

Catalogue no. 12-539-X
ISSN 1708-6256

Statistics Canada Quality Guidelines

Sixth Edition – December 2019

Release date: December 4, 2019



Statistics
Canada

Statistique
Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at STATCAN.infostats-infostats.STATCAN@canada.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "[Standards of service to the public](#)."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

© Her Majesty the Queen in Right of Canada as represented by the Minister of Industry, 2019

All rights reserved. Use of this publication is governed by the Statistics Canada [Open Licence Agreement](#).

An [HTML version](#) is also available.

Cette publication est aussi disponible en français.

Table of contents

Preface	5
Abbreviations	6
Important definitions	6
Executive summary	7
Introduction	8
Statistical information and credibility of data producers.....	8
Quality management in the Canadian statistical system	8
Purpose and scope of the guidelines.....	8
Guidelines for ensuring data quality	9
Relevance.....	10
Definition, importance and objectives.....	10
Guidelines for ensuring data relevance.....	10
Indicators of quality for relevance	10
Timeliness and punctuality.....	10
Definition, importance and objectives.....	10
Guidelines for ensuring data timeliness	11
Indicators of quality for timeliness	11
Accuracy and reliability	11
Definition, importance and objectives.....	11
Guidelines for ensuring data accuracy.....	12
Indicators of quality for accuracy.....	12
Accessibility and clarity	13
Definition, importance and objectives.....	13
Guidelines for ensuring data accessibility.....	13
Indicators of quality for accessibility.....	13
Interpretability.....	14
Definition, importance and objectives.....	14
Guidelines for ensuring data interpretability	14
Indicators of quality for interpretability.....	14
Coherence and comparability	14
Definition, importance and objectives.....	14
Guidelines for ensuring data coherence	14
Indicators of quality for coherence.....	15

Process quality guidelines	16
Purpose and scope of the process quality guidelines	16
Quality and phases of the statistical business process	16
1. Specify needs	18
2. Design	20
3. Build	24
4. Data acquisition	30
5. Profile and prepare data.....	34
6. Integrate, estimate and compile	38
7. Analysis	41
8. Disseminate	45
9. Evaluate	47
Bibliography	48

Preface

Led by Statistics Canada, Canadians and their government have a powerful national statistical system (NSS) they can count on to access quality statistical information. The credibility of the system and the confidence in the information it produces can only be established if that information meets the country's needs, represents the environment in which we live and which it attempts to describe, all while making sure to protect citizens' privacy. In other words, the information must be relevant, of high quality and obtained ethically.

Quality is therefore a cultural element that is fundamental in fulfilling the mandate of Canada's statistical system, which is to produce information. It also has a characteristic that is important to understand: it automatically deteriorates in the absence of proactive measures. This is why the methods and tools used to maintain the production of quality data must evolve as the environment around us evolves.

Such evolution has resulted in rapid changes to data acquisition methods, as well as growing awareness of the privacy concerns that this new environment brings. The Canadian statistical system is under increasing pressure to justify these activities in the context of ethics and privacy. Therefore, it is necessary that the NSS adhere to a transparent and coherent approach to addressing such issues as they pertain to the statistical business process.

In light of these realities it must face, the Canadian statistical system, under Statistics Canada's leadership, has long provided guidance on the principles to apply in the various statistical business processes. To this end, the experiences and conclusions which constitute "best practices" have been compiled into a set of quality guidelines. The first edition of these guidelines was published in 1985; it was subsequently revised four times, in 1987, 1998, 2003 and 2009. This edition aligns with the need to update the *Quality Guidelines* and has three objectives: a) to provide all other data producers in the Canadian statistical system with a reference document; b) to adapt to the new reality of administrative data by covering the main statistical business processes; and c) to facilitate compliance with current quality assurance methods.

It is important to understand that the guidelines presented in this document do not all apply systematically to each data production process. Therefore, their relevance and importance must be carefully considered based on the particular requirements and constraints of each program and each of their phases.

The guidelines in this document are not intended to replace the expertise and judgment of the staff responsible for producing data. Moreover, all activities of every production process must show a concern for quality. All staff involved in statistical activities are responsible for ensuring that quality has high priority when designing and implementing statistical methods and procedures.

I would like to thank and congratulate the many experts who contributed to the preparation of this Sixth Edition of the *Quality Guidelines*, as well as the Quality Secretariat, which led and coordinated this effort.



Anil Arora

Chief Statistician

Abbreviations

ADI	Administrative Data Inventory
EDCDS	Economic Disclosure Control and Dissemination System
GSBPM	Generic Statistical Business Process Model
ICMIC	International Cooperation and Methodology Innovation Centre
IMDB	Integrated Metadatabase
NSO	National Statistical Office
NSS	National Statistical System
OECD	Organisation for Economic Co-operation and Development
PUMF	Public use microdata file
QAF	Quality Assurance Framework
UN	United Nations
UNECE	United Nations Economic Commission for Europe

Important definitions

Concept	Definition
Statistical business process	Any activity to acquire, collect or manipulate information to produce statistical data.
Crowdsourcing	The use of specialized networks or the general public to acquire data at geographical levels that are inaccessible because of an insufficient number of researchers or because they cannot be in more than one place at the same time.
Web Mapping	A technique for extracting content from websites to transform it into useful data in a completely different context.
Data profiling	A process by which available data in an existing source (e.g., database, administrative file, web, etc.) are examined to collect statistics (mean, median, frequency, variation, etc.) and information on its characteristics (structure, content, format, classifications, etc.).
Confidentiality	Protection against disclosure of identifiable personal information about an individual, a business or an organization.
Policy	Official guidelines that impose specific responsibilities on departments. The Policy is binding.
Directive	Official instruction that requires departments to take (or avoid) a specific action. The Directive is binding.
Standard	A set of operational or technical measures, procedures or practices for government-wide use. The Standard is binding.
Guideline	A document providing guidance, advice or explanations to managers or functional specialists. The Guideline is a strongly suggested recommendation.
Canadian national statistical system	In practice, all federal and provincial departments and agencies that produce and publish statistical data. It has no legal existence in that section 3 of Canada's <i>Statistics Act</i> refers only to Statistics Canada.
Privacy	The right to be left alone, to be free from surveillance and to preserve the anonymity of personal information. In this document, privacy refers to information about individual persons.
Need	A gap in statistical products that must be filled.

Executive summary

The evolution in the world of statistics and data in recent years has created a need for transformation in both the way data are acquired and processed in the production of official statistics. To maintain a high quality level in statistical products and processes, it has become essential to look into existing methods and processes to evaluate their relevance, and modify them to preserve it. In fact, this new and fast-paced data revolution has required the development of new techniques and methods in order to sustain the ability to make informed decisions based on these data. Therefore, the new version of the Quality Guidelines recommends new guiding principles and best practices on topics such as the use of alternate data sources to traditional surveys, as well as data integration.

Given that traditional collection methods and data processing for probabilistic surveys are still supported, they are well covered in the new Quality Guidelines. With the addition of guiding principles and best practices on alternate data, as well as a stronger emphasis on ethical practices, privacy and proportionality, the sixth edition provides a better picture of today's reality. As an example, instead of referring to data collection, the broader topic of data acquisition is discussed.

Statistics Canada has also taken advantage of this opportunity to update the format in which the information is presented by adopting a structure similar to that of the Generic Statistical Business Process Model (GSBPM). This new way of displaying the information guides the reader more easily through each and every phase and sub-process of the statistical business process.

Introduction

Statistical information and credibility of data producers

Statistical information is essential for any organized human society to function. A lack of quality data would seriously jeopardize decision-making processes, the allocation of resources and the ability of governments, businesses, institutions, and the general public to understand the country's social and economic reality. A NSS plays a crucial role in producing and publishing statistical information.

The credibility of an organization that produces official statistics depends on a number of factors, the most important being the production of quality statistical data, cost-effectiveness, protection of personal information, confidentiality, transparency and a competent, ethical and motivated workforce with great expertise in statistical methods.

The dimensions of quality, particularly relevance, are of utmost importance for a statistical organization. If a statistical organization were unable to produce quality data, both data users and suppliers would soon lose confidence in it, making its mission more difficult to achieve. Consequently, a set of rigorous mechanisms to manage all quality matters is vital in a NSS.

Quality management in the Canadian statistical system

In the Canadian statistical system, the [Quality Assurance Framework \(QAF\)](#) outlines the strategies and mechanisms in place to facilitate and ensure effective quality management in all statistical programs and corporate initiatives. It includes a description of the management structure, policies and guidelines, consultation mechanisms, as well as the project delivery and management approach based on the scientific approach.

The effectiveness of this framework depends not on one single mechanism or process, but on the combined effect of many interdependent measures based on employees' professional interests and motivation, which reinforce each other as means to meet client needs. These measures emphasize employees' professionalism and reflect a concern for data quality. An important feature of this strategy is the synergy resulting from the many players in the programs operating within a framework of coherent processes and consistent messages.

These quality guidelines, which fall within this framework, are a companion document that describes the best practices throughout all the "stages" of the statistical business process. They are intended for members of teams responsible for developing and implementing statistical activities.

Purpose and scope of the guidelines

The main objective of the *Quality Guidelines* is to provide a fairly comprehensive list of guiding principles and best practices to apply during a statistical business process.

The document focuses on ensuring quality through effective and appropriate design or implementation of statistical projects or programs, from the beginning to data publication and evaluation of the process. These guidelines draw on the collective knowledge and experience of many individuals who have worked in the Canadian NSS, particularly Statistics Canada.

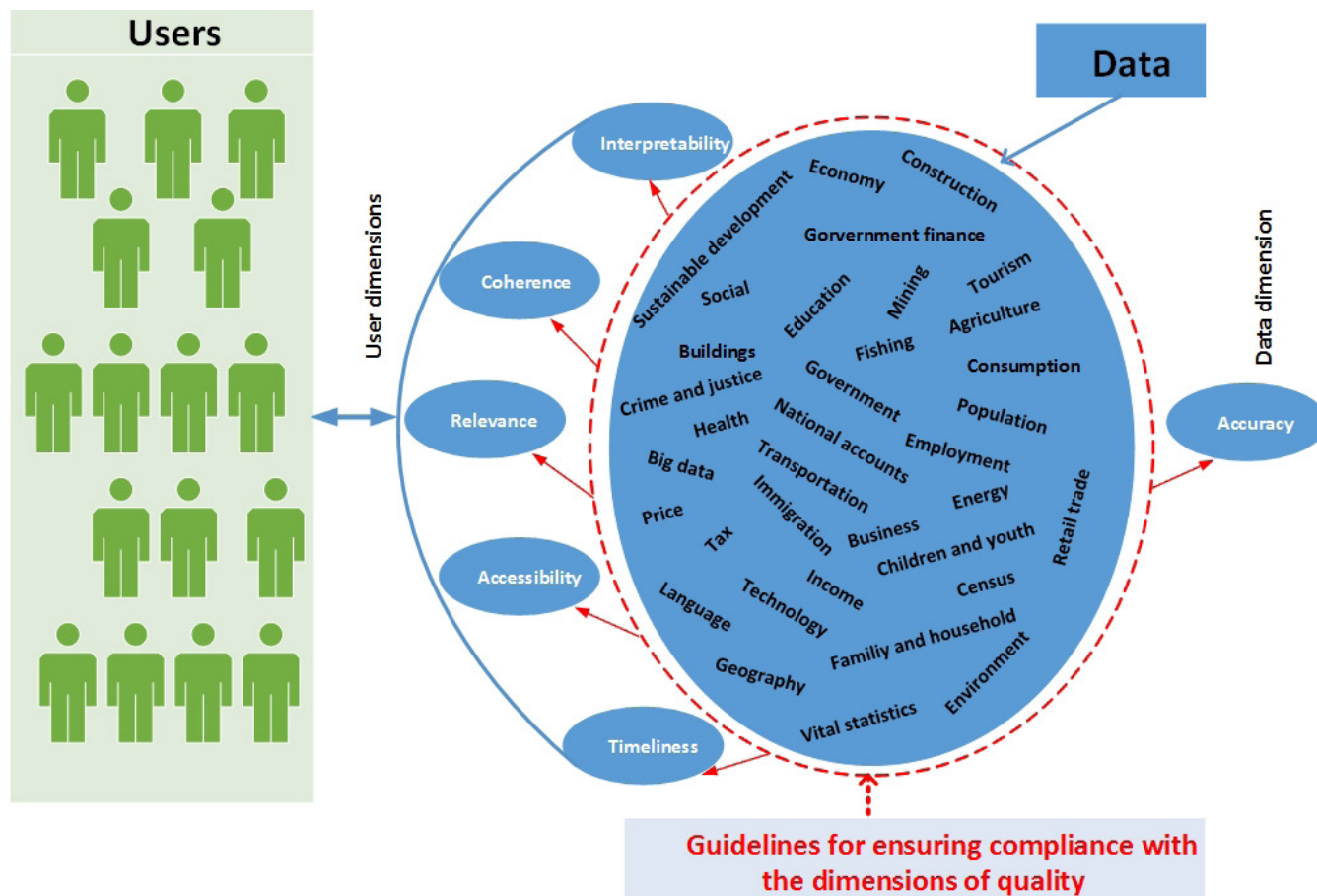
The document has two main sections: the first covers product quality and the second addresses the quality of processes in a statistical business activity. It is important to remember that the context in which a statistical business activity is prepared carries some constraints. Each new statistical activity not only aims to satisfy immediate information needs, but also adds information to a statistical database that can be used for a much wider range of purposes than initially identified. Therefore, it is important to ensure that the output from each activity can, to the extent possible, be combined and used with data on related subjects derived from other activities.

