity and the registrar must be familiar with the data provided as text values or numerical ranges, which must be harmonized in an accessible file; there must also be specific guidance on what action to take in the event that the data from the information source are not consistent with those of the statistical register being compiled.
- Validation: benchmarks by which to assess whether the data obtained meets the expected requirements.

The input data from the sources must coincide with the those actually observed at a given time from a sample of selected representative units or from all units with available information. Sufficient data should be collected from the units under review to allow for a valid statistical analysis that can provide a trend or pattern according to the sample analysed.

Therefore, variable data must be obtained to extract indicators that are capable of showing trends over time and the degree to which the different components of the variables that make up the time trend are affected.

In turn, the way in which timeless results are obtained must be identifiable as a constant feature of the analysis.

C. Design of collection

The national statistical offices work together with the institutions to establish the design for the collection of the information contained in their administrative records, the content of which has information and can be linked to the statistical business register.

As part of the collection design, the best-suited methods and instruments are determined according to the characteristics established for the register. Also included is the design of any formal arrangements related to the provision of data, such as inter-institutional agreements or memoranda of understanding, as well as confirmation of the legal basis for collecting the data.

Important steps in the design are:

- (i) Review of the legal frameworks of the different institutions that have administrative records with statistical value.
- (ii) Definition of confidentiality agreements on the proper use of information from administrative records, particularly when they include sensitive information.
- (iii) Definition of communication arrangements between national statistical offices and governmental and private institutions to ensure the exchange of information.
- (iv) Design of protocols for data storage and custody at a centralized repository for internal use by different departments of national statistical offices.
- (v) Use of a software platform for efficient and secure transmission of information.
- (vi) Communication mechanisms within national statistical offices to share information provided by public or private institutions, guaranteeing the confidentiality of the information.

It is recommended to check which regulations cover the administrative register, in order to confirm the legal basis for data collection and whether it includes metadata, the update frequency and whether it is auditable, among other aspects.

Based on this, agreement templates should be established to help standardize the terms and conditions for the exchange of information between institutions and national statistical offices. Inter-institutional agreements or memorandums of understanding should include legal and methodological considerations, such as the legal framework for the exchange, the responsibilities of each institution in terms of delivery dates, periodicity, technology and procedures for data transfer, reconciliation and quality control processes, confidentiality clauses, information treatment and dissemination.

The contents of this legal arrangement (whether an agreement or memorandum) should include all the agreed provisions regarding the delivery of information, the persons responsible and the dates and periodicity of delivery and receipt of information. In addition, they should detail the variables and data required, the mechanisms for information reconciliation, advisory agreements, the institution's responsibility for the timeliness and accuracy of the information, commitments to information audits by the national statistical office, confidentiality treatment, conditions for the dissemination of data and action clauses in the event that any of the terms established in the arrangement are breached.

The agreed delivery format depends on the characteristics of the administrative record. The same format can be used or only part of the information it contains. The administrative record can be delivered as databases or through questionnaires, in digital or printed format. The means of transmission or delivery should also be indicated, or provided via authorized links on the institutional website, depending on the degree to which institutions are digitalized and automated. The characteristics and terms of the exchange of information must be clearly defined ahead of the information exchange agreement.

In the event that both the institution that delivers the information from its administrative record and the national statistical office have a software platform that can be aligned in a network and supports consultation of previously agreed databases, this can be a mechanism that allows the two-way capture and verification of information. If possible, at the time the administrative record is updated, it is advisable to allow the data previously linked in the records of both parties to be consulted or validated automatically to avoid duplication or false links between records. This can be a long-term working objective between both parties.

The national statistical office should have infrastructure that allows the consultation of official statistical records in a database common to all areas that work with the information in the records. The updates made to statistical business registers from the sources that provide the institutional records with certain periodicity go through a period of review and validation before being considered useful information for statistical processing. These must be published after approval by the national statistical office.

D. Design of processing and analysis

Data that help in the creation of base registers for units or objects of observation of strategic interest, among others.

In the design of the system of economic administrative records, information can be obtained at three levels:

- (i) Administrative records selected as sources of information divided into:
 - a) Administrative business registers.
 - b) Economic administrative records.
 - c) Other administrative records useful for classifying and linking information among the above.
- (ii) Statistical business register characterized by:
 - a) Unique identifier key.
 - b) Unit name.
 - c) Databases of identifying variables (identifier database) (BDI1) (BDI2) (BDI3) ... (BDIn).
- (iii) Databases built through direct collection of information.
 - a) Economic censuses (if conducted in the country).
 - b) Sampling frames.
 - c) Field surveys.
 - d) Telephone surveys.
 - e) Internet surveys.
 - f) International statistical frameworks.
 - g) Other information sources.

The statistical business register is mainly fed by the administrative business register and is complemented or validated by other sources, which can be censuses and surveys designed to configure a statistical business register that guarantees coverage and pre-established ranges to cover all statistical observation units in the country.

In putting together the administrative record with statistical value, the units of the business register obtained are linked by means of their identifying key to the identifying numbers of the statistical classifier databases that form part of the international statistical frameworks, with which the national statistical offices operate.

Once these information bases have been created, with the periodic updating of the records and databases, the current and valid characterization of the country's business units can be obtained in a single query. In addition, by making several queries with a certain periodicity (for example, annually), it is possible to study how the business economic environment changes across successive periodic measurements.

This system must be linked and interact by exchanging information and data with the statistical records that contain variables and indicators of economic statistical analyses that already exist in the institution, such as systems of accounts, tax payments, employee payroll, production reports, trade and customs, among others.

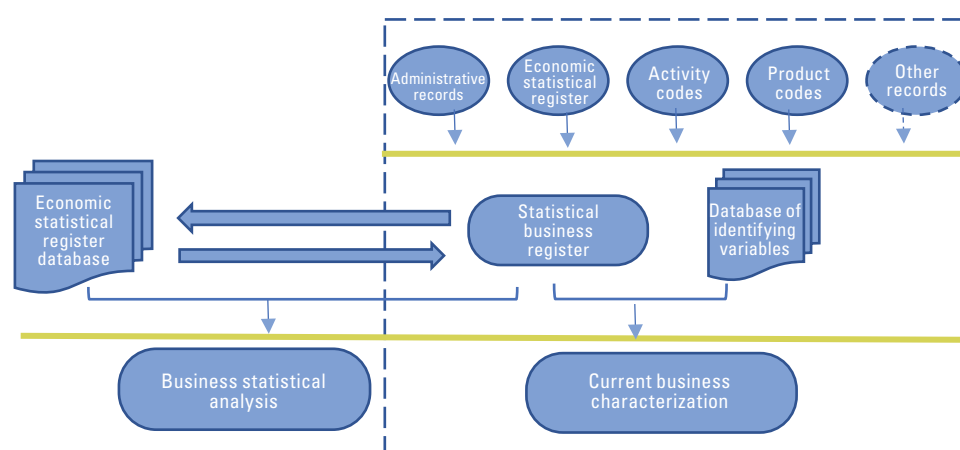
It is possible to link them based on four premises to be considered:

- (i) Data entry to these statistical registers can be linked to the unique number that identifies the unit registered in the statistical business register without modifications to the current register databases.
- (ii) Data entry to these statistical registers is linked to another identification number that can be validated with the unique number that identifies the unit registered in the statistical business register.
- (iii) Economic statistical records are not designed to be linked to the unique number that identifies the unit recorded in the statistical business register, but by making minor modifications to the databases a field can be established to link them.
- (iv) It is not possible to link economic statistical data without first making considerable modifications to the current computer systems.

Each of these premises has a different working system to be assessed and corrected in order to meet the objective of establishing a link between the statistical business register and the economic statistical registers currently in use. The structure should look like the diagram shown in diagram IV.3.

Diagram IV.3

Schematic outline of linkage of the statistical business register with economic statistical registers



Source: National Office of Statistics and Information (ONEI) of Cuba.

The input data provided for the statistical business register may change significantly over a period of a year due to the dynamism of the business economy itself, as well as the way the economic environment is affected, which relatively frequently causes businesses to change their purpose, technology, structure and even their territorial location. For this reason, the updating of the statistical business register from the sources chosen should be continuous.

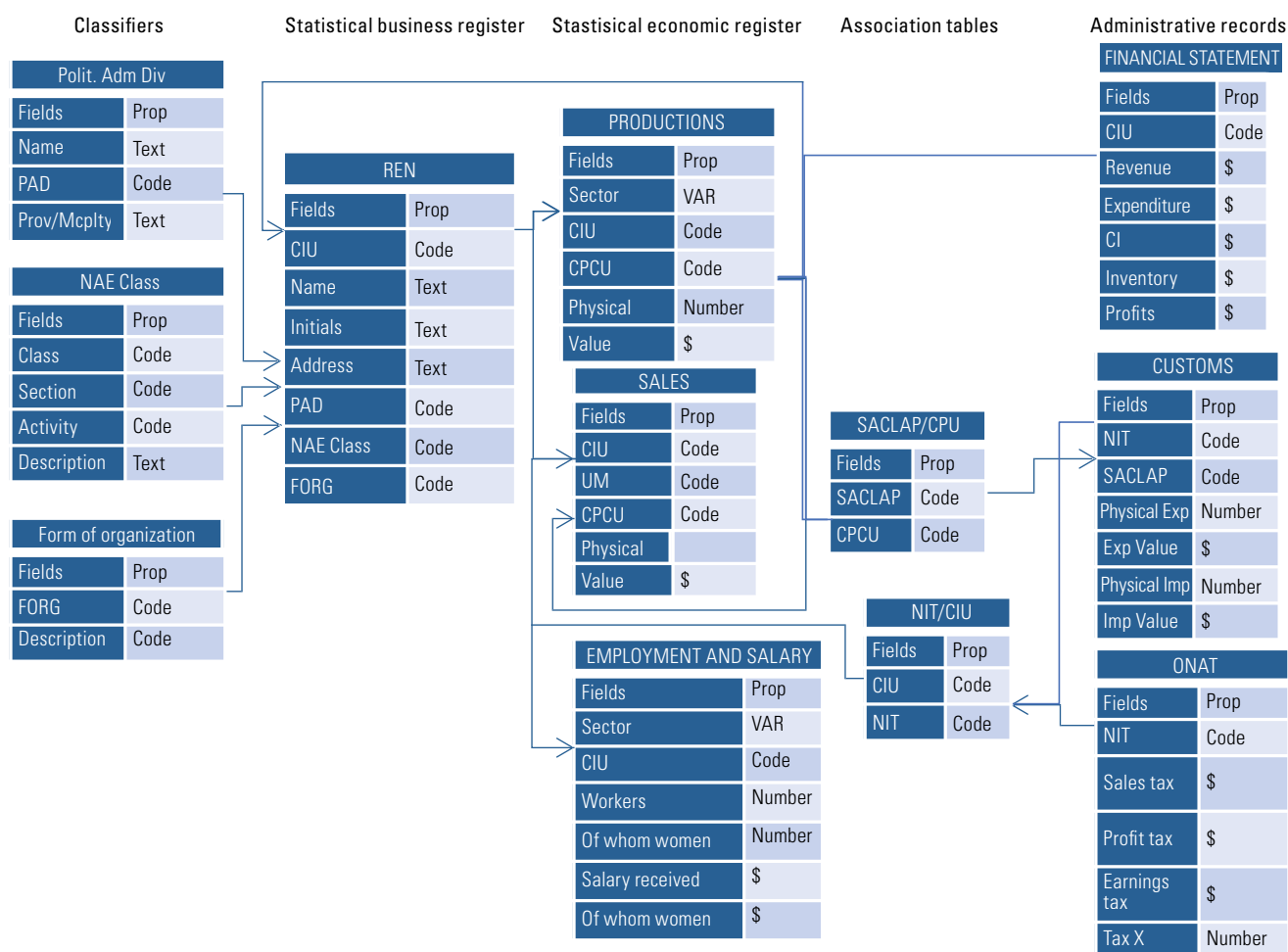
Ideally, there should be a warning system in case of a modification of any of the data obtained for the basic unit. This system would work through a notification issued by the owners of the administrative register or other sources used, to the national statistical offices, with the content and description of the modifications that have been made. For this notice to be complied with the requisite effectiveness for the analyses issued by the statistical business register, the notice system should be mandated by law or established in contracts setting out the obligations of the contracting parties.

Another way to keep the data current would be to take on the task of regularly updating the statistical business register, which would result in an additional workload for the registrars. An agreement is reached to deliver the updated source data at certain intervals and, after careful examination to determine what modifications have been introduced in the sources since the last update, the data are identified and updated in the statistical business register.

As for maintenance of the links of the statistical business register and its BDIs with the databases of the statistical economic register, future work on their maintenance would be governed by an alert taking into account the link already established between the two registers whenever there are any updates: new inclusions, modifications or deletions. The alert indicates whether the link with the statistical business register is maintained or added, so that they work through the same linkage system without errors. This ease of database maintenance is due to the fact that the national statistical offices own both the statistical business register and the statistical economic registers to which it can be linked (see diagram IV.4).

Diagram IV.4

Example of a model database for the system of economic administrative records



Source: National Office of Statistics and Information (ONEI) of Cuba.

E. Design of production systems and workflows

This study should be conducted by national statistical offices. They will be responsible for the preparation, maintenance and updating of administrative statistical records.

The figures within the national statistical office who are responsible for this function may be appointed in accordance with the office's own organizational system or, alternatively, by an ad hoc working team, either temporary or permanent, comprising specialists needed to perform these functions.

1. Steps in the selection and use of sources

The objectives of this working group in the first stage of system's design would also determine the initial content of the functions to be carried out, which would be:

- Collect all the legal information that supports the process of formation and maintenance of the records. Evaluate whether current legislation is sufficient, or whether it is necessary to reform or adopt new laws and resolutions to ensure that the needs that arise are met.
- Define the selection and acceptance criteria at each stage of the records development process.
- Make an inventory of the sources within the country that can provide the information necessary for the preparation of the statistical business register. Subsequently, evaluate which ones meet the predefined selection criteria.
- Direct and advise the computer processing for the insertion of the necessary database creation and updating processes within the existing systems or the creation of a new system to be linked with the existing one.

Once the first stage has been completed—or at the same time, as definitions on how to proceed emerge—a process begins of preparing the legal and methodological framework on which the exchange of information, the creation and maintenance of the databases, as well as security in the handling and publication of the information collected will be based. This process features the following:

- Approval of the reforms and laws guaranteeing the legal framework for the entire process and, as part of that process, the legal and methodological framework for the exchange of information with sources, managers and users of the information.
- Preparation of the record metadata and variable and indicator fact sheets to be used.
- Management of the first data delivery, validation of names and identifiers of data and variables, while measuring compliance with the coverage and range specified in the source evaluation, discarding information that will not be used and making links to databases as they are formed (see more details in section IV.E.2, on steps in the design of linkages between source data and statistical register databases).
- Measurement and final evaluation of the register formed. Perform possible analysis exercises, starting from the databases created.
- Final design report, including strengths and weaknesses, with a view to making improvements.
- Preparation of a system operating manual for guidance and functionality queries in database maintenance and update work modes.
- Training of registrars and operators who will be in charge of the operation of the system with the final design basis.
- Training of statistical specialists and users of all kinds in the analysis possibilities of the information provided by the published databases that can be used for consultations or as a basis for new possibilities for statistical reports for governmental purposes, financial consulting, economic studies, educational studies, and so on.

At all stages of the processes, it must be ensured that only global or sectoral analyses are made public, and that there is no way of accessing databases that could disclose classified information on basic units in the register.

2. Steps in the design of linkages between source data and statistical register databases

The first step is to review the quality and coverage of administrative data, checking several points, including the following:

- The file is of the expected size in relation to the number of records and variables.
- The main variables have a valid format in terms of range, designation of “text” or “numerical values,” etc.
- The coverage of the main variables includes values for all basic units.
- The main variables are reliable, intends of numerical variables corresponding to a range of admissible values and the codes used are valid, among other things.

If this first quality analysis reveals that the source variables are of sufficient quality and can help increase the quality of the statistical business register, it is necessary to devise a set of procedures to dictate how to pre-process the variables in the administrative register in order to meet the requirements for their transfer to the statistical business register. For this purpose, the regulated definitions in the register metadata regarding statistical definitions and application, among other things, must be followed.

This set of procedures should include the necessary steps, such as the creation of algorithms and secure databases to convert variables and classifications, such as the conversion of codes in the activity classification, the details of the geographic location code, etc.

Chapter V

Construction

Introduction

The purpose of this phase is to build and test the IT infrastructure, components, applications and software services, in order to create a complete operational environment for the production of information, as well as the tests supporting it.

According to process standardization logic, the inputs provided by the design process facilitate the construction of different components of an operating environment from which the statistical operation can be generated. Those components should be flexible, since they require continuous improvement to adapt to changing information needs, and the incorporation of information sources defined in the design process.

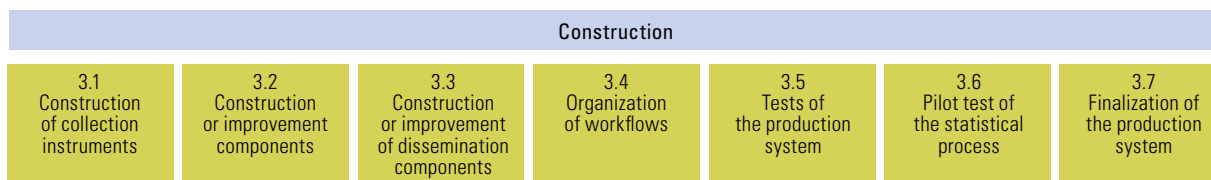
A. Definitions at the process level

The GSBPM indicates that within this process the production solution is built and tested until it is ready for use in a “live” environment. In this way, the inputs yielded by the design process should facilitate the building of an operating environment in which the statistical operation can function. In keeping with logic of process standardization, the components developed at this stage should be reusable if possible and enhanced in line with the requirements identified in the design process.

This process is divided into seven subprocesses, as shown in diagram V.1.

Diagram V.1

Subprocesses included in the construction phase



Source: Economic Commission for Europe (UNECE), “Generic Statistical Business Process Model GSBPM (Version 1.0 in Spanish, September 2016)”, Geneva, 2016.

In order to understand the organizational logic with regard to the administrative records defined in this document, a detailed description of each of the GSBPM subprocesses is presented.

1. Construction of collection instruments

Within the GSBPM, this subprocess is defined as describing the activities for constructing the instruments to be used during the “collection” phase.

While the collection instrument can be understood as referring to a questionnaire, the GSBPM also refers to different collection methods, and mentions: “Collection instruments may also be data extraction routines used to gather data from existing statistical or administrative data sets”

In that regard, the national statistical office, in agreement with data providers, must establish an environment for the exchange of administrative records. That environment should consider security and storage aspects so that all the necessary requirements are met.

In addition, extraction routines can be automated, semi-automated or manual, depending on how advanced the various data providers are in IT terms. Nevertheless, the national statistical office must adapt to all possible scenarios, in order to capture the data necessary for the construction of the statistical business register, as well as all those that will allow economic administrative registers.

Thus, data exchange may occur in an IT environment furnished by data providers, or one made available by the national statistical office.

Where the administrative records to be collected are considered small in volume, they could be transferred to the national statistical office by other digital media, such as e-mail or magnetic disks. Likewise, steps should be taken towards greater automation and a secure environment to protect any sensitive information received.

Therefore, in this subprocess technological tools are developed to support the activities of transmission and reception of the variables in the administrative records, derivative variables, files or statistical or non-statistical databases that will be the object of study. This means developing the technological means of transmission and interoperability of data obtained from administrative records, files or statistical or non-statistical databases, developing the techniques and tools to verify that the information sent or received satisfies the quality conditions established for data exchange; and preparing the procurement plan, as well as the bids, contracts, agreements or arrangements necessary to perform the statistical operation.

The importance of developing technological means of transmission and interoperability must be considered. This is becoming a challenge for data exchange, in the sense that there is an increasing volume of information available for use in statistical operations. Thus, national statistical offices must be prepared to face these challenges in their different realities.

National statistical offices must be prepared to capture different types of data. These can be classified according to origin, rank, structure, and so on. According to their structure, three types of data can be defined: structured, semi-structured and unstructured. Structured data are typical for most relational databases. These databases typically have a given system that determines what form the tables in which the data are stored take, what type of fields they have and how they relate to each other. They are operated using SQL (Structured Query Language), which is designed precisely to manage and retrieve information from relational database management systems.

Semi-structured, non-relational or NoSQL data do not fit into a tabular format but are organized using tags that allow them to be grouped and hierarchies to be created. Thus, NoSQL databases make it possible to store information that do not fit well into a register or tabular format—such as variable-length text—and also facilitate the exchange of data between different databases. An example of semi-structured data is delimited files, which contain elements that allow data to be divided into independent hierarchies.

In the case of unstructured data, their volume and growth are radically higher than those of structured data. They are diverse in origin; for instance, information extracted from the Internet using web scraping techniques. Due to its structure, it is not possible to use the relational architecture, making it necessary to work with big-data tools, and aspects related to scalability and parallelism become critical in these architectures.

Historically, national statistical offices have only focused on extracting and analysing information from structured data. However, with the current growth of semi-structured and unstructured information, national statistical offices must look for solutions that can help them analyse all three types of data. This represents an analytical advantage and, in fact, the ability to extract value from them is one of the main drivers of the rapid growth of big data.

2. Construction or improvement of components

This subprocess describes the activities to build new components and enhance existing ones, as well as services required for the processing and analysis phases, depending on what has been determined in the design phase.

The GSBPM references core elements for work involving administrative records in terms of computer system features, which may “include dashboard functions and features, information services, transformation functions, ... workflow frameworks [and] provider and metadata management services” (UNECE, 2019).

Given that part of the focus of the construction phase is on the development, updating, enhancement and testing of the computer systems that would support statistical operations, this section presents a series of modules that should be considered as an important part of the management of economic administrative records and the statistical business register.

It is important to understand that, in terms of systems, not only aspects related to the capture, storage and processing of administrative records must be considered, but also considerations to do with provider performance, data evaluation, data system integration and tools for the documentation of records, among others.

Therefore, it should be understood that IT solutions should provide solutions for administrative record-based statistical production with a cross-sectional view of the process of those that execute it.

The following are different considerations to be kept in mind in the development of the computer systems that support the system of economic administrative records.

(a) Technology

The technology must be flexible enough to evolve in line with new requirements. It is therefore recommended that the system to be developed fit in with the national statistical office’s information technology architecture and be as compatible as possible with other systems, such as those for administrative data acquisition and commercial survey collection.

Also, if the national statistical office has set standards, they should be followed. If not, it is suggested to think of a swift methodology for showing specific results in a regular way.

In addition, the document tensions that the system’s requirements can be divided into two categories: functional and non-functional. The former refers to what the system must do or execute, while the latter relate to the organization’s compliance with technical standards, performance, scalability, security, availability, accessibility, quality and usability. Addressing the technology solutions that will support the system of economic administrative records requires posing a series of questions that must be answered so that ad hoc solutions matching the level of development attained by the various national statistical offices can be considered.

The evaluation of these considerations and their solutions will directly influence how the system should be designed. Wallgren and Wallgren (2012) mention factors such as the storage technology used by the register, update frequency, amount of data, security and communication technologies.

Regarding storage technology, it should be considered whether it is possible to transfer administrative records directly to the national statistical office. Regarding the update frequency, it is indicated that the more frequent the transfer, the higher the level of automation that should be considered.

Related to the previous point, if the amount (volume) of data is large, a solution should be developed accordingly. In addition, with respect to security, aspects to protect data confidentiality and system vulnerability must be established.

Regarding communication technologies, possible solutions for the factors mentioned above must be established in order to choose the best available alternatives for the national statistical office.

(b) Databases

As part of the development of a system that stores administrative records and all the data that support the system of economic administrative records, specifications for how databases relate to each other should be included in the core definitions.

There are various available solutions for this purpose, including Microsoft SQL Server, Oracle and others. Although the use of unstructured data has increased in recent years, initially a relational data model must be set up so that links can be established between different records, using identifiers defined for that purpose. In addition, developing the system of economic administrative records and the statistical business register requires having unique identifiers to establish links between the statistical units and the data available in the system. This is fundamental for linking the statistical business register and economic administrative records.

As regards the construction of the system, the database requirements identified may have cost implications, for example, as regards software licensing, the type of storage required, the quantity and capacity requirements of the database servers and so on. Regarding the database item description, it is important to consider how the database model should be built, which tables and variables to store and how the tables will interact. In line with this, as previously mentioned, the use of unique identifiers to establish links between administrative records makes it possible to relate them to each other and enhance their statistical use. It could also facilitate integration between systems, as long as it eases linkages between the data to be analysed.

A relational database model has positive aspects, in that model integrity requirements can be established to ensure consistency between related tables. Restrictions and criteria can also be established for loading data into the system. Therefore, a relational data model allows the establishment and fulfilment of quality criteria for the data to be used.

Wallgren and Wallgren (2012) allude to the importance of database design for effective system operation. They indicate that “database design for an IT system for register-based statistics does not differ fundamentally from database design for other IT systems for statistical production.” (p. 202). Thus, here again is the idea that national statistical offices should give consideration to their architecture and system definitions in developing IT requirements in relation to administrative records.

Regarding databases, they mention three main types: input, processing and output databases.

The input database should “be able to receive all types of updates, including completely incorrect input data” (Wallgren and Wallgren, 2012, p. 202). Basic consistency rules can be applied to input databases.

The processing database is the one that is standardised and has stricter consistency rules applied to it. The output database, for its part, should not be changed and is optimized for the most commonly used outputs.

In addition, Wallgren and Wallgren (2012) point to a need to define aspects related to database performance within the model. For example, if there are large volumes of information, introducing certain changes may worsen performance, so other types of databases can be established to optimize the design.

Wallgren and Wallgren (2012) also say that “it is not possible to implement the register system in one physical database, probably not even on one server” (p. 204). While the server can ensure faster operation, when it comes to connecting different servers, decisions must be made as to how best to distribute the data and how necessary it is to have data distributed on different servers.

Finally, the databases must be correctly indexed to ensure the good performance of the system. This means that appropriate indexes must be established to make data cross-referencing more efficient. In this way, the primary and external keys can be indexed, in order to locate the elements that will be used in data cross-referencing more quickly. This is another aspect to be defined within a database model.

(c) Documentation

The documentation related to the system of economic administrative records needs to be established to ensure its long-term operation. This enables the history to be developed and the changes that have occurred, both at the system level and in the data being managed, to be guided and understood. This means that documents must be documented and managed to clearly establish how the modules and processes developed work. Changes within the administrative records that exist in the system should also be documented.

It is also recommended to monitor updates in administrative sources. For example, an economic administrative register that is automated should include the date and source of updates. This information can be useful for troubleshooting specific maintenance or update processes.

In line with the UNECE approach, Wallgren and Wallgren (2012) advances the logic of a complete documentation of the system and the data it contains as a core consideration. Those authors propose a set of formalized metadata and other documents.

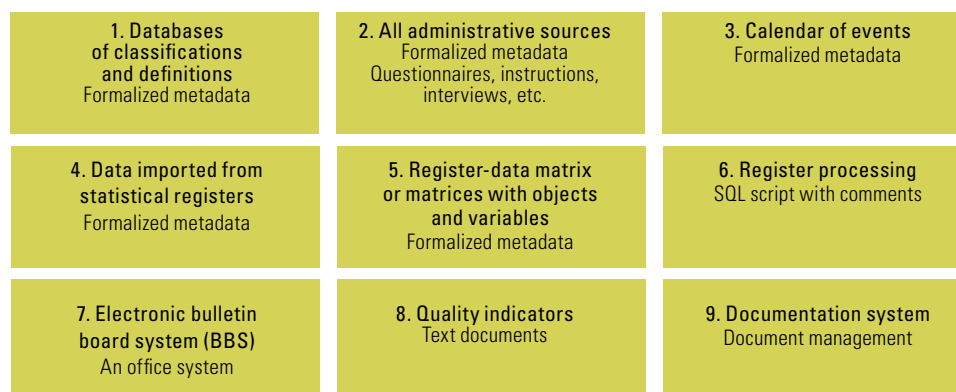
The first group includes the databases of classifications and definitions, documentation of databases from administrative sources, calendar of events (of changes in the register), data imported from statistical registers and documentation of data matrices in the statistical register (the register population, type of objects and variables).

Other documents include information from administrative systems, administrative questionnaires with instructions, minutes and notes from meetings with record providers and SQL scripts with comments describing record processing.

Although the authors mention SQL, national statistical offices can work with other computer languages for storage, processing and analysis of data within a system of economic administrative records (see diagram V.2).

Diagram V.2

Different types of metadata and tools in the documentation of a register



Source: A. Wallgren and B. Wallgren, *Estadísticas basadas en registros: aprovechamiento estadístico de datos administrativos*, Aguascalientes, National Institute of Statistics and Geography (INEGI), 2012.

Therefore, the system must allow management of these documents, which facilitate understanding of the functionalities and records contained for statistical use.

3. Construction or improvement of dissemination components

This subprocess describes the activities to build new and improved components and services necessary for the dissemination of the statistical outputs such as those designed in subprocess 2.1 (design outputs).

Therefore, all elements of dissemination involving statistical operations related to economic administrative records should be considered.

This subprocess includes the development of all the tools that will support the statistical outputs defined on the basis of the system of economic administrative records and of the statistical business register.

Among the dissemination components to be developed are databases, analysis reports, documents and metadata, as well as geospatial tools. Dissemination tools should be in line with those available to each institution for its different statistical operations, whether based on administrative records, censuses or surveys.

4. Organization of workflows

This subprocess “configures the workflow, systems and transformations used within the statistical business processes, from data collection through to dissemination” (UNECE, 2016).

Proper management of economic administrative records requires national statistical offices to have defined within their organizational structure the roles and responsibilities of all the units involved in the management of economic administrative records and the entities that provide them.