This implies a need to consider and respect the statistical standards on content or fields that have been put into place to achieve data coherence and harmony in the NSS. These standards include statistical frameworks (such as the System of National Accounts), statistical classification systems (such as for industry or geography), as well as other concepts and definitions that describe the statistical variables to measure. New statistical data are more useful when they can be used together with existing data.

Guidelines for ensuring data quality

In the Canadian statistical system, data quality is defined by six quality dimensions: relevance, accuracy, timeliness, accessibility, interpretability, and coherence. The QAF, which describes strategies to aid and ensure effective quality management in all statistical programs and organizational initiatives, identifies these dimensions in its guiding principles. These six dimensions are addressed in the QAF to describe what statistical organizations must do to ensure quality across their organization. In these guidelines, they help to assess whether the data produced are fit for use. Figure 1 illustrates the scope of the guidelines. It shows that, regardless of the type of data, ensuring compliance with the dimensions serves to ensure data quality.

Figure 1



The following sections describe how the Canadian statistical system ensures quality data by managing the six dimensions effectively. For each dimension, the first subsection presents a definition, followed by a brief explanation of the importance of the dimension (the “why”) and the objectives it addresses. The second subsection provides guidelines to guide the statistical organizations’ personnel in completing their tasks. Considering that how quality dimensions are managed directly affects the quality of the statistical products, the third subsection proposes quality indicators that not only measure the dimension, but also evaluate it against the intended objectives.

Relevance

Definition, importance and objectives

The **relevance** of statistical information refers to the degree to which it meets real user needs. Data are relevant when they relate to the issues users care about most. This dimension allows national statistical offices (NSOs) to adjust the direction of their programs, as needed. However, assessing relevance is subjective because it depends on various user needs.

The human, material and financial resources for each program are generally considered under this dimension. Therefore, the greatest challenge for any NSO is to gauge the conflicting needs of current and potential users in order to design a program that best meets their key needs, given the resource constraints. The mechanisms for ensuring relevance not only provide basic control over the other quality dimensions, but also assess user needs in relation to the other dimensions.

We must therefore gauge the relative importance of the data needs and explore different alternatives. This notion is important because it determines the optimal balance between users' need for data and respondents' need for privacy protection. Although it may be difficult to evaluate this aspect, it is essential to devote an appropriate effort to it.

Guidelines for ensuring data relevance

- ▶ Create and maintain a list of current and potential users of all the statistical organization's statistical products.
- ▶ Maintain bilateral and multilateral contact with the key users.
- ▶ Periodically evaluate the statistical programs.
- ▶ Realign program priorities, as needed, as part of a strategic planning process. This can be done by reviewing emerging information needs and the relevance of current programs.
- ▶ Consider how changes to a statistical program will affect other related programs.
- ▶ Clearly state the goals and objectives of the statistical activities.
- ▶ Determine the data uses and level of user satisfaction.
- ▶ Determine parameters to establish the optimal balance between the need for data and the impact on the privacy of Canadians, and choose the best option among those considered.

Indicators of quality for relevance

- ▶ An updated list of program data users
- ▶ The level of program data compliance with regional and international standards
- ▶ The number of online consultations of program products
- ▶ The rate of user satisfaction with program products
- ▶ The percentage of users who actually use the statistical product.

Timeliness and punctuality

Definition, importance and objectives

The **timeliness** of statistical information refers to the delay between the information reference point (or the end of the reference period) and the date on which the information becomes available. Punctuality refers to the difference between planned and actual availability.

Effectively managing this dimension ensures that the data produced and released remain relevant. The timeliness of information must be planned from the design phase of the statistical activity. Planning is often based on compromises between the precision sought, cost and data availability. Timeliness is therefore important.

Regular follow-ups should be scheduled and activity progress monitored to prevent problems and to detect poor practices, which can lead to extreme delays.

Guidelines for ensuring data timeliness

- ▶ Check if there is a legislative mandate for data release. If there is, it may reduce the flexibility of the implementation schedule and the ability to modify it.
- ▶ Plan and develop a schedule for each production phase and ensure they are carried out with regular follow-up mechanisms.
- ▶ Negotiate sharing agreements with data providers for planned and timely recurrent transfers.
- ▶ Agree with data providers on an availability calendar for their databases.
- ▶ Consider external factors when planning and developing the release schedule.
- ▶ Announce the release schedule in advance, as well as the planned date changes.
- ▶ For certain programs, adopt a preliminary data dissemination strategy, following up with revised and final figures which are more current.
- ▶ Use new technologies to shorten release times.
- ▶ Assess the feasibility of releasing preliminary data, followed by revisions.
- ▶ Inform users about timeliness constraints.
- ▶ Set a new release date for any discrepancies with the planned schedule in advance.
- ▶ Assess the importance of users' needs relative to the length of time required to meet them.

Indicators of quality for timeliness

- ▶ A start and end date for each phase of the production process
- ▶ A dissemination schedule
- ▶ A follow-up and contingency plan
- ▶ A document on the needs assessment
- ▶ The degree of punctuality according to the planned release schedule (the number of days, months or years of delay from the original release date)
- ▶ The period between the reference date and the release date
- ▶ The period between the release dates of the preliminary and final versions.

Accuracy and reliability

Definition, importance and objectives

The **accuracy** of statistical information is the degree to which the information correctly describes the phenomena it was designed to measure. Accuracy is often characterized by errors in statistical estimates and is typically broken down into bias (systematic error) and variance (random error). It may also be described in terms of the major sources of error likely to result in inaccurate data (e.g., coverage, sampling, non-response). Related to accuracy, reliability reflects the degree to which statistical information, consistently over time, correctly describes the phenomena it was designed to measure.

With this dimension, the main objective is for the data produced to reflect reality as accurately as possible. Managing accuracy requires special attention during the design, implementation and evaluation phases of the statistical activity. Therefore, the level of accuracy attained is a function of the methods and processes established to detect and control potential errors in the various phases of the statistical activity.

The acceptable level of accuracy is determined and justified based on the optimal balance between user needs and the level of protection of privacy required to satisfy these needs, budgetary, material and human resource constraints, as well as contextual factors related to the execution of the statistical activity.

Guidelines for ensuring data accuracy

- ▶ Use appropriate quality control methods and quality assurance processes to detect and control potential errors in the various phases of the production process.
- ▶ Consult with administrative data providers to validate concepts and understand data processing in their original business context (e.g., use, coverage, collection and reconciliation cycle, missing or low-quality values) and make the necessary adjustments.
- ▶ Follow up with data providers if variables and/or data appear to be missing or incorrect.
- ▶ Determine the quality of record linkage from different files (e.g., matching administrative data to statistical registers).
- ▶ Minimize and assess coverage and classification errors of the statistical units in the frame.
- ▶ Pre-test and test all versions of the data collection instruments (e.g., questionnaires, data capture platforms, imputation and estimation programs or software, transfer modules).
- ▶ Choose unbiased and minimally biased sampling procedures, such as probabilistic sampling procedures.
- ▶ Consider using non-probabilistic methods only in a valid inference framework (Beaumont 2018) or if there is no other alternative.
- ▶ Select a sample design that optimizes the accuracy of the data produced.
- ▶ Check the survey sample against the original sample design.
- ▶ Check the representativeness of the data.
- ▶ Have a collection follow-up plan to improve response rates and data quality.
- ▶ Minimize and assess processing errors (data capture, edit and imputation, file manipulation, etc.).
- ▶ Check that the calculation of weights corresponds to a known total.
- ▶ Identify and describe the main reasons for non-response (non-contact, refusal, ineligible, etc.).
- ▶ Assess the extent of errors due to non-response.
- ▶ Provide a variance estimation that takes into account the sample design and various manipulations performed on the survey data (non-response adjustment, imputation, calibration, etc.).
- ▶ Validate the models used and the underlying assumptions.
- ▶ Use other sources to compare, evaluate and validate calculated estimates.
- ▶ Use new technologies to automate procedures as much as possible, thus minimizing errors from manipulation and data inconsistency.
- ▶ Evaluate the relative importance of user needs and respondent privacy needs to determine the optimal balance.

Indicators of quality for accuracy

The first five of the following seven indicators are generally related to the accuracy of data from sample survey processes. Research is underway at Statistics Canada to develop indicators for data from other statistical business processes. This research will also verify the suitability of using some of the indicators used to measure survey data. The results of this research will be added to the next edition of these guidelines.

- ▶ Coverage rate
- ▶ Response rate
- ▶ Variance (coefficient of variation)
- ▶ Imputation rate and ratio
- ▶ Imputation variance
- ▶ Measurement error
- ▶ Precocity or closure errors (for population estimates).

Accessibility and clarity

Definition, importance and objectives

The **accessibility** and **clarity** of statistical data refers to the ease with which users can learn that the information (including metadata) exists, find it, view it and import it into their own work environment. It also refers to whether the information or medium is appropriate. For some users, the cost of information may also affect accessibility.

As part of this dimension, users should be able to determine what information they are requesting and if it meets their needs. Accessibility is closely tied to relevance, where user satisfaction with a statistical product must be measured not only in terms of the product's content, but also with respect to how it is disseminated.

In terms of dissemination, the Canadian statistical system aims to maximize usage of the information it produces while ensuring that dissemination costs do not reduce the organization's ability to collect and process data. The dissemination policies of statistical organizations and delivery systems determine most aspects of accessibility.

Guidelines for ensuring data accessibility

- ▶ Establish a system for documenting and archiving statistical data.
- ▶ Design statistical products that are well integrated with the statistical organization's cataloguing systems.
- ▶ Periodically ensure archived files are accessible to users.
- ▶ Establish procedures early in the life course of repetitive surveys to ensure that all important files are retained.
- ▶ Follow procedures for distributing restricted-use files, as needed.
- ▶ Apply a disclosure control methodology that minimizes suppressions.
- ▶ Provide multiple access and extraction methods.
- ▶ Take advantage of new data visualization technologies.
- ▶ Make data files accessible for future analytic activities.
- ▶ Ensure that the products comply with the requirements of the organization's policies and standards for accessibility, or refer to Statistics Canada's Policy on Access to Information, Policy on Microdata Access and Policy on Highlights of Publications.
- ▶ Ensure that the products comply with the organization's requirements, or refer to Statistics Canada's [Policy on Informing Users of Data Quality and Methodology](#).
- ▶ Inform users of available publications, publication frequency and main output modes.
- ▶ Organize official release sessions for results of statistical processes.
- ▶ Write and publish clear and concise press releases about the statistical product.
- ▶ Conduct user satisfaction surveys on ease of access and dissemination modes.

Indicators of quality for accessibility

- ▶ A system for archiving and storing the statistical product
- ▶ Availability of the statistical product in several formats
- ▶ A reference for the statistical product within the archiving and storage system
- ▶ The organization of official release sessions for the statistical product
- ▶ A press release on the statistical product
- ▶ User satisfaction measures for ease of access and dissemination modes.

Interpretability

Definition, importance and objectives

The **interpretability** of statistical information refers to the availability of supplementary information and metadata needed to interpret and use statistical information appropriately. This information normally covers the underlying concepts, variables and classifications used, the methodology of data collection and processing, and indications of the accuracy of the statistical information.

The Canadian statistical system is responsible for providing sufficient information for users to be able to interpret statistical information effectively. Communicating metadata, or information about information, is key to managing interpretability.

Guidelines for ensuring data interpretability

- ▶ Consult the organization's data interpretability policy, or refer to Statistics Canada's [Policy on Informing Users of Data Quality and Methodology](#) and the Directive on Documenting Statistical Metadata.
- ▶ Ensure that the organization's metadatabase is up to date. A template for completing and updating records in Statistics Canada's Integrated Metadatabase (IMDB) is available.
- ▶ Provide access to information on the uses and objectives associated with the data to ensure transparency.
- ▶ Ensure that quality indicators are available so that users are fully informed and can understand the strengths and limitations of the data.

Indicators of quality for interpretability

- ▶ A copy of the methodological documents about the data, available to users at all times
- ▶ A metadatabase that is up to date, available to users, and includes quality indicators
- ▶ An assessment report for the statistical process, available internally.

Coherence and comparability

Definition, importance and objectives

The **coherence and comparability** of statistical information refer to the degree to which it can be reliably combined and compared with other statistical information within a broad analytical framework over time. To improve coherence, it is important to use standard concepts, classifications and target populations, as well as a common methodology across surveys. Coherence does not necessarily imply full numerical consistency.

Guidelines for ensuring data coherence

- ▶ Keep concepts, definitions, classifications and methodologies up to date.
- ▶ Keep an up-to-date document of methodologies and frameworks for acquiring administrative data and collecting data from field operations.
- ▶ Use a common frame for all surveys of the same type.
- ▶ Use common wording for questions.
- ▶ Document how new classifications and concepts affect statistical series.
- ▶ Comply with established standards. Otherwise, clearly document the reasons why the standards were not used and how this affects the comparability and coherence of the product.
- ▶ When necessary, specify the differences between concepts and methods used at different times for the same variables.
- ▶ Assess and document the effects of using different methods at different times.
- ▶ Evaluate the differences between local practices, regional practices and international practices, if they exist.

Indicators of quality for coherence

- ▶ Concepts, definitions and classifications that comply with established standards
- ▶ Compliance with regional and international standards for statistical methods
- ▶ A description of the inconsistency when it cannot be fully explained by accuracy.

Process quality guidelines

Purpose and scope of the process quality guidelines

This section provides guidelines and checklists on issues to consider when establishing quality objectives for statistical activities. It focuses on how to ensure quality data production processes from the start through to evaluation, documentation and dissemination. These guidelines draw on the collective knowledge and experience of many employees of Statistics Canada and other data-producing organizations in the public sector.

The core objective of these guidelines is to provide a comprehensive list of guiding principles and best practices regarding the data production process. To better understand the scope of these guidelines, it is important to define “statistical business process” in the Canadian statistical system.

“Statistical business process” is a generic term for all activities that involve collecting and manipulating information to produce statistical data, including the following:

1. Data collected from administrative records, in which data are created or derived from records originally kept for non-statistical purposes
2. The census and sample survey, which collect data on all the population or on a sample (usually random) of the population, respectively
3. Record linkage, which consists of identifying records that are associated with the same person or entity in one or more statistical files or registers
4. Derived statistical activities, in which estimates are produced by integrating data from several different sources
5. Crowdsourcing and web mapping, growing types of data production processes that are not covered in this document.

Unlike previous versions of the *Quality Guidelines*, which focused heavily on the Census and sample surveys, this version of the guidelines focuses on the first four types of processes, the fifth type being still relatively young.

In addition, following guidelines in the statistical business process makes it possible to standardize production practices, thus guaranteeing consistent execution over time despite staff mobility; as well as providing a way to analyze, evaluate and improve those production practices.

Quality and phases of the statistical business process

In any production process, quality must be considered at all phases. This is no different in statistics. It encourages statistical organizations to model their statistical process in a way that not only operates smoothly, but also integrates and applies the quality dimensions at each phase. Since the April 1987 edition of the *Quality Guidelines*, the Canadian NSS has used a schematic diagram to represent the process. Recently, the Canadian NSS adopted the Generic Statistical Business Process Model (GSBPM) as a standard for its statistical business processes. The GSBPM is an international standard created by a joint working group of the United Nations Economic Commission for Europe (UNECE), Eurostat and the Organisation for Economic Co-operation and Development (OECD). Adopted and adapted by many national statistical institutes and international organizations, the GSBPM, now in its fifth edition, proposes a structured model for the statistical business process.

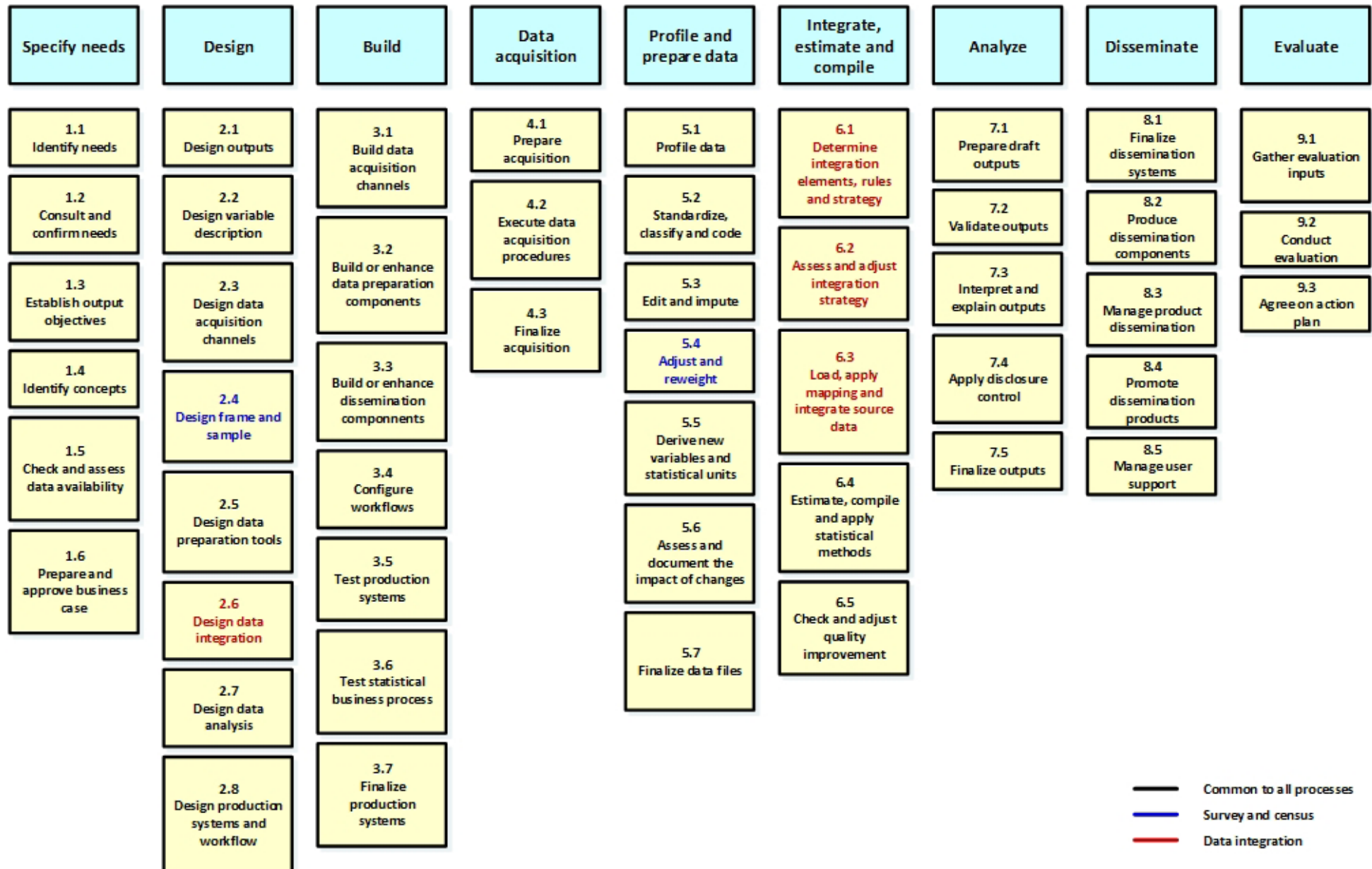
This production model has been adapted not only to reflect the current production environment in the Canadian statistical system, but also to be as inclusive as possible. This adaptation considers some distinctions between the different types of statistical processes mentioned above.

This adaptation of the GSBPM keeps the model’s original scheme of dividing the statistical process into a number of phases. This model for the quality assurance management process, defined both by aspects related to the quality dimensions, and by the phases of the statistical process, arises naturally from the ideas that quality is multidimensional, must be integrated at each phase and communicated in a transparent manner. A comprehensive approach to quality management demands that all cells of this matrix be considered.

This section contains nine sections corresponding to the nine phases that these guidelines propose for carrying out any statistical business process within the Canadian statistical system. Each section is divided into

Figure 2
Phases in the statistical business process

Adaptation of Generic Statistical Business Process Model



sub-processes with the same structure: first, they provide a “Description of this phase and its sub-processes,” then “Guidelines for ensuring quality at this phase” and finally, the “Quality dimensions and indicators at this phase.” Moreover, the guidelines illustrate exceptional cases.

1. Specify needs

Description of this phase and its sub-processes

The first phase in any statistical business process is to determine its main objectives. This phase begins when a need for new statistics emerges or when stakeholders provide feedback on current statistics.

When identifying and stating objectives, include specific requirements for the data and their uses, key quality expectations, issues related to the protection of privacy, budget constraints, business cases and expected delivery dates. Define relevant concepts, the unit of analysis and the target population during this phase to allow the intended users and potential users to determine if, and to what extent, the expected project results meet their expressed needs.

Typically, certain departments or subject-matter divisions of the statistical organization, along with other NSS stakeholders, with their expertise, are responsible for identifying needs.