It is important to highlight that the workflow should include focused profiles for a statistically relevant use of all existing data in the system of economic administrative records. For that purpose, professionals with expertise in statistics, scientific methods and data analysis should be considered. Furthermore, taking into account the importance of working with large volumes of data, it is important that constant support for the statistical office’s computer equipment is included in the workflow.

5. Tests of the production system

This subprocess “is concerned with the testing of assembled and configured services and related workflows” (UNECE, 2016).

In addition, this subprocess defines the scope of the test to be performed, determines costs of the tests and establishes the expected outputs, defines the instruments that should be tested (instructions, manuals, control formats or computer applications), specifies the conditions under which the tests will be carried out, as well as the formats used, their time frames and those responsible for them, and prepares the report documenting the results, conclusions and recommendations of the tests performed (DANE, 2020, p. 42). Taking into account that the work with administrative records has an important IT component, this subprocess should include all tests of the components and services involved, such as the interaction of file transfer systems with information providers.

6. Pilot test of the statistical process

This sub-process describes “the activities to manage a field test or pilot of the statistical business process. Typically, it includes a small-scale data collection, to test collection instruments, followed by processing and analysis of the collected data, to ensure the statistical business process performs as expected” (UNECE, 2016).

Although this description adopts a statistical operation logic that uses questionnaires for data capture, within the framework of the work to be carried out with economic statistical records, all necessary tests must be carried out to establish a continuous process starting from data collection whenever work involving administrative records is necessary.

7 Finalization of the production system

This subprocess includes “the activities to put the assembled and configured processes and services, including modified and newly-created services into production ready for use by business areas” (UNECE, 2016).

B. System modules

It should be considered that the computer system or systems supporting the system of economic administrative records and the statistical business register should contain different elements that enable proper management, taking into account the challenges of handling large volumes of information.

1. Outreach to data providers

This is defined as the set of activities to promote and foster an effective relationship between providers of public and private administrative records and national statistical offices, the creation of agreements, the execution of exchanges of administrative records and their transformation into a statistical register. This also applies to economic administrative records.

Considering that national statistical offices face the challenges of managing volumes of data from different providers, mechanisms must be put in place to better manage this work. Thus, this module should make it possible to establish basic information for data provider entities, such as name of the provider, details of the data provision agreement, counterparts within the institution, form of data transfer, and so on.

Ideally, each data provider should be identified with a unique identifier, as should each administrative record it provides. These identifiers will facilitate linkage to the documentation supporting this module.

In addition, this module can be used to extract the inventory of data providers, which consists of a characterization of the data sources available in the system of economic administrative records and makes it easier for the teams producing statistics to know all the background information they have available for the statistical use of the economic administrative records.

2. Collection of administrative records

As is mentioned in Chapter V regarding the subprocess of building collection instruments, the technological means of transmission and interoperability of data obtained from administrative records, files or statistical or non-statistical databases must be developed, as must the techniques and tools to verify that the information sent or received satisfies the quality conditions established for data exchange.

It is very likely that the technological means can be furnished with different support tools, including magnetic media, file transfer protocol (FTP), web services, plain file, Excel spreadsheets databases and so on. However, it is important for the management of the system of economic administrative records to be able to manage within one system all the background information related to file transfers, including administrative record captured, file size, capture medium, identify if any permission is required or if there are any restrictions on access to the administrative record, etc.

3. Transformation of administrative records

Once the economic administrative record has been captured, national statistical offices must carry out a set of activities for its transformation into a statistical register. Thus, at the computer system level, there must be established procedures for developing data consistency, standardization and normalisation activities, and the development of unique keys.

As mentioned throughout the chapter, it is possible that some of the activities to be developed may not be in a single computer system but may also use other services available at national statistical offices. For example, for the standardization of addresses, a service established in the statistical office that meets these purposes is used. Also, for anomaly detection, routines could be developed in statistical software, such as SAS or R.

Therefore, it is important to understand that for the development of the computer system it is possible to establish interactions between specific modules of the system with statistical software applications.

The development of unique keys is fundamental for the operation of the system of economic administrative records. It is within this module that each statistical unit within the system of economic administrative records is assigned a unique identification variable, which allows the unit to be univocally linked to the different data sources where it is found.

It may be the case, as Wallgren and Wallgren (2012) propose, that, in order to guarantee the privacy of each statistical unit within the system, dummy unique keys will be generated to safeguard sensitive unit information. In this way, linkage within the system of economic administrative records, that is both within the statistical business register and the economic administrative record, is achieved by means of these dummy unique keys.

4. Data documentation

As mentioned in point 2 of section A of this chapter (Construction or enhancement of components), establishing the documentation of the system of economic administrative records for statistical use is essential.

Therefore, a repository should be established within the system to store all documents related to the statistical business register and economic administrative records and to the execution of the processes associated with their transformation into statistical records. This module should support the uploading of technical documents, agreements, quality diagnostics and so on.

5. Administration

The computer system should allow those responsible for the management of the system of economic administrative records to have maintenance and control elements to be able to manage users, parameters (such as classifiers and codifiers), etc.

Chapter VI

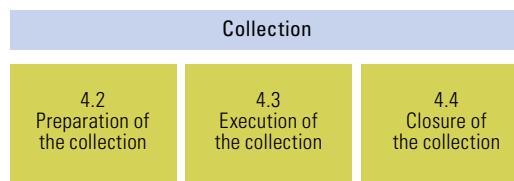
Collection

Introduction

According to the structure of the GSBPM statistical process model, this phase comprises the following subprocesses that frame the activities that must be carried out in that process to produce official statistics (see diagram VI.1).

Diagram VI.1

Subprocesses included in the collection phase



Source: Economic Commission for Europe (UNECE), "Generic Statistical Business Process Model GSBPM (Version 1.0 in Spanish, September 2016)", Geneva, 2016.

Note: Subprocess 4.1 "Formation of the sampling framework and selection of the sample", which does not apply to administrative records, is omitted, since it is oriented towards considerations covered by other traditional methods for compiling statistics, such as sample surveys.

The purpose of this phase is to capture the data to be used statistically. The activities that were defined, designed, built and tested during the previous phases are carried out.

In the collection phase, data can be obtained from various sources, such as traditional ones from statistical operations. Also collected are data generated by information-producing units based on administrative records, such as geographic information, images and so on derived from databases or microdata.

Given the reality of each country, and of each unit that keeps administrative records, there are several scenarios for requesting data and their statistical use; However, it is important to distinguish two situations or levels that occur in this process, namely the collection of data by the units holding administrative records, and access by the national statistical offices to the data or information for statistical use.

Considering the issues discussed in the previous phases, the information collection process should be underpinned by a methodology established for the purpose that covers the processes, the availability of appropriate technology, materials and other relevant elements for the preparation and execution of the collection.

It is essential to establish guidelines for the process of collecting the necessary data for the production of economic statistical information. Contacts should be arranged with the agencies producing administrative records and the national statistical offices for the timely delivery of data in good order, as well as basic validation of the structure and integrity of the data received, so as to comply with the required technical specifications and confidentiality protocols.

Many government agencies have administrative records generated on the basis of methodologies and information needs linked to the scope of operation of national statistical offices. The recommended mechanisms for the exchange or collection of information are described on the basis of that premise.

A. Regulatory or legal considerations

Technological advances and their incorporation into the different administrative processes foster the massive generation of information from administrative records. Despite not having been created for statistical purposes, they can be used to that end. In fact, some countries have adjusted and utilized their respective laws in order to make regular use of them and have them as a strategic source in the production of their official statistics. Based on these experiences and depending on the needs detected in terms of the definition of technical teams, the countries of the region should consider strengthening their regulatory frameworks or, failing that, making gradual or specific adjustments to them. Those frameworks should be flexible and guarantee access by national statistical offices to the administrative records of the public, private or mixed register-generating entities that are part of the NSS. Those offices, in turn should have access to useful feedback that contributes to continuous improvement in technical and even conceptual terms, in order to strengthen the generation of useful statistical information for the country's public policy decision-makers.

As for adapting individual frameworks, should a country consider it necessary, the region has an updated, harmonized tool based on the countries' needs and requirements. We are referring to the Generic Law on Official Statistics for Latin America, which provides a reference model for Latin American countries that need to develop or reformulate the legal basis governing the operation of the NSS and the production of official statistics.

Article 17 of the Generic Law deals with access to administrative records and emphasizes the obligation of the entities holding administrative records. Article 17 provides: "All administrative data providers shall provide the Producers of Official Statistics, free of charge, with the data or records in their possession at the level of detail necessary for the production of official statistics and, if possible, with the metadata that enable assessment of data quality. Confidentiality or secrecy provisions in other legislation cannot be invoked unless the legislation explicitly excludes the use of data for statistical purposes" (ECLAC, 2020, p. 22).

In countries where this process is not contemplated in the law, or is so tacitly or in an unregulated manner, legal mechanisms that can be effectively applied for accessing information include inter-institutional agreements or memorandums of understanding, which should consider the interests of the parties and seek to ensure that all obtain benefits.

1. Inter-institutional agreements or memorandums of understanding

In designing collection operations, it is necessary to consider the initial agreements between the unit in possession of administrative records and the national statistical office, especially the points relating to the responsibilities assumed in terms of participation with human, material and financial resources. Consideration should also be given to procedures for the registration and transfer of data to the national statistical office, participation in the fulfilment of functions related to the quality control of the data collected and the exchange of technical assistance for the generation of statistics and so on. As far as possible, these agreements should be established for the long term, regardless of administrative or staff changes.

Agreements and support arrangements may be made between areas of the same agency or organization, or between areas of different and independent institutions.

In any of these situations, agreements are reached and signed at the management level, where interests are reconciled and conditions are established for subsequent decision making during the development of the statistical project and, in particular, for the collection operation, taking into account the following:

- Agree on a functional structure and programme of work.
- Determine responsibilities regarding the participation of human, material and financial resources.
- Identify the technology and procedures for data transfer from the units in possession of administrative records to the national statistical office.
- Determine the participation in the fulfilment of functions relating to the quality control of the data collected for the generation of statistics, among others.
- Establish commitments and control arrangements for the collection and timely delivery in good order of information batches by units in possession of administrative records.

The regulations, policies and standards that provide the legal basis for the records that public agencies and organizations hold as part of their functions are the ones that keep the record-keeping process going and, therefore, make it possible to implement statistical projects of this type.

This legal framework basically consists of:

- the countries' constitutions;
- the organic law creating the institution;
- the national development plan, and
- the law on transparency.

It is also complemented by other types of laws and internal regulations of the agencies, which govern the operations of the units involved.

In general, strategies for more-efficient data collection depend on:

- Recognition of a governing body to regulate, coordinate and monitor compliance with national statistical law or regulations to ensure that information of public interest is made available to statisticians.
- Initiatives promoted by the statistical unit, where appropriate, aimed at modifying relevant national laws or legal or administrative regulations, in order to establish a solid basis for the improvement of the quality and timely production of statistics.
- Compliance with these laws implies, on the one hand, the obligation that commits record-keeping units to delivering their information in the agreed time and manner; as regards the statistical unit, its obligation is to safeguard the principle of confidentiality of information.
- The support of certain government agencies or secretariats linked to the record-keeping units that condition the authorization of procedures for their activities in exchange for the delivery of their information.

Legislative changes affect the statistical process, particularly collection, so it is necessary to clearly define the areas of competence —both state and municipal— and to identify the legal instances that can be used to raise awareness among stakeholders about the importance of the statistics in question and the delivery of information with the agreed quality and confidentiality requirements.

B. Methodological considerations

In the exchange format, the national statistical office must receive the administrative records systematically and, in that context, the following must be complied with:

- Based on an agreed schedule, the information is exchanged. The unit in possession of the administrative records transfers the necessary data to the national statistical office, which, in turn, provides feedback.
- Likewise, as long as the national statistical office is in a position to do so, it should provide support, through software development or technical advice, to facilitate the work of units in possession of administrative records and improve the quality of the information. Therefore, it is essential to promote and encourage the advance and development of the IT infrastructure of the latter.
- In the event that the prior diagnostic assessment finds that the records do not meet the necessary requirements, the technical, conceptual and technological infrastructure framework delimitations will be defined by mutual agreement. This entails making the necessary adjustments to the collection instrument used by the unit in possession of the administrative records, as well as their respective conceptual and operational backups.

1. Structure and necessary information

The inter-institutional agreement should determine the structure and the minimum or basic information that needs to be available for statistical purposes. This should be based on the findings of the diagnostic assessment of the required information with respect to the information available from the various sources.

The structure will be defined based on the evaluation of the use of the administrative register. It includes aspects relating to metadata and, crucially, technological support, since a match between formats and variables must be achieved. This structure also prioritizes the fundamental information needs of the statistical business register. If they are not fully met, alternative or complementary sources are identified to complete or supplement needs.

Once the appropriate structure for all parties has been defined, the unit in possession of the administrative records that will supply the information to the national statistical office will do so with the corresponding metadata and in the agreed format.

Finally, the defined structure and the information to be shared must be expressly set down in the document containing the inter-institutional agreement, which will also specify the format and modality of the exchange, considering the laws to which both parties are subject.

2. Periodicity

It is necessary to arrange a work schedule that sets out the periodicity of the following:

- The gathering of the data needed for statistical purposes by the entity in possession of the administrative record, and the reference period for the data.
- The transfer of the data to the national statistical office (collection) and, consequently, of the feedback from that office under the agreed terms.

3. Timeliness

To ensure timeliness, it is essential to comply with the agreements on periodicity and the dates established for the release of data. This should be clearly set down in the inter-institutional agreement or memorandum of understanding. It is also important clearly to establish the technical criteria required, from both parties, with respect to compliance with the pre-availability controls that must include the review and coding of variables. These criteria should be previously agreed upon in the framework agreement between the parties.

4. Medium or channel of exchange

With respect to means or channels of exchange of administrative records, the following considerations should be borne in mind:

- Format, equipment and peripherals: this refers to harmonization of format and mode of delivery, since it must be ensured that the institutions involved can access the format, the source or form of delivery and, above all, the technological capability to store administrative records while safeguarding security and confidentiality. In this context, it will be necessary to define, as a minimum, the following requirements:
 - i) Operating system, including the specified version thereof.
 - ii) Permissions to install or uninstall software.
 - iii) Amount of hard disk space required (for installation).
 - iv) Amount of hard disk space required for sending and receiving information.
 - v) Type of Internet connection and bandwidth required.
- Legal: each country has its own legislation, while internally institutions also have specific confidentiality rules. In this context, strategies to bridge discrepancies (if any) should be considered, and clarifications and reconciliations that support the exchange of information should be explicitly included in the agreement.

5. Confidentiality

This point should address the conditions and levels of clearance for staff at the unit in possession of the administrative records and at the national statistical office. A letter of commitment to confidentiality must be signed, recognizing responsibilities and stating the commitment to abide by the law in terms of respecting the confidentiality of the information accessed. This letter should be signed in triplicate with individual copies sent to the staff member, the entity or unit in possession of the administrative records and the national statistical office. The letter will be renewed every calendar year counted from the date of signature.

Once the authorized personnel for the transfer of data have been designated by the unit in possession of the administrative records, they must be registered as such with the national statistical office and the above procedure will be included in the agreement or memorandum of understanding. Once this procedure is completed, they will be issued a certificate or card signed by the highest authorities of the institutions that are parties to the agreement.

These authorized personnel will receive a notification by e-mail, informing them of their designation as the persons responsible for sending and receiving the information. The notification will contain the following:

- a link to download the computer system installer;
- the activation key (required during system installation), and
- the user's guide (this document) to guide them through the download, installation, registration and operation of the system.

C. Technological mechanism

In order for the units in possession of administrative records to provide the national statistical office with the information within their scope of competence, the national statistical office will have the necessary and sufficient technological tools to enable this process to be carried out.

The national statistical office responsible for developing the IT tool for the exchange, to which end, the metadata of the unit in possession of the administrative records must be taken into account in order to ensure compatibility and practicality at the time of implementation. To the extent possible, this tool should be part of a platform that covers the primary collection of information by the unit in possession of the administrative records and the transmission of the data compiled by that unit to the national statistical office, as well as the protocols for its storage, access and security.

The system must guarantee the security of the information exchanged and, consequently, must be subjected to encryption and compression processes using a secure password and channel encryption. This prevents loss or unauthorized access, thus ensuring the confidentiality of the information.

The implementation should include a training plan for the staff involved in this task, taking into account the roles and profiles of the personnel, the corresponding topics, the hourly intensity and method (virtual, face-to-face, semi-face-to-face, or blended learning).

D. Training

NSOs should provide ongoing training for the staff responsible in entities in possession of administrative records, so as to improve the information processing and technical procedures.

A training strategy should be adopted and a schedule for its execution programmed in a coordinated manner at different levels of the organizational structure, comprising technical personnel assigned in accordance with the functions to be performed.

Chapter VII

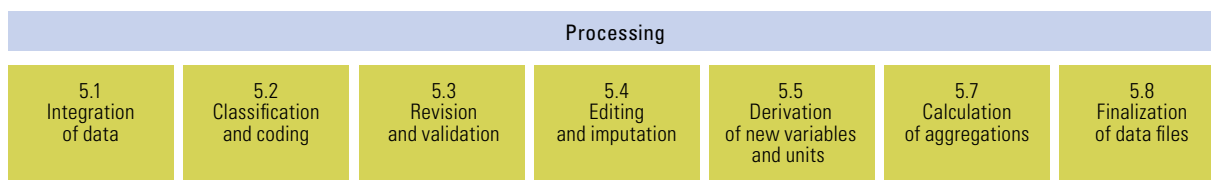
Processing

Introduction

This is the phase that follows the collection of data from economic administrative records. Its aim is to prepare the data for analysis and to enable their conversion into usable information for the production of statistics. It comprises a number of subprocesses through which input data are verified, cleaned and transformed, classifications and codes are standardized, and data are reviewed, validated, edited imputed and, finally, integrated. It will also be possible to create new variables, units, as well as weights and aggregates, and then organize those processed data in a database for analysis and dissemination as statistical outputs (see diagram VII.1).

Diagram VII.1

Subprocesses included in the processing phase



Source: Economic Commission for Europe (UNECE), "Generic Statistical Business Process Model GSBPM (Version 1.0 in Spanish, September 2016)", Geneva, 2016.

Note: Subprocess 5.6 "Calculation of weights", which does not apply to administrative records, is omitted, since it is oriented towards considerations covered by other traditional methods for compiling statistics, such as sample surveys.

It is recommended that, prior to the collection and processing of data from the economic administrative register identified with the inventory, a quality assessment be carried out in order to determine its potential statistical use within the framework of the target information needs.

As discussed in the preceding chapters, economic administrative records come from various public and private sources. They are generated with objectives, variables, codes and processes for the control of their activities, which differ from the statistical purposes pursued, and their status and quality level should be known.

The statistical quality of administrative records should be assessed bearing in mind the dimensions or attributes that international organizations have acknowledged as necessary to fulfil in statistical activities. The Organisation for Economic Co-operation and Development (OECD) prioritizes eight dimensions or attributes in its document *Quality Framework and Guidelines for OECD Statistical Activities*, as shown in box VII.1.

Box VII.1**Key dimensions in the quality of statistics**

According to the OECD's Quality Framework and Guidelines for OECD Statistical Activities (2012), the main dimensions for ensuring the quality of statistics are:

1. Relevance: statistics meet users' needs.
2. Accuracy: statistics accurately and reliably represent reality.
3. Credibility: users have confidence in the statistical outputs.
4. Timeliness and punctuality: statistics are published in a timely manner.
5. Accessibility: data are stored in secure locations known to the users who request them.
6. Interpretability: data have all the complementary information to be understood by users.
7. Coherence and comparability: statistics are consistent internally, in time and in space, and it is possible to combine and make joint use of related data from different sources.
8. Cost-efficiency: although it is not a quality dimension per se, it is a factor that should be considered in quality analyses.

Source: A. Rojo, "Del dato administrativo de firmas al dato estadístico: desafíos a vencer en América Latina", CAF Working Paper, No. 2018/17, Caracas, Development Bank of Latin America (CAF), 2018; Organisation for Economic Co-operation and Development (OECD), Quality Framework and Guidelines for OECD Statistical Activities (STD/QFS(2011)1), Paris, 2012.

It is proposed to assess the statistical quality of administrative records in two parts:³

- The first part, for the purpose of conceptual analysis or analysis of the register management process, in order to observe compliance with the OECD dimensions or attributes concerning relevance, accessibility, credibility, timeliness, coherence and comparability. The use of an ad hoc questionnaire may be considered in this part. Documentation for the register should be attached to this questionnaire. The person in charge of the register may also be interviewed.

The principal elements to be reviewed in this analysis include definitions, main variables, register coverage, collection methods, types of classification and coding, use of technology, validation criteria and users.

- The second part is to analyse the database of the administrative register. The use of statistical tools is recommended due to the magnitude of the data. Its purpose is to review compliance with the dimensions of accuracy and consistency.

The analysis begins with a review of internal consistency in order to look for errors, duplicates, coding changes and missing information, among other things. This is essential when administrative records have little or no documentation. Next, external consistency is reviewed, i.e., comparisons are made with external data on structure, quantities and compositions with other available statistical sources.

The results of the quality assessment allow the development of an improvement action plan, which should be shared with those responsible for the administrative register in order to correct errors or inconsistencies in the short term. In addition, improvements to strengthen the administrative register will be coordinated and carried out over a longer term.

Currently, there are questionnaires or instruments designed for the evaluation of the statistical quality of administrative registers, including the following:

- (i) The Quality Evaluation Tool for Administrative Registries (HECRA), an instrument developed with the support of the World Bank and used by Mexico's INEGI, which has made improvements to it.

³ For further reading on statistical quality assessment of administrative records, see Rojo (2018).

HECRA offers users the possibility of elements of self-diagnosis to establish their areas of timeliness in order to improve the quality of administrative registers that can be used to generate statistical and geographic information.⁴

- (ii) The Questionnaire to Evaluate the Quality of Administrative Registers (Cuestionario para Evaluar la Calidad de los Registros Administrativos – CECRA) is an instrument developed as part of the project “Population and real estate statistics based on the use of official administrative records in the Andean Community” (Estadísticas de población e inmuebles a partir del uso de registros administrativos oficiales en la Comunidad Andina) with the support of the Inter-American Development Bank (IDB), within the framework of its regional public goods (RPG) programme.

CECRA is intended to serve as an instrument for collecting information from administrative sources in order to carry out a diagnostic assessment of those administrative registers at the national and subregional levels. The stages in the process are design, production, analysis, delivery and dissemination.

The following is a description of the subprocesses that make up the processing phase, based on the GSBPM.

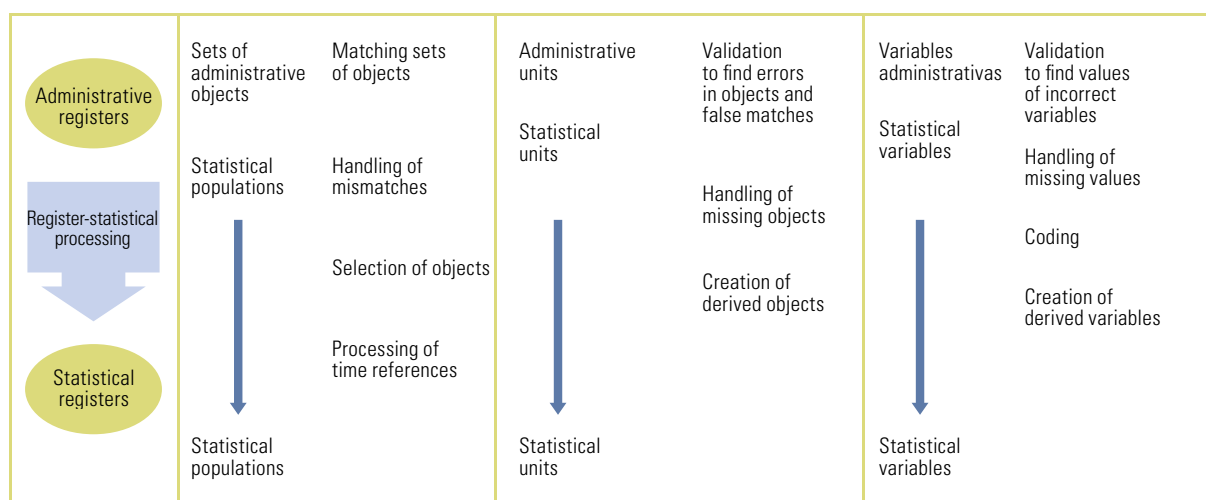
A. Integration of data

Prior to the integration of administrative records, the transformation or conversion from administrative records to statistical records must be performed.

This involves preparing the administrative register for integration with other records. Diagram VII.2 describes the process of transforming administrative records, which may be one or more, into a statistical register.

Diagram VII.2

Statistical records from administrative records



Source: A. Wallgren and B. Wallgren, *Estadísticas basadas en registros: aprovechamiento estadístico de datos administrativos*, Aguascalientes, National Institute of Statistics and Geography (INEGI), 2012.

In the transformation of administrative records, variables, classifiers and codes are standardized and normalized, after which they are reviewed and validated. Unique keys are also developed, classifiers and codes are reviewed, and input data are cleaned and transformed to be used for statistical purposes. In other words, there must be pre-established rules on consistency (see section B.3 of chapter V, on the transformation of administrative records).

⁴ See National Institute of Statistics and Geography (INEGI), “Herramienta para la evaluación de la calidad de los registros administrativos (HECRA)” [online] <https://www.inegi.org.mx/mejorespracticas/hecra/>.

Once the database is clean, it can be integrated with other administrative or statistical records.

Integration is the subprocess that begins once the data cleaning stage has been completed. It requires that economic administrative records have a unique identification key to facilitate their linkage.

Almost all countries that have tax administration systems identify companies with a unique tax code, through which the tax authority monitors the performance of those entities. This unique code assigned to companies is extremely important, as it will allow the integration of administrative records in order to have more and better economic information available to meet the demands of users, once the information has been processed and analysed.

In Peru, this code is the tax identification number (RUC), generated by the National Tax and Customs Administration (SUNAT), which is already standardized. This source of information is currently used in the INEI Central Directory of Companies and Establishments (Directorio Central de Empresas y Establecimientos).

For the integration of one or more administrative records, it is recommended that a good evaluation of their statistical quality has been carried out. It should be borne in mind that joining records will yield a combination of results from the subprocesses in the collection phase. For that reason, as the input data originates from different registers, they may have different collection methods.

The integration of administrative records may produce a system of economic administrative records or incorporate them into the statistical register. As the case may be, it will also cause new variables and units to be created.

It is advisable to anonymize the records to guarantee the confidentiality of the information before updating the statistical production database. This can be done after integration.

Segui (2016) says that deterministic and probabilistic methods are used to integrate records (unify or match them).

Unification of records has three stages: pre-unification, unification and post-unification. Pre-unification involves the normalization of text variables and the standardization of other variables (carried out in phase 5, Construction). Unification of records is done by applying standardized statistical algorithms widely used in various disciplines. Lastly, comes post-unification or the results evaluation process (next subprocess).

The author mentions the following steps to take after the integration of records:

- (i) evaluate the results of the integration;
- (ii) derive new variables and units;
- (iii) data protection;
- (iv) create or update the statistical register in the data production database, and
- (v) create or update the base register.

B. Classification and coding

This subprocess serves to review the classification and coding system used for administrative records in the economic statistical register in order to determine whether they meet the standards for statistical production or are ad hoc codes or tables developed for their own use. This review will assess whether the OECD dimensions or attributes of accuracy, credibility, timeliness and punctuality are met.

The review of the classifications and coding used in the administrative records that feed the statistical register is of utmost importance because it will make it possible to identify whether records are related to national statistical offices' standard classifications, enabling comparison with the information that those institutions produce. If that is not the case, tables must be prepared to convert administrative records. In statistical production these tables are known as conversion tables, and they facilitate their harmonization for conversion for statistical purposes.

A unique classification and coding system allows information to be stored in an orderly fashion, enabling data to be obtained from part or all of the record. For example, in the case of the Central Directory of Companies and Establishments, information can be obtained for a given region of the country or for a specific economic activity in the International Standard Industrial Classification of All Economic Activities (ISIC) rev. 4. Information can also be retrieved on gender, i.e., the proportion of women who are business owners or in management positions.

A statistical classification can have a flat or linear structure, for example, when information is required about the type of business organization, whether it is a natural person or a legal entity.

A statistical classification can also have a hierarchical structure, i.e., with several levels. For example, the International Standard Industrial Classification of All Economic Activities (ISIC) (revision 4) provides information at the section, division, group and class levels. This group also includes territorial distribution classifiers at the country level, which provide economic information by geographic area. Peru's system of political-territorial division is structured into departments, provinces and districts.

When literal classifications are transformed into numeric ones, this is done by means of a coding system, in order to facilitate the tabulation that will allow the presentation of statistical data in the form of tables or charts. In administrative records it is common to find variables classified into categories, which are usually coded with numeric values. In some records sex is coded as 1 if male, and 2 if female, but in other administrative records the letters of the alphabet are used (M and F). Hence the need to check the coding used in this process to ensure that it is correct. In addition, if several administrative records are involved, they should be checked to ensure that they are harmonized with those used by national statistical offices. Coding faults show through in the results.

The most commonly used forms of coding are as follows:

- Manual coding: generally used in administrative records that are not automated and assigned by staff in charge of the register.
- Automatic coding: assigned by means of a previously set up coding programme using electronic catalogues.
- Semi-automatic coding: assigned with both manual and electronic catalogues.

C. Revision and validation

This subprocess reviews the input data to detect potential problems in the economic statistical register put together with administrative records.

Errors, inconsistencies and discrepancies that may exist are identified. The latter can be outliers, responses with no data, missing data and coding errors. This review is carried out on data entered from administrative records.