The Specify needs phase includes six sub-processes:

1.1 Identify needs	1.2 Consult and confirm needs	1.3 Establish output objectives	1.4 Identify concepts	1.5 Check and assess data availability	1.6 Prepare and approve business case
------------------------------	---	---	---------------------------------	--	---

- 1.1. Identify needs:** This sub-process involves identifying and reviewing the necessary data, as well as the needs the data must satisfy. Examine the practices, particularly standards and methods, of other regional and international statistical organizations that produce similar data. Consider the needs of specific user groups (e.g., persons with disabilities, ethnic groups) as well as the sensitivity of the information needed to meet the needs at this sub-process.
- 1.2. Consult and confirm needs:** For this sub-process, consult broadly with the various stakeholders of the statistical activity. These consultations clarify and confirm stakeholder needs in detail and indicate when, how and why the data must be produced. For production processes needing several types of input data, such as macroeconomic accounts and population estimates, consultations must also include other similar and related programs in order to meet national and international legislative requirements.
- 1.3. Establish output objectives:** This sub-process focuses on identifying statistical products that will meet user needs. Production objectives are generally established by considering the balance between quality measures and the resources available to the statistical organization for the activity.
- 1.4. Identify concepts:** In this sub-process, the concepts to be measured by the statistical activity take shape. Concepts do not need to be aligned with existing statistical standards at this time. Aligning standards, as well as choosing and defining statistical concepts and variables, is carried out in sub-process 2.2.
- 1.5. Check and assess data availability:** This sub-process involves verifying the availability of data sources that can meet user needs. Before making any decisions, assess the available data to determine whether they are appropriate for statistical purposes and whether they meet the expressed needs and objectives. Evaluate the legal framework that allows data collection and the use of alternative sources to ensure rules of ethics and protection of privacy are respected.
- 1.6. Prepare and approve business case:** This sub-process involves documenting and analyzing all possible types of business processes for these statistical data and determining, based on available resources, which is best able to respond to user and client needs. Typically presented as a business case, the chosen statistical process must meet a number of pre-established standards and be approved by senior management before the chosen statistical process can be put into place.

Guidelines for ensuring quality at this phase

► All statistical processes

- Identify and analyze the information needs of internal and external users in relation to new information requests or the required environmental changes.
- Compare similar statistical operations, particularly standards and methods, used in other regional and international statistical organizations.
- Identify the needs of specific user groups, such as people with disabilities or ethnic groups.
- Examine information needs for the best short and long-term solutions.
- Consult users systematically and extensively to clarify content and generate support from project partners.
- Establish and maintain relationships with data users in all sectors to improve information relevance as well as the dissemination of products and services.
- Identify and define operational constraints, such as the reference period, costs, resources and data collection methods.
- Establish production goals together with users and key stakeholders.
- Include measurable quality dimensions in the statement of objectives.
- Consider the objectives and needs of subsequent or parallel statistical activities when determining production objectives.
- Develop measurable concepts according to the statistical information to be produced (target population, statistical unit, etc.). The concepts do not need to be aligned with existing statistical standards at this phase.
- Analyze available and accessible sector data in terms of relevance, frequency, quality, timeliness, etc.
- Determine whether record linkage is a viable option by starting to identify which datasets could be linked.
- Assess the legal framework of all possible sources concerning collection and use.

► Statistical processes involving administrative data

- Consult the organization's data inventory to check if the data are already available. Statistics Canada has a vast inventory of administrative data. Information on accessing these data can be found at Administrative Data Inventory (ADI).
- Contact other public and private organizations in the field to determine if administrative data are available for producing statistics. Statistics Canada's guidelines on using administrative data are provided in the [Policy on The Use of Administrative Data obtained under the Statistics Act](#) and the Administrative Data Handbook.
- Conduct an exploratory data evaluation before collection to ensure that the data meet user needs in terms of concepts and quality. Specifically, ensure that the acquisition of these data is necessary (that is, the other options evaluated are not adequate to meet the needs) and proportional to the intensity of the expressed needs. Statistics Canada produced The Administrative Data Evaluation Guide – Exploration and initial acquisition phases for this purpose.
- Analyze the context in which the administrative data were created (e.g., legislation, objectives, needs, etc.).
- Assess the discriminatory power of the linkage variables in the identified data to determine whether statistical information can be produced by record linkage.

► Statistical processes involving data integration

- Producing statistics by record linkage:

- Consult the organization’s policies on record linkage. At Statistics Canada, these are defined in the [Directive on Microdata Linkage](#).
- Assess whether information needs will be met by linking administrative data with survey data, considering methodological constraints.
- Identify all the data that could be linked and the linkage unit.
- Examine the availability and quality of the identifying variables that are common to all source datasets for linkage.
- Assess the discriminating power of the available variables to determine whether linkage is possible.
- Assess the adjustments required if the variables do not fully meet user needs.
- Producing National Accounts:
 - Consult the information on national and international legislative requirements regarding public finances.
 - Consult international macroeconomic accounting manuals and international compilation guides to ensure better process management.
 - Check if the data required to create the accounts are available and under which conditions they will be available, including any usage restrictions.
- Producing population estimates and projections:
 - Consult the information on national and international legislative requirements regarding population estimates.
 - Check whether the data (surveys, linkages and administrative files) needed to create population estimates are available and under which conditions they will be available, including any usage restrictions.

► **Statistical process for a new survey**

- Determine the extent to which the survey meets user needs, while striking a reasonable balance between these needs and issues related to the response burden and the protection of personal information.
- Examine the balance between existing statistics that can meet user needs, and the cost, the time required and the value added of creating a new survey.
- Prepare a business case for the business process to be approved. This analysis should include:
 - a. A description of the business process, highlighting the gaps and problems to resolve
 - b. Solutions detailing how the statistical business process will be developed to produce new or revised statistics
 - c. An assessment of costs, benefits, and external constraints.

Quality dimension and indicators at this phase

a. **Quality dimension:** Relevance

b. **Indicators of quality:**

- A description of user needs and of how users intend to use the data
- Analysis plans that include a description of tables to be released
- A business case establishing the gap between user needs and intended outcomes.

2. Design

Description of this phase and its sub-processes

The design phase of the statistical business process involves describing the design, creation, development and research activities needed to create statistical outputs, concepts, variables, methodologies, data acquisition

methods and operational processes. It includes everything needed to define or enhance the statistical products or services indicated in the statistical business process. In this phase, the metadata, which are essential for continuing the business process and for interpreting the statistical products, are defined. It is also at this stage that a careful assessment of needs allows one to compare the need for information with the sensitivity of the data required to meet these needs. It is only after a thorough analysis that this optimal balance is found, which ensures that the effort devoted to data collection is proportional to the needs expressed.

Regional, national and international standards are widely used in design activities. This not only produces comparable data, but also reduces the length and cost of the design process. Statistical organizations are therefore encouraged to reuse or adapt design elements from existing processes.

The design phase is broken down into eight sub-processes that can be done in order from left to right, or at the same time, and can be iterative. Sub-processes 2.4 and 2.6 are specific to the survey process and the data integration process, respectively. The sub-processes are:

2.1 Design outputs	2.2 Design variable description	2.3 Design data acquisition channels	2.4 Design frame and sample	2.5 Design data preparation tools	2.6 Design data integration	2.7 Design data analysis	2.8 Design production systems and workflow
------------------------------	---	--	---------------------------------------	---	---------------------------------------	------------------------------------	--

- 2.1. Design outputs:** This sub-process involves providing a detailed description of the statistical outputs to produce as well as all associated services. It also involves defining and designing the data integration strategy and method, where applicable, as well as the systems and tools for the production and the dissemination phases, the disclosure control methods and the processes governing access to confidential data.
- 2.2. Design variable descriptions:** This sub-process involves defining the statistical variables that describe the phenomena for which the source data will be collected, as well as the data that will be derived from them in sub-process 5.5 (Derive new variables and statistical units). It also involves determining how these source data will be transformed to be consistent with the concepts and conventions associated with the statistical activity in question. During this sub-process, it is strongly recommended that regional, national and international standards and classifications be followed as much as possible when defining statistical variables.
- 2.3. Design data acquisition channels:** This sub-process involves determining and describing the data collection and acquisition methods. Activities under this sub-process may vary depending on the type of statistical process, but could include computer-assisted interviews, paper questionnaires, self-enumeration, administrative data transfer interfaces, the Internet, etc. This sub-process also involves drafting any formal data acquisition agreements, such as memoranda of understanding, and confirming the legal basis for data collection.
- 2.4. Design frame and sample:** This sub-process only applies to business processes that require data collection based on sampling. It involves identifying and specifying the population of interest, developing a sample frame and identifying the most appropriate sampling criteria and methodology for the phenomena to be measured. During this sub-process, an analysis will be carried out to determine whether the target population is covered by the chosen frames.
- 2.5. Design data preparation tools:** This sub-process involves determining the most appropriate statistical methods for profiling and preparing the data. It typically involves outlining the processes for compliance verification, error detection and correction, imputation, seasonal adjustment, modelling, deflation or calibrating, validating and finalizing data.
- 2.6. Design data integration:** This sub-process applies only to processes that require multiple data sources to be integrated. It involves designing the integration strategy and method, assessing quality, and establishing constraints, identities and access requirements. For record linkage, the most appropriate method to use is determined based on the availability of certain variables of interest.

Furthermore, if there are no source data, it may be necessary to design models for statistical processes intended to produce macroeconomic accounts, population estimates or microsimulation data.

2.7. Design data analysis: This sub-process involves determining the most appropriate statistical methods to apply during the Analyze phase.

2.8. Design production systems and workflow: This sub-process involves designing the production systems and the workflow of all operations from data collection to dissemination. It is important to check whether existing systems are compatible before designing new ones. All phases of the business process must be taken into account when carrying out this sub-process to maintain the structure and to avoid duplication.

Guidelines for ensuring quality at this phase

► All statistical processes

- Provide a detailed description of the statistical data and products to be produced.
- Provide a detailed description of the systems and tools that will be used during the Disseminate phase.
- Define and describe the metadata that will accompany the published data in the various formats they can be accessed.
- Define and describe the disclosure control methods that will be used and the processes that will govern access to confidential data.
- Define the target population, including all statistical units for which information is being collected.
- Clearly state the concepts and variables of the phenomena to be measured as well as their purpose.
- Define concepts, variables, classifications, statistical units and populations using, where appropriate, the standard definitions outlined in Statistics Canada's [Policy on Standards](#).
- Use the most recent version of all variables to ensure they are relevant.
- Check the list of harmonized content standards for concordance between certain standardized international classifications and the Canadian statistical system first, unless the term is not listed.
- When choosing naming conventions, think about the difference between standards and usage. In other words, save standardized titles for terms defined in the classification systems used.
- If no official standards exist, rely on the concepts, variables and classifications used in related statistical productions or consult Statistics Canada's Standards Division.
- Define and operationalize derived variables.
- Aggregate data at a higher level to meet the particular needs of the analysis or to satisfy data confidentiality or reliability constraints.
- If possible, use the higher classes or aggregations dictated by the standard. Otherwise, opt for a common grouping strategy, then document the discrepancies between the standard and the selected classification or aggregation levels.
- Use classifications that reflect both the detailed and aggregated levels. Always explain to users how classifications fit into the higher level.
- Define and identify the most appropriate instruments for acquiring data and metadata.
- Identify and design the statistical processing and analysis method for the data. This involves outlining processes for coding, imputation, estimation, modelling, seasonal adjustment, deflation, validation and finalization of the data.
- Design, purchase or adopt production systems and establish the workflow of all operations, from data collection to dissemination.
- Check the compatibility of existing systems before designing new ones.
- Make sure all phases of the business process are taken into account to maintain the structure and to avoid duplication.

► **Statistical processes involving administrative data**

- Design, purchase or adopt, together with the data provider, a capture platform as well as a coding manual when the data exist only on paper. At Statistics Canada, the main directions on this matter are provided in the [Policy on The Use of Administrative Data obtained under the Statistics Act](#).
- Design, purchase or adopt the most appropriate method of data transmission when the data exist in electronic format. Always refer to the organization's guidelines for designing or adopting data transmission methods. At Statistics Canada, the main directions on this are in the Directive on the Transmission of Protected Information.
- Make sure there is a document that clearly defines each variable in the file.
- Maintain communication with the provider to stay informed of any changes to the file.
- Work with the provider organization's designers responsible for upgrading the administrative systems or designing new ones so that the statistical requirements will be integrated into the systems from the beginning of the project.

► **Statistical processes involving data integration**

- Design the most appropriate data integration and quality assessment strategy to ensure that the integrated dataset is appropriate for the intended use.
- Design the integration strategy based on the availability of known variables classified as potential integration identifiers.
- When designing the integration strategy, take into account the project objectives, the end use of the integrated data and the resources available.
- When selecting an integration strategy, review the methods and processes used in similar projects.
- Select internal and external validation measures to assess the quality of the integration process and the accuracy of the integrated data, respectively.
- Design adjustments to improve the overall quality of the integrated dataset.
- Design data creation models if certain source data do not exist.

► **Statistical process for a new survey**

- Define and determine the most appropriate collection instruments and modes for the survey.
- Carefully examine and assess different data collection modes.
- When designing the questionnaire, consult with the main data users to ensure that their intended use is understood.
- Design a collection process that reduces response burden and costs, provides the most accurate data possible and speeds up the reception of the data.
- Try to combine collection and capture as much as possible (e.g., by collecting data in electronic format).
- Use standardized collection tools and methods (e.g., standard screens and standardized questions) to facilitate work and minimize the risk of capture errors.
- Review the question libraries and development tools of existing questionnaires.
- Establish the sample design and select the sampling selection method and criteria most suitable for the phenomenon to be measured. Consider using any supplementary information available to make the sample design more effective.
- Assess and determine whether the target population is covered by the selected or designed sample frame.
- If no single frame can provide coverage for the target population, use a methodological approach that relies on multiple frames (combination of two or more sample frames).
- Ensure that the frame is as up to date as possible relative to the survey reference period.

- Think about using a multi-stage or indirect sampling method if there is no cost-effective frame for populating the units of interest for the survey.
- Use supplementary data from other sources to offset the coverage error of the sample frame.
- Implement procedures to detect and correct potential sample frame coverage errors.
- Use the same sample frame for surveys with the same target population to improve consistency, avoid contradictions, make it easier to combine survey estimates and reduce the cost of frame maintenance and assessment. Think about using the organization’s existing statistical registers as a frame.
- Use global positioning systems (GPS) as much as possible when creating geographic subsets in the frame.
- Save information on sampling and data collection to coordinate surveys and better manage respondent relations and response burden.

Quality dimensions and indicators at this phase

a. **Quality dimensions:** Relevance, accuracy, coherence and interpretability

b. **Indicators of quality:**

- Detailed description of the main statistical concepts, including statistical measures, population, variables, units, domains and reference period
- Accurate references for the concepts, variables and standard classifications used
- Report to check adequate coverage of the target population
- Detailed description of the instruments and methods for data collection, preparation and analysis. A detailed description of the methodology used will also be available for surveys or data integration.

3. Build

Description of this phase and its sub-processes

This phase involves building and testing the complete production operating environment to the point where it is ready for use in a “live” environment. Outputs of the Design phase guide the assembly and configuration of various components in this phase to create the operating environment. New components should only be built as an exception, created in response to gaps in the existing catalogue and should be widely reusable. For statistical outputs produced on a regular basis, this phase only takes place the first time or following a change in the methodology or technology used.

This phase is broken down into seven sub-processes that are generally done in a sequential order, but can also be done at the same time. The sub-processes are:

3.1 Build data acquisition channels	3.2 Build or enhance data preparation components	3.3 Build or enhance dissemination components	3.4 Configure workflows	3.5 Test production systems	3.6 Test statistical business process	3.7 Finalize production systems
---	--	---	-----------------------------------	---------------------------------------	---	---

3.1. Build or enhance data acquisition channels: This sub-process involves preparing or building the channels through which the data will be acquired. This must be done in accordance with the specifications created in the Design phase for the selected acquisition method. This sub-process also involves testing the content and operation of that instrument. There are many types of data acquisition channels, such as regular extraction or data transmission platforms used to collect existing sets of statistical or administrative data, or face-to-face or telephone interviews, and paper, electronic or online questionnaires for surveys or censuses.

Build collection instruments: For a sample survey-based statistical process, this sub-process involves drafting the questionnaire, creating or updating the sample frame and selecting the sample. To select a sample, create a subset of units from the frame that is representative of the target population. This requires coordination with other business processes using the same frame or frames to avoid overlap and help

distribute respondent burden. Quality assurance of the frame created or updated is also a part of this sub-process.

- 3.2. Build or enhance data preparation components:** This sub-process involves selecting the mechanisms, tools and methods for reviewing and collecting information on the data acquired, and for detecting and correcting existing and potential problems. It is therefore important to select the appropriate tools and methods for profiling, cleaning and converting the data based on the type of statistical process used. For data integration processes, this sub-process establishes and codifies the logical rules for transforming the source information obtained in sub-process 3.1 in order to be consistent with the required concepts and presentations. It also describes the activities and tools necessary for mapping transformed source data.
- 3.3. Build or enhance dissemination components:** This sub-process involves building and developing components and new services or enhancing existing ones in order to publish the results of the statistical activity. Accessibility must be taken into account when creating or enhancing dissemination components and services by providing users with multiple options to access both open data outputs and microdata.
- 3.4. Configure workflow:** This sub-process involves configuring the operations and systems workflow used throughout the statistical business process, from data collection to process evaluation. This sub-process is important because it ensures that the workflow specified in sub-process 2.8 works in practice.
- 3.5. Test production systems:** This sub-process involves the technical testing and approval of new computer programs and routines. It also confirms that existing routines from other statistical business processes are suitable for use in this case. The activities carried out in this sub-process also include testing the services and operations configured in sub-process 3.4.
- The business architecture in many NSOs, including Statistics Canada, has the advantage of being operational at all times and adaptable to different business processes. Therefore, testing production systems is one of the simplest sub-processes to carry out after some adjustments have been made.
- 3.6. Test statistical business process:** This sub-process involves conducting a pilot of the statistical business process. It typically involves a small-scale data collection (20 to 100 records or more) to test the instruments and mechanisms for collecting data, following up on processing and analyzing the collected data to ensure that the statistical business process performs as expected.
- 3.7. Finalize production systems:** This sub-process includes all activities for putting the assembled and configured tools, mechanisms and services, including those that were modified, into production so that they are ready to be used when needed. Although some business process components are finalized in sub-process 3.6, it is important to test the performance of all components in the production environment in order to ensure that they perform as expected.

Guidelines for ensuring quality at this phase

► All statistical processes

• Database coverage

- Assess the reliability and relevance of the various databases available and accessible at the planning stage before making a decision.
- Negotiate required changes with the managers of the selected databases for derived statistical activities where coverage changes may be beyond the control of the immediate manager.
- Adjust data in the selected frames or use supplementary data from other sources to compensate for the coverage error.
- In the statistical activity documentation, include descriptions of the target and survey populations, the differences between the target population and the survey population, and the description of the databases and coverage errors.

• Data capture and coding

- Consider using an automated system or automated machine learning techniques to assign codes to descriptions. Statistics Canada developed an automated system named G-Code which assigns codes to descriptions.

- Build a data capture platform or adapt the existing one to the needs of the statistical activity in order to reduce costs, speed up data acquisition and ensure quality.
 - Integrate edit rules into the electronic and online collection system to prevent invalid data from being digitized.
 - Integrate algorithms and their parameters into automated data capture systems to reduce error rates.
 - Use flexible entry and coding processes that can be adapted to make the appropriate changes if necessary from an efficiency point of view.
 - Use manual or automatic coding if pre-coding cannot be used, for example if the data acquired includes answers to open-ended questions.
 - Prepare training documents and activities for data capture and coding staff.
 - Test automated data capture systems that are based on intelligent character recognition from scanned images prior to implementation.
 - Link the metadata and the data to be captured.
 - Use statistical quality control methods to assess and improve the quality of data acquisition, capture and coding operations.
 - Ensure that the automated coding system have been tested, modified if necessary and validated and is ready to be used in the process.
- **Data profiling**
 - Based on the type of statistical process, determine the most appropriate methods, techniques and tools for data profiling.
 - Use generalized and reusable software for data profiling, although manual intervention may be necessary in some situations.
 - Whenever possible, avoid increasing the volume of controls if it has little impact on the final results.
- **Imputation strategy**
 - Include algorithms for imputing data in the system preparation component.
 - Consider using the generalized systems available in the organization when developing imputation methodologies, as they offer a variety of pre-programmed methods. Alternatively, refer to Banff, Statistics Canada's generalized edit and imputation system, which offers a variety of methods for continuous and categorical data.
 - Consult LogiPlus, a program in Microsoft Windows to create, edit and verify logic and to process and impute by decision tables.
 - Identify variables in various sources (e.g., current survey data, historical data, administrative data, paradata) that could act as auxiliary variables for imputing missing data.
 - Assess the quality and relevance of the available variables to determine which ones to use as auxiliary variables.
 - Take into account the type of characteristics required when choosing auxiliary variables and the imputation strategy in order to preserve the relationships of interest between the variables.
 - If a generalized system is used, ensure that the chosen methods and parameters have been tested, modified if necessary and validated; and that the system is ready to be used in the process.
- **Confidentiality rules**
 - Consult the organization's confidentiality resources before making technical decisions, or refer to Statistics Canada's Policy on Privacy and Confidentiality.
 - Consult the organization's disclosure control resources, or contact Statistics Canada's Disclosure Control Resource Centre.

- Use well-established generalized disclosure control software, such as G-Confid, instead of customized systems.
- Evaluate the feasibility of using corporate platforms, such as the Economic Disclosure Control and Dissemination System (EDCDS).
- **Data analysis**
 - Determine the appropriate analytical method for the data before investigating the software choices available to apply it.
 - Use organization-approved commercial or non-commercial software that is suitable for the types of analyses selected.
 - Determine whether the data must be reformatted in order to use the chosen software.
 - If necessary, consult the Data Analysis Resource Centre (DARC), which is a team of statistical consultants and researchers in Statistics Canada’s Methodology Branch.
 - If necessary, consult the Time Series Research and Analysis Centre (TSRAC), which is a team of time series specialists who ensure that time series analysis methods used at Statistics Canada are accurate and in line with recent developments.
- **Data dissemination**
 - Consult the organization’s communications and dissemination specialists for any questions on this topic.
 - If necessary, consult Statistics Canada’s policy and directive on data dissemination components and standards: a) Policy on Official Release b) Directive on the Release of Microdata Files.
- **Operations**
 - Create a multidisciplinary committee to approve operations.
 - Develop a communication plan with the team members.
 - Define and share responsibilities based on expertise, experience and specific activities or components.
 - Review the component plans to identify the inputs, outputs and dependencies for each component.
 - Create and maintain a schedule of deliverables and dependencies with the various corporate services.
 - Develop mechanisms for follow-ups and updating activities or timelines.
 - Formulate the final version of operations and the testing plan.
- **Testing**
 - Test and approve the new programs and routines developed in the previous steps.
 - Confirm that existing routines used in other statistical business processes are appropriate for this one.
 - Test the interactions between the assembled and configured services to ensure that the production solution works as a coherent set of processes, information and services.
 - Acquire data (e.g., through collection, transfer, etc.) on a small scale to test the channels, as well as the data processing and analysis, to ensure the statistical business process performs as expected.
 - If necessary after the pilot, go back to a previous phase and make adjustments to instruments, systems or components.
 - Produce documentation about the process components, including technical documentation and user manuals.

- Transfer the process components into the production environment and make sure they perform as expected.

► **Statistical processes with administrative data**

- Review the organization’s policies, directives and guidelines on acquiring administrative data, or refer to those in effect at Statistics Canada:
 1. [Policy on the Use of Administrative Data Obtained under the *Statistics Act*](#)
 2. Directive on Obtaining Administrative Data under the Statistics Act
 3. Directive on the Security of Sensitive Statistical Information
 4. Guidelines on Data Available to the Public.

► **Statistical processes involving data integration**

- Consult the organization’s guidelines on administrative data integration.
- If necessary, consult Statistics Canada’s Record Linkage Resource Centre, which provides record linkage and pre-processing services for such linkage, and G-Link, Statistics Canada’s probabilistic linkage system designed primarily to address record linkage problems when there are no unique identifiers.