This subprocess is also known as "input data validation" and is initiated when data enters the system. Generally, this validation is performed iteratively, i.e., continuously and as many times as necessary.

It is recommended that, prior to review and validation, a manual be prepared with basic consistencies describing the rules that the data must comply with in relation to information flows, data ranges, logical validation rules, coverage rules, coding verification controls, generation of tables for economic activity and for geographic distribution, verification control for extreme values and identification of missing values, among other things. Prior analysis of the statistical quality diagnostic will be very useful for that purpose.

Data review and validation can be done manually or automatically. It should be done before and after the integration of administrative records and on any data source feeding the statistical register.

When it is automated, reports will be produced with the records that were not validated, which must be reviewed by the national statistical offices responsible by means of an exploratory analysis to identify the specific treatment that the information requires.

In Peru, the review and validation of data from SUNAT administrative records is automated. There are basic consistency manuals whose rules have been programmed with the support of IT personnel. The inconsistency reports are then reviewed by a team of people trained for that task.

D. Editing and imputation

The purpose of this subprocess is to improve the accuracy and credibility of the data in the economic register. It consists of the review of input data to detect erroneous, missing, incomplete, inconsistent or unreliable data that need to be corrected; new values can also be inserted.

It should be noted that the statistical register has been created with data compiled from several administrative registers, which have a common identification key and whose data conform to the needs of the statistical topic being addressed. These data have already undergone editing processes carried out by the generating entity, which prioritized this information for administrative use.

For this reason, it is recommended that, when editing the statistical register, national statistical offices give attention to reviewing the data that are most useful for statistical purposes.

The aim of editing is to ensure that the statistical record is clean and consistent, so that there are no disparities between individual records. If inconsistencies arise it is necessary to review them because they could be due to system errors in the administrative records.

As national statistical offices usually perform editing in their statistical operations, including for the economic statistical register, each record will first be screened by internal logic checks to detect any errors or duplicates. Those checks may refer to information flow rules, logical rules for data validation and rules for coverage control or outlier verification, among others.

The results of applying these consistency rules will guide the measures needed to correct any errors or duplicates that may be found. If systematic errors are detected in the economic administrative records from which the data were obtained, it is essential that they be corrected, and it will be necessary to contact those responsible for such records in the generating entities, making it important to have good communication with them.

It is also common to find missing data in economic statistical records. This may be due to a number of situations, such as: (i) when there are inconsistent data coming from administrative records, in which case it is necessary to review the documentation and the analysis of the management process performed on the record to ensure that the data really does not have a value and an imputation is necessary; (ii) when the data does not have a value, so it will also be necessary to carry out an exploratory analysis to determine how to correct it; and (iii) when, upon unifying several administrative records, it is found that there is no linkage between them, which may be due to the fact that some economic units are missing. In such cases, it is recommended that the exploratory analysis be carried out at this level.

Depending on each situation, if it is found that there are indeed missing data, which could not be corrected through those responsible for the administrative records in the generating entities or by means of editing processes to correct inconsistencies, imputations may be made to the missing data in order to complete the statistical record.

Some possible imputation methods are mentioned below (Useche and Mesa, 2006):

- Deterministic techniques, when repeating the imputation on several units under the same conditions will produce the same responses (mean, mode, regression and time model imputation methods).
- Fundamental techniques on external information, when they are based on variables related to a survey belonging to other databases or previous rules (deductive methods, look-up tables).

E. Derivation of new variables and units

The purpose of this subprocess is to create new variables that are not found in any of the administrative or statistical records that were integrated and were not programmed in the data collection but are necessary to meet the demands of users.

These variables and units are created from other variables in the administrative records or statistical registers that have been linked. They are intended to improve accuracy and credibility, as well as consistency and comparability, which are dimensions or attributes prioritized by the OECD for quality.

The derivation of new variables and units can be repetitive during the process because they will be based on other derived variables, so it is recommended to pay very close attention to ensure that it is done in a correct order.

Since statistics based on administrative records are obtained from the design of the data collection by the offices that generate them, it is not possible for national statistical offices to define the variables needed for statistical purposes. Hence the need to resort to the creation of aggregate variables based on variables available in the administrative records.

Seguí (2016) mentions that derived or aggregate variables are created in six different ways:

- (i) By means of arithmetic calculations or logical procedures.
- (ii) By grouping values.
- (iii) Through coding.
- (iv) By association with other variables.
- (v) By aggregation of other variables.
- (vi) By means of statistical models.

He also points out that derived variables can be generated by combining these methods or even by applying more complex calculations with procedures based on certain rules, and he recommends that they be documented according to the standard defined for such purposes. It is important to describe the formula or calculation rule used when creating the variable.

F. Finalization of data files

This subprocess includes the closing activities performed in each of the subprocesses of this phase. The following should be considered:

- Microdata files and administrative record data collection forms must be submitted for legal safekeeping.
- Microdata files and databases for the results presentation phase will be submitted for data analysis.

- Centralize the documentation managed in this phase and also the processing report, which will be useful for data analysis.
- The documentary files should contain the programmes that have served as sources, as well as the catalogues for the classifiers, for the identification and integration keys, the imputation methods, the creation of the derived variables that have been generated, and everything related to the loading of the data and metadata collected in an electronic environment.
- Backups of computer systems and information on computer media for archiving.
- All this information should be put into a data folder (usually macro data) that will be used as the basis for the analysis phase or, in the case of a project, to provide continuity.

Chapter VIII

Analysis

Introduction

The object of this phase is to ensure that the information produced is fit for purpose, i.e., ready for use and dissemination.

The microdata should be examined in its basic dimensions, both in the uptake into the administrative register and in the processing of the data to obtain reliable results in accordance with the vision of the statistical business register and the system of economic administrative records.

A. Analysis of the business register

The main objective of this chapter is to examine the microdata in their basic dimensions, at both the data capture and the data processing phases, in order to obtain reliable statistical results in accordance with the vision of the statistical business register and the system of economic administrative records (see chapter II). The intention is also to establish the basic parameters of the structure of the documents to be disclosed.

In this phase, the processes related to the information collection instrument used by the institution responsible for the business register are analysed to obtain a holistic view in order to evaluate the quality of the register, which may have a problem with under-registration, for example.

The aim is to validate the use of classifiers and quality standards, such as ISIC and ISO 27000 certification, to verify the security of the information in order to generate confidence in those responsible for the source and in the owners of the companies that are required to provide the information according to the administrative record. The implementation and subsequent certification of these quality standards for the production processes should help to preserve the integrity and confidentiality of company data and all business processes.

It is imperative to have available the documentation of the statistical operation of the administrative register according to the inventory carried out by the governing institution of the NSS, in which the legal framework that supports the source, the standards or classifier used, the periodicity or cut-off of the database, among other things, are displayed. The type of company and the administrative records associated with the company to be registered must also be specified, namely:

- Limited liability companies (LLC), corporations and simplified joint-stock companies (SAS)
This involves an analysis of the following administrative records:
 - Affidavit Form for the Registration and Updating of Company Data (Formulario de Declaración Jurada para el Registro y Actualización de Datos de Sociedades) (RC-02).
 - Certificate of commercial registration.
 - Receipt of payment for the incorporation of a company.
 - Certificate of trade name issued by the National Office of Industrial Property (Oficina Nacional de la Propiedad Industrial – ONAPI).

- Limited liability sole proprietorships (EIRL)
 - Affidavit Form for the Registration and Updating of Company Data (RC-02).
 - Articles of incorporation registered at the chamber of commerce.
 - Certification of bank deposit.
 - Certificate of commercial registration.
 - Certificate of trade name issued by ONAPI.
- Joint ventures
 - Affidavit Form for the Registration and Updating of Company Data (RC-02).
 - Contract signed and stamped by the Civil Register or City Hall of the National District.
 - Receipt of payment of authorized capital stock.
- General partnerships, limited partnerships and limited joint-stock companies
 - Affidavit Form for the Registration and Updating of Company Data (RC-02).
 - Certificate of commercial registration.
 - Certificate of trade name issued by ONAPI.
 - Receipt of payment of authorized capital stock.
- Sole proprietorships
 - Signed letter of application, requesting the registration of the sole proprietorship.
- Foreign corporations
 - Affidavit Form for the Registration and Updating of Company Data (RC-02).
 - Certificate of commercial registration.
 - Articles of incorporation registered at the chamber of commerce.
- Consortiums
 - Affidavit Form for the Registration and Updating of Company Data (RC-02).
 - Consortium contract registered by the Civil Register and Mortgage Register (Conservaduría de Hipotecas).
 - Certificate of trade name issued by ONAPI.
 - Receipt of payment of authorized capital stock.
- Entities under incentive laws
 - Affidavit Form for the Registration and Updating of Company Data (RC-02).
 - Resolution or certification from the governing body.
 - Certificate of commercial registration.

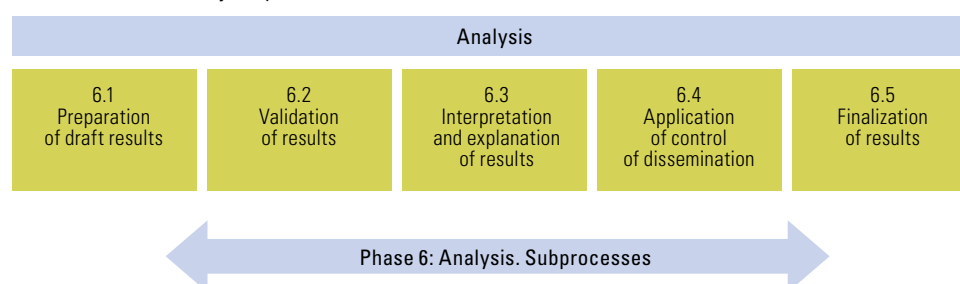
Coherence between each administrative register will contribute to the construction of the statistical business register and the analysis to be carried out in each case. For these purposes, phase 6 of the GSBPM has been divided into five subprocesses, which are described in the following section.

B. Subprocesses of the analysis phase

The analysis phase is divided into five interrelated subprocesses, organized sequentially from one to five. However, they can be carried out concomitantly and made to interact with each other, considering that each generates an output that is independent from the others, as shown in diagram VIII.1.

Diagram VIII.1

Subprocesses included in the analysis phase



Source: Economic Commission for Europe (UNECE), "Generic Statistical Business Process Model GSBPM (Version 1.0 in Spanish, September 2016)", Geneva, 2016.

1. Preparation of draft results

Based on the demand identified in the national planning instruments, the national statistical plan, the sectoral statistical plan or the country's statistical prioritization, and considering the rules and procedures for the production of economic and environmental statistical products, this subprocess identifies the indicators of interest for decision-making in both the public and private sectors. The variables for the tabulation, table design, graphics and geospatial integration of the data are also identified, as is, if necessary, the level of disaggregation of business data: territorial, business size, number of employees by gender, economic activity, type of company and type of taxpayer, among others.

With aspects such as these, the data captured and processed in phase 5 (process) can be generated and statistical results can be obtained from them. However, these results need to be analysed, and their quality checked.

2. Validation of results

This subprocess analyses the behaviour of the processed data, looking for patterns and implicit trends. It also evaluates whether the quality of the data is sufficient for release.

In that regard, it is important to determine the level of underreporting and the response rate according to forms or administrative records. A comparison is done with the summaries of other institutions that have served as sources for the period under examination, the verification of the result of look-ups and cross-checking of variables from different tables according to the administrative records that make up the company's register (see the type of company in section 8.1).⁵ Lastly, a check is performed of the company's declaration regarding number of employees, type of company and so on.

A fundamental aspect is to verify that the company is located at its stated address. This mainly concerns medium-sized and micro-enterprises, which tend to change addresses if their owners move from one community to another. This has an adverse impact on the statistical record, for example, when cross-referencing these data with environmental variables and the impact on the territory, contribution to GDP by community, employment by territory, mobility policies and educational offerings at the academic degree or technical training level, among others.

⁵ An object in a database that is organized into rows and columns, similar to an Excel spreadsheet.

Geographical location and economic activity should be classified according to official catalogues. In the data capture process, the companies' administrative records must undergo instant validation of geographic location characteristics, by means of a catalogue system that ensures the hierarchy of geographic entities. For these purposes, the national statistical office must have access to the geographic code of companies and establishments from the official version of the country's territorial division.

Separately, classifiers must be verified using the descriptions of economic activity, in order to check that they have the necessary elements for classification in any of the ISIC economic activity classes.

Once the inconsistencies or errors have been detected, it must be determined how they should be corrected based on duly established methodologies, and the documentation of the entire process must be prepared until optimum results are obtained.

3. Interpretation and explanation of results

Clearly, a correct interpretation of the results depends on the analytical capacity of the technical personnel responsible. However, leaving a task of immense importance to personal judgment is a risk, so a documented methodology is needed that includes the most appropriate economic methods to explain the data and that identifies the media and its reference, or the source used for interpretation.

Each result leads to inferences based on reports, tables and graphs, with analysis of the behaviour of variables by type of enterprise, linked to macroeconomic indicators and business development. The growth, strengths and weaknesses of the business sector should be visualized according to the branch of the economy.

By integrating the data with geospatial information, it will be possible to obtain a holistic view of the behaviour of the variables that make up the statistical register.

4. Application of control of dissemination

Two fundamental factors for guaranteeing the continuity of the statistical business register are: (a) confidentiality and (b) anonymity.

When determining whether the document will be released by printed or digital means, as well as the availability of the database and the level of disaggregation (either in microdata or aggregated values), the document to be disclosed must be free of any information that could identify the company. In that regard, the national statistical office must process the data corresponding to each company and remove any such information from the results. It is a right of the informants, unless otherwise specified in writing.

Since the information to be published must be free of company identification, the data must go through a process of anonymization, including territorial disaggregation and information according to the type of company.

Accordingly, the legal framework of the national statistical office and the regulatory provisions on access to information are instruments that can empower that office to guarantee these fundamental considerations.

5. Finalization of results

An annual report is prepared according to the previously defined structure, conceptual and methodological framework, background, definition and objectives, coverage by economic activity, definitions and variables included, territorial division used, validation guide, classification and coding of economic activities and auxiliary activities and so on.

Regarding enterprises, the economic activity, the manufacturing sector, the size of the enterprises and their geographic distribution, among other factors, must be defined.

Chapter IX

Dissemination

Introduction

This phase concerns the release of statistical products to users. It includes all activities associated with the assembly and publication of a range of dynamic and static products via different channels. These activities help users to access and use the data published by the statistical organization.

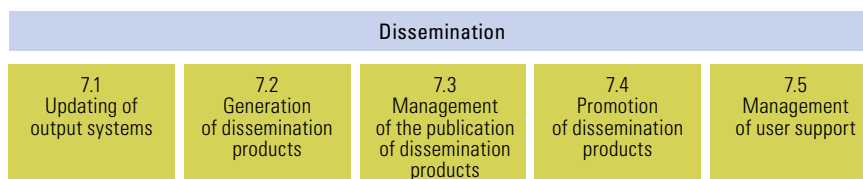
This phase represents the interlocution between the national statistical office and the end users of the statistical products, which can be:

- public agencies at different levels of government;
- companies and entities representing economic sectors;
- researchers, and
- society at large.

This phase consists of five subprocesses that are usually sequential but can also occur in parallel and be iterative. These subprocesses are detailed in diagram IX.1.

Diagram IX.1

Subprocesses included in the dissemination phase



Source: Economic Commission for Europe (UNECE), "Generic Statistical Business Process Model GSBPM (Version 1.0 in Spanish, September 2016)", Geneva, 2016.

A. Updating of output systems

This subprocess has to do with the final production of the results tables for dissemination and updating of metadata. It includes:

- formatting data and metadata ready for placement in output databases;
- load data and metadata into the output databases, and
- ensure that the data are linked to relevant metadata.

It is preferable that the procedures of formatting, uploading and linking metadata be performed upstream; however, this subprocess includes a final review to ensure that all necessary metadata are in place and ready for dissemination.

Statistical results must be accompanied by adequate metadata to enable users to retrieve them, interpret them correctly and form an opinion on their quality.

According to the Data and Metadata Reporting and Presentation Handbook (OECD, 2007, p. 126), it is important to include the following metadata: (i) the name of the source of the administrative data; (ii) a description of the purposes for which the data were originally compiled and collected by the administrative agency; (iii) an analysis of the strengths and weaknesses of the data in terms of their statistical application (coverage, differences in concepts from international recommendations and the use of non-standardized classifications, among others); (iv) a description of processing by the national statistical office following receipt of the administrative data; and (v) descriptions of the reliability of the data, including adherence to international norms and standards and caveats/limitations on the statistical use(s) of the data.

B. Generation of dissemination products

This subprocess yields the relevant products, as designed in subprocess 2.1, to meet user needs (e.g., printed publications, press releases and websites). Products can take many forms, including interactive graphs, tables, public-use microdata sets and downloadable documents. Typical steps in this process include:

- preparing the product components (explanatory texts, tables, graphs, quality statements, etc.);
- assembling components into products, and
- editing the products and reviewing them for compliance with publication standards.

C. Management of the publication of dissemination products

This subprocess ensures that the elements to be disclosed are in place. It includes managing the timing of their release, informing specific groups such as the press or ministers, and coordinating arrangements for any embargoes prior to launch; provision of products to subscribers and management of access to confidential data for authorized groups of users (such as researchers). Occasionally, a national statistical office must recall a product (e.g., if an error is discovered). That possibility is also included in this subprocess.

D. Promotion of dissemination products

While marketing can be considered a cross-cutting process, this subprocess focuses on the active promotion of statistical products produced in a global statistical process so that it can reach the greatest number of information users. It involves the use of relationship management tools to effectively identify potential users of the product, as well as other tools such as websites, wikis and blogs to facilitate the process of communicating statistical information to the user.

This subprocess incorporates the activities that go into the active promotion of statistical products of structural enterprise surveys. It includes the development of virtual banners, the use of infographics in social media posts, the creation of podcasts and short videos for media such as radio and television, in addition to the choice of the main highlights of the launch and other social communication elements of the national statistical office. This process, of course, presupposes direct dialogue with the Institute's journalists, as well as staff from the design and marketing areas.

E. Management of user support

This subprocess ensures that customer queries and requests related to services, such as access to microdata, are logged and that timely responses are provided within a set time frame. Queries and requests should be reviewed periodically to provide input to the cross-cutting quality management process, as they may indicate new needs or changes in client needs.

Chapter X

Evaluation

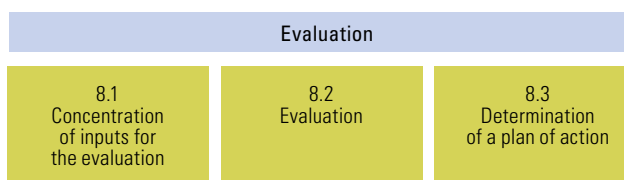
Introduction

The purpose of this phase is to decide whether the next information production cycle should be carried out using the same specifications as to needs, design and construction, or whether any enhancements are required.

The evaluation of the processes and subprocesses carried out in the construction of the statistical product should cover everything from the identification of the statistical need to a satisfactory final product and its dissemination. This will highlight weaknesses and strengths in the process and provide useful information to establish an action plan for continuous improvement of the statistical production process (see diagram X.1).

Diagram X.1

Subprocesses included in the evaluation phase



Source: Economic Commission for Europe (UNECE), "Generic Statistical Business Process Model GSBPM (Version 1.0 in Spanish, September 2016)", Geneva, 2016.

A. Characteristics of the evaluation

The evaluation of the process to obtain a statistical product involves, in the first instance, consolidating the input and output inputs connected with the conceptual, technological and analytical aspects on which the statistical product was built.

Likewise, the evaluation should consider the following:

- the organizational and political conditions of the entity that generated the statistical product, and
- resources and logistics with which to perform the evaluation.

It is important to highlight these considerations and to be aware that the conditions for an evaluation may be lacking.

Accordingly, an evaluation based on the dimensions described below is proposed.

B. Feasibility of evaluation

An important factor in the feasibility of the evaluation is a suitably explained, structured and standardized formulation of the process.

Evaluation feasibility is defined as the possibilities for a public intervention to be evaluated (MIDEPLAN, 2012, p. 22; INAP, 2019).

Therefore, in addition to the conceptual and analytical aspects of the proposal to be evaluated, the following should be taken into account:

- the organizational and political conditions for carrying out the evaluation, and
- the resources, logistics and organization necessary to perform the evaluation.

It is important to highlight these considerations and bring them to the attention of those in charge of this task, since the right conditions for the evaluation may not be in place.

1. Institutional dimension

Line 1 of the evaluation aims to analyse and explain the institutional and inter-institutional conditions that promote interoperability that gives meaning to the use of a product or input, such as data, from the public and private sectors. The processes for a richer, broader and less costly statistical culture for the actors involved, are favoured and enhanced.

2. Specific objectives

- Identify the conditions that favour or hinder the development of institutional and inter-institutional interoperability.
- Inquire into the perceptions that different institutional actors have about the work methodology.
- Analyse whether the process has encouraged the enhancement of pre-existing capabilities by creating effective learning in the work teams that work together to develop interoperability, and whether it has contributed to a better relationship between the national statistical office and the entities in possession of administrative records.
- Describe the inter-institutional configurations that support and consolidate synergy processes favourable to the development of national statistical offices' and administrative record-keeping entities' capacities in terms of the quality of their administrative records.
- Study the cost or workload implications for reporting entities and national statistical offices.

3. Technical-methodological dimension

Line 2 aims to analyse aspects of the technical-methodological implementation strategy of the source of administrative records as inputs for official statistics relating to a statistical business register and the administrative practices responsible for the basic input for that register.

4. Specific objectives

- Review the subprocesses for analysing coverage of the relevant statistical or geographic information needs.
- Study whether the procedures used are suitable for maintaining comparability and conceptual consistency with results obtained from other sources of information and even by other methods.
- Validate compliance with the international standards and subprocesses implemented for source selection and quality assessment, data standardization, record linkage and transformation of administrative records into statistical ones.

- Verify whether the processes carried out complied with the principles of confidentiality and information security.
- Inquire as to satisfaction of information needs among internal and external users.
- Verify the implementation of an IT solution aligned with established standards and system interoperability.

5. Methodological approach to the evaluation

(a) Baseline

In order to achieve the proposed objectives, a baseline should be established:

- To describe the current operational dynamics of the data capture processes of administrative records and the construction of statistical data from those processes.
- To construct the relevant indicators.
- To identify the sources of information from which data will be obtained.

An example of this is a study of the costs and workloads for reporting entities and for national statistical offices in terms of generating inputs and data, through traditional methods such as censuses or surveys. So, too, are the end products offered to users and the pre-existing information on economic units.

(b) Universe of study

The universe of study comprises state agencies that compile administrative records from which statistical data are or can be constructed, as well as the national statistical office.

The following questions must be asked in constructing the sample:

- (i) Who will be covered by the evaluation?
- (ii) What institutional levels will be considered?
- (iii) Which actors will be selected?
- (iv) What type of users will be included?
- (v) Who will be the users of the evaluation results?

The users identified in paragraph (e) will be the primary source of information. As for secondary sources, the experience of other national statistical offices in this area will be drawn upon to establish the baseline, as will existing information at national statistical offices or internal records.

(c) Data collection instruments

To collect the necessary information, qualitative and quantitative research techniques and instruments that are relevant to the purpose of the evaluation will be used.

Interviews and field notes are proposed. The interviews will be semi-structured, with a topic guide and questionnaires containing questions that can be open-ended or pre-coded (see annex A8).

(d) Interview

It is recommended that interviews be conducted with:

- National statistical office executives and technical personnel involved in the design of inter-institutional agreements and the technical work regarding the statistical registration of businesses, respectively.

- Executives at entities in possession of administrative records that work with administrative records and are part of the agreement.
- External users of the information produced by the national statistical office.
- Internal users of the national statistical office that use the products from the statistical business register as inputs.

(e) Field notes

To supplement the interviews, field notes can be used to record facts observed directly by the interviewer. This is an interpretation that describes the activity and settings, who talks to whom, etc. Potentially useful comments should be included, for example, on the characteristics of the setting in which the data were collected and how their consolidation is addressed with.

(f) Evaluation team

The specificity of the process to be evaluated requires a basic knowledge of its phases and subphases. Therefore, it would be advisable that its members be part of the national statistical office or the NSS, although not the same ones who work on generating statistics from a statistical business register.

6. Characteristics to be considered in evaluating dimensions

Tables X.1 and X.2 present the evaluation dimensions that express the structural aspects of the process. Since this is a general document and not a specific case, the proposed operationalization in the column “Indicators on” refers to components on which indicators should be built, not to defined indicators.

7. Agreements or arrangements

Access to the source of administrative records of public agencies is irregular due to changes in government, as well as management changes at an agency within the same government. Evaluating their implementation is important because it will make it possible to identify advantages and limitations in making the necessary adjustments to promote planning and the institutional framework. In other words, their language should take into account the long-term needs for official statistics, and clauses that allow agreements to be interrupted at the discretion of the officials on duty must be avoided.

8. Needs

Making new or renewed statistical products available to different users can satisfy the needs raised, but they can also lead to the emergence of new needs. Thus, evaluating the behaviour of users’ needs will provide feedback and improve the process, while at the same time determining the needs-supply and supply-needs sequences. This will enable national statistical offices not only to adhere to user requests, but also to expand their statistical offerings from underexploited data sources.

Table X.1
Institutional dimensions matrix

Line I. Institutional				
Dimension	Indicators on	Source of information	Technique or method	Remarks
Agreements or arrangements	<ul style="list-style-type: none"> – Number of agreements or arrangements – Validity of agreements or arrangements – Agencies with agreements or arrangements – Type of clauses – Degree of compliance with agreements or arrangements 	<ul style="list-style-type: none"> – Institutional relations area of national statistical offices – Executives of national statistical offices – Executives of agencies in possession of administrative records – Technical personnel of substantive areas of national statistical offices – Managers of agencies in possession of administrative records 	<ul style="list-style-type: none"> – Internal records of national statistical offices – Questionnaires 	<ul style="list-style-type: none"> – The agreements/arrangements are expected to favour institutionalization – and cover long-term official statistics needs
Needs	<ul style="list-style-type: none"> – Previous requirements – Compliance with previous requirements – New requirements – Inventory of sources – Relationship between design and previous needs 	<ul style="list-style-type: none"> – Internal users of NSS – External users 	<ul style="list-style-type: none"> – Internal records – Questionnaires – Meetings 	
Receipt of administrative records	<ul style="list-style-type: none"> – Costs and maintenance 	<ul style="list-style-type: none"> – Substantive area of national statistical offices – Substantive area of agencies in possession of administrative records – Technical personnel of substantive areas of national statistical offices – Managers of agencies in possession of administrative records 	<ul style="list-style-type: none"> – Internal and exchange records of national statistical offices and agencies in possession of administrative records 	
Confidentiality	<ul style="list-style-type: none"> – Statistical secrecy – Administrative secrecy – Other secrecy – Agreement clauses – Identification of subphases with non-anonymized data 	<ul style="list-style-type: none"> – Legal area of national statistical offices – Legal area of agencies in possession of administrative records – Technical personnel of substantive areas of national statistical offices – Managers of agencies in possession of administrative records 	<ul style="list-style-type: none"> – Legal frameworks of national statistical offices and agencies in possession of administrative records – Internal records of national statistical offices – Internal records of agencies in possession of administrative records – Questionnaires 	<ul style="list-style-type: none"> – Consideration should be given to the rigidity and flexibility of legal frameworks that hinder or favour access to administrative sources
Costs and benefits	<ul style="list-style-type: none"> – Cost of general statistical operations – Cost of statistical operations of administrative registers – Frequency of consultation with reporting entities by areas of national statistical offices – Dispersion or centralization of consultations – Users of national statistical offices, the NSS, external users – Data access times – Statistical operations with administrative records in relation to the total 	<ul style="list-style-type: none"> – Accounts area of national statistical offices – Area managers – Technical personnel of substantive areas of national statistical offices – Managers of agencies in possession of administrative records 	<ul style="list-style-type: none"> – Internal records of national statistical offices – Questionnaire 	<ul style="list-style-type: none"> – The cost should be considered from the point of view of the overall cost-benefit for national statistical offices, not only as a unitary monetary equation

Source: National Institute of Statistics and Censuses (INDEC) of Argentina.

Table X.2
Matrix of technical-methodological dimensions

Line II. Technical-methodological				
Dimension	Indicators on	Source of information	Technique or method	Remarks
Conceptual and analytical design	<ul style="list-style-type: none"> – Inventory of sources – Metadata defined in the design – Outputs designed – Percentage available in relation to the design and needs national statistical offices – Conceptual harmony – Geographic coverage – Identifier used – Classifiers 	<ul style="list-style-type: none"> – Substantive area: national statistical office – Substantive area of agencies in possession of administrative records – Technical personnel at substantive areas of national statistical offices – Technical personnel at substantive areas of agencies in possession of administrative records 	<ul style="list-style-type: none"> – Internal records and records of exchanges between national statistical offices and agencies in possession of administrative records – Questionnaires 	
Receipt of administrative records	<ul style="list-style-type: none"> – Information and communications technology (ICT) software or platform – Capture instruments – Identifier(s) – Storage – Volume of data – Update – Information security – Confidentiality – Data providers – Intake quality control – Documentation – Manuals – Classifiers – Times used 	<ul style="list-style-type: none"> – IT areas of national statistical offices – IT areas of agencies in possession of administrative records – Substantive areas of national statistical offices – Technical personnel in substantive and IT areas of national statistical offices and agencies in possession of administrative records 	<ul style="list-style-type: none"> – Internal records of national statistical offices – Questionnaire 	– Aspects that favour interoperability should be considered
Integration of administrative records	<ul style="list-style-type: none"> – Quality control – Conceptual analysis – Processing consistency – Internal validation – External validation – Proposed changes in administrative records – Duplicates – Conversion of classifiers – Errors – Imputations – Times used 	<ul style="list-style-type: none"> – IT areas of national statistical offices – IT areas of agencies in possession of administrative records – Substantive areas of national statistical offices – Technical personnel in substantive and IT areas of national statistical offices and agencies in possession of administrative records 	<ul style="list-style-type: none"> – Internal records of national statistical offices – Questionnaire 	
Analysis and dissemination	<ul style="list-style-type: none"> – External validation – Consistencies – Population coverage – Series – Metadata – Classifications – Anonymization – Output data and metadata – Product components – Dissemination schedule – Interlocution with users and the press – User support 	<ul style="list-style-type: none"> – Technical personnel of substantive areas of national statistical offices – Technical personnel in charge of dissemination at national statistical offices – External users – Accredited journalists 	<ul style="list-style-type: none"> – Internal records of national statistical offices – Questionnaire 	
Technical assistance and registration	<ul style="list-style-type: none"> – Metadata of national statistical offices – Metadata of agencies in possession of administrative records – Training offerings – Specific training – Exchange meetings – Requests for changes in administrative records – Time frames 	<ul style="list-style-type: none"> – Institutional relations area of national statistical offices – Executives of national statistical offices – Executives of agencies in possession of administrative records – Technical personnel of substantive areas of national statistical offices – Managers of agencies in possession of administrative records 	<ul style="list-style-type: none"> – Internal records – Questionnaires 	

Source: National Institute of Statistics and Censuses (INDEC) of Argentina.

9. Conceptual and analytical design

This is a key phase of the entire statistical process where statistical products, concepts, methodologies, collection instruments, operational processes and so on are designed. It is put to the test when working with inputs designed by other actors and for other purposes. The evaluation of these aspects will make it possible to determine the quality of an external source and its adaptation to the needs of the national statistical office.

10. Receipt of administrative records

The flow of data from external sources to the national statistical office has an impact on the quality of its work, the accurate and timely delivery of its products, and the internal dynamics of the work process.

Monitoring and evaluation of this phase are essential for good performance, improvement and planning of the production of official statistics.

11. Integration of administrative records

Based on the design proposal of the national statistical office, this phase deals with the actual data and designs different from the statistical needs. Consistency, quality and validation control will be more intense compared to that carried out during the data capture phase. Evaluating this stage will offer the possibility of fine-tuning to achieve statistical products in accordance with the original design.

12. Analysis and dissemination

The change of source or the incorporation of, and combination with, pre-existing sources may require interpretative and explanatory efforts, both with the internal work team and with users. By evaluating the impact on the work process and the final results, it will be possible to identify aspects to change for improving results.

13. Confidentiality

One of the limitations on the availability of data from administrative record sources has to do with legal frameworks that establish different types of secrecy depending on the phenomenon concerned: tax, administrative, medical, statistical and others.

This aspect is critical to the possibility of evaluating its impact on a process that seeks to move past the current situation and become public policy.

14. Costs and benefits

The economic approach that accords primacy to economic factors over all other factors is insufficient to evaluate a process such as the one described in this document.

The overall cost-benefit approach should be considered for the NSS and, therefore, for users as a whole. This refers to cost-benefit as a greater and better supply of official statistics based on the possibility of having a greater volume of data, both quantitatively and in terms of a greater variety of phenomena and geographical scope.

15. Technical assistance and registration

NSOs generate their own data through traditional sources: censuses and surveys. These statistical operations involve, among other methodological considerations, the design of collection instruments and metadata.

In the case of administrative records, as mentioned throughout the document, their design is up to the agencies with administrative records in their possession.

If those administrative records are used as a source, it is necessary that the mechanisms for converting administrative instruments to the methodological designs of national statistical offices be designed as part of the dynamics of the exchange with the agencies in possession of administrative records. This should be done without affecting or interfering in the specific functions of the administrative entities and always considering mutual interests.

Evaluating these considerations is crucial for the ongoing performance of the process and its improvement.

Chapter XI

Conclusions and recommendations

Introduction

National statistical offices in Latin American and Caribbean countries face similar difficulties when trying to utilize administrative statistical records that lead to problems in their statistical use.

The purpose of this document is to share recommendations with the countries of Latin America and the Caribbean in order to look for alternatives that will make the best possible use of the statistical potential of administrative records, since they are a very good alternative for generating statistics.

This document's vision is the construction of the four basic registers identified by Wallgren and Wallgren (2007 and 2012). However, it is limited to the economic issue and its proposal is to build a system of economic administrative records.

It is very important to keep in mind the vision of a system of economic administrative records, as proposed by Wallgren and Wallgren (2007 and 2012). This will allow national statistical offices in Latin American and Caribbean countries to take firm steps in that direction, regardless of their level of statistical development or the availability and quality of administrative records, since identifying problems is the first step in attempting to fix them.

Each country, depending on its resources, will have the possibility of addressing some of the proposed recommendations and examples of good practices will enable them to learn how other countries have solved similar problems.

Without question, the participation of the region's countries was of great importance in the preparation of this document, as it reflects the reality and challenges that they face in producing economic statistics, which can largely be solved by developing a system of economic administrative records.

A. Towards a system of economic administrative records

This document presents a vision of a system of economic administrative records, in which the statistical business register is the main component.

Having a robust statistical business register will make it possible, among other things, to create sampling frames for economic surveys, generate business demographics indicators, extract directories with information on economic units corresponding to probabilistic and non-probabilistic surveys, and obtain economic statistics that combine information from various sources.

The creation of a single system of economic administrative records based on a statistical business register and administrative registers maintained by each country's national statistical office will provide countries with an up-to-date economic information infrastructure, allowing them to obtain sampling frames and directories, while offering the possibility of producing new or more timely economic statistics.

The fundamental idea is that the countries of Latin America and the Caribbean will have elements that will enable them to move forward in the consolidation of a statistical business register. This can be updated periodically with economic administrative records and, in countries where they are carried out, with economic censuses.

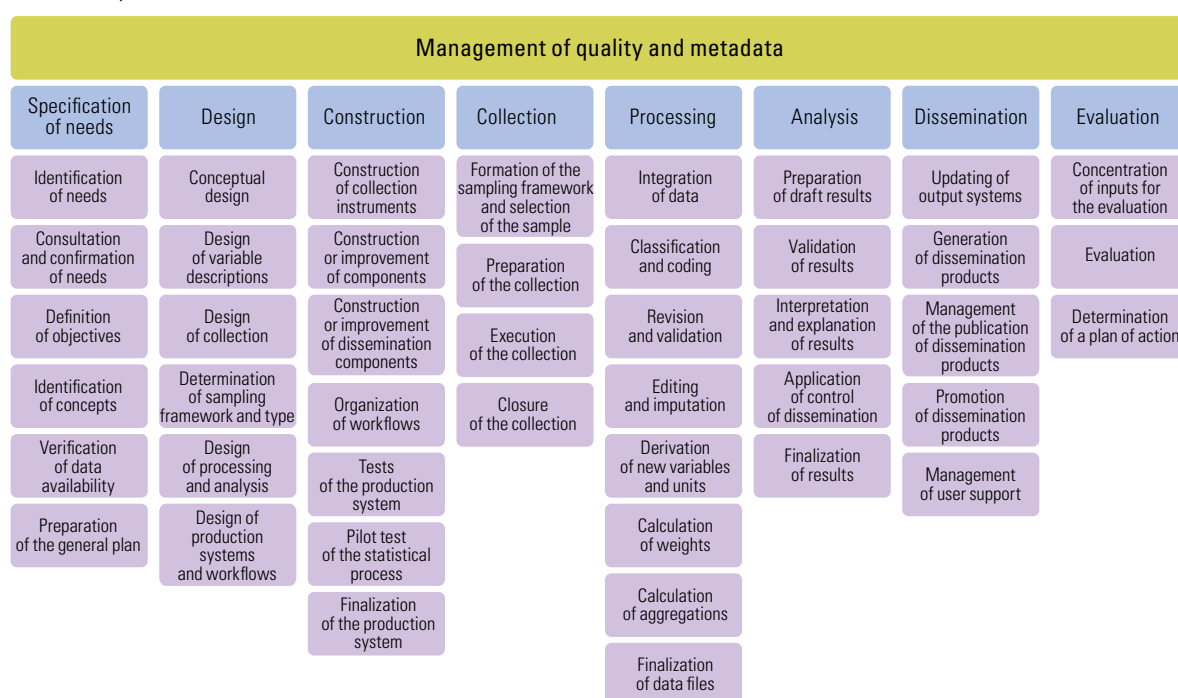
B. Statistical process model (GSBPM)

The recommendations were prepared using the model developed by the UNECE to standardize statistical processes, known as GSBPM, particularly version 1.0 (Spanish).

Diagram XI.1 shows the eight phases of the GSBPM and their corresponding subprocesses.

Diagram XI.1

Phases and subprocesses of the Generic Statistical Business Process Model (GSBPM)



Source: Economic Commission for Europe (UNECE), "Generic Statistical Business Process Model GSBPM (Version 1.0 in Spanish, September 2016)", Geneva, 2016.

C. Detection and analysis of needs

In order to build a system of economic administrative records, it is necessary to analyse demand for information at the country level, at the institutional level or in the form of an explicit request from an international organization or a governmental entity. It is important to include the information needs of all users and avoid duplication of efforts. These needs must be duly documented and prioritized in order to identify available sources of information to address them effectively and efficiently.

It is recommended not to seek from the outset to try to meet all information demands. It is necessary to consider achievable goals in the short term —on the basis of information from accessible administrative and statistical records and data— that meet the expected quality requirements, in order then to devise an improvement plan based on objectives and goals to be met in the long term.

It is very important to include as part of the information needs that all possible variables be disaggregated by sex, which will make it possible to visualize the role of women in different areas of the country's economy.

It is suggested that all the subprocesses identified in the GSBPM model be followed: (a) identification of needs; (b) confirmation of requirements; (c) definition of objectives; (d) identification of concepts; (e) verification of data availability, and (f) development of the general plan.

D. Design

Once the needs of users, both internal and external, have been documented and prioritized, and the availability of data to meet them has been identified, the system of economic administrative records is designed, with the main objectives as follows:

- Attention to the needs of users based on priorities and availability.
- Identification of statistical units and use of a unique identifier for each one.
- Creation of a statistical business register that is updated periodically from administrative records and, if necessary, economic censuses.
- The system of economic administrative records is only created to be used for statistical purposes. This is due to the need to protect the confidentiality of individual information collected by the information systems.

At this stage, the strategies for carrying out the construction, collection, processing, analysis and dissemination phases are outlined, so that goals that can be achieved in short periods of time should be prioritized, after which a longer-term improvement plan should be devised.

E. Construction

The GSBPM indicates that it is in this phase that the production solution is developed and tested to ensure that it can be used in a real-world environment.

The type of solution will depend on the needs of users and the possibilities in each country. Therefore, the national statistical office, in agreement with data providers, must establish an environment for the exchange of administrative records. That environment should consider security and storage aspects so that all the necessary requirements are met.

As previously mentioned, the main component of the system of economic administrative records is the statistical business register. It is recommended to follow the UNECE (2015) guidelines for its construction.

The extraction routines can be automated, semi-automated or manual, depending on how advanced the different data providers are in IT terms. However, the national statistical office must be able to adapt to all possible scenarios in order to capture the data necessary for the construction of the statistical business register.

In addition, the national statistical office should develop the technological means of transmission and interoperability of data obtained from administrative records, files or statistical or non-statistical databases, as well as the techniques and tools to verify that the information sent or received satisfies the quality conditions established for data exchange.

F. Collection

Given the reality of each country, and of each unit that keeps administrative records for the national statistical office, information can be collected from various sources, such as traditional ones from statistical operations (censuses or surveys). Data generated by information-producing units based on administrative records, geographical data, images and so on derived from databases or microdata, can also be used.

Where information is collected from administrative records, it is very important to strictly respect existing agreements between the institutions that have administrative records and the schedules for the delivery of the information. Information should be collected as regularly as possible.

The national statistical office should promote the statistical use of administrative records in its respective sphere. For this reason, it is highly recommended that permanent working groups be formed within the framework of the NSS to encourage the statistical use of data collected by different government and private institutions. To that end, it is recommended that the legal frameworks of both the national statistical office and the agencies involved be revised to allow for the conclusion of collaboration agreements to facilitate exchange of information for statistical purposes.

G. Processing

Data from the administrative register must be processed for analysis so that they can be converted into usable information for statistical production.

Processes must be carried out to verify, clean and transform input data, as well as to standardize classifications and codes, with a view to their integration. In addition, the data must be reviewed and validated; also, if necessary, editing and imputation processes must be carried out. The processes, which can be manual, semi-automatic or automatic, must always be carried out to ensure the quality of the information.

It is worth highlighting the importance of a common identifier in administrative records for their integration. After being identified, the processes must undergo a statistical quality assessment.

It is recommended that a tool be used to support the evaluation of the quality of administrative records. These include HECRA and CECRA.

H. Analysis

The information produced must be relevant and timely, i.e., it must meet the needs of both external and internal users in a timely manner and in good order. It is important to review the documentation of users' needs for that purpose.

In addition, the quality of the information produced must be assured, as must the confidentiality of the national statistical office. It is necessary to carry out the following activities for this purpose: (a) preparation of draft results; (b) validation of results; (c) interpretation and explanation of results; and (d) application of control of dissemination.

I. Dissemination

This subprocess generates the products as previously designed to meet the needs of users, which may include printed or digital publications, press releases and websites. Products can take many forms, including interactive graphs, tables, public-use microdata sets and downloadable documents. Other typical steps in this process include the following:

- Preparing the product components (explanatory texts, tables, graphs, quality statements, etc.).
- Assembling components into products.
- Editing the products and reviewing them for compliance with publication standards.

J. Evaluation

The purpose of this phase is to determine whether the next information production cycle should be carried out using the same specifications as to needs, design and construction, or whether any enhancements are required.

The evaluation should be carried out for all phases of the GSBPM cycle. The following dimensions of the process should be evaluated:

- Collaboration agreements or arrangements
- Needs
- Conceptual and analytical design
- Collection
- Integration of administrative records
- Analysis and dissemination
- Confidentiality
- Cost-benefit
- Technical assistance

Bibliography

- Brazil, Presidency of the Republic of (2011), *Lei N° 12.527, de 18 de novembro de 2011*, Brasília.
- Cabinet of Ministers (2015), "Sistema Integral de Seguimiento y Evaluación de la Gestión (SISEG)," Buenos Aires.
- DANE (National Administrative Department of Statistics) (2020), *Lineamientos para el proceso estadístico en el Sistema Estadístico Nacional*, Bogotá.
- ECLAC (Economic Commission for Latin America and the Caribbean) (2020), *Generic Law on Official Statistics for Latin America* (LC/CEA.10/8), Santiago.
- ____ (2019), "Taller: Big Data, Economía Digital y Desarrollo Sostenible," Santo Domingo [online] <https://www.cepal.org/es/eventos/taller-big-data-economia-digital-desarrollo-sostenible>.
- INAP (National Institute of Public Administration) (2019), "Seguimiento y evaluación de políticas públicas (IN31860/19)," Buenos Aires [online] <https://capacitacion.inap.gob.ar/actividad/seguimiento-y-evaluacion-de-politicas-publicas/>.
- INEGI (National Institute of Statistics and Geography) (2020a), *Norma para el Aseguramiento de la Calidad de la Información del Instituto Nacional de Estadística y Geografía*, Aguascalientes.
- ____ (2020b), *Norma Técnica del Proceso de Producción de Información Estadística y Geográfica para el Instituto Nacional de Estadística y Geografía*, Aguascalientes.
- ____ (2019), "Norma Técnica para la incorporación y actualización de información en el Registro Estadístico de Negocios de México", *Diario Oficial de la Federación*, Aguascalientes, 27 August.
- ____ (2018a), *Norma para la Difusión y Promoción del Acceso, Conocimiento y Uso de la Información Estadística que Genera el Instituto Nacional de Estadística y Geografía*, Aguascalientes.
- ____ (2018b), *Norma Técnica para la Clasificación Nacional de Delitos para Fines Estadísticos*, Aguascalientes.
- ____ (2018c), *Ley del Sistema Nacional de Información Estadística y Geográfica*, Aguascalientes.
- ____ (2017), *Presentación de datos Estadísticos en Cuadros y Gráficas: actualización 2017*, Aguascalientes.
- ____ (2015a), *Glosario del SNIEG 2015*, Aguascalientes.
- ____ (2015b), *Lineamientos para Incorporar la Perspectiva de Género en el Sistema Nacional de Información Estadística y Geográfica*, Aguascalientes.
- ____ (2015c), *Norma Técnica para la Elaboración de Metadatos para Proyectos de Generación de Información de Estadística Básica y de los Componentes Estadísticos Derivados de Proyectos Geográficos*, Aguascalientes.
- ____ (2012a), *Planeación y control administrativo en proyectos estadísticos*, Aguascalientes.
- ____ (2012b), *Procesamiento de la información*, Aguascalientes.
- ____ (2012c), *Proceso estándar para el aprovechamiento de registros administrativos*, Aguascalientes.
- ____ (2011a), *Diseño de la muestra en proyectos de encuesta*, Aguascalientes.
- ____ (2011b), *Presentación de datos estadísticos en cuadros y gráficas*, Aguascalientes.
- ____ (2010a), *Norma Técnica para la Generación de Estadística Básica*, Aguascalientes.
- ____ (2010b), *Diseño conceptual para la generación de estadística básica*, Aguascalientes.
- ____ (2010c), *Proceso estándar para encuestas por muestreo*, Aguascalientes.
- MIDEPLAN (Ministry of National Planning and Economic Policy) (2012), *Manual gerencial para el diseño y ejecución de evaluaciones estratégicas de gobierno*, San José.
- Ministry of Culture and Youth (2019), *Glosario de términos del Ministerio de Cultura y Juventud en el marco del proyecto Sistema de Registros Administrativos de Cultura y Juventud*, San José.
- Ministry of Education (2011), *Nuevas voces, nuevos escenarios: estudios evaluativos sobre el Programa Conectar Igualdad*, Buenos Aires.
- National Assembly of Costa Rica (2019), *Ley del Sistema de Estadística Nacional*, San José.
- Nirenberg, O., J. Brawerman and V. Ruiz (2003), *Programación y evaluación de proyectos sociales: aportes para la racionalidad y la transparencia*, Buenos Aires, Paidós.
- ____ (2000), *Evaluar para la transformación: innovaciones en la evaluación de programas y proyectos sociales*, Buenos Aires, Paidós.

- OECD (Organisation for Economic Co-operation and Development) (2012), *Quality Framework and Guidelines for OECD Statistical Activities* (STD/QFS(2011)1), Paris.
- _____(2007), *Data and Metadata Reporting and Presentation Handbook*, Paris.
- Rojo, A. (2018), "Del dato administrativo de firmas al dato estadístico: desafíos a vencer en América Latina," *CAF Working Paper*, No. 2018/17, Caracas, Development Bank of Latin America (CAF).
- Seguí, F. (2016), "Marco conceptual y metodológico que sustenta el diseño, desarrollo e implementación de un sistema integrado de registros estadísticos de población e inmuebles," Washington, D.C., Inter-American Development Bank (IDB).
- _____(2012), "Guía de la herramienta para la evaluación de la calidad de registros administrativos a ser usados con fines estadísticos," Aguascalientes, National Institute of Statistics and Geography (INEGI).
- SIEMPRO and others (Social Programme Information, Monitoring and Evaluation System and others) (1999), *Gestión Integral de Programas Sociales Orientada a Resultados: Manual Metodológico para la Planificación y Evaluación de Programas Sociales*, Buenos Aires, Fondo de Cultura Económica (FCE).
- UNECE (Economic Commission for Europe) (2019), "Generic Statistical Business Process Model GSBPM (Version 5.1, January 2019)," Geneva.
- _____(2016), "Generic Statistical Business Process Model GSBPM (Version 1.0, September 2016)," Geneva.
- _____(2015), *Guidelines on Statistical Business Registers*, Geneva.
- _____(2011), *Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices*, Geneva.
- _____(2000), "Terminology on statistical metadata," *Conference of European Statisticians Statistical Standards and Studies*, No. 53, Geneva.
- Useche, L. and D. Mesa (2006), "Una introducción a la imputación de valores perdidos," *Terra Nueva Etapa*, vol. 22, No. 31, Caracas, Central University of Venezuela.
- Vale, S. (2013), "Quality and administrative data," Geneva, Economic Commission for Europe (UNECE) [online] <https://statswiki.unece.org/display/adso/Quality+and+Administrative+Data>.
- Wallgren, A. and B. Wallgren (2012), *Estadísticas basadas en registros: aprovechamiento estadístico de datos administrativos*, Aguascalientes, National Institute of Statistics and Geography (INEGI).
- _____(2007), *Register-based Statistics: Administrative Data for Statistical Purposes*, Hoboken, Wiley.

Annexes

Annex A1

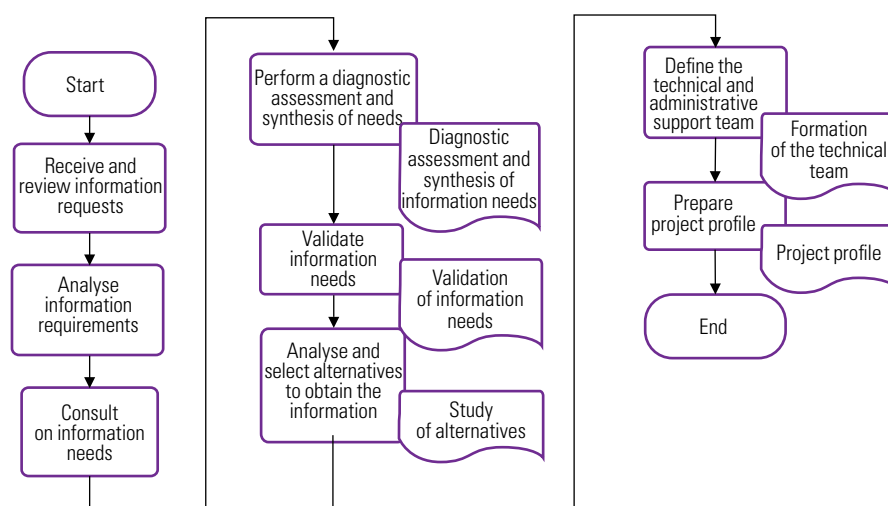
Costa Rica: examples of case studies for each subprocess of needs specification

The identification of needs includes not only consultation with potential users of the information, but also with potential information providers, as well as a review of experiences with similar projects at both the national and international levels, so that they can serve as a reference in designing the system of economic administrative records.

Diagram A1.1 offers an example of a flowchart with the activities that can be carried out in the needs specification phase and that show in a practical way how each subprocess could be approached for the conformation of a system of economic administrative records (an explanation of what each consists of is provided below).

Diagram A1.1

Needs specification flowchart



Source: National Institute of Statistics and Censuses (INEC) of Costa Rica, "Identificación de necesidades del Instituto Nacional de Estadística y Censos de Costa Rica (INEC)", San José, unpublished.

1. Receive and review information requests

The national statistical office receives information requests from users (internal or external) to be reviewed.

Such requests may be the result of a need at the country or institutional level, or an express request from an international organization or governmental body.

This need may be directly related to the use of administrative business registers or some other research need, given that it does not involve the use of records or entail combination with some other collection method.

Ideally, there should be a document that shows all the initial requirements of the information request, a study that evidences the need or some other document that serves as a basis to support the need and the use that would be made of the information, among other aspects.

A matrix such as the one in table A1.1 can be shared with users to provide greater clarity as to needs and the basis of those needs.⁶

Table A1.1

Consultation matrix for identification of information needs

Need	Institutional policy, study or plan

Source: National Institute of Statistics and Censuses (INEC) of Costa Rica, "Identificación de necesidades del Instituto Nacional de Estadística y Censos de Costa Rica (INEC)", San José, unpublished.

2. Analysis of information requirements

The national statistical office analyses the request for information, including each requirement, in order to have more clarity as to the needs and their basis. In any case, it is an input that will be subsequently reinforced and validated in the consultation.

This analysis may include a more detailed investigation of the information needs specified by the different institutions. For example, a more thorough search can be conducted on the websites of each, in order to determine, among other things, whether or not they produce any information related to the needs or if there is any other institution that generates information on the subject. This helps to avoid a duplication of efforts and to ensure that the needs have not been addressed by another institution or initiative.

An example of how to document this requirements analysis is presented in table A1.2.

Table A1.2

Matrix of information needs and existing information

Institution	Information needs	Does it produce information related to its needs?	What information does it produce?	Other information it produces related to companies and establishments	Does another institution produce information to meet this need?	Name of the institution producing the information	What does it produce?	Remarks

Source: National Institute of Statistics and Censuses (INEC) of Costa Rica, "Identificación de necesidades del Instituto Nacional de Estadística y Censos de Costa Rica (INEC)", San José, unpublished.

3. Consultation on information needs

Once the national statistical office has the request for information and the matrix of information needs, it must organize a consultation. This may consist of meetings or work sessions with the information users.

⁶ This matrix can be shared internally in the institution or organization, so that, when it is sent to the national statistical office, it includes the needs in a comprehensive way.

As regards users, both domestic and international actors as well as internal users (of the national statistical office itself) must be taken into account. Internal needs must be considered since the objective is to avoid any duplication of efforts and having to make the same requests for administrative business records to a given institution.

The objective of the meetings should be to explain the system of economic administrative records and to allow users to explain and detail their individual information needs included in the needs matrix. They should be taken into account provided that they included in the project objectives and have a basis that allows a more robust argument for the inclusion of their requirements in the design of the statistical operation.

4. Diagnostic assessment and synthesis of information needs

Once the information users have been consulted, the national statistical office should carry out a diagnostic assessment to summarize the main requirements.

For this purpose, a document with the following sections can be prepared:⁷

- (i) Introduction.
- (ii) Rationale (where does the requirement originate from).
- (iii) Description of the initial request (this is a transcription of the document mentioned in the section “Receive and review information requirements”).
- (iv) Background (detailing the results of the thematic and conceptual bibliographic review, as well as of the review of national laws and of national and international experiences and recommendations).

An initial identification of concepts should be made in this step. As mentioned previously, these preliminary concepts will be an input to be used as part of the subject matter of the system of economic administrative records that needs to be measured. An analysis of the information submitted by users will also be taken into account in order to expand on the topics covered by the information needs, among other things.

A matrix such as the one presented in table A1.3 can be used to identify the concepts.

Table A1.3

Preliminary concept identification matrix

Preliminary main topics	Concepts (preliminary) associated with the main topic	Definition of concepts (preliminary) associated with the main topic ^a	Link or source

Source: National Institute of Statistics and Censuses (INEC) of Costa Rica, “Identificación de necesidades del Instituto Nacional de Estadística y Censos de Costa Rica (INEC)”, San José, unpublished.

^a This can be an input for the conceptual design phase.

The following is a description of each section of the matrix:

- Preliminary main topics: these are the topics that were defined in the identification of needs and that could be addressed in the formulation of the project.
- Concepts (preliminary) associated with the main topic: concepts related to the preliminary topics that can be used to explain them or converted into variables according to each theme.
- Definition of concepts (preliminary) associated with the main theme: this involves the definition of each concept identified.
- Link or source: a backup from where each definition was obtained.

⁷ Taken from National Institute of Statistics and Censuses (INEC), “Diagnóstico y síntesis de las necesidades de información”, San José, unpublished.

As mentioned above, it is important to keep in mind that there may be several concepts from different sources. Later, in the design of the conceptual framework, these could be an input for determining which of all the definitions to implement in the system of economic administrative records.

- (i) Information requirements or needs (define potential users at the internal (institutional), domestic and international levels, as well as stakeholder requirements (needs to be met with respect to the request).
- (ii) Methodological aspects of the request (address the methodological scope (in general terms) that the statistical business register should have, based on the identification of needs).
- (iii) Conclusions and recommendations.

5. Validation of information needs

A fundamental part of the process of identifying information needs is to be clear about the requirements and, most important, that they are consistent with the objectives of the system of economic administrative records.

To that end, the national statistical office needs to validate the information needs with each of the actors involved (users who previously outsourced their information needs) by having them complete the needs identification matrix and subsequently holding meetings with them. The objective is to confirm the scope of the requirements or make appropriate adjustments based on the consultation with users.

One strategy proposed is to carry out a process of repeat consultations with each user at a second or even a third meeting or work session where each of the needs submitted in the matrix can be analysed and fleshed out in greater detail, doubts clarified and so on.

For that purpose, a document with the following structure can be prepared:⁸

- (i) Acronyms and abbreviations.
- (ii) Executive summary.
- (iii) Introduction.
- (iv) Working methodology. This consists of explaining what steps were carried out for the validation of needs.
- (v) Description of institutions and their requirements. This consists of documenting in detail the information from each of the actors consulted in the identification of needs, for example:
 - Name of the institution.
 - Background information (ideally related to the request for information).
 - Substantiation of requests and information needs (details of the basis or grounds provided by the institution at the time of the request (this is part of the proposed matrix)).
- (vi) Analysis of information requirements. Basically, this consists of a joint review of the information needs of the actors consulted.
 - Information requirements that could be obtained through administrative records (information needs related to the objectives of the system of economic administrative records, which are identified and could be obtained from administrative records).
 - Discarded information requirements (information needs that are outside the objectives of the system of economic administrative records, incongruous with its thematic and methodological coverage or whose definition poses an obstacle that precludes their inclusion in the work to be carried out; mainly this is because they cannot be obtained from an administrative register, or that they do exist in registers but the level of detail involved is not part of the intentions, etc.).

⁸ See National Institute of Statistics and Censuses (INEC), "Caso práctico del proceso de identificación de necesidades de estadísticas de empresas", San José, unpublished.

- Identification of providers of administrative records (the institutions or organizations that could be information providers and what administrative records could be requested from each).