► **Statistical processes involving a new survey**

• **Questionnaire design or redesign**

- Consult with Statistics Canada’s Questionnaire Design Resource Centre (QDRC) regarding plans for questionnaire development. QDRC is a team of research consultants who specialize in the design and evaluation of questionnaires. Refer to Statistics Canada’s Policy on the Development of Questionnaires to help guide the questionnaire design process.
- Maintain ongoing communication with other survey stakeholders and data users throughout the questionnaire design process to ensure a clear understanding of how the data are to be used.
- Refer to Appendix B of Statistics Canada’s Directive on Informing Survey Respondents to ensure that best practices on the minimum requirements for communicating information to respondents have been adopted.
- Ensure that the questionnaire contains only questions that are relevant to achieving the stated objectives.
- Ensure the opening questions apply to all respondents, are easy to understand and establish whether the respondent is part of the survey population.
- Use words and concepts that have the same meanings for both respondents and the questionnaire designers. For businesses, choose questions, reference periods and response categories that are compatible with the establishments’ record-keeping practices.
- Use words and terminology that encourage respondents to answer questions as accurately as possible. The questionnaire must focus on the survey topic, be as brief as possible and flow smoothly from one question to the next to facilitate recall and direct respondents to the appropriate information source.
- To the extent possible, use existing concepts and terminology, or harmonized content for consistency. When appropriate, use existing questions to ensure consistency and comparability with the results of similar surveys that have adequately measured the same concepts.
- Check that the English and French versions of the questionnaire correspond to each other, as well as versions in other languages, if applicable. Surveys are sometimes conducted in other languages, such as when they are about Indigenous peoples.
- Establish a direct link between the questionnaire and the statistical metadata system so that the metadata can be more easily captured during collection.

- Design self-completed questionnaires to be professional in appearance and easy to complete. If the questionnaire will be administered by an interviewer, make it interviewer-friendly.
 - With respect to the question layout, provide titles or headings for each section in the questionnaire. Include instructions and answer spaces that facilitate accurate answering of questions.
 - Use symbols, graphical features and spatial arrangement to attract attention and guide respondents or interviewers to the parts of the questionnaire that are to be read, and to indicate where answers should be entered.
 - Ensure that the instructions to respondents and interviewers are clear, concise and positioned where the information is needed. Provide definitions at the beginning of the questionnaire or in specific questions, as required.
 - Use boldface selectively to emphasize important items in questions. Ensure that reference periods (date or period) and response units are clear to the respondent.
 - At the end of the questionnaire, provide a space for additional comments from respondents, and include a statement of thanks for the respondent.
 - Choose among a wide range of methods to test and evaluate the questionnaire before production. For example, qualitative or quantitative studies such as focus groups discussions, cognitive interviews, pilot surveys.
- **Coverage and creation of the sample frame**
 - Consider the generalized systems available at the organization. If necessary, consult G-Sam, Statistics Canada’s generalized system that provides probabilistic sampling functionalities.
 - Use the same frame for surveys with the same target population to avoid inconsistency, make it easier to combine survey estimates, and reduce frame updating and evaluation costs.
 - Consider using the records that are already available and maintained by the organization.
 - Try to get as much information as possible on the frame in order to make the various survey phases (sampling, collection, edit, imputation and estimation) more efficient. If necessary, pair the frame with other sources to obtain more relevant information.
 - Retain and store information on sampling, rotation and data collection to coordinate surveys and better manage respondent relations and response burden. For example, record how often each unit is selected by each survey using the same frame.
 - Ensure there are current edit and validation procedures to minimize errors in the frame.
 - For area frames, use geospatial data and similar statistics to create reasonable geographic areas when developing sampling and specific to the survey needs.
 - Determine whether the survey design information can be incorporated into the analysis and, if so, how this should be done.
 - Use software that is specifically designed for survey data analysis, and accounts for weighted point estimates and variance for survey-weighted estimates.

Quality dimensions and indicators at this phase

- a. **Quality dimensions:** Relevance, accuracy, coherence, timeliness, accessibility and interpretability (all six dimensions of the survey production process).
- b. **Indicators of quality:**
 - Data acquisition mechanisms, tools and the components of the production and dissemination process are functional and ready to be used as intended.
 - A dashboard is used to record operations, systems and transformations used in the statistical business processes, from data collection to dissemination.

4. Data acquisition

Description of this phase and its sub-processes

The Acquisition phase in the statistical business process involves acquiring or facilitating the acquisition of all necessary information (data and metadata) through different channels. Once acquired, the information is loaded into an appropriate environment to prepare it for use and analysis. Data acquisition is more than a source of information; it is also the main link between statistical organizations and the general public who must be convinced to participate in survey processes. Data acquisition is also the main link to other statistical organizations, important partners for statistical or administrative databases, registers or other non-statistical databases.

The Acquisition phase includes three sub-processes, which are generally sequential. The sub-processes are:

4.1 Prepare acquisition	4.2 Execute data acquisition procedures	4.3 Finalize acquisition
----------------------------	--	-----------------------------

4.1. Prepare acquisition: Activities under this sub-process are conducted to ensure that staff, mechanisms and tools are fully prepared to acquire data and metadata according to the established strategy. When the business process recurs, certain activities are not explicitly required.

In a survey process, this sub-process includes the following activities: developing a collection strategy; training staff; checking resource availability; configuring information request and reception systems; preparing collection instruments; and preparing respondent materials (depending on the collection strategy, letters to respondents, survey brochures, follow-ups in cases of non-response or refusal, information for survey participants).

For statistical business processes other than surveys, this sub-process includes ensuring that the necessary confidentiality processes, systems and procedures are in place to receive or extract the necessary information from the main source.

4.2. Execute data acquisition procedures: This sub-process involves acquiring the required information by executing data collection procedures, using the various instruments designed and developed in phases two and three. This sub-process involves establishing relationships with the data providers to ensure effective data acquisition. It also includes managing communication with data providers to maintain positive and constructive relationships between them and the statistical organization.

4.3. Finalize acquisition: This sub-process consists of downloading or capturing the acquired data and paradata in a specifically designed or modified electronic environment to prepare the data for the next processing phase.

Guidelines for ensuring quality at this phase

► All statistical processes

- **Preparing and carrying out acquisition procedures**

- Where possible, use protected, integrated systems to obtain information from external sources (respondents, users, data recipients or data providers).
- Ensure that integrated systems are used to transmit protected information—any exemption must be approved by the organization’s information security unit. Refer to the organization’s directive on the subject, or refer to Statistics Canada’s Directive on the Transmission of Protected Information.
- Inform the organization’s information security unit of any IT security issues encountered when transmitting or receiving data.

- **Data capture**

- Ensure that data capture officers receive appropriate training and have adequate work tools, including instruction manuals and all the necessary individual materials.

- Leverage accessible technology to enhance the efficiency and quality of the data capture process. For example, new technologies offer new possibilities, such as optical data capture and scanners.
- Implement effective control systems to ensure the security of data capture, transmission and manipulation.
- Ensure that all units follow data capture procedures consistently in order to minimize errors.
- Incorporate automatic checks for errors the data capture operator can correct (i.e., checks that will identify keyboard entry errors) and record these cases for further analysis and review.
- Centralize data capture to reduce costs and to benefit more easily from expert knowledge.
- Review and analyze quality control measures and results to identify the root causes of errors.
- Ask data capture officers to check the accuracy of automatic capture in a sample; the results of this assessment will improve the automatic capture process.
- Develop procedures for the destruction of data that are no longer needed.
- Evaluate all data capture operations and document the results for future use.

► **Statistical processes involving administrative data**

- Use the memorandum of understanding signed with the data provider(s) to confirm the terms for transmitting data and associated metadata.
- Use existing mechanisms and tools to verify that all data have been transmitted and meet the standards established during negotiations with the provider(s).
- Verify that the transmitted data are stored in the designated secure directory.
- Ensure that access to administrative data for other individuals within the organization respects the need-to-know principle.
- Quickly contact data provider(s) if any problems arise during data transmission.

► **Statistical processes involving a new survey**

• **Sampling**

- Develop an effective sample design using the auxiliary variables available on the frame that are most closely correlated with the survey variables of interest (e.g., by stratifying the frame or by carefully allocating the sample).
- To increase the accuracy of estimates, create strata that are as homogeneous as possible with respect to the survey variables of interest.
- Whenever possible, consider areas of interest when constructing strata and distributing the sample to the strata (Singh, Gambino and Mantel, 1994) to ensure an adequate sample size in the areas of interest and to reduce the need to use small area estimation methods.
- Reduce the number of sampling phases and the selection of clusters as much as possible to reduce the design effect on estimates and decrease the possibility of observing empty areas of interest.
- Select strata based on survey objectives, availability of variables in the frame, distribution of the variable of interest and the level of precision sought for the estimates.
- For longitudinal surveys, select stratification variables that correspond to characteristics which are stable over time.
- For populations with a highly skewed distribution, create a large unit stratum that contains units which will definitely be included in the survey.
- To determine the sample size and allocation among the strata, take into account predicted classification error rates in units, non-response, and other anomalies in the frame.

- For each type of unit, check the following: availability of an adequate frame or the possibility of creating one, ease of communication and of data collection and measurement, quality of data provided by the unit, cost of collection.
- To determine the sample size, consider:
 - Levels of precision required to produce survey estimates
 - Type of design and estimator to use
 - Availability of auxiliary information
 - Budgetary constraints
 - Sampling factors (e.g., clustering, stratification) and non-sampling factors (e.g., non-response, out-of-scope units, attrition in longitudinal surveys).
- Use the results of previous or similar surveys to calculate the design effect in order to determine the sample size, when required by complex designs. See Gambino (2001), Kish (1965) and Gabler et al. (2006).
- **For periodic surveys**
 - Consider expected births and deaths related to the units in the changing survey population.
 - Develop procedures to monitor the quality of the sample design over time.
 - Develop a method to maintain a stable sample size and therefore to keep collection costs stable as the sample size increases with population size.
 - Ensure the survey design is as flexible as possible to cope with future changes, such as sample size increases or decreases, restratification, resampling and updating selection probabilities.
 - Implement an update strategy for selective redesign of strata that have deteriorated.
 - When estimates are required for specific areas of interest (e.g., small geographic areas), form strata by combining small stable units related to the identified areas.
 - At the implementation stage:
 - Compare the size and characteristics of the actual sample against expectations
 - Compare the accuracy of the estimates against the expected objectives.
- **Longitudinal surveys**
 - For longitudinal or panel surveys, determine the panel's duration (time in sample) while maintaining a balance between meeting survey needs (data on duration) and the effects of sample attrition and conditioning.
 - Use an overlapping panel design (i.e., panels with overlapping durations) when cross-sectional estimates and longitudinal estimates are required.
- **Prepare and execute collection**
 - Identify roles and responsibilities for all aspects of collection, including the communication strategy, implementation, assessment, monitoring, contingency planning and security.
 - Design the collection process to alleviate respondent burden, reduce costs and accelerate obtaining the most accurate data possible.
 - Implement measures to ensure confidentiality of the data to collect.
 - Leverage available technology to enhance the efficiency and quality of collection processes (e.g., online electronic data collection and the ability to obtain information from a variety of media, such as tablets, cellphones or other type of personal digital assistant).
 - Establish sample control procedures and measures for each stage in data collection (e.g., delivery and return of paper or electronic questionnaires, follow-up on gaps or inconsistencies, and non-response follow-up).