Once a preliminary identification of concepts has been made, the availability of the data must be checked through an investigation of existing administrative records that could be key inputs for obtaining them.

For this purpose, it is suggested to review the technical fact sheets of the National Inventory of Statistical Operations (Inventario Nacional de Operaciones Estadísticas), which are generally available in the NSS of each country. It is also suggested to search the website of each user or entity providing records that have been identified and to schedule meetings with information producers to learn about availability and the legal framework, among other things.

Table A1.4 provides an example of how to document the verification of data availability.

Table A1.4

Example for verifying data availability

NSS institution	Statistical operation	Technical fact sheet summary ^a	Legal framework

Source: National Institute of Statistics and Censuses (INEC) of Costa Rica, "Identificación de necesidades del Instituto Nacional de Estadística y Censos de Costa Rica (INEC)", San José, unpublished.

^a Includes aspects such as name of the operation or register, objective, topic, geographic coverage, study unit and periodicity.

In this step it is also important to analyse the legal framework, not only of the institution producing the administrative record, but also of the national statistical office.

- Conclusions and recommendations.
- Bibliography.
- Annexes (persons contacted in each institution or any document supporting any of the requests).

6. Analyse and select alternatives to obtain the information

In order to identify alternatives, the national statistical office must propose one or more solutions for each of the identified needs to be met.

This consists of evaluating the different alternatives or strategies proposed. Precise criteria should be used to select the best intervention strategy.

The selection of the best (optimal) alternative is based not only on whether it is the most feasible in economic, technical, legal and environmental terms, but also on whether it is relevant, efficient and effective. For example, aspects such as budget, cost, duration of implementation, dependence on external resources, contribution to the solution of the problem, or, if necessary, other criteria may be included.

Table A1.5 shows an example of how an analysis of alternatives can be carried out taking into account the considerations mentioned above.

Table A1.5
Example of alternatives analysis

c. Selection criteria and categories		d. Analysis of each alternative		
		Alternative A	Alternative B	Alternative C
		Score	Score	Score
The cost is ...	Low = 3 Medium = 2 High = 1			
The duration is ...	Low = 3 Medium = 2 High = 1			
Dependence on external resources is ...	Low = 3 Medium = 2 High = 1			
The technical feasibility (competent human resources) of the alternative is ...	Low = 3 Medium = 2 High = 1			
The operational complexity of the alternative is ...	Low = 3 Medium = 2 High = 1			
The political viability of the alternative is ...	Low = 3 Medium = 2 High = 1			
The probability of securing financing for the alternative is ...	Low = 3 Medium = 2 High = 1			
The contribution to the problem is ...	Low = 3 Medium = 2 High = 1			
Total score		0	0	0
b. Provide a brief description of each alternative				
a. Name the solution alternatives.	Alternative A			
	Alternative B			
	Alternative C			

Source: National Institute of Statistics and Censuses (INEC) of Costa Rica, "Identificación de necesidades del Instituto Nacional de Estadística y Censos de Costa Rica (INEC)", San José, unpublished.

The alternative with the highest score may be the most feasible to implement. However, it may not be the one that is finally chosen. Another could be selected that is considered more feasible based on the selection criteria assessed, for example, from the perspective of its contribution to solving the development problem.⁹

7. Define the technical and administrative support team

This consists of establishing the core team at the national statistical office that will be required to address the needs specification phase and, if possible, what their participation would be.

This should include not only technical staff, but also the support team (financial, human resources, ICT, etc.).

8. Develop a profile for the system of economic administrative records

The document can be structured as follows:

- (i) Identification of development problem or needs. Here, the development problem or need must be documented, describing the adverse situation (need) perceived by those involved, which must be analysed and solved.

⁹ From National Institute of Statistics and Censuses (INEC), "Instructivo para la identificación, formulación, ejecución y evaluación de proyectos", San José, unpublished.

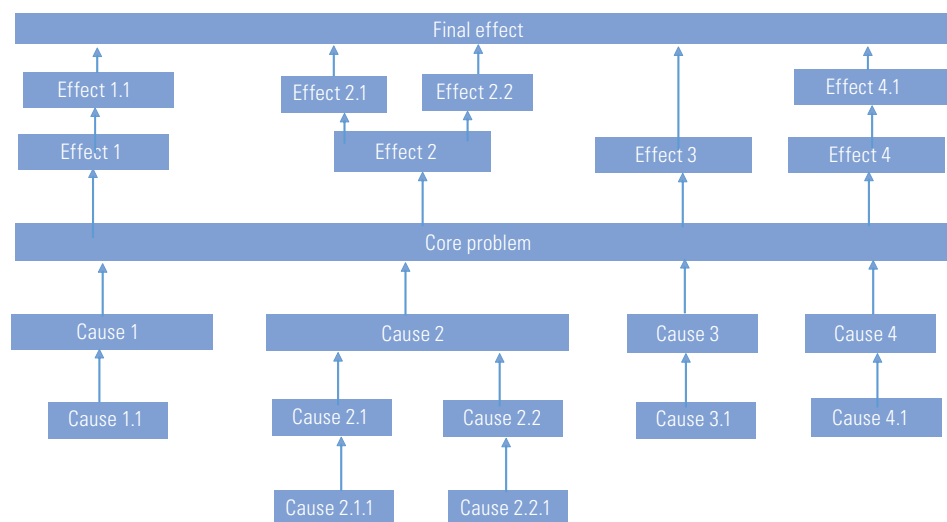
It should be remembered that the development problem or need may derive from previous diagnostic assessments or studies, or from the opinion or judgment of the persons involved.

A good way to document these needs is by determining the causes (remember that these causes will become the possible alternatives to be identified to solve the need) and the corresponding effects (results or consequences of that need).

One way to visualize the problems (needs) detected is to arrange them as shown in diagram A1.2.

Diagram A1.2

Example of a schematic outline for problem or needs analysis



Source: National Institute of Statistics and Censuses (INEC) of Costa Rica, “Identificación de necesidades del Instituto Nacional de Estadística y Censos de Costa Rica (INEC)”, San José, unpublished.

- (ii) Formation of the core team. Here, the team identified in the flowchart process “Define the technical and administrative support team” must be documented.
- (iii) Identification and analysis of stakeholder groups. During the identification of needs, the stakeholder groups (those who have some link to the development problem or need for the system of economic administrative records and which could be part of the solution) should be identified and analysed.

The format shown in table A1.6 can be used to identify stakeholder groups.

Table A1.6

Example of a table identifying the stakeholder group

Stakeholder groups linked to the development problem or need	Link(s) of the stakeholder group to the problem or need	Position of the stakeholder group for each expectation raised		Remarks
		Favourable	Adverse (risks)	

Source: National Institute of Statistics and Censuses (INEC) of Costa Rica, “Identificación de necesidades del Instituto Nacional de Estadística y Censos de Costa Rica (INEC)”, San José, unpublished.

^a Examples: authorization of public budgets; approval of regulations, decrees, laws; authorization of positions, loans of assets, trade union support, political support, logistical support, assertive communication with the public, international cooperation, training, advisory services, etc.

This analysis identifies the possibility of support for the solution of the development problem or need. In addition, an assessment can be made of the feasibility of the system of economic administrative records.

In the event that negative perceptions are identified, these should be considered potential risks to the future intervention. Therefore, they should be promptly addressed, so that measures can be taken to reverse or counteract them and gradually create the technical, economic, political and operational feasibility of the intervention.

- (iv) Analysis of objectives. Defining objectives consists of converting the adverse conditions in the needs identified into solutions.

Steps in building the objectives tree:¹⁰

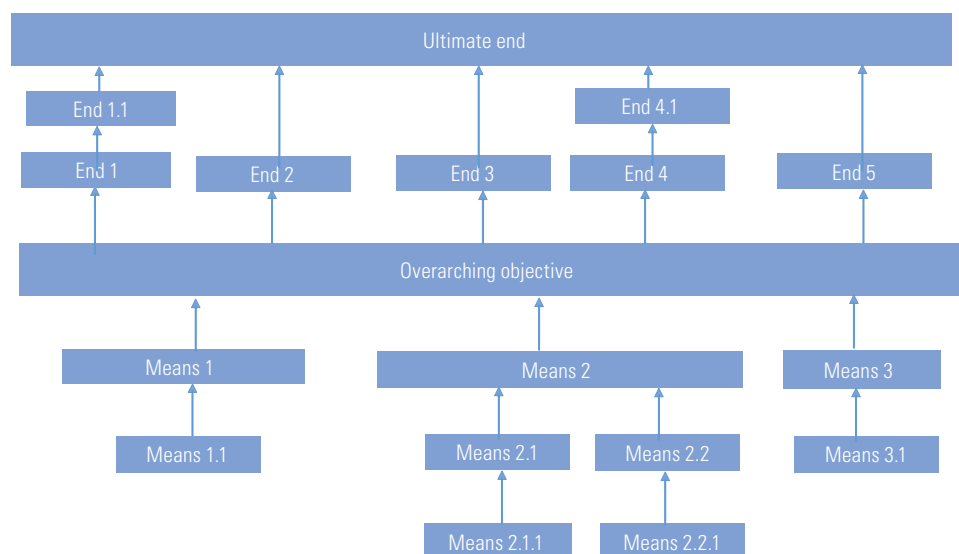
- Step 1. Determine the overarching objective. This is achieved by turning the identified development problem or need into a positive situation.
- Step 2. Identify the ends. In other words, the effects determined in the identification of needs must be converted into ends, in such a way that the expected positive state can be visualised.

The ends of the overarching objective are the positive consequences that occur when the identified problem is solved.

- Step 3. Identify the means. This consists of replacing the causes that give rise to the problem or need with opposing events that contribute to solving it.
- Step 4. Validate the objectives. Once the objectives have been constructed, the relationships between the established means and ends are examined to ensure that the analysis outline is valid and complete (see diagram A1.3).

Diagram A1.3

Example of an outline for determining objectives



Source: National Institute of Statistics and Censuses (INEC) of Costa Rica, "Identificación de necesidades del Instituto Nacional de Estadística y Censos de Costa Rica (INEC)", San José, unpublished.

- Identification of solution alternatives and selection of the most advisable one. Here, the team identified in the flowchart process "Analyse and select alternatives to obtain the information" must be documented.
- Constitutive act of the system of economic administrative records. Once the information on the problem or need and the alternative chosen for its solution have been gathered, the basic elements required to submit it for consideration by the corresponding authority are available (see table A1.7).

¹⁰ From National Institute of Statistics and Censuses (INEC), "Instructivo para la identificación, formulación, ejecución y evaluación de proyectos", San José, unpublished.

Table A1.7

Example of a constitutive act

Constitutive act of the system of economic administrative records	
Name (selected alternative)	System of economic administrative records
Name of coordinator of the system of economic administrative records	
Overarching objective	
Development problem or need	
Rationale or importance	
Components	
Estimated implementation time for the statistical business register	
Cost (if possible, an approximation of the cost, based on similar projects, could be provided)	
Stakeholder groups	
Risks	

Submitted by: _____
 Approved by: _____

Date: _____
 Date: _____

Source: National Institute of Statistics and Censuses (INEC) of Costa Rica, "Identificación de necesidades del Instituto Nacional de Estadística y Censos de Costa Rica (INEC)", San José, unpublished.

Annex A2

Cuba: practical experience with a statistical business register

1. Legal framework for the creation of the statistical business register in Cuba

Decree Law No. 335 on the Public Records System of the Republic of Cuba stipulates that the National Office of Statistics and Information (ONEI) participates in the organization of the public records system, as its members are a source of information for the government.

Decree Law No. 6 on the Government Information System (SiGob) defines SiGob as an integrated set of systems organized to meet information needs related to the government's objectives and plans at all organizational levels of society and citizens. It also designates the ONEI as the national entity whose mission is to direct the NSS and to be responsible for the methodological direction of SiGob, including its organization, coordination, integration and control.

Decree No. 9 on the Implementing Regulations of Decree Law No. 6, establishes that, among other of functions, the ONEI defines and implements, for statistical purposes, methodologies, classifications and nomenclatures, as well as advising on their use; coordinates the consistent integration of public registers and administrative records that constitute the sources of SiGob and identifies, defines and controls, as appropriate, for statistical purposes, the reporting centres that contribute to the NSS.

The above ensures a legal framework for the use of registers, the impact on their design, the implementation of standards and codifiers defined by the ONEI, the registration of SiGob actors and the allocation of a unique identifier as part of those codifiers.

The principle of both systems is interoperability, so that both registers and systems communicate with each other, an aspect envisaged in the computerization process.

2. Process of creation of the statistical business register in Cuba

The statistical business register in Cuba is the Directory of Institutional Units and Establishments (Directorio de Unidades Institucionales y Establecimientos – DUIE); the records of legal entities in the country are its main source of data.

Due to low computerization in the country's registers of legal entities, the ONEI introduced direct registration, requiring legal units to submit their incorporation documents to ONEI.

Based on the incorporation documents, the register collects not only information on identification and location, but also codes a set of attributes, including the type of economic activity according to the National Classification of Economic Activities (Clasificador Nacional de Actividades Económicas – CNAE), institutional sector, form of financing, location in special development zones, business groups with which it is linked or to which it is subordinate, establishments, and so on.

When the unit is registered in the DUIE, it is assigned a unique identification number issued by ONEI and an official document of that registration is issued to the registered unit. This unique identification number is not repeated, even if the entity is extinguished. The entry in the register is made verbatim on the basis of the documents provided by the economic entities. The unique identification number issued to the unit is widely used in the country for opening bank accounts, registering with the Tax Office and entering into commercial and service contracts.

Coordinated use of that code by entities with the bank and the tax office as the main partners, which use the same identifier number in their databases, ensures mandatory registration of entities in the ONEI and coverage of close to 100% of legal entities.

The code issued by ONEI identifies SiGob actors, so that, in principle, through the unique code, statistical information can be related to existing information in the government's administrative records.

The identification code does not provide information about the characteristics of the entity. However, the attributes of each legal unit are identified in the administrative records database, which allow different types of registered units to be related to the statistical information reported and to other associated entities, such as business groups and establishments.

The DUIE registers government units, state-owned companies, mixed and foreign capital companies and commercial companies with Cuban capital, as well as agricultural and non-agricultural cooperatives; it also includes some non-profit associations. Apart from legal entities, the DUIE also contains establishments (still in the process of identification), which are updated based on a survey of legal entities in the locality where subordinate structures organized in strata are established. Based on a study of each legal unit, the scope of the unit is determined down to its subordinate units at the local level, by type of activity or establishment, as appropriate.

An important datum is the main economic activity of the legal unit, as indicated in its incorporation documents, based on the corporate purposes or functions established in principle for each legal unit. All this is contrasted with the economic information provided in the different statistical operations that the unit reports based on the assigned identification number.

In addition to the unique identification number of the legal units, there is also a tax identification code or Tax Identification Number (Número de Identificación Tributaria – NIT), which is mainly used as an identifier in the country's customs register. However, the ONEI has been working on correlating both codes in order to use the customs register for statistical studies based on it.

Annex A3

Cuba: inventory of administrative records for statistical purposes

National Office of Statistics and Information of Cuba
Directorate of Information Methodology
March 2017

Objective

To establish an instrument to update the inventory with the characteristics (source and metadata) of the administrative registers.

Administrative register inventory guide

(1) General data

- 1.1 Name of institution: the name of the institution where the inventory information is collected.
- 1.2 Mission of institution: the *raison d'être* of the institution, which must be written as stated in official documents.
- 1.3 Geographical address: the main and secondary streets and the number of the building where the institution is located are recorded.
- 1.4 Register contact person: data of the contact person for the administrative register, i.e., first and last names, address and administrative unit where they work, telephone number and e-mail address.

(2) Characteristics of the administrative register

- 2.1 Name of the administrative register: the name of the administrative registers managed by the institution must be recorded.
- 2.2 Type of register: for the purposes of the survey, registers will be classified as "population", "activity", "business" and "real estate".
- 2.3 Maximum legal regulation supporting the creation of the register: the maximum legal regulation that allows the operation of the administrative register. Only one answer allowed.
- 2.4 Objective of the administrative record: the overarching objective of the administrative record in the institution should be stated. Target population: individuals or objects for which data are recorded in the administrative register, which does not correspond to the institution's own administrative data, but to the people who use the services.
- 2.5 Coverage: the broadest geographic coverage reached by the administrative register.
- 2.6 Year of creation: the year in which the administrative register was officially created.
- 2.7 Means of capture: the means by which the institution captures the information in the administrative register, which may be physical, electronic or other means. Indicate the primary means by which information is captured directly from the user.
- 2.8 Periodicity of capture: the periodicity with which information is captured (collected) from users (individuals, objects).
- 2.9 Format in which the administrative register data are available: the computer programme (statistical software) in which the data collected in the administrative register are available.
- 2.10 Centralized or decentralized databases: refers to the availability of a centralized or non-centralized database in a single repository, when records are collected at different geographical locations.

- 2.11 Year of availability of the distributed database (DDB): the year from which the data captured from the administrative record is available in a database.
- 2.12 Cut-off period for storage: the frequency with which information is cut off for storage in the DDB.
- 2.13 Unique identifier: the availability or not of a unique identifier for each object (individual) in the administrative register's database.
- (3) Dictionary of variables
- 3.1 A dictionary of variables, explanatory fact sheet or some document describing the structure of the variables in the administrative register: this refers to the availability or not of a dictionary of variables, explanatory fact sheet or document containing information on the characteristics of the data (variables) captured by the administrative register (see tables A3.1 and A3.2).

Table A3.1

Administrative register inventory form

1. General data	
1.1. Name of institution	
1.2. Mission of institution	
1.3. Geographic address	
1.4. Register contact person	Names: Surname: Administrative directorate: Unit where they work: Telephone: ext.: E-mail:
2. Characteristics of the administrative register	
2.1. Name of administrative register	
2.2. Type of register	
2.3. What is the highest legal regulation supporting the creation of the administrative register?	Constitution <input type="checkbox"/> 1 Treaty <input type="checkbox"/> 2 International law <input type="checkbox"/> 3 Decree <input type="checkbox"/> 4 Resolution <input type="checkbox"/> 5 Other <input type="checkbox"/> 6 Please explain None <input type="checkbox"/> 7
2.4. What is the overarching objective of the administrative register?	
2.5. What is the target population being registered?	
2.6. What is the coverage of the administrative register?	National <input type="checkbox"/> provincial <input type="checkbox"/> Other <input type="checkbox"/> please explain
2.7. Year of creation of the register	
2.8. What is the means of capture?	Physical form <input type="checkbox"/> 1 attach form E-mail form <input type="checkbox"/> 2 attach form Other <input type="checkbox"/> 3 Please explain
2.9. What is the periodicity of capture?	Annual <input type="checkbox"/> 1 Half-yearly <input type="checkbox"/> 2 Four-monthly <input type="checkbox"/> 3 Quarterly <input type="checkbox"/> 4 Monthly <input type="checkbox"/> 5 Weekly <input type="checkbox"/> 6 Daily <input type="checkbox"/> 7 Other <input type="checkbox"/> 8 Please explain

Table A3.1 (concluded)

2. Characteristics of the administrative register	
2.10. In what format is the administrative register data available?	SPSS <input type="checkbox"/> 1
	Excel <input type="checkbox"/> 2
	Stata <input type="checkbox"/> 3
	REDATAM <input type="checkbox"/> 4
	Other <input type="checkbox"/> 5
	Please explain
	None <input type="checkbox"/> 6
2.11. Are the register data centralized or decentralized?	Decentralized database <input type="checkbox"/> 1
	Centralized database <input type="checkbox"/> 2
2.12. Year of availability of the ddb	
2.13. What is the main cut-off periodicity for ddb storage?	Annual <input type="checkbox"/> 1
	Half-yearly <input type="checkbox"/> 2
	Four-monthly <input type="checkbox"/> 3
	Trimestral <input type="checkbox"/> 4
	Monthly <input type="checkbox"/> 5
	Weekly <input type="checkbox"/> 6
	Daily <input type="checkbox"/> 7
	Other <input type="checkbox"/> 8
	Please explain
2.14. Does the ddb have a unique identifier for each object (individual)?	Yes <input type="checkbox"/> 1
	Please explain
	No <input type="checkbox"/> 2
2.15. Does the institution generate statistics with the administrative register's database?	Yes <input type="checkbox"/> 1
	Which?
	No <input type="checkbox"/> 2
3. Dictionary of variables	
3.1. Does the administrative registry have a dictionary of variables, explanatory fact sheet or any other document describing the structure of the variables in the register?	Yes <input type="checkbox"/> 1
	Attach files and complete technical fact sheet
	No <input type="checkbox"/> 2
	Complete technical fact sheet

Source: National Office of Statistics and Information (ONEI) of Cuba.

Table A3.2

Administrative register technical fact sheet

Name of variable (write out in full what appears in the form)	Definition of the field or variable (if applicable, indicate the unit of measure)	Type					Field	Open			Closed			Variable categories		
		Numeric	Numeric with decimals	Alphabetical	Alphanumeric	NS/NR		Open	Closed	NC/NR	Coded	Not coded	NC/NR		Exclusive	Multibrand

Source: National Office of Statistics and Information (ONEI) of Cuba.

- (a) Variable categories according to function
- Identification variable (key)
 - Communication variable (contact)
 - Reference variable (external keys)
 - Time reference
 - Technical variables (internal use)
 - Statistical variables (age, sex, marital status)

Annex A4

Chile: the experience of the INE

As part of the formulation of strategic projects arising from the development process of the 2018–2022 Strategic Plan of the National Institute of Statistics (INE) of Chile (which is framed around three pillars: public value of statistical data, national statistics and organizational excellence), the institution identified the need for a system to integrate the administrative records available in the country and enhance their use in statistical production. Thus, it was decided to design the strategic project “Integrated system for the statistical use of administrative records”.

The project aims to devise, design, implement and document an efficient system that integrates the statistical use of administrative records into the INE’s production in accordance with international quality and security standards. This does not mean that the INE does not currently carry out statistical operations based on administrative records; however, it is necessary to give greater impetus to this source of data, in line with the information requirements of today’s world.

To build the system of statistical records, INE must have access to administrative records available at different administrative offices. Therefore, agreements and projects should be established with different ministries and authorities at the managerial level, with the aim of promoting constant cooperation between the INE and other institutions. Ways should be explored for administrative offices to have access to basic and statistical registers for the purposes of statistical studies and analysis. It is important to mention that the privacy of the data must be protected throughout this process.

The large volume of information required for decision making makes the work of national statistical offices increasingly costly, which, in turn, entails a significant burden for reporting units.

The statistical burden is directly related to the various questionnaires received by reporting units. In this context administrative records become a key tool for supplementing countries’ statistical output.

The project “Integrated system for the statistical use of administrative records” seeks to lay the methodological foundations for creating a sustainable system of statistical records. This requires a transition from a traditional statistical system to a new system based on administrative records. This entails new methods for how to use administrative data, new forms of institutional cooperation and long-term work for the development of new statistical production methodologies.

Within the project, some preliminary definitions have been introduced regarding the conceptual framework for the statistical use of administrative records, as has a model for the management and statistical production of administrative records. The statistical production and management model presents a scheme for organizing administrative records, from the delivery of records by external institutions to the release of data for statistical use by producers.

In line with the construction process are two experiences developed by the INE within the framework of the project. They are associated with the subprocesses of construction of collection instruments and construction or improvement of components.

1. Construction of collection instruments: file transfer

To build the system of statistical records, the INE must have access to administrative records from different administrative sources. Therefore, agreements and projects should be established with different ministries and authorities at management level, with the aim of promoting constant cooperation between the INE and other institutions. The latter should be involved in statistical operations carried out by the INE, so that if changes occur in administrative office data, they can understand the impact on the statistics.

- (i) Whenever the INE concludes a provision agreement with a data provider that involves the receipt of administrative records, the following channels of exchange should be considered for data transfer:
- (ii) VPN (virtual private network) provided by data providers, where specific INE staff must connect to external servers to extract information and then store it in a shared directory (file server).

The INE provides a VPN, where the data provider transfers information to a specific folder. INE credentials for those responsible for transferring data are delivered by certified mail.

These two ways of capturing information make it possible to establish a secure environment for the transfer of large volumes of data onto a server established for that purpose, with controlled personnel who can access and extract the files.

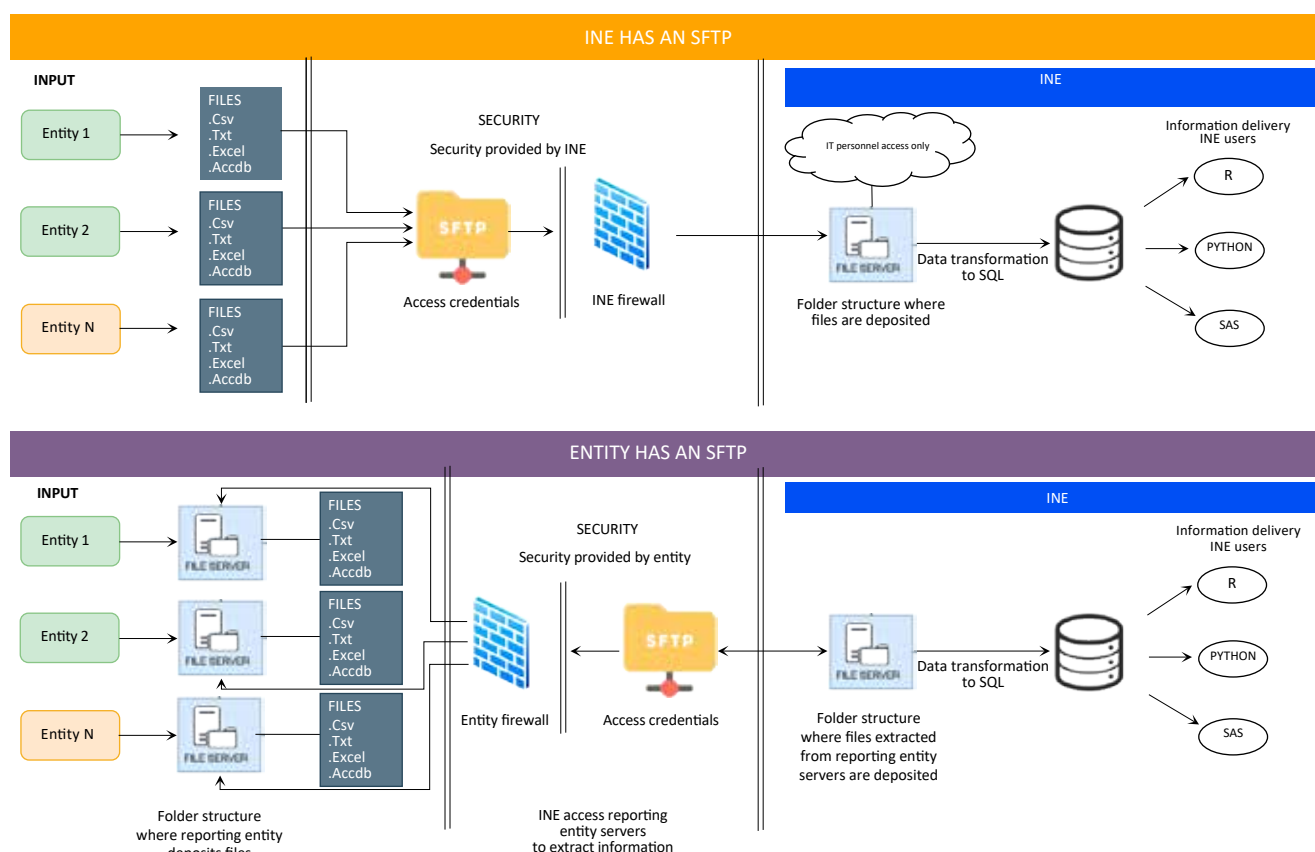
With these alternatives, the INE can adapt to the needs of the data provider and also establish the conditions for data exchange. These methods of exchange support all file formats (plain text, Excel spreadsheets, etc.).

While files can be transferred in a structured format, the focus of the INE definitions relates to the exchange of files without any formatting, given its advantages in terms of reduced file volume.

To safeguard data confidentiality, direct access to download these files within the institution should be limited to a clearly defined team in the information technology unit. This information will later be transformed into statistical records to be made available to production teams. The scope of that transformation is explained in the following section (see diagram A4.1).

For the collection of the background information associated with the channel of exchange, the following items must be documented: channel of exchange, name of the data backup server, path of the data backup server, systems contact at the organization and systems contact at the INE. Table A4.1 presents details on each item. This background information facilitates data exchange management.

Diagram A4.1
File transfer scheme



Source: National Institute of Statistics (INE) of Chile.

Table A4.1
Items of the channel of exchange

Item	Description
Channel of exchange	Indicate the channel of exchange through which the information is transferred. These may be internal SFTP, external SFTP, magnetic media, etc.
Name of the data backup server	Indicate the name of the server where the data from the organization is backed up.
Location of the data backup server	Path where the backup of the administrative records is located.
Systems contact at the organization	Provide the name, position, e-mail and telephone number of the person responsible for the organization's information systems associated with the information transfer.
Person responsible for INE systems	Provide the name and position of the contact person responsible for information systems at the INE.

Source: National Institute of Statistics (INE) of Chile.

2. Construction or improvement of components: data warehouse for work with administrative records

Based on the above, when plain text files (in .csv or .txt format) are received in the exchange of information with data providers, the files must be extracted from the server.¹¹ Then, the activities for delimitation and transformation from an administrative register into a statistical register are carried out.¹²

¹¹ Other types of plain text files can also be received.

¹² As can other structured formats.

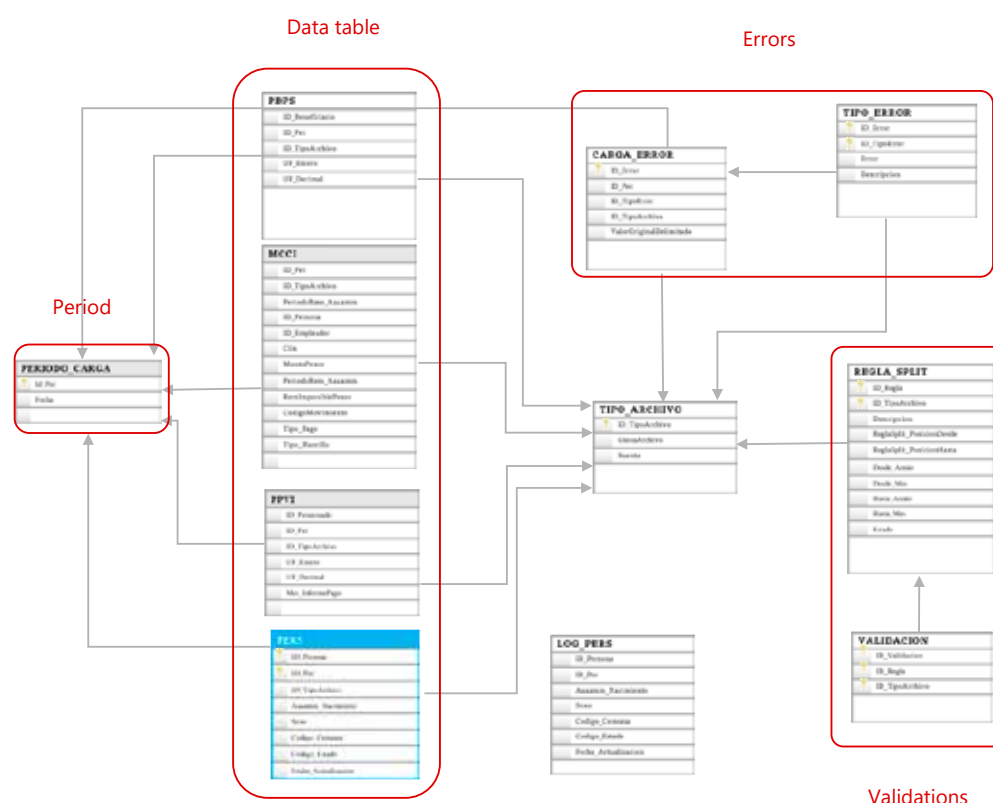
First, in each file, the separation of variables according to the defined delimitation and attributes must be considered. Once the files are correctly delimited, they are stored in a single database model (data warehouse) where four groups are distinguished: data tables, period, errors and validations. The data tables correspond to the files provided by the data providers; the period corresponds to the defined loading period; the validations refer to all the business rules applied to the files; the errors correspond to the storage of all the inconsistencies detected on the basis of the defined business rules.

This new way of modelling the work with administrative records responds to the need to address the requirements for working with large volumes of information and to develop a data system that takes advantage of the current competencies of the institution's IT teams, which have extensive knowledge of working in SQL.

Diagram A4.2 shows the data warehouse model applied to data from the Office of the Superintendent of Pensions.¹³ With different administrative records, each working model will have different table names in the box containing the data table; however, this is the working model that is being applied to each administrative register for its transformation into a statistical register.

Diagram A4.2

Data warehouse modelling scheme with data from the Office of the Superintendent of Pensions



Source: National Institute of Statistics (INE) of Chile.

Part of the delimitation activities involve validating the defined design. This means that, if design envisages the file as containing five variables, this structure must be maintained whenever files are received.

¹³ This corresponds to one of the administrative records that has been processed using this model, given their monthly volume. Data has been available since December 2009, with data tables containing around 7 million rows per month.

If there is a different delimitation structure, the data provider should be contacted and informed of the inconsistencies found, and it may be necessary to modify the design and construction of the script in previous stages, in order to re-execute the delimitation. To that end, a documentation protocol is established for each transferred file containing the information in tables A4.2 and A4.3.

Table A4.2

Data structure items to characterize files (tables)

Item	Description
File title	Indicate the title by which the file is known.
File name	Indicate the file name.
File format	Indicate the file format (.txt, .csv, .xlsx, etc.).
File delimitation	Indicate the type of delimitation (fixed width, delimited by character, or if already structured). If a delimiter exists, indicate which character is used.
Number of variables	Indicate the number of variables contained in the file.
Periodicity of file submission	Indicate the periodicity of file receipt by the INE. If possible, the receipt period should be stated; for example, last business day of the month, year-end, etc.
File size (approximate)	Provide an estimate of the file size in megabytes (MB) in order to calculate the space to be used for its backup and later disposition in the work servers.
File description	Describe the main aspects of the principal characteristics of the variables contained in the file.

Source: National Institute of Statistics (INE) of Chile.

Table A4.3

Data structure items to characterize the contents of each file

Item	Description
File name	Indicate the file name.
Field information	Indicate the description of the field or variable, together with its content.
Field size	Indicate the field size.
Field format	Indicate the field format, whether it is alphanumeric, numeric, date, etc.
Additional field notes	Include additional notes regarding the content of the field, such as scope of the variable, list of categories or values that the variable may take, etc.
Field position	Indicate the position of the variable in the file. It must be a number.

Source: National Institute of Statistics (INE) of Chile.

In addition, as part of the transformation of the administrative registry into a statistical registry, standardizations are carried out and variables are coded and converted to facilitate unit integration and value comparisons using unique codes. These activities include the creation of new variables. Validations are also performed within the model, either with respect to logical intervals or to consistency within tables, as defined in the design.

It should be indicated, within the transformations to be carried out, if steps are taken to safeguard the identity of statistical units. All administrative records containing identifier variables, where it is specified that the latter will be used to create links (which must be defined for each variable of each administrative record in the data dictionary), must be transformed into an internal identifier, which is ultimately the variable that will enable the integration of records.

In addition, all those variables that contain explicit information and where statistical units can be directly identified are defined as sensitive variables and will be encrypted to maintain their confidentiality and indeterminacy. This is done so that all sensitive information identifying the statistical unit is kept statistically secret within the system.

Finally, once the transformation of the administrative registry into a statistical registry has been completed, the data can be used by INE producers. Each user who requires it must request access to the server where the data is located. They can connect using statistical software, such as SAS or R, in order to conduct their analysis.

Annex A5

Peru: experience of the INEI in the processing phase

The National Institute of Statistics and Informatics (INEI), through the Executive Directorate of Censuses and Surveys of Enterprises and Establishments of the National Directorate of Censuses and Surveys, updates the Central Directory of Enterprises and Establishments (DCEE) every year.

The DCEE is an information system whose main source of information is the Taxpayer Register of the National Tax and Customs Administration (SUNAT), which is itself updated from studies periodically carried out by INEI and from the records of other institutions.

It is important to note that the DCEE is a core instrument for the construction of sampling frames for surveys of enterprises and establishments in the NSS. Its maintenance or updating is important for improving the quality of economic statistics.

1. Uses of the directory

- It is used to obtain sampling frames for business surveys.
- It provides information on business demographics (creation and death of companies).
- It enables analysis of the country's business structure.
- It provides information on the location of enterprises by means of georeferencing.

2. Processing

The DCEE is processed on a quarterly and annual basis. The activities carried out for information processing encompass five phases, as described below.

2.1 Data integration

This phase entails the integration of different administrative records. In the case of Peru, the basic register is the Taxpayer Register provided by SUNAT, the tax authority. The information provided by SUNAT is for all companies formed or incorporated by natural persons and legal entities. Sources such as economic surveys and other administrative records are also included, with the Single Taxpayer Register (RUC) number used as the variable or enterprise identifier code that enables the integration of all records or surveys in the DCEE.

SUNAT, which is the main source, provides two types of information: one is annual (information from the Taxpayer Register) and the other daily. The latter refers to new developments and is called "additions and deletions of businesses".

(a) Receipt and processing of information (with daily information)

The information is issued by the tax authority on a daily basis. Six types of .txt files are stored online via SFTP. However, the update is performed quarterly. Updating involves processes that include receipt of information, extraction, cleansing, imputations and coding, until a standardized database is obtained.

(b) Receipt and processing of information (with annual information)

The information is issued by the tax authority on an annual basis under an agreement with that institution. The information submitted enables the updating of variables, including legal organization, ISIC, addresses, and so on.

It is important to note that the coverage of the DCEE includes all SUNAT-registered enterprises that have a RUC code and are active or in operation. Enterprises formed as natural persons without a business are excluded. Informal economic units are also not considered.

The coding system used is based on the variables contained in the DCEE, such as, for example, ISIC, Rev.4; geographic location code (UBIGEO) and others.

There are also internal data upload validation criteria, such as the requirement that the RUC code have 11 digits; if not, the record is considered invalid. Regarding variable data, to ensure consistency in the information, it is checked, for example, that the ISIC code has only four digits and the UBIGEO, only six, and that dates are in a specific format, among other things. However, there are errors or inconsistencies that do not conform to rules and must be reviewed manually before the information is integrated.

As regards consistency analysis, reports are prepared on distributions, minimums and maximums when variables are numerical. Likewise, each variable has a code that is verified in this process. Consistency analysis is not only performed with internal data, but also with external sources that have a particular weight depending on the type of variable. Economic surveys for updating ISIC, addresses, sales and employed personnel carry the greatest weight.

2.2 Classification and coding

The directory uses international classifiers such as ISIC, Rev.4 to code economic activities, as well as the geographic location code (UBIGEO). Text variables are also coded in order to have data or concepts that allow national or international comparability.

Coding can be done automatically or manually. It is assigned by IT and in the information validation process.

2.3 Review and validation

The validation process verifies that enterprises have a valid RUC number. The code must be 11 digits long and must start with the number 1 or 2. Only active statistical units are also considered, while natural persons without a business, marital partnerships without a business, undivided estates without a business and units in the process of liquidation are excluded. There should be no duplication of data.

2.4 Editing and imputation

Editing in the DCEE is performed, for example, when the description of variables is cut off or abbreviated. Accordingly, what is done is to standardize and edit.

Imputations are only done for the sales and employed personnel variables of the enterprise, since information is not available for all economic units. For that reason, minimum values have been created to assign to such variables when information is not available.

In the DCEE, enterprises without that information are generally recently created enterprises or very small business that do not file tax returns with the tax authority.

2.5 Derivation of new variables and units

This subprocess is carried out to obtain new variables in line with the results to be disseminated. They are generally aggregations of variables, that is, groupings of variables that facilitate the analysis of the information, or variables derived from other variables.

(a) Variables derived by grouping

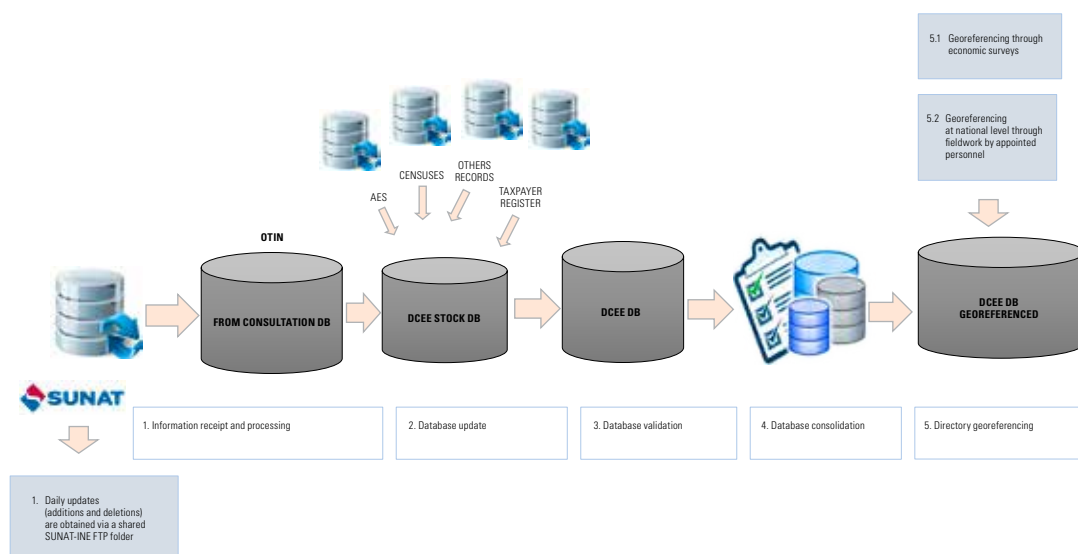
This category refers, for example, to the range of sales or the range of personnel employed. In this case, numerical variables are divided into intervals for better analysis.

(b) Variables derived by aggregation

The derived variable is created from a record, using another record that has more disaggregated information. For example, in the case of the DCEE, the ISIC variable is used and new variables such as division, group or section are created. More aggregated variables can even be created for better analysis (see diagrams A5.1 and A5.2).

Diagram A5.1

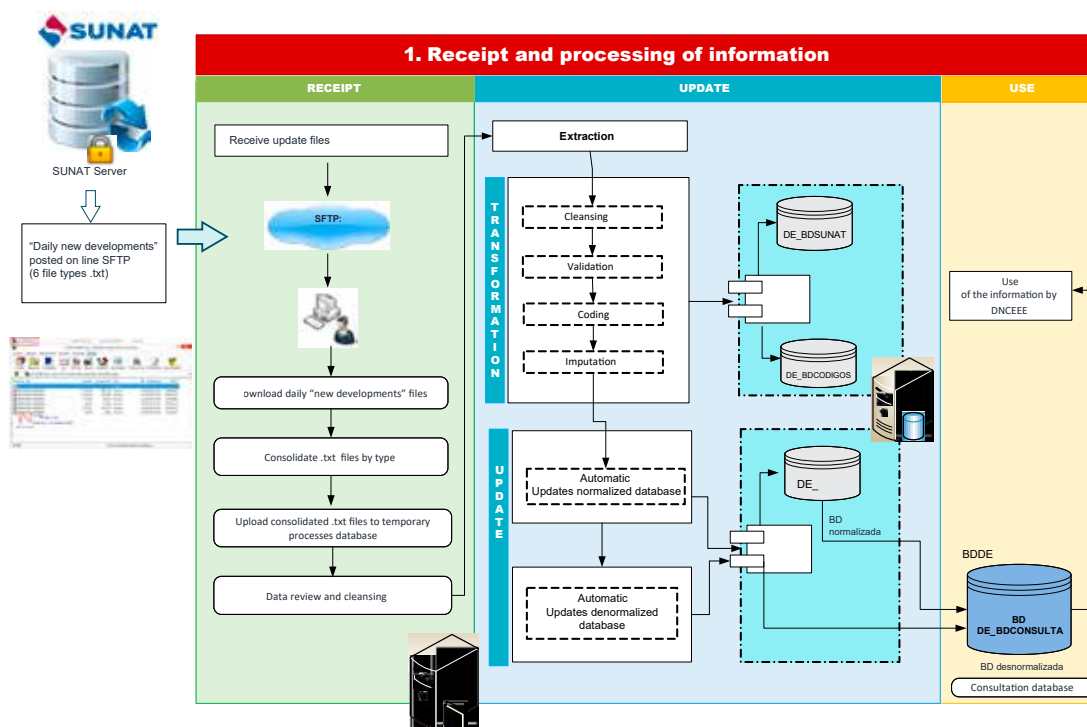
Process for updating the Central Directory of Enterprises and Establishments



Source: National Institute of Statistics and Informatics (INEI) of Peru.

Diagram A5.2

Receipt and processing of information



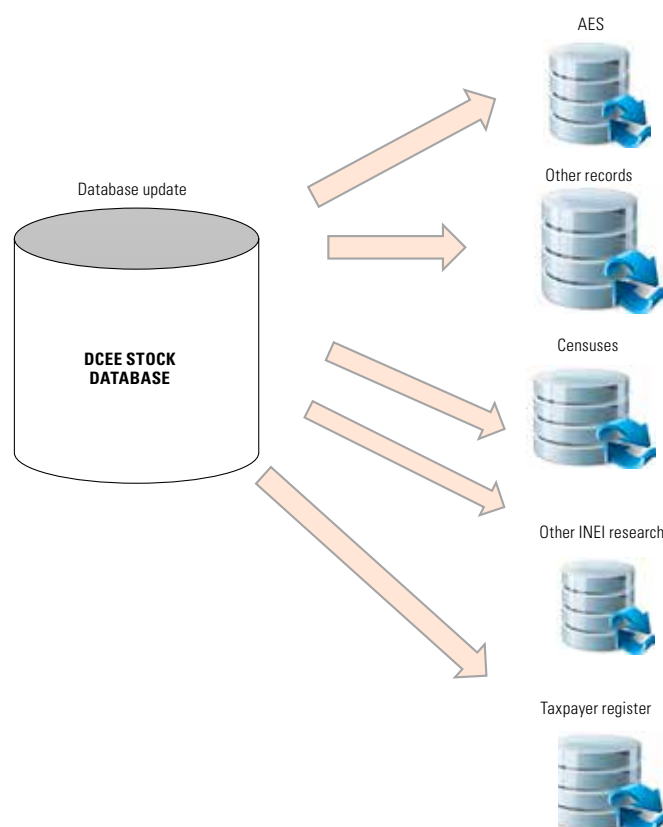
Source: National Institute of Statistics and Informatics (INEI) of Peru.

3. Database update

After the new information has been received and processed, the DCEE database is updated using different sources of information, such as economic surveys, censuses, administrative records, other research carried out by the INEI and the Taxpayer Register sent by SUNAT each year. Updates are done quarterly (see diagram A5.3).

Diagram A5.3

Database update



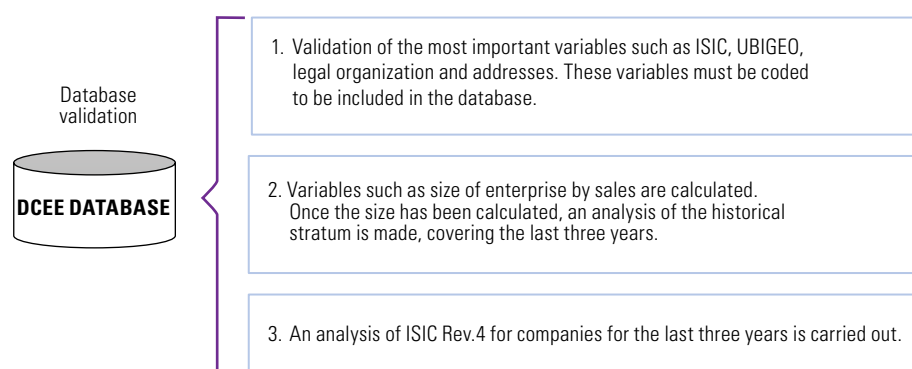
Source: National Institute of Statistics and Informatics (INEI) of Peru.

4. Database validation

The validation process verifies that enterprises have a valid RUC number. The code must be 11 digits long and must start with the number 1 or 2. Only active statistical units are also considered, while natural persons without a business, marital partnerships without a business, undivided estates without a business and units in the process of liquidation are excluded.

Then, variables such as ISIC, UBIGEO, legal organization and addresses, among the most important ones, are validated. However, the information provided by SUNAT, which contains more than 8 million records, is stored in a database managed by the OTIN server. Of the total, only 2.4 billion records are used for DCEE purposes (see diagram A5.4).

Diagram A5.4
Database validation



Source: National Institute of Statistics and Informatics (INEI) of Peru.

4.1 Validation and update by means of economic surveys (Annual Economic Survey (EEA), National Employment Survey (ENE), innovation, etc.).

This validation is performed using selected samples from the DCEE. Data on address, ISIC, company manager's name, telephone, e-mail, UBIGEO, etc., are validated by the staff of each survey by telephone, e-mail or field inspections during the information collection period.

5. Consolidated database

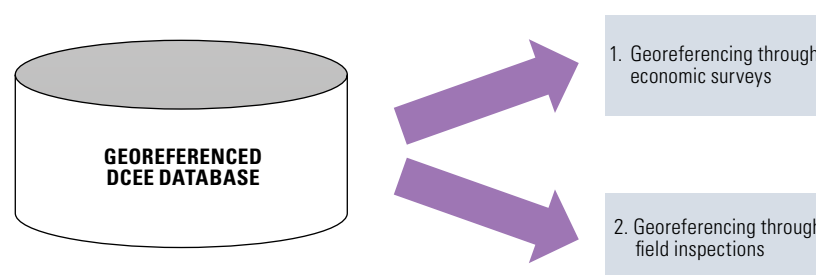
Once the information has been validated, a consolidated and consistent database is obtained, which is used in the preparation of publications, such as those on business structure and business demographics.

Likewise, this updated DCEE is delivered quarterly to the team in charge of georeferencing enterprises.

6. Georeferencing

Georeferencing is done in two ways, as shown in diagram A5.5.

Diagram A5.5
Directory georeferencing



Source: National Institute of Statistics and Informatics (INEI) of Peru.

6.1 Georeferencing through economic surveys

(a) Through the Annual Economic Survey

This is done annually by the reporting entity using an online form. This process georeferences nearly 10,000 business, of which more than 6,000 are large enterprises. In this way, most large enterprises are georeferenced.

(b) Through other economic surveys (National Business Survey, innovation, among others).

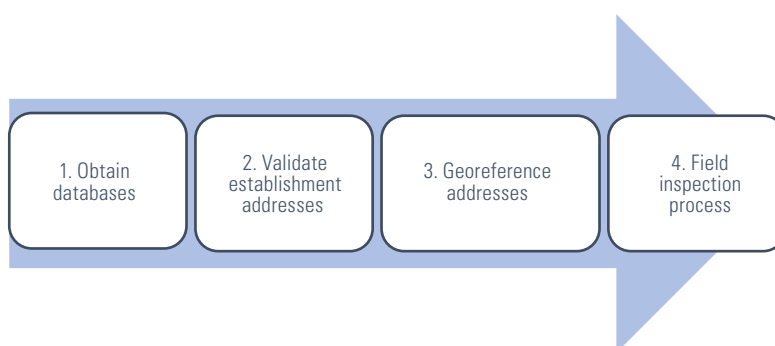
Georeferencing is done by the surveyor in the field.

6.2 Georeferencing through field inspections

Georeferencing through field inspections is carried out as shown in diagram A5.6.

Diagram A5.6

Field georeferencing process



Source: National Institute of Statistics and Informatics (INEI) of Peru.

7. Obtain databases

7.1 DCEE databases

This consists of obtaining the DCEE database, current to the last quarter.

7.2 Cartographic data

Cartographic information on zones, blocks and roads, as well as on municipal addresses, is obtained from other projects that have already been georeferenced and have the database of the Executive Directorate of Cartography and Geography of the INEI. Cartographic databases in shapefile or DWG format are also requested and sought from public or private entities, in order to assist the georeferencing work.

7.3 Validate establishment addresses

The establishments by district are extracted from the directory using an Excel spreadsheet detailing the information for each enterprise, such as RUC number, business name, type of establishment, address and other relevant information that helps characterize businesses.

The validation process begins with the road name obtained from the cartographic database of the Executive Directorate of Cartography and Geography of the INEI, where different details related to the street or road are listed, including category, name of category, name of road, alternate name, district and other data, which help in standardising the relevant addresses obtained from the SUNAT database, which is submitted with multiple errors or different characteristics.

Diagram A5.7
Cartographic information

Information received		Information validated	
Description	Road name	Description	Road name
Street	A. Aramburu	Street	Andres Aramburu

Source: National Institute of Statistics and Informatics (INEI) of Peru.

Establishments are located using the name of a road (avenida, calle, jirón, alameda, etc.) and a door number (number, block, lot, apartment, suite, floor, block, etc.) and corroborated using the RUC number and the address published by SUNAT on its web page.

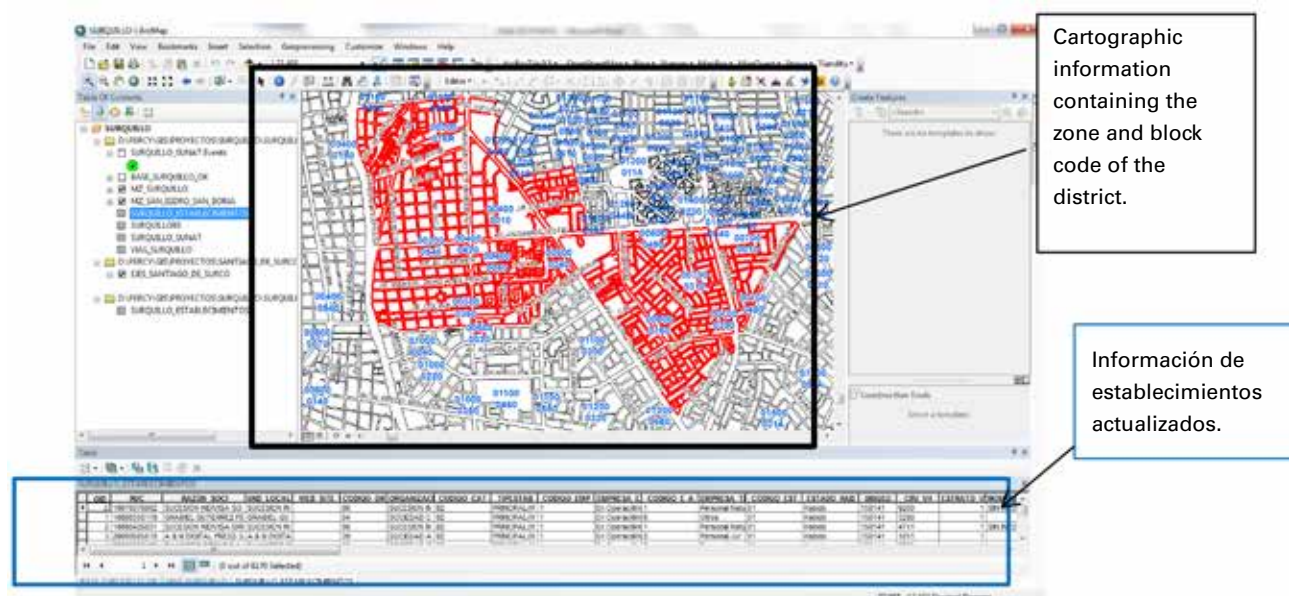
7.4 Georeference addresses

(a) Automatic process

The addresses of establishments that already have coordinates are located using geospatial information provided by the Executive Directorate of Cartography and Geography of the INEI and the Excel file of updated and harmonised establishments.

Image A5.1

Cartographic information with zone and block codes

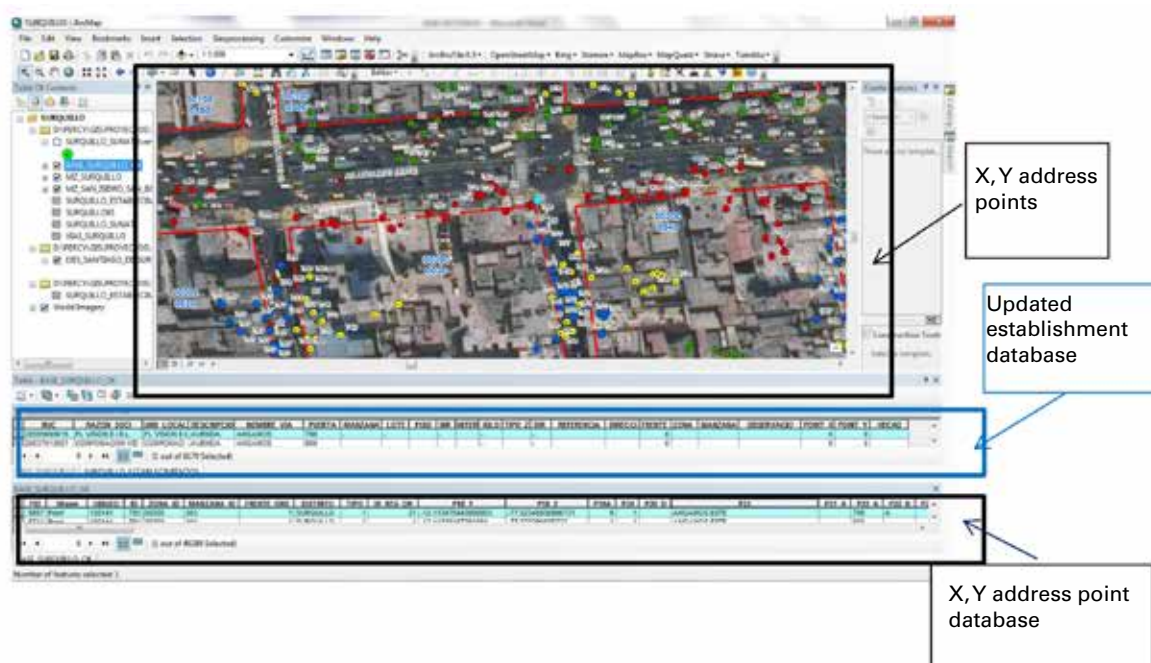


Source: National Institute of Statistics and Informatics (INEI) of Peru.

With the cartographic information for addresses (X and Y points) provided, subject to existing information, an intersection of equal data is made between the DCEE updated establishment database and the spatial information previously obtained from the Executive Directorate of Cartography and Geography of the INEI.