- Develop a respondent communication strategy to maximize response rates. This includes promoting the survey, informing respondents in advance that they have been selected to participate in the survey, providing an email address and a toll-free telephone number for any questions, publishing key statistics that may encourage participation, facilitating communication of information to the public (e.g., websites, questionnaire user guides or information lines), and thanking respondents for participating.
- Ensure interviewers receive appropriate training and have adequate work tools, including instruction manuals and all necessary individual materials.
- Develop a communication method where interviewers can interact with the head office during collection.
- Carefully monitor operations to distribute and retrieve paper or online questionnaires during collection while ensuring that each unit selected for the survey receives the appropriate questionnaire.
- Provide plans and tools to actively manage data collection as it progresses, using productivity measures and cost indicators.
- Choose the appropriate time to contact respondents or designated persons in selected households or organizations, using paradata produced in previous survey cycles or in similar surveys.
- Contact respondents when it is most convenient for them, and ensure the number of calls, visits or reminders does not exceed acceptable limits.
- Establish an order of priority for contacting selected units, and follow up in a way that properly balances desired quality with collection costs.
- Search for respondents whose contact information seems to be outdated.
- Establish mechanisms so that respondents can update their contact information between survey cycles.
- Let respondents use their preferred method and format to report data.
- Relax reporting arrangements to reduce respondent burden and facilitate data collection (e.g., specific arrangements could be suggested for businesses participating in several surveys at once so they report only once for all surveys).
- For completed questionnaires, check the accuracy of coverage information and data quality. Ideally, edits and validations should be integrated into the questionnaire so corrections can occur during collection. Otherwise, a reverse record check may be necessary to check the quality of the responses.
- When questionnaires are not received, follow-up with respondents is required by telephone, in person or by email to determine why they did not complete it (refusal, disappearance of the respondent, closure, etc.).
- If necessary, make an appointment with the respondent to convince them to participate in the survey or to conduct an interview.
- To avoid unnecessary follow-up, promptly inform collection staff when questionnaires are returned.
- At the end of data collection, contact a subsample of non-respondents or all non-respondents to verify their eligibility for the survey. If eligible, it is recommended that some essential information be collected to adjust for non-response. To obtain this information, refer to the administrative data on hand.
- Conduct a retrospective assessment of all collection operations and document the results for future use.
- Assess the methods used and document lessons learned to improve each component.
- Collect information on whether collection tools and procedures require quality improvements for future survey cycles.

Quality dimension and indicators at this phase

a. **Quality dimension:** Accuracy

b. **Indicators of quality:**

- Coverage rate of the databases used
- Rate of proxy reporting
- Total and partial response rate
- Refusal rate
- Impact of follow-up strategies
- Errors attributable to survey eligibility or ineligibility
- Distribution of interviews according to average duration
- Impact of the collection mode (suggestion: mode effect)
- Edit failure rate
- Capture or coding error rate.

5. Profile and prepare data

Description of this phase and its sub-processes

Data profiling and preparation involves using profiling to identify incomplete, incorrect, inaccurate or irrelevant data in the original source and then modify, correct or suppress it when needed. This phase can also involve suppressing typographical errors and correcting values in relation to a list of known entities. Another common practice of data preparation is improving the data by adding related or complementary information. The goal of this phase is to ensure that the acquired source information is consistent and ready for the next phases.

Data profiling and preparation includes seven sub-processes, which may occur sequentially or in parallel and can be iterative. Profiling the data is the first priority. The seven sub-processes are:

5.1 Profile Data	5.2 Standardize, classify and code	5.3 Edit and impute	5.4 Adjust and reweight	5.5 Derive new variables and statistical units	5.6 Assess and document the impact of changes	5.7 Finalize data files
-------------------------------	---	-------------------------------	-----------------------------------	---	---	-----------------------------------

5.1. Profile data: This sub-process involves reviewing and gathering as much information as possible about the data that have been acquired to determine and profile their different characteristics. The main objectives of the collected information are:

- a. to determine whether the information contained in the data can easily be used for other purposes
- b. to identify record and variable subsets from administrative sources, and confirm which are in scope for the statistical program
- c. to develop measures on data quality and on their compliance with the statistical organization’s standards
- d. to assess the risks of integrating these data with other sources, including the challenges of record linkage
- e. to assess whether the metadata correctly describe the information in the data source
- f. to assess the distribution models of values and the functional dependencies of information contained in the data source
- g. to gain a good understanding of the impact of the information on future projects in order to anticipate future challenges
- h. to have an overall view of the information that will be used for database management or data stewardship in order to improve its quality.

- 5.2. Standardize, classify and code:** This sub-process involves standardizing the structure, format, and sets of codes and abbreviations of source data variables to ensure their standardization, comparability and stability for all files. This sub-process also involves assigning numerical or alphanumeric values to text responses according to a predetermined classification system. Although activities performed during this sub-process are frequently automated, sometimes they may require complex decision making and must therefore be done manually. Data classification and coding help produce formatted data that will be used in subsequent phases.
- 5.3. Edit and impute:** This sub-process is a logical continuation of data profiling and occurs when data are reported as incorrect, missing or unreliable. Imputation is the process of assigning replacement values to missing, invalid or inconsistent data rejected in the data validation phase. This process must follow certain rules and includes the following phases:
- a. determining the relevance of adding or modifying data
 - b. choosing the method to use
 - c. adding or modifying the values of the data to impute
 - d. producing metadata on the imputation process.
- 5.4. Adjust and reweight:** This sub-process involves adjusting survey weights, and occurs mainly in statistical sampling survey processes when the observed sample is smaller than the sample initially selected. This occurs in cases of total non-response or because out-of-scope units were selected. Adjustment factors should be based on the response probability of each sampled unit, in cases where non-response is related to the measured variables.
- 5.5. Derive new variables and statistical units:** This sub-process involves creating new variables and statistical units from the values of those obtained from source data. These variables and units are usually necessary to create the desired statistical product. Derivation involves applying mathematical formulas to one or more of the variables present in the dataset, or applying different modelling assumptions. It is sometimes necessary to proceed iteratively, since some variables used to derive others may themselves have been created by derivation. It is therefore important to ensure that variables are derived in the correct order.
- 5.6. Assess and document the impact of changes:** This sub-process involves assessing and documenting the impact of various transformations carried out in the previous five sub-processes. This exercise is important because it provides not only an overall picture of the improvements made to the source data, but also provides information about how the transformations affected data quality.
- 5.7. Finalize data files:** In this sub-process, results from the previous six sub-processes are combined to produce the data file to be used for the next phase.

The final processing file is created in this sub-process. This file is the dissemination database for all data files, namely the master file, the public use microdata file (PUMF) and the shared file.

The dissemination files will be used inside and outside Statistics Canada to analyze the survey data. The PUMF is an abridged version of the master file that has been thoroughly assessed by statisticians to eliminate all reasonable risks of identifying respondents who may have somewhat unique characteristics. The shared file contains only the records of respondents who consented to have their data shared.

Guidelines for ensuring quality at this phase

► All statistical processes

• Data profiling

- Identify profiling method(s) based on the intended use of the data (query optimization, cleaning, integration, analysis, scientific management, etc.) at subsequent phases of the statistical business process.
- Select the appropriate software based on type of profiling to perform.
- Perform individual analyses of the different database columns with emphasis on cardinality, data type and value models, value distribution and allocation, etc.

- Conduct multiple analyses of database columns focusing on uniqueness (key discovery—conditional, partial), inclusion dependencies (discovery of foreign keys—conditional, partial) and functional dependencies (conditional, partial).
 - Conduct overlap analyses for multiple data sources.
 - Develop a report with profiling results and use the results to update metadata as needed.
- **Data coding**
 - Centralize coding operations to reduce costs and benefit more easily from expert knowledge.
 - Ensure that coding personnel have adequate training and tools for successful coding operations.
 - Ensure that coding procedures are followed consistently for all units in the file to minimize errors.
 - Use quality control methods to check the level of precision of data coded by the system or by coders against pre-established criteria.
 - Use a team of data coding employees to handle special cases that cannot be automatically coded.
 - Create and update reference files to maximize the number of sentences recognized by the system and to minimize errors.
 - Evaluate a sample of the data that were coded automatically and check their accuracy.
 - Use the evaluation results to increase and enhance the content of reference files used for data coding.
 - **Data examination, editing and cleaning**
 - Provide adequate training to all staff involved in examining and editing the acquired data.
 - Provide a guide of edit rules.
 - Monitor the work to ensure that edit rules are interpreted identically and consistently by all editors.
 - Re-apply edits to corrected units to ensure that no errors were introduced directly or indirectly during the correction process.
 - Analyze characteristics of respondents and non-respondents to determine the non-response model, and then choose the appropriate method or approach to compensate for missing data (imputation or reweighting).
 - Remove typographical errors found during the profiling sub-process and correct values against a list of known entities, if any.
 - Use formatting tools to standardize variables in order to make it easier to compile and classify them.
 - **Edit and impute**
 - Explore various data sources to identify variables that could be used as auxiliary variables to impute missing data.
 - Assess the quality and relevance of available variables to determine which ones can be used as auxiliary variables or to establish imputation classes.
 - Identify the auxiliary variables that can explain the non-response mechanism(s) in order to enrich the imputation method.
 - Consider the type of characteristics to be estimated when choosing the auxiliary variables and imputation strategy in order to maintain the relationships of interest.
 - Some surveys require multiple imputation methods, depending on the availability of auxiliary information. In these cases, it is necessary to:

- o Establish a limited number of method hierarchies using pre-defined rules
- o Develop and test the methods at each hierarchical level
- o Retest the methods with the new imputation classes, in cases where imputation classes have been combined.
- In the case of donor imputation:
 - o If possible, impute all missing data from the record with a single donor to ensure consistency between data and to maintain the relationships between variables
 - o Consider the nearest-neighbour method to find a donor that is as close as possible to the record to be imputed and select appropriate variables to identify the nearest neighbour
 - o Choose a random donor from all potential donors that are at the same or a similar distance from the recipient to be imputed.
- Ensure imputed records closely resemble the records rejected at the check and edit phase.
- Consider excluding certain units when calculating imputed values if they are very different from the units to be imputed.
- Identify imputed values and clearly flag the imputation methods and sources.
- Retain the imputed and un-imputed values of the record's fields for assessment purposes.
- Assess the degree and effects of imputation.
- Measure the added variance introduced by imputation.
- **Adjust and reweight**
 - Adjust survey weights to reduce non-response, coverage and sampling errors or to ensure consistency with other data sources.
 - Consider calibration when the auxiliary data are correlated to the variables of interest. However, auxiliary data must be available for the sampled units, and the corresponding population totals must be known or accurately estimated.
 - Use weight range control methods to avoid extreme or negative weights (see Deville and Särndal [1992], for example).
 - Adjust longitudinal weights and cross-sectional weights to account for sample erosion in longitudinal analysis.
 - If it is of adequate quality and correlated with key survey variables, use auxiliary information if possible to improve the consistency and precision of estimates.
 - Account for non-response by selecting auxiliary variables that are related to the response probability and to the main survey variables.
 - Weight for non-response by modelling response indicators and validating the model through diagnosis.
 - Form non-response weighting classes to strengthen against failure of the non-response model and to avoid extreme weights.
 - Discuss weighting classes, generalized regression adjustment factors, units to include and exclude in the calculation of adjustment factors, and adjustment factors based on response probability when non-response is related to the measured variables.
- **Assessment and documentation**
 - Document each process (data profiling, coding, examination and cleaning, edit and imputation, and adjustment and reweighting), including methods and tools used, results found, impact on data quality and recommendations made.
 - Ensure the accessibility of all documentation produced.

- **Final processing**

- Add survey weights to the individual records, including the weights to be used with the shared file, if any.
- Create record identification codes for the master file, the PUMF and the shared file.
- Add new variables that may have been created as a result of suppressions required to create a PUMF or a shared file.

Quality dimensions and indicators at this phase

a. **Quality dimensions:** Accuracy, interpretability, coherence

b. **Indicators of quality:**

- A description and justification of the methodology used for each phase of data profiling and preparation is available, with supporting results
- Quality indicators, accuracy measures and/or quality assurance measures are available for the various phases
- Where models have been used, a description of the models’ assumptions and an assessment of their likely effects on data quality is available
- The generalized systems and the parameters used have been tested, modified as necessary, and validated
- A data dictionary and a user guide are available, as needed.