Image A5.2

Cartographic information of establishments by X and Y coordinates

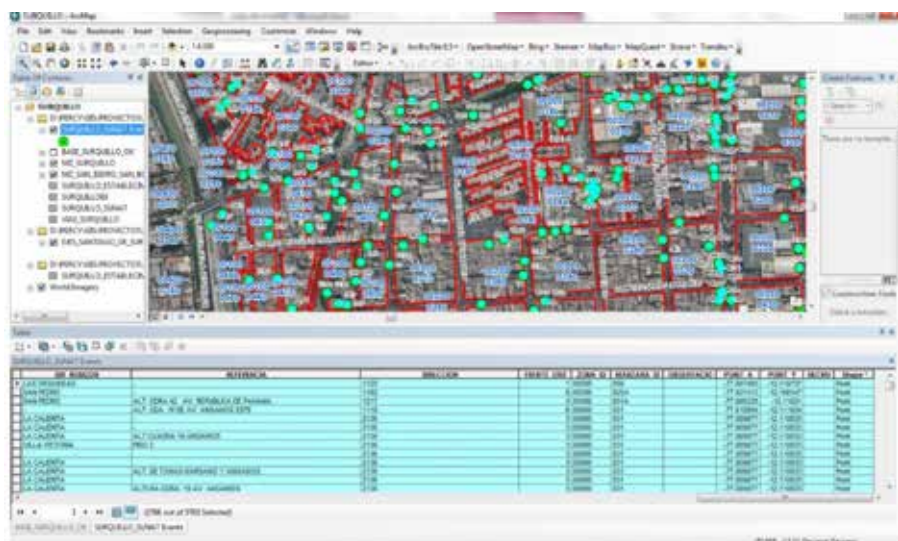


Source: National Institute of Statistics and Informatics (INEI) of Peru.

Based on the intersection of matching addresses, codes for the census area, census tract, census block and X and Y coordinates are assigned to the DCEE establishment database.

Image A5.3

Cartographic information by census area, census tract, census block, and X and Y coordinates



Source: National Institute of Statistics and Informatics (INEI) of Peru.

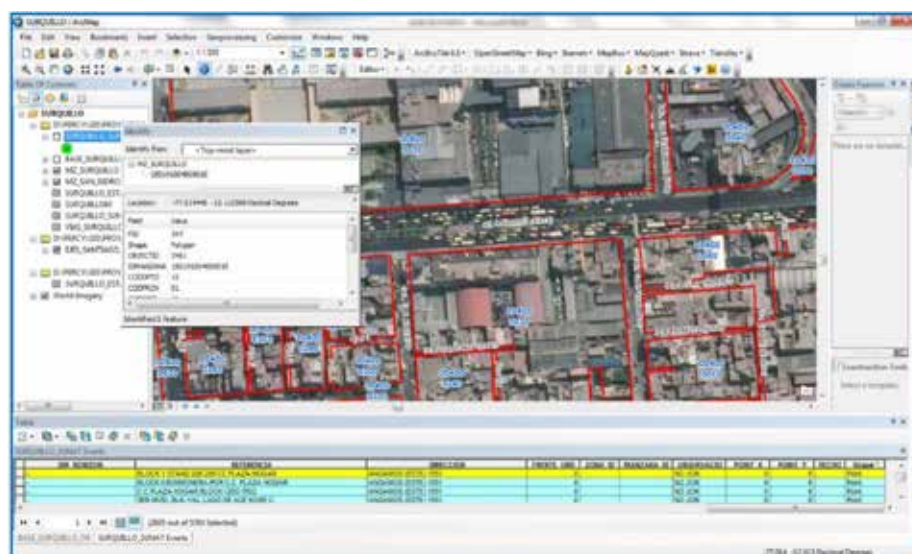
(b) Inspection process

When the address fields do not match, the codes for census area, census tract, census block and X and Y coordinates codes are filled in. The cartographic database of cities of the Household Targeting System (SISFOH),

which contains the census tract and census block codes, is used for their location. In the case of addresses where the street name or number is not specified (only block, lot or reference number), the address is located with the aid of address websites, such as Google Maps, Geodir Maps, street guides, etc.

Image A5.4

Cartographic information of establishments using the city cartography database of the Household Targeting System (SISFOH)

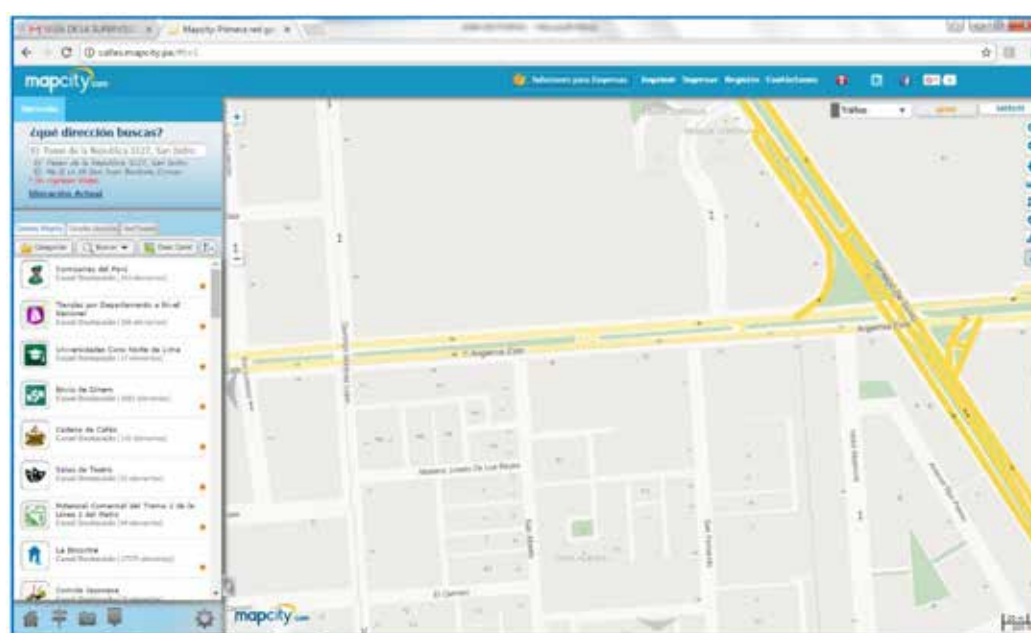


Source: National Institute of Statistics and Informatics (INEI) of Peru.

Image A5.5 shows the use of Mapcity for addresses with lot and block numbers.

Image A5.5

Use of Mapcity for addresses with lot and block numbers



Source: National Institute of Statistics and Informatics (INEI) of Peru.

7.5 Field inspection process

(a) Printing of work assignments

Once the DCEE has been georeferenced, headers are standardized via the following steps:

- The fields of interest are transferred from the georeferenced database according to the specifications for field inspections to the businesses.
- The first columns contain information on the census tract, census block and the name of the personnel in charge of verifying the georeferenced addresses in the DCEE.
- For the field verification of each company, the economic activity column is filled in with the ISIC Rev.4 codes.
- An order of assignment to field personnel is detailed, who will obtain their respective assignments from the area to be verified.
- Finally, the PDFs of the Lima and department statistics and informatics offices (ODEI) PDFs are generated using Access with a view to their printing and physical transfer to cartography, where the geographic map and instructions for distribution to field verification personnel in Lima and provinces are attached.

(b) Registration of information from field inspections

Verification of assignment. This will depend on the number of businesses to be verified, which on average should be 25 each. The assignments are received and checked to ensure that they have arrived complete and are properly filled out.

Image A5.6

Field supervision record

SEDE	DIRECCIÓN_EJECUTIVA	DEPARTAMENTO	PROVINCIA	DISTRITO	VERIFICADOR	N_FICHAS
LIMA	OTAJ	LIMA	LIMA	PUEBLO LIBRE	CARMEN ELISA ALE DUEÑAS	4
LIMA	OTAJ	LIMA	LIMA	PUEBLO LIBRE	JULIA ELENA AYALA CUYA	3
LIMA	DNCE	LIMA	LIMA	PUEBLO LIBRE	FELIX RUIZ FLORES	1
LIMA	DNCE	LIMA	LIMA	PUEBLO LIBRE	HECTOR SAAVEDRA MORA	1
LIMA	DNCE	LIMA	LIMA	PUEBLO LIBRE	MARIA I. VALLADARES ALCANTARA	2
LIMA	DNCE	LIMA	LIMA	PUEBLO LIBRE	LEONEL VILCHEZ SANCHEZ	2
LIMA	DNCE	LIMA	LIMA	PUEBLO LIBRE	SIMEON QUISPE MUÑOZ	4
LIMA	DNCE	LIMA	LIMA	PUEBLO LIBRE	WALTER ROMERO CASTILLO	3
LIMA	DNCE	LIMA	LIMA	PUEBLO LIBRE	SILVIA ANGULO VERA	1
LIMA	DNCE	LIMA	LIMA	PUEBLO LIBRE	VICTOR LEU GUTIERREZ	1
LIMA	DNCE	LIMA	LIMA	PUEBLO LIBRE	EVA AGUADO MOLINA	1
LIMA	DNCE	LIMA	LIMA	PUEBLO LIBRE	LUIS ARIAS MANRIQUE	3
LIMA	DNCE	LIMA	LIMA	PUEBLO LIBRE	MAXIMO ASURZA OLACHEA	1
LIMA	DNCE	LIMA	LIMA	PUEBLO LIBRE	MARIO BELLIDO CASTILLO	3
LIMA	DNCE	LIMA	LIMA	PUEBLO LIBRE	MARLENY BLANCAS DE LA CRUZ	2
LIMA	DNCE	LIMA	LIMA	PUEBLO LIBRE	JUAN ABAD ROMERO	6
LIMA	DNCE	LIMA	LIMA	PUEBLO LIBRE	CARMEN ASSANTE DIAZ	3

Source: National Institute of Statistics and Informatics (INEI) of Peru.

Data entry. The information is recorded in an Excel file, with the header provided by the person responsible for updating the DCEE. The information collected is entered simultaneously.

Image A5.7
Digitization record

RUC	RAZON_SOC	NOMB_COMER	ACTIV_ECON
20544437063	INMOBILIARIA & CONSTRUCTORA JOSANED S.R.L.		ACTIVIDADES INMOBILIARIAS REALIZADAS A CAMBIO DE U
20557466992	CARDIO-LAC PRESTACI#N DE SERVICIOS MEDICO DE ALTA COMPLEJIDAD E.I.R.L.		ACTIVIDADES DE M#DICOS Y ODONT#LOGOS
20562792881	EMPRESAS & NEGOCIOS CORP E.I.R.L.		CONSULTOR#A DE INFORM#TICA Y DE GESTI#N DE INSTALA
20600518870	KORY WASI ENGINEERS SAC		CONSTRUCCI#N DE EDIFICIOS
20600609981	T-REGISTRO S.A.C.		ACTIVIDADES JUR#DICAS
20600757467	CENTRO DE CONCILIACION, ARBITRAJE Y ESTUDIO	CCAES VEREDICTUM	ACTIVIDADES JUR#DICAS
20601256011	CONTROLE S.A.C. - CONTROLE		ACTIVIDADES DE CONSULTOR#A DE GESTI#N
20509383368	GRUPO MOLINERO VILLA EL SALVADOR S.A.C.		OTRAS ACTIVIDADES DE SERVICIOS PERSONALES N.C.P.
20510879920	CLAUDI SERVICIOS E INVERSIONES GENERALES S.A.C.		OTRAS ACTIVIDADES DE SERVICIOS PERSONALES N.C.P.
20535642321	365 CONSULTORIA Y COBRANZAS S.A.C. - 365 COBRANZAS S.A.C		ACTIVIDADES DE CONSULTOR#A DE GESTI#N
20477894861	CONSORCIO Y SUMINISTROS LOGISTICOS NERALM	CONSULNER SAC	OTRAS ACTIVIDADES DE SERVICIOS PERSONALES N.C.P.
20508698659	AMFARCO SOCIEDAD ANONIMA CERRADA		ACTIVIDADES DE CONSULTOR#A DE GESTI#N
20512503897	CORPORACION CIRILO ANTONIO MAURO ESPERIL	COPORACION CAMERV SAC	VENTAS DE PARTES, PIEZAS Y ACCESORIOS PARA VEH#CULO
20478197331	AUTO SANCHEZ ASESORES SOCIEDAD ANONIMA CERRADA - AUTOSAN S.A.C.		OTRAS ACTIVIDADES DE SERVICIOS PERSONALES N.C.P.
20392969773	TALLER TORETTO SOCIEDAD ANONIMA CERRADA		MANTENIMIENTO Y REPARACI#N DE VEH#CULOS AUTOMO
20532366268	IMPORTACIONES EXPORTACIONES COMERCIAL JACKELYN EMPRESA INDIVIDUAL DE RESP		VENTA AL POR MAYOR NO ESPECIALIZADA
20545492963	IMPORTACIONES LJM SOCIEDAD ANONIMA CERRADA - IMPORTACIONES LJM S.A.C.		VENTA AL POR MENOR DE PRENDAS DE VESTIR, CALZADO Y
20545922305	GESTOCAR SOCIEDAD ANONIMA CERRADA-GESTOCAR S.A.C.		OTRAS ACTIVIDADES DE SERVICIOS DE APOYO A LAS EMPRE
20552670494	GER MEDICAL ESTUDIOS Y SERVICIOS MEDICOS ESPECIALIZADOS E.I.R.L.		ACTIVIDADES DE M#DICOS Y ODONT#LOGOS
20254302041	ASOC PARA EL DESARRO INTEG DE LA COMUNID	ADINCO	ACTIVIDADES DE OTRAS ASOCIACIONES N.C.P.

Source: National Institute of Statistics and Informatics (INEI) of Peru.

Likewise, codes are assigned to the variables “business status” and “registration result.” Both must be linked and related to each other.

Image A5.8
Status and visit result

Estado		Resultado de Visita	
Ubicada.....1	Otra Empresa.....5	Direcci#n Verificada.....1	
No Ubicada	Vivienda	Ubicaci#n Censal No	
Especifique.....2	Particular.....6	Corresponde.....2	
Ausente.....3	Rechazo.....7	Direcci#n Inexacta.....3	
Desocupado.....4	Liquidaci#n.....8	Otros	
	Otros.....9	Especifique.....4	

Source: National Institute of Statistics and Informatics (INEI) of Peru.

- Database validation. This step is performed for the updated addresses with the following criteria: “code_type”, “road_name”, “door”, “block”, “floor”, “suite”, “block”, “lot”, “kilometer”, “zone_type”, “zone_name” and “reference”.
- Quality control. The finished database is checked to ensure that all variables are present and the information is consistent.

Subsequently, quality control of the information obtained from the field inspections is carried out, both for Lima and for the departmental offices of statistics and informatics (ODEI). In particular, checks are performed on forms not completed with the information provided or not correctly filled out. This is done by calling the verifiers at their extensions, or by searches at SUNAT.

Finally, the verified information is assigned X and Y coordinates for its subsequent incorporation into the DCEE database server.

7.6 Improvements made in the verification and updating of the Central Directory of Enterprises and Establishments

- A mandatory DCEE update and verification is carried out in the last quarter of each year for each database obtained prior to the address validation work performed by each district and ODEI.
- Addresses in the establishment censuses cartographic database updated to 2016, as obtained from the Executive Directorate of Cartography and Geography of the INEI, is validated with the SISFOH road database. This allows a greater number of data matches with the addresses of establishments that already have coordinates.
- Website of the Commission for the Formalization of Informal Properties (COFOPRI) (GEO LLAQTA): use of the virtual geographic platform for georeferencing addresses through the field inspections process. This allows the nationwide location of addresses that are more complicated to verify and for which there is only information such as block, lot, name of population centre, human settlement, and so on.

8. Software used in the update

The software applications used for processing, updating and georeferencing are shown in table A5.1

Table A5.1

Description of the software used for updating

Software	Uses
SQL Server	Used to analyse, review, merge and update the DCEE database with inputs provided by statistical research and the SUNAT Taxpayer Register. Also used to generate tables and reports requested by the Executive Director.
SPSS	Used to process, analyse, review and generate tables for quarterly and annual publications and Infoinei requests from public and private institutions and the general public regarding enterprises in the DCEE.
Access	Used to make print templates to generate lists of businesses for field inspections.
Excel	Used to review and analyse DCEE information.
ArcGIS	The main software used for updating and georeferencing the directory of enterprises. Activities include migration of the database and other vector files, verification and analysis of geographic coverage, database cross-referencing, geolocated vectorization of businesses and preparation of thematic maps, among others.
AutoCAD	Used to visualize files in DWG format, such as available municipal maps and others, which complement the work of georeferencing and updating of the DCEE.
QGIS	Complementary software that assists updating and georeferencing activities.

Source: National Institute of Statistics and Informatics (INEI) of Peru.

9. Determination of variables

9.1 Determination of business size

The range of sales provided by SUNAT under a cooperation agreement is used to determine the size of businesses. In other cases, sales are obtained from statistical research carried out by the INEI.

Another variable used to determine the size of a business is the range of employees.

It should be noted that some imputations are also made, mainly for micro-enterprises, since they are not registered or do not declare sales, based on special arrangements with the tax authority (see table A5.2).

Table A5.2

Determination of the business segment

Business segment	Tax Units (UIT)
Microenterprise	Up to 150 UIT (up to 630,000)
Small enterprise	More than 150 and up to 1,700 UIT (more than 630,000 and up to 7,140,000)
Medium-sized enterprise	More than 1,700 and up to 2,300 UIT (more than 7,140,000 and up to 9,660,000)
Large enterprise	More than 2,300 UIT (more than 9,660,000)

Source: National Institute of Statistics and Informatics (INEI) of Peru.

Note: UIT (2019) = 4,200 (soles).

9.2 Determination of geographic area

The geographic location (UBIGEO) code approved by Resolution No. 149–2001-INEI is used to determine political-administrative and geographic areas. This is mandatory for agencies belonging to the NSS.

9.3 Determination of economic activity

The determination of economic activity is based on the International Standard Industrial Classification of All Economic Activities (ISIC), Rev.4, which was adopted in Peru by resolution No. 024–2010-INEI in January 2010. For presentation purposes, they are grouped into main categories of economic activity.

The SUNAT register uses ISIC Rev.3 to classify economic activities, while the DCEE uses ISIC Rev.4. A forced match table developed by the INEI was used to make the ISIC change (see table A5.3). SUNAT has been gradually transitioning to ISIC Rev.4 since July 2015.

Table A5.3

Determination of legal organization

Legal organization
Natural person
Stock corporation
Civil company
Limited liability company
Limited liability sole proprietorship
Associations
Other ^a

Source: National Institute of Statistics and Informatics (INEI) of Peru.

^a Includes limited partnerships, limited joint-stock companies, general partnerships, foundations, cooperatives and others.

Annex A6

Brazil: the experience of the Coordination Office for Services and Trade with structural business surveys

The GSBPM was adopted as a methodological model in the project to modernize the production of economic statistics at the Brazilian Institute of Geography and Statistics (IBGE). In 2018 and 2019, the Coordination Office for Services and Trade implemented that model in structural business surveys:

- Annual Survey of Trade
- Annual Survey of Services
- Annual Survey of Industry - Product
- Annual Survey of Industry - Enterprise
- Annual Survey of Construction Industry

The process, which involved managers and technicians of the surveys used in the stages and subprocesses, was intended to analyse and describe the processes responsible for generating official statistics.

This process listed all the activities associated with the construction and dissemination of the final products through various channels. At the IBGE, the main channels are:

- Its website (www.ibge.gov.br),¹⁴ where the public can find the description, metadata and final results of all the Institute's surveys. Publications with an analysis of the results can be consulted at the Institute's online library (official publications and technical notes) or purchased at the portal's online store.

¹⁴ See [online] <https://www.ibge.gov.br/>.

- SIDRA (IBGE Automatic Retrieval System): SIDRA is the IBGE's central repository of statistical tables, where survey results tables throughout the historical series can be reproduced and customised.
- Assistance channels: telephone or e-mail assistance for queries on IBGE surveys or publications, as well as to request customized results tables. Each request is logged and routed to the dissemination team, which evaluates the feasibility and issues a final decision. This channel is also used for orders concentrated on the Fala.BR platform (Office of the President of the Republic of Brazil, 2011).¹⁵ The most frequent are requests for additional information on surveys under the Access to Information Act. Finally, requests for access to microdata for use in the Restricted Access Room are also registered in the IBGE assistance channels for evaluation by the dissemination team.

In the case of regularly produced statistical data, this phase occurs with each iteration. As there is an annual periodicity, the statistical products of Structural Business Surveys are produced regularly (annually). In other words, the "Dissemination" stage is recommended once a year for each of the five surveys of the Coordination Office for Services and Trade.

In updating its output systems, the IBGE seeks to use best practices in statistics production and to bring its products into line with internationally recognized standards. In keeping with the international trend towards increasing use of metadata systems in the role of guiding statistical production, the implementation of a metadata system for the institution has garnered special attention. A single, centralized system was developed to store and manage all statistical metadata in an integrated manner. Apart from concepts and methodology, it covers the description of work processes and information on production and product quality.

This system is expected to improve the quality of the statistical documentation provided to data users by means of a structured basis for organizing and disseminating information on the content and structure of the statistical information produced, based on international standards and models and reflecting best practices in the modernization of statistical production.

The IBGE Statistical Metadata System uses the Data Documentation Initiative (DDI) standard as its basis for documentation. This international standard allows the documentation of all aspects of a statistical operation, be they surveys, censuses or administrative records.

The generation of dissemination products creates the necessary products to meet the needs of the users of each survey. The Coordination Office for Services and Trade opted for a flexible model that takes account of the specificities of the main types of users of structural business surveys. Thus, in addition to direct consultation of results in the IBGE dissemination channels, text products were created that include both a general analysis of results and a detailed analysis of specific results.

Therefore, associated with this subprocess is the set of tasks for preparing the components of this product (preparation of tables, graphs and explanatory texts), as well as aggregating them in the form of specific products for each communication channel and editing them according to the Institute's standards. This stage also includes dialogue with the technical areas of each survey and is intended to produce analytical texts on the results.

- Survey brief: this product, which contains between four and eight pages, is produced separately for four of the five structural business surveys, namely the Annual Survey of Trade, Annual Survey of Services, Annual Survey of Industry - Enterprise and Annual Survey of Construction Industry. It includes some aggregate analysis by CNAE 2.0 section, in addition to the main highlights at the CNAE 2.0 division level (or activity groupings). Due to space restrictions, the analyses are generic and are

¹⁵ At this site, Brazilian citizens can register a request for access to information, report, praise, complaint, suggestion or application for federal government services, which includes services connected with IBGE statistical production. See Brazilian Institute of Geography and Statistics (IBGE) [online] <https://sistema.ouvidorias.gov.br/publico/Manifestacao/SelecionarTipoManifestacao.aspx>; "Sala de Acesso a Dados Restritos" [online] <https://www.ibge.gov.br/acesso-informacao/sala-de-acesso-a-dados-restritos.html>.

intended for users interested in a systematized overview of the main structural changes observed in some of the sectors covered by each survey, based on a comparison of results over a decade. The text is edited and laid out by designers, and infographics can be inserted to illustrate results. After the launch of each survey, this brief is sent by e-mail to the contact list of the reporting enterprises for that cycle (newsletter).

- Highlights and clarifications: this product is prepared for each structural business survey and is about 20 pages long. Each text presents the main short-term results in order to situate the reader and clarify the macroeconomic scenario that permeated the survey results, the main results of the survey's dissemination cycle and the most recent years, the main structural changes over the past 10 years and the main regional structural changes over the past 10 years. The text is produced by the survey dissemination team itself and has a freer format.
- Support tables for the production of releases: This product, prepared in each survey, is a set of customized results for each of the 27 IBGE state units, corresponding to the 27 units in the country's federation. Its purpose is to provide subsidies for its technicians to write releases and prepare local material to support dissemination.
- Presentation of results: this product summarizes survey results in slide format.
- Dissemination management is centralized in a specific team. This team is responsible for ensuring that all elements for dissemination are correct and for coordinating with other areas of the Institute that participate in the preparation and dissemination of results. In this way, the team organizes the information sections with journalists responsible for IBGE's institutional communication. Embargo agreements with the press and the pre-release of surveys are also included in this item.
- User support is provided year-round and includes direct assistance to users of structural business surveys. Users come in a variety of forms, from public agencies in different areas of government to academic researchers. Requests for access to confidential data are monitored by technicians responsible for preparing technical feasibility resolutions and coordinating with other areas of the Institute on operational feasibility and conditions of access to data. This subprocess ensures that queries, requests for special tabulations and access to microdata are registered and addressed within the agreed time frames. On average, the Coordination Office for Services and Trade prepared 40 special tabulations per year in 2019 and 2020. These requests included tables that could not be constructed through the institution's other official data deposit channels due to specificities in the disaggregation of economic activities, geographic coverage or reduction in some of the survey variables. These records are consulted throughout the year in order to review the survey dissemination process and identify new user needs.

Annex A7

Brazil: subnational initiatives for using tax and administrative records in statistical production

In order to identify initiatives for the use of administrative records for statistical purposes, in May and June 2020, the IBGE conducted, with support and financing from the World Bank, a review of the use of tax and administrative records in regional accounts looking for statistical production efforts, mainly involving the use of value added tax (VAT) records, by statistical agencies or finance secretariats in the 27 units of the federation.

Of the 27 units in the federation, 23 responded to the IBGE review. Of those, 14 regularly used tax records in a structured manner for statistical purposes or produced some type of information for the public from those records. Of the other nine that responded to the review but were not yet using tax records in a structured manner, it was found that two had ongoing studies to that end.

The most frequent use identified (18 units in the federation) concerned “lowest price” online consultation tools, in which local tax authorities, based on data extracted from sales invoices to end consumers, inform the public of the lowest price of a certain product in a given period. Aside from such consultation tools, other initiatives were also identified, such as:

- monitoring of developments in the civil construction sector based on sales invoices for construction materials;
- monitoring of the impact of local festivities on the tourism economy of municipalities, based on variations in the volume of tax documents issued, and
- production and availability of health indicators from invoices of drug sales.

The experience of IBGE with the use of tax records shows that the greater emphasis on the use of VAT-type tax records, in the case of Brazil, is due to the following factors:

- the electronic nature of the register;
- standardization of records and coding of operations at the national level;
- the timeliness of records, which are normally available a few days after the conclusion of transactions, and
- standardization of the classifications used to record activities and products.

In Brazil, VAT-type tax records related to merchandise use the MERCOSUR Common Nomenclature (NCM) for product coding. This is a regional classification, initially developed for use in foreign trade and based on the Harmonized System (HS). For activity coding, all establishments and enterprises are classified according to the National Classification of Economic Activities (CNAE), which, in turn, is based on ISIC Rev.4.

Annex A8

Argentina: evaluation questionnaires

Considering that this document describes the process in general and ideal terms, and that each national statistical office has its own particularities, tables A8.1 to A8.9 present the guiding questions for each dimension presented, which each national statistical office will adapt to its actual reality by deleting or adding questions or items to this tool.

Box A8.1

Questionnaire on agreements or arrangements

Considering that access to the source of administrative records of public agencies is irregular due to changes in government, as well as changes in management at an agency within the same government, the validity should be considered in terms of elapsed time, but also considering changes in management of the provider entity as in periods of government. It is to be hoped that these arrangements benefit the institutional framework, that is, that their wording takes account of the long-term need for official statistics, avoiding clauses that allow interruptions at the discretion of current officials.

1. What is the term of the agreement? Measured in years and government administrations.
2. Were technical staff of the national statistical office involved in drafting the contents?
3. Does it allow the national statistical office to plan in the medium or long term on the availability of the source?
4. Were all agreed clauses complied with?
5. Does it contain clauses on the data exchange platform?
6. Which ones were not fulfilled? If possible, state the reasons.
7. Was the agreement subsequently modified? If possible, state the reasons.
8. Was it disseminated among the technical staff of the agency? National statistical office and agency in possession of administrative records.
9. Did it entail any change in work dynamics at the agency in possession of administrative records?
10. What is the technical staff's perception of arrangements of this type? Agencies in possession of administrative records.

Evaluator's field notes.

Source: National Institute of Statistics and Censuses (INDEC) of Argentina.

Box A8.2

Questionnaire on needs

1. To what degree have the above needs been fulfilled?
2. Are the required variables available?
3. Is there an inventory of statistical operations at the national statistical office?
4. What is its geographical scope? National and subnational.
5. Is there an inventory of sources at the national statistical office?

Evaluator's field notes.

Source: National Institute of Statistics and Censuses (INDEC) of Argentina.

Box A8.3

Questionnaire on conceptual and analytical design

1. Is there an inventory of sources of administrative records?
2. Does it meet the needs expressed by users (internal, NSS, external)?
3. Is the identifier used by the national statistical office the same as that used by the agency in possession of administrative records?
4. If not the same, how was that resolved?
5. Was the statistical use of those administrative records possible?
6. Was it possible to plan for the use of a unique identifier in future deliveries?
7. Is there conceptual harmony between the design of the national statistical office and the source of the agency in possession of administrative records?
8. If not, how was that resolved?
9. How did the incoming quality control work?
10. Did it reduce the needs of the national statistical office?
11. Was it possible to move forward with the harmonization process without it?
12. Is there a working environment between the teams of the national statistical office and the agency in possession of administrative records that allows for medium- or long-term planning?
13. Were there changes in the frequency of production and dissemination?

Evaluator's field notes.

Source: National Institute of Statistics and Censuses (INDEC) of Argentina.

Box A8.4

Questionnaire on capture of administrative records

1. Did the chosen storage cope with the volume of capture?
2. Was the capture from suppliers homogeneous in terms of the technology used or did it vary?
3. Is the data capture design versatile, considering the differences between providers and media?
4. Was it possible to integrate these different forms of uptake into the system of economic administrative records?
5. Was the inventory of administrative records and suppliers sufficient?
6. Were the data received structured or unstructured?
7. If unstructured, how was their processing addressed?
8. Did the design and construction allow for them?
9. Did the database linkage occur as planned?
10. Was there a single identifier?
11. If not, how was that resolved?
12. Does the national statistical office have the capacity to capture and process unstructured data?
13. If the capture platform required changes, was the adaptation done promptly?
14. Were the standards set out in the design versatile or rigid?
15. Does the documentation module consider source updates or changes and their impact on the system of economic administrative records?
16. Is it possible to identify changes or updates in a timeline?
17. How was the work with the classifications done?
18. Did metadata registration keep pace with data capture?
19. Was the work team sufficient in terms of quantity and quality?
20. Was training necessary?
21. Did the incorporation of administrative records result in greater involvement of other areas of the national statistical office?
22. Was the management of the system of economic administrative records based on different modules agile?

Evaluator's field notes.

Source: National Institute of Statistics and Censuses (INDEC) of Argentina.

Box A8.5

Questionnaire on integration of administrative records

1. How did the statistical conversion of administrative records according to the source media proceed? Digital, hardcopy.
2. Was it possible to make progress with the computerization of non-digital records?
3. Based on the conceptual analysis proposed in the design phase, what proportion of the administrative records was statistically usable?
4. Was the coverage of the needs identified reduced?
5. How were unusable administrative records handled?
6. Was it possible to make progress with the agencies in possession of administrative records on harmonization?
7. Which items in the conceptual analysis were most contentious?
8. Was it possible to match classifications?
9. Are there any estimates of non-registration in the area?
10. Was the creation of aggregate variables possible?
11. How were duplicates detected?
12. What percentage of duplicates was detected?
13. Are updates to the system of economic administrative records automatic and timely?
14. Did the work times vary from the usual times?
15. Was an assembly or linkage of administrative records and census or survey data possible?
16. Did the geographic coverage vary? If so, was it higher or lower?
17. Did the updates of administrative records source data lead to improvements in the timeliness of statistical products?

Evaluator's field notes.

Source: National Institute of Statistics and Censuses (INDEC) of Argentina.

Box A8.6

Questionnaire on confidentiality

One of the limitations when it comes to the availability of source data from administrative records concerns legal frameworks that establish different types of secrecy depending on the matter in question: tax-related, administrative, medical, statistical, and so on.

1. Were there conflicts of confidentiality or secrecy?
2. Could they be avoided?
3. Was it possible to access the non-anonymized microdata?
4. Did it result in reduced access to the source?
5. Were mechanisms set up to resolve this in the future?
6. Was confidentiality maintained?

Evaluator's field notes.

Source: National Institute of Statistics and Censuses (INDEC) of Argentina.

Box A8.7

Questionnaire on costs and benefits

The cost should be considered from the point of view of overall cost-benefit for the NSS and not only as a single monetary equation per statistical operation.

The cost-benefit is understood as a greater and better statistical offering even as the contributions of sources to the national statistical office diminish or remain unchanged.

Access to sources such as administrative records can result in the availability of a greater volume of data, both quantitatively and in terms of a wider variety of phenomena and geographical scope.

1. How many times is an economic unit consulted by different areas of the national statistical office?
2. Is there a central data repository in the national statistical office?
3. Did the incorporation of administrative records generate more statistical products?
4. Were administrative records able to substitute for respondent queries?
5. Were there any changes in the frequency of respondent queries?
6. Were new users of the national statistical office detected?
7. Are there records of unsatisfied user queries?

Evaluator's field notes.

Source: National Institute of Statistics and Censuses (INDEC) of Argentina.

Box A8.8

Questionnaire on analysis and dissemination

1. How does non-registration and under-registration impact the analysis of the products generated on this basis?
2. If there are differences, can they be identified and explained?
3. Do they cause changes in the series?
4. Are they significant?
5. Do they entail discontinuities?
6. Do you necessitate additional explanations or methodological notes?
7. Does the team have experience in analysing and interpreting products based on administrative records?
8. Economic activity in administrative tax records in some countries is self-declared, unlike in national statistical offices, where it is assigned to economic units by a team of experts. How does this difference impact the statistical products analysed?
9. What about census and survey linkages?
10. In relation to the design, to what degree was it complied with?
11. Were there deviations?
12. What were the reasons?
13. Did the publications based on administrative records require additional explanations?
14. Were there any changes in the metadata?
15. Did it have an impact on users?
16. Were additional explanations required?
17. Did it give rise to new users?
18. Did the promotion of products based on administrative records entail any changes in strategy?
19. Is there a promotion strategy based on the source and its overall advantages, or only for specific products?
20. How did the press react to products sourced from official administrative records?
21. Were there any misgivings or demands regarding confidentiality?

Evaluator's field notes.

Source: National Institute of Statistics and Censuses (INDEC) of Argentina.

Box A8.9**Questionnaire on technical assistance and registration**

1. Were agencies in possession of administrative records able to make any progress with the proposed changes to the administrative records?
2. Was the metadata set transferred from the national statistical office?
3. Was it possible to hold training sessions, meetings or gatherings to move the process of conceptual harmonization forward?
4. Is there a dynamic and collaborative relationship between the teams of the national statistical office and the agencies in possession of administrative records?
5. Are queries and comments from the national statistical office resolved in a timely manner?
6. Are there any aspects of the agreement that should be modified, added to, excluded, etc.?

Evaluator's field notes.

Source: National Institute of Statistics and Censuses (INDEC) of Argentina.

Glossary

Accessibility: set of conditions under which users can obtain statistical or geographic information.

Source: National Institute of Statistics and Geography (INEGI), *Glosario del SNIEG 2021*, Aguascalientes, 2021.

Administrative datum: the value of a variable corresponding to an act, fact or event that is part of an administrative record.

Source: F. Segui, "Guía de la herramienta para la evaluación de la calidad de registros administrativos a ser usados con fines estadísticos", Aguascalientes, National Institute of Statistics and Geography (INEGI), 2012.

Administrative data source: public, private or mixed public/private agency or body responsible for the administrative register and its offices where the register data collection and maintenance processes are carried out.

Source: F. Segui, "Guía de la herramienta para la evaluación de la calidad de registros administrativos a ser usados con fines estadísticos", Aguascalientes, National Institute of Statistics and Geography (INEGI), 2012.

Administrative register: set of data that were generated for operational purposes or as part of the functions of a public or private institution on a type of object, subject, action, fact or event, and systematically obtained based on a specific printed, digital or other form and under a framework of functions and powers formally established in legal or regulatory instruments.

Source: National Institute of Statistics and Geography (INEGI), *Norma Técnica del Proceso de Producción de Información Estadística y Geográfica para el Instituto Nacional de Estadística y Geografía*, Aguascalientes, 2020.

Administrative register manager: administrative unit or units responsible for the management, capture, maintenance and updating of administrative register data.

Source: F. Segui, "Guía de la herramienta para la evaluación de la calidad de registros administrativos a ser usados con fines estadísticos", Aguascalientes, National Institute of Statistics and Geography (INEGI), 2012.

Administrative regulation: administrative procedures and formalities through which governments collect information and intervene in individual economic decisions.

Source: F. Segui, "Guía de la herramienta para la evaluación de la calidad de registros administrativos a ser usados con fines estadísticos", Aguascalientes, National Institute of Statistics and Geography (INEGI), 2012.

Availability: information security attribute that consists of the information remaining accessible for use when required by individuals, public servants or authorized processes.

Source: National Institute of Statistics and Geography (INEGI), *Política para la Seguridad de la Información Estadística y Geográfica del Instituto Nacional de Estadística y Geografía*, Aguascalientes, 2020.

Basic register: a statistical register of great importance for the whole system of records. It must define important object types and sets of important objects or standardized populations; it also contains links to objects in other basic registers.

Source: A. Wallgren and B. Wallgren, *Estadísticas basadas en registros: aprovechamiento estadístico de datos administrativos*, Aguascalientes, National Institute of Statistics and Geography (INEGI), 2012.

Big data: The term "Big Data" is used to describe data sets of increasing volume, velocity and variety; the three V's. Sources described as 'Big Data' are often largely unstructured, meaning that they have no pre-defined data model and/or do not fit well into conventional relational databases.

Source: Economic Commission for Europe (UNECE), "In-depth review of big data. Prepared by the temporary Task Team on Big Data and the Secretariat", Geneva, 2014.

Category: a set subject to quantification and characterization.

Source: National Institute of Statistics and Geography (INEGI), *Presentación de datos Estadísticos en Cuadros y Gráficas: actualización 2017*, Aguascalientes, 2017.

Cell: intersection of a column and a row.

Source: National Institute of Statistics and Geography (INEGI), *Presentación de datos Estadísticos en Cuadros y Gráficas: actualización 2017*, Aguascalientes, 2017.

Class: each of the nominal modes or numerical intervals admitted by a variable.

Source: National Institute of Statistics and Geography (INEGI), *Presentación de datos Estadísticos en Cuadros y Gráficas: actualización 2017*, Aguascalientes, 2017.

Classification: a set of homogeneous, exhaustive and mutually exclusive observations that can be assigned to one or more variables to be measured in the collection or presentation of data. A statistical classification is one that is developed for the collection and presentation of systematically collected numerical data (i.e., statistics). The classifications are divided into two:

- (i) Adopted classifications: these are classifications taken from the international reference without any adaptation to the particular application context.
- (ii) Adapted classifications: these are classifications that are adapted to the economic and social structure of the country, with greater importance given to phenomena that are more relevant and specific to the national reality.

Source: United Nations, “UN Glossary of Classification Terms” [online] https://unstats.un.org/unsd/classifications/bestpractices/glossary_short.pdf.

Classificatory disaggregation: level of detail in the listing of a classification.

Source: National Institute of Statistics and Geography (INEGI), *Presentación de datos Estadísticos en Cuadros y Gráficas: actualización 2017*, Aguascalientes, 2017.

Classification variable: a type of variable used to classify the units of the population of interest according to certain classification criteria.

Source: F. Segui, “Guía de la herramienta para la evaluación de la calidad de registros administrativos a ser usados con fines estadísticos,” Aguascalientes, National Institute of Statistics and Geography (INEGI), 2012.

Class interval: a segment on a measurement scale conventionally bounded by two distinct points on that scale.

Source: National Institute of Statistics and Geography (INEGI), *Presentación de datos estadísticos en cuadros y gráficas*, Aguascalientes, 2011.

Coding: procedure for assigning numeric or alphanumeric identifiers to concepts in an established order.

Source: National Institute of Statistics and Geography (INEGI), *Glosario de Estadística Básica*, Aguascalientes, 2012.

Coherence: refers to the degree to which the concepts used, the methodologies applied, and the results produced by the operation are logically connected.

Source: Organisation for Economic Co-operation and Development (OECD), “Glossary of Statistical Terms” [online] <https://stats.oecd.org/glossary/index.htm>.

Collection instrument: a form, either printed or electronic, designed to record the data to be obtained from observation units in a project to generate basic statistics.

Source: National Institute of Statistics and Geography (INEGI), *Glosario de Estadística Básica*, Aguascalientes, 2012.

Comparability: the degree to which definitions and classifications of common concepts and data from different sources, points in time or geographical units are equivalent.

Source: National Institute of Statistics and Geography (INEGI), *Glosario del SNIEG 2021*, Aguascalientes, 2021.

Confidentiality: information security attribute that indicates that information is only disclosed to authorized individuals or processes.

Source: National Institute of Statistics and Geography (INEGI), *Política para la Seguridad de la Información Estadística y Geográfica del Instituto Nacional de Estadística y Geografía*, Aguascalientes, 2020.

Congruence: a measure of data quality relating to the degree to which data can be combined in different sets and for different purposes with statistical information from other sources.

Source: Prepared by the authors on the basis of National Institute of Statistics and Geography (INEGI), *Proceso estándar para el aprovechamiento de registros administrativos*, Aguascalientes, 2012.

Consistency: soundness or stability of data in terms of their consistency with other periods and with related information.

Source: National Institute of Statistics and Geography (INEGI), *Glosario de Estadística Básica*, Aguascalientes, 2012.

Crossover of variables: combination of each of the classes in a classification with each of the classes in another, with respect to the variables concerned.

Source: National Institute of Statistics and Geography (INEGI), *Presentación de datos Estadísticos en Cuadros y Gráficas: actualización 2017*, Aguascalientes, 2017.

Databases: a set of related data that is systematically structured and stored and can be quickly consulted according to a desired choice of characteristics.

Source: National Institute of Statistics and Geography (INEGI), *Procesamiento de la información*, Aguascalientes, 2011.

Data editing: procedure to detect and correct data that are blank or do not comply with certain information consistency rules.
Source: F. Segui, "Guía de la herramienta para la evaluación de la calidad de registros administrativos a ser usados con fines estadísticos", Aguascalientes, National Institute of Statistics and Geography (INEGI), 2012.

Data imputation: statistical procedure to assign values to missing data or to substitute invalid or inconsistent response values. The purpose is to replace missing, erroneous or inconsistent values using auxiliary variables by means of standardized statistical procedures.

Source: F. Segui, "Guía de la herramienta para la evaluación de la calidad de registros administrativos a ser usados con fines estadísticos", Aguascalientes, National Institute of Statistics and Geography (INEGI), 2012.

Data set: organized data. They include structural metadata linking and grouping such data in the same thematic unit.

Source: National Institute of Statistics and Geography (INEGI), *Norma Técnica del Proceso de Producción de Información Estadística y Geográfica para el Instituto Nacional de Estadística y Geografía*, Aguascalientes, 2020.

Data transfer: The sending of data from the administrative register to the primary user by the administrative source.

Source: F. Segui, "Guía de la herramienta para la evaluación de la calidad de registros administrativos a ser usados con fines estadísticos", Aguascalientes, National Institute of Statistics and Geography (INEGI), 2012.

Data verification: a detailed review by the analyst, who relies on the validation criteria established in the conceptual design, depending on the subject matter addressed in the questionnaire. Its objective is to identify inconsistencies immediately after the information has been collected and to correct them through direct follow-up consultations, if afforded the possibility of returning to the source of information. In the case of administrative records, follow-up consultation refers to the source.

Source: National Institute of Statistics and Geography (INEGI), *Glosario de Estadística Básica*, Aguascalientes, 2012.

Datum: the value assumed by a statistical variable or a geographic spatial object in an observation unit.

Source: National Institute of Statistics and Geography (INEGI), *Glosario del SNIEG 2021*, Aguascalientes, 2021.

Economic unit: an observation unit about which economic information is requested and published. Depending on the activity, it may be an agricultural unit or a non-agricultural unit or establishment, the latter being a single establishment, parent company or branch, according to its type of organization; or fixed or semi-fixed, depending on its type of installation.

Source: National Institute of Statistics and Geography (INEGI), *Glosario del SNIEG 2021*, Aguascalientes, 2021.

End-user: the person who ultimately receives the statistical product. This is the person who consumes the statistical information contained in the final publication on the subject of the study based on the administrative register. In some cases, it may be a user in the institution responsible for the administrative register and in other cases it may be an external user.

Source: F. Segui, "Guía de la herramienta para la evaluación de la calidad de registros administrativos a ser usados con fines estadísticos", Aguascalientes, National Institute of Statistics and Geography (INEGI), 2012.

Electronic questionnaire: a document prepared by means of computer programmes, with questions organized in a sequential and structured manner based on a specific information need.

Source: National Institute of Statistics and Geography (INEGI), *Proceso estándar para encuestas por muestreo*, Aguascalientes, 2010.

Geographic coverage: territory to which data collection refers in a statistical project.

Source: National Institute of Statistics and Geography (INEGI), *Glosario de Estadística Básica*, Aguascalientes, 2012.

Geographic information: a set of data that has a geometric or spatial component, which describes the location of objects in space and the spatial relationships between them. Geographic information is also understood as the product of georeferencing thematic databases with geographic attributes.

Source: National Administrative Department of Statistics (DANE)/Instituto Geográfico Agustín Codazzi (IGAC), "Consolidación de la política nacional de información geográfica y la infraestructura colombiana de datos espaciales - ICDE", *Documento Conpes*, No. 3585, 2009.

Geographic reference: a territorial area to which a data set corresponds.

Source: National Institute of Statistics and Geography (INEGI), *Presentación de datos Estadísticos en Cuadros y Gráficas: actualización 2017*, Aguascalientes, 2017.

Graph: illustration that represents statistical data by means of points, lines and figures associated with measurement scales.

Source: National Institute of Statistics and Geography (INEGI), *Glosario de Estadística Básica*, Aguascalientes, 2012.

Identification variable: a type of variable that allows the unique identification of each unit or case in the administrative register.

Source: F. Segui, "Guía de la herramienta para la evaluación de la calidad de registros administrativos a ser usados con fines estadísticos", Aguascalientes, National Institute of Statistics and Geography (INEGI), 2012.

Indicator: measurement that relates one or more concepts through the application of a methodology on phenomena of interest, enabling their analysis and serving as a basis for the establishment of objectives and goals, as well as for their follow-up, in terms of their magnitude, distribution or behaviour in time and space.

Source: National Institute of Statistics and Geography (INEGI), *Reglas para la integración, difusión y administración del Catálogo Nacional de Indicadores*, Aguascalientes, 2015.

Information management: the development, implementation and monitoring of plans, policies, programmes and procedures that secure, control, protect and enhance the value of information throughout its life cycles.

Source: International Data Management Association (DAMA International), "Resumen Webinar: la importancia del gobierno del dato en el framework de DAMA," Madrid, 27 May 2020 [online] <https://anjanadata.com/resumen-la-importancia-del-gobierno-del-dato-en-el-framework-de-dama/>.

Legal entity or entities: subjects with rights and obligations. They are usually created by a group of people who come together for a specific lawful purpose.

Source: National Institute of Statistics and Geography (INEGI), *Glosario del SNIEG 2021*, Aguascalientes, 2021.

Legal framework: a set of laws, regulations, policies and standards that provide the legal basis for the records that public administration agencies and organizations keep as part of their functions.

Source: National Institute of Statistics and Geography (INEGI), *Glosario de Estadística Básica*, Aguascalientes, 2012.

Metadata: structured data describing the characteristics of the content, collection, processing, quality, condition, access and distribution of statistical or geographic information.

Source: National Institute of Statistics and Geography (INEGI), *Norma Técnica del Proceso de Producción de Información Estadística y Geográfica para el Instituto Nacional de Estadística y Geografía*, Aguascalientes, 2020.

Medium on which the record is made: a type of physical instrument used to contain forms; there are two types: printed and electronic.

Source: National Institute of Statistics and Geography (INEGI), *Captación en registros administrativos*, Aguascalientes, 2010.

Microdata: the data on the characteristics associated with the observation units that are consolidated in a database.

Source: National Administrative Department of Statistics (DANE), *Norma técnica de la calidad del proceso estadístico: requisitos de calidad para la generación de estadísticas*, 2020 [online] <https://www.dane.gov.co/files/sen/normatividad/NTC-Proceso-Estadistico-PE-1000-2020.pdf>.

Natural person or persons: an individual or individuals with the capacity to incur obligations and exercise rights.

Source: National Institute of Statistics and Geography (INEGI), *Glosario del SNIEG 2021*, Aguascalientes, 2021.

Nesting: presentation of a classification in each of the classes of another.

Source: National Institute of Statistics and Geography (INEGI), *Presentación de datos Estadísticos en Cuadros y Gráficas: actualización 2017*, Aguascalientes, 2017.

Observation unit: a single element from which data are obtained for statistical purposes on the set to which it belongs.

Source: National Institute of Statistics and Geography (INEGI), *Glosario del SNIEG 2021*, Aguascalientes, 2021.

Primary data: data obtained from the system's sources on facts that are relevant to the knowledge of economic, government, public security and justice delivery, demographic and social phenomena, as well as their relationship with the environment and territorial space.

Source: National Institute of Statistics and Geography (INEGI), *Norma Técnica del Proceso de Producción de Información Estadística y Geográfica para el Instituto Nacional de Estadística y Geografía*, Aguascalientes, 2020.

Principle of veracity or quality of the records or data: The information contained in the databases must be truthful, complete, accurate, current, verifiable and understandable. The recording and dissemination of partial, incomplete, fractional or misleading data is prohibited.

Source: Congress of the Republic of Colombia, *Ley 1266 de 2008*, Bogotá, 2008.

Process: a set of activities, human resources, data and infrastructure logically related to produce a result.

Source: National Institute of Statistics and Geography (INEGI), *Norma Técnica del Proceso de Producción de Información Estadística y Geográfica para el Instituto Nacional de Estadística y Geografía*, Aguascalientes, 2020.

Processing: a series of activities by which the files containing the information collected are organized, stored and prepared, thus ensuring their congruence, in order to proceed with their use for the presentation of statistical results.

Source: National Institute of Statistics and Geography (INEGI), *Glosario de Estadística Básica*, Aguascalientes, 2012.

Property: urban or rural land containing buildings or not, which is subject to a property regime with a recognized area and physical limits, owned and administered by a single entity, whether private, collective, social or public, and endorsed by the competent authority.

Source: National Institute of Statistics and Geography (INEGI), *Glosario del SNIEG 2021*, Aguascalientes, 2021.

Quality: the degree to which a set of inherent characteristics of processes and products meets certain attributes.

Source: National Institute of Statistics and Geography (INEGI), *Principios y Buenas Prácticas para las Actividades Estadísticas y Geográficas del SNIEG*, Aguascalientes, 2015.

Register: that which contains the geographical names of spatial objects throughout the country.

Source: National Institute of Statistics and Geography (INEGI), *Glosario del SNIEG 2021*, Aguascalientes, 2021.

Relevance: refers to a qualitative assessment of the value provided by the data, which is characterized by the degree to which the data serve the purposes for which they are sought by the users. It depends on both the coverage of the required topics and the use of appropriate concepts.

Source: Organisation for Economic Co-operation and Development (OECD), *Quality Framework and Guidelines for OECD Statistical Activities* (STD/QFS(2011)1), Paris, 2012.

Reliability: the proximity of the initial estimated value to the subsequent estimated value.

Source: National Institute of Statistics and Geography (INEGI), *Glosario del SNIEG 2021*, Aguascalientes, 2021.

Spatial datum: a digital record with combinations of attribute values that make it unique and unmistakable relative to other spatial data, and which must be georeferenced and contain an attribute that relates it to time for comparison purposes. A spatial datum corresponds to an occurrence of a spatial object.

Source: National Institute of Statistics and Geography (INEGI), *Glosario del SNIEG 2021*, Aguascalientes, 2021.

Statistical data: data that are collected or generated by statistics, either in the statistical observations process, or by processing statistical data.

Source: Statistical Office of the European Union (Eurostat), "Eurostat's Concepts and Definitions Database" [online] http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=DSP_GLOSSARY_NOM_DTL_VIEW&StrNom=CODED2&StrLanguageCode=EN&IntKey=24031736&RdoSearch=CONTAIN&TxtSearch=DATA&CboTheme=&IsTer=&IntCurrentPage=6&ter_valid=0; Statistical Commission/Economic Commission for Europe (UNECE), Terminology on Statistical Metadata, United Nations, Geneva, 2000 [online] http://ec.europa.eu/eurostat/ramon/coded_files/UNECE_TERMINOLOGY_STAT_METADATA_2000_EN.pdf.

Source: entity or person who provides information on the action, event or fact being recorded.

Source: F. Seguí, "Guía de la herramienta para la evaluación de la calidad de registros administrativos a ser usados con fines estadísticos", Aguascalientes, National Institute of Statistics and Geography (INEGI), 2012.

Statistical indicator: quantitative variable whose values are susceptible to interpretation in a field of knowledge, with respect to certain reference values established theoretically or empirically.

Source: National Institute of Statistics and Geography (INEGI), *Presentación de datos Estadísticos en Cuadros y Gráficas: actualización 2017*, Aguascalientes, 2017.

Statistical information: the set of quantitative results or data obtained from statistical and statistics-related geographical activities, based on primary data obtained from the system's sources on facts that are relevant to the knowledge of economic, demographic and social phenomena, as well as their relationship with the environment and territorial space.

Source: National Institute of Statistics and Geography (INEGI), *Norma para el Aseguramiento de la Calidad de la Información del Instituto Nacional de Estadística y Geografía*, Aguascalientes, 2020.

Statistical metadata: information about the statistical data. Metadata consist of data and other documentation that describes objects in a formalized way.

Source: United Nations, "Terminology on statistical metadata", *Conference of European Statisticians Statistical Standards and Studies*, No. 53, Geneva, 2000.

Statistical operation: set of processes and activities comprising the identification of needs, design, construction, collection or compilation, processing, analysis, dissemination and evaluation, leading to the production of statistical information on a subject of national or territorial interest.

Source: National Administrative Department of Statistics (DANE), *Norma técnica de la calidad del proceso estadístico: requisitos de calidad para la generación de estadísticas*, 2020 [online] <https://www.dane.gov.co/files/sen/normatividad/NTC-Proceso-Estadistico-PE-1000-2020.pdf>.

Statistical or geographic project: a set of organized activities carried out to produce, integrate, analyse and disseminate statistical or geographic information that enables the quantification and characterization of a study universe or specific aspect. Source: National Institute of Statistics and Geography (INEGI), *Reglas para la Prestación del Servicio Público de Información Estadística y Geográfica*, Aguascalientes, 2012.

Statistical or geographic register: a structured, complete and exhaustive list of the units of a population whose maintenance, evaluation of its quality and updating is carried out for statistical or geographic purposes. Source: National Institute of Statistics and Geography (INEGI), *Norma Técnica del Proceso de Producción de Información Estadística y Geográfica para el Instituto Nacional de Estadística y Geografía*, Aguascalientes, 2020.

Statistical product: one or more indicators, statistical tables and microdata files prepared for use for statistical purposes, whether for decision making, policy shaping or time series analysis. Source: F. Seguí, "Guía de la herramienta para la evaluación de la calidad de registros administrativos a ser usados con fines estadísticos", Aguascalientes, National Institute of Statistics and Geography (INEGI), 2012.

Statistical quality: the fulfillment of the properties that the statistical process and product must have in order to satisfy the information needs of the users.

Statistical table: matrix ordering of quantitative values and their conceptual description. Source: National Institute of Statistics and Geography (INEGI), *Presentación de datos Estadísticos en Cuadros y Gráficas: actualización 2017*, Aguascalientes, 2017.

Study population: the set of observation units, human beings, establishments, or other types of elements to be studied in a particular location in time and space. In a probabilistic scheme, in order to calculate the sample size, it is necessary to have an estimate of its variance, the coefficient of variation and the effect of the design, all of which is obtained from previous similar surveys or from a pilot study. Source: National Institute of Statistics and Geography (INEGI), *Glosario de Estadística Básica*, Aguascalientes, 2012.

System of administrative records: set of administrative records, as well as their metadata, methodologies or specific specifications used in their generation, which have been captured by the owner of the administrative register and which the national statistical office, ensuring their quality, has integrated into a system for their use. Source: National Institute of Statistics and Geography (INEGI).

Timeliness: the length of time between the occurrence of the phenomenon being studied and the dissemination of the statistics, in such a way that they are useful for decision-making. Source: European Commission, *European Commission Guide to Statistics in Development Cooperation*, Luxembourg, 2013 [online] <http://ec.europa.eu/eurostat/documents/3859598/5926561/KS-RA-13-002-ES.PDF/c0174f3d-c08c-4f26-aa80-bf5ee6678eb3>.

Timeliness: the time between the actual delivery of the data and the date established in the publication schedule. Source: National Administrative Department of Statistics (DANE), *Norma técnica de la calidad del proceso estadístico: requisitos de calidad para la generación de estadísticas*, 2020 [online] <https://www.dane.gov.co/files/sen/normatividad/NTC-Proceso-Estadístico-PE-1000-2020.pdf>.

Time reference: a point in time, date or period to which data correspond. Source: National Institute of Statistics and Geography (INEGI), *Presentación de datos Estadísticos en Cuadros y Gráficas: actualización 2017*, Aguascalientes, 2017.

Time series: a set of statistical data referring to different points in time or intervals of a period and to the same indicator, presented in chronological order. Source: National Institute of Statistics and Geography (INEGI), *Presentación de datos Estadísticos en Cuadros y Gráficas: actualización 2017*, Aguascalientes, 2017.

Use of administrative records: method of generating statistical information from records on existing facts or elements used by government units as part of their functions. Source: National Institute of Statistics and Geography (INEGI), *Norma Técnica para la Elaboración de Metadatos para Proyectos de Generación de Información de Estadística Básica y de los Componentes Estadísticos Derivados de Proyectos Geográficos*, Aguascalientes, 2015.

Validation: a set of activities to identify, in the information captured, the data that meet the requirements of logical and arithmetic consistency, completeness and integrity, in order to apply, to those that do not meet them, a solution under specific criteria that ensure the elimination of inconsistencies without affecting the original valid data. Source: National Institute of Statistics and Geography (INEGI), *Glosario de Estadística Básica*, Aguascalientes, 2012.

Validation criteria: set of rules of a conceptual nature that serve as a basis for the identification and solution of problems that arise in statistical data.

Source: National Institute of Statistics and Geography (INEGI), *Glosario de Estadística Básica*, Aguascalientes, 2012.

Value: each of the nominal numbers or modalities that a variable can admit.

Source: National Institute of Statistics and Geography (INEGI), *Presentación de datos Estadísticos en Cuadros y Gráficas: actualización 2017*, Aguascalientes, 2017.

Variable: characteristic of statistical units that can assume a set of values, which can be quantitative or qualitative.

Source: National Administrative Department of Statistics (DANE), *Norma técnica de la calidad del proceso estadístico: requisitos de calidad para la generación de estadísticas*, 2020 [online] <https://www.dane.gov.co/files/sen/normatividad/NTC-Proceso-Estadistico-PE-1000-2020.pdf>.

Veracity: the degree of approximation of calculations or estimates to the exact values of the statistical or geographic information they are intended to represent.

Source: National Institute of Statistics and Geography (INEGI), *Glosario del SNIEG 2021*, Aguascalientes, 2021.

Web scraping: the process of collecting information automatically from the World Wide Web, based on tools (called scrapers, Internet robots, crawlers, spiders and so on) that browse, extract content from websites and store extracted data in local databases for further processing.

Source: G. Barcaroli and D. Summa, "Using Internet as a data source for official statistics: a comparative analysis of web scraping technologies" [online] https://ec.europa.eu/eurostat/cros/system/files/Barcaroli-etal_WebScraping_Final_unblinded.pdf.



Economic Commission for Latin America and the Caribbean (ECLAC)
Comisión Económica para América Latina y el Caribe (CEPAL)
www.eclac.org



LC/CEA.11/18