6. Integrate, estimate and compile

Description of this phase and its sub-processes

Data integration involves extracting data from different sources (databases, files, applications, web services, emails, etc.), applying certain transformations to the data (joins, de-duplication, concatenation, calculations, etc.), and finally, consolidating the data into a single database.

Data compilation involves creating aggregate data and population counts from microdata or lower-level aggregates. As part of statistical business processes involving multiple input data sources, data compilation involves mapping multiple types of statistics, concepts, classifications, and conventions through an iterative process leading to aggregate data. In this case, integration and compilation are performed simultaneously.

In sample survey processes, the accuracy of estimates can be improved if auxiliary data are available. The integration of auxiliary data into these estimation processes is called calibration.

Data integration and compilation includes five sub-processes, which may occur sequentially or in parallel, and can be iterative. The sub-processes are:

6.1 Determine integration elements, rules and strategy	6.2 Assess and adjust integration strategy	6.3 Load, apply mapping and integrate source data	6.4 Estimate, compile and apply statistical methods	6.5 Check and adjust quality improvement
--	--	---	---	--

6.1. Determine integration elements, rules and strategy: This sub-process involves determining integration elements—essentially, variables or combinations of variables that allow data to be integrated from multiple sources. These variables are the database identifiers or keys. Some data files have unique keys for each record, which can facilitate their integration when these keys are the same for all integration files. In statistical processes by record linkage, this sub-process involves identifying a set of record pairs that correspond perfectly to a given key or a particular criterion. These record pairs represent potentially matched pairs, also called potential pairs.

This sub-process also involves developing and defining the rules and strategy that will be used for data integration. As part of a record linkage process, this sub-process begins with the comparing the fields and records where, for each pair, the attributes of the linked records are compared. Attribute comparison is also based on comparison functions that differ depending on how they process missing values. Results of the

comparison are then used to choose the linkage strategy. This sub-process generally ends with producing preliminary linkage keys.

For processes involving multiple types of input data, such as macroeconomic accounts and population estimates, this sub-process is often limited to validating that there is nothing to modify. However, in some cases, new types of transactions may emerge for which no structure or rule has yet been created or developed. In these cases, the rules and process will have to be determined before the components or accounts can be compiled.

6.2. Assess and adjust integration strategy: This sub-process involves assessing the quality of integration elements.

In a record linkage process, this sub-process involves assessing the linkage quality (internal validation) and fitness for use of the linked dataset (external validation). The goal of this phase is to ensure that the quality of the linked dataset is appropriate for the dataset's intended use. Any limitations of the linked data noted during this sub-process should be taken into account when deciding whether to use the data.

This sub-process also contribute to the improvement of the record linkage strategy, if necessary, based on the results of the internal and external validation. Improvement may involve further refinement of the linkage rules or thresholds used to identify pairs as matches and non-matches. This sub-process generally ends with the production of final linkage keys.

For macroeconomic processes, this sub-process involves the analysis of the accounts to check their coherence and reasonableness. A variety of perspectives or dimensions are considered, including the dimensions relating to time coherence and structure. Accounting constraints and balances are also reviewed during this sub-process to ensure overall coherence. The knowledge gained in this stage may result in a change to the account design, which could lead to a return to the design phase. Once the design adjustments are made, return to sub-process 6.1 to update the accounting structures, rules and integration strategy.

For population estimates and projections, this sub-process involves the analysis of trends and population growth to check their coherence and reasonableness. The age-sex structure and spatial distribution should also be included in the analysis.

6.3. Load, apply mapping and integrate source data: This sub-process involves using basic identifiers to create new data from two or more source datasets.

For statistical processes such as macroeconomic accounts, population projections or estimates, this sub-process involves loading the acquired source data and transforming them so that they conform to the concepts and mapping of macroeconomic accounting or of demographics. Any information acquired in the early phases is also considered, since it is important to the general process of creating macroeconomic accounts or population components.

For statistical processes involving record linkage, integration usually includes systematic linkage, the use of linkage keys to create a linked dataset, and reconciliation of variables when two or more source datasets contain the same variables. To review the merging process, frequency or record checks should be conducted on each individual source dataset prior to the merge process, and again following the merge. Record counts are then compared to identify any discrepancies.

For processes involving sample surveys, integration is conducted with auxiliary data and is referred to as calibration (see sub-process 5.4 – Adjust and reweight). Calibration consists of readjusting the initial weights so that the estimates of the auxiliary variable(s) correspond to known totals.

6.4. Estimate, compile and apply statistical methods: This sub-process involves creating aggregate data and population counts from microdata or low-level aggregates. It includes summing data for records that share certain characteristics, calculating averages and dispersion, and applying weights from sub-process 5.4 (Adjust and reweight) to calculate appropriate totals. For sample surveys, sampling errors can also be calculated in this sub-process and associated with the relevant aggregates. Compilation is an iterative process that can be a good method for assessing data quality because it highlights certain inconsistencies. This sub-process also includes applying statistical methods such as seasonal adjustment, deflation, benchmarking.

6.5. Check and adjust quality improvement: This sub-process involves verifying that the integrated or compiled data meet the production objectives and can be used reliably. Verification activities include comparing the statistical product with other relevant data (internal and external), comparing the statistics with expectations and field knowledge, or studying inconsistencies between statistics. This sub-process is important for record linkage processes because it provides additional information on the fitness for use of the linked dataset. Errors or limitations identified in this sub-process may require further adjustments to the record linkage strategy, quality assessments, or quality improvement adjustments. Adjustments may require certain previous sub-processes to be repeated.

Guidelines for ensuring quality at this phase

▶ **Data integration involving record linkage**

- For large source datasets, generate and assess all potential pairs using indexing or blocking to reduce the number of possible pairs generated by cross-tabulation products.
- Choose one of the two main methods (probabilistic and deterministic) to compare record pairs according to rules based on an assessment of the content of the different regions of the pairs.
- Conduct a thorough review (which could include manual edits) of the selected matched pairs, unmatched records, and the overall linkage rate, and adjust the initial linkage strategy until an optimal strategy is achieved. This process can be iterative until the preliminary linkage keys are produced.
- Generate error estimates for the linked data, such as false positive and false negative linkage rates, specificity and sensitivity.
- Conduct an internal validation by comparing the overall linkage rates with expected levels based on past experience or external sources, or by analyzing linkage rates for sub-groups or populations in order to detect possible biases or to confirm expected trends.
- Conduct an external validation by comparing results from the linked data with external data. This validation must be conducted with subject matter experts, where applicable.
- Adjust the rules and final linkage strategy according to the results of the internal and external validations in order to produce the final linkage keys.
- Write a report on the linkage process that includes data limitations and is presented in a format that enables data users to understand the basic concepts of the linkage strategy and the results of the quality assessment.

▶ **Integration of data from multiple sources**

- Establish, update or validate certain rules or assumptions regarding the components or accounts to be integrated. Sometimes, this step consists in confirming that there is nothing to change within a recurring process.
- Ensure that conversions are performed on all input data so that they conform to the concepts and mappings of the statistical activity in question (national accounts, population estimates or projections, etc.).
- Examine, from different perspectives and dimensions, the accounts or components of the statistical activity in question, such as of time consistency and structure.
- Examine accounting restrictions and account balancing to ensure overall coherence when dealing with a process for macroeconomic accounts.
- Make the necessary modifications and adjustments if certain inconsistencies or imbalances are noted in the previous points.
- Update the account structure, rules, and source data for a macroeconomic accounts process.

▶ **Estimation and compilation**

- Use estimation weights (adjusted survey weights) to calculate descriptive and analytical statistics of the fields of interest.

- Use special methods to estimate small areas in case they were not included in the sampling plan.
- For each survey estimate, estimate its sampling error as sampling variance, standard error, coefficient of variation, margin of error or confidence interval.
- Use composite estimation for periodic surveys with significant sample overlap between instances (Gambino, Kennedy and Singh 2001).
- Use the organization’s generalized estimation software, where applicable; or refer to G-Est, Statistics Canada’s generalized estimation system, a modular generalized system for domain estimation in sample surveys.

► **Check and improve quality**

- For processes involving several types of input data, check the dimensions relating to time consistency and the component structure.
- Apply the quality improvement adjustments as per sub-process 2.6 and the revisions described in sub-process 6.2.

Quality dimensions and indicators at this phase

a. **Quality dimensions:** Relevance, accuracy, interpretability, coherence, accessibility.

b. **Indicators of quality:**

- An integrated data validation report
- A data dictionary
- Quality indicators on integrated data.

7. Analysis

Description of this phase and its sub-processes

The analysis phase involves examining, interpreting and preparing the data for release, as well as developing answers to specific questions. The activities in the Analysis sub-processes also help statistical analysts understand the data. These activities include identifying the topics of analysis, determining the availability of appropriate data, selecting and applying the methods to use in order to answer questions of interest, and evaluating, summarizing and conveying the results.

This phase also plays a key role in the evaluation of data quality because it is here that specific problems can be identified, which can lead to future improvements to the process.

Generally speaking, the data analysis results are often published or summarized in the statistical organization’s official releases. Moreover, some programs consider the analytical results to be the main data product when microdata cannot be published due to confidentiality.

The analysis phase includes five sub-processes, which are generally sequential, but can also occur in parallel, and can be iterative. The sub-processes are as follows:

7.1 Prepare draft outputs	7.2 Validate outputs	7.3 Interpret and explain outputs	7.4 Apply disclosure control	7.5 Finalize outputs
-------------------------------------	--------------------------------	---	--	--------------------------------

7.1. Prepare draft outputs: This sub-process is where the data collected and processed in the previous phase are transformed into statistical outputs. It also includes activities that produce other related measures.

7.2. Validate outputs: This sub-process is where the quality of the data produced is validated in accordance with the QAF and the production objectives defined at the outset of the statistical process. It also involves gathering as much information as possible about the topic being studied and comparing it with the data to identify any divergence from expectations and to ensure that the analysis is evidence-based.

7.3. Interpret and explain outputs: This sub-process is where analysts interpret data in order to understand current and emerging relevant issues and to determine how to present the results to the general public. The

analytical studies carried out in this sub-process also explain the behaviour of selected characteristics and the possible relations that may exist between them.

7.4. Apply disclosure control: This sub-process ensures that the provisions for protecting confidentiality are respected and that the usefulness of the data produced is preserved as much as possible. It includes checks for primary and secondary disclosure, as well as the application of data suppression or perturbation techniques.

The principles underlying disclosure control activities are almost exclusively governed by paragraph 17(1) (b) of the *Statistics Act* (1970, R.S.C., 1985, c. S-19). The rigorous disclosure control and confidentiality protection program in the Canadian statistical system maintains public confidence, which is essential to obtaining relevant, high-quality data, both from respondents and data providers.

7.5. Finalize outputs: This sub-process ensures that the statistics produced, and their related metadata and services, reach the required quality level and are suitable for use.

Guidelines for ensuring quality in this phase

► All statistical processes

• Validate outputs

- Refer to Statistics Canada’s Directive for the Validation of Statistical Outputs and the Guidelines for the Validation of Statistical Outputs.
- Check the coherence of outputs against similar internal and external data sources, such as surveys, other survey iterations and administrative data.
- Check internal coherence by calculating ratios that are expected to fall within certain known limits (e.g. proportions of men to women, average value of assets, etc.).
- Examine the individual contribution of large units to the totals (generally applied to business surveys).
- Compare the data quality indicators calculated in the previous phases with the production objectives.
- Check cross-tabulations to ensure consistency and accuracy of the key variables and important fields.
- Organize feedback meetings with the staff involved in data acquisition and processing.
- Mandate external specialists who are familiar with the field in question to check whether the outputs are plausible and to write a report on the work in progress, before publishing the results.
- Assess coverage and sampling errors, non-response errors, and measurement and processing errors, based on the analyses in other phases of the production process.
- Analyze data coherence based on recent events in the news.

• Interpret and explain outputs

- Choose a suitable analytical approach for the topic and data at hand.
- Check whether sources are consistent; and if more than one data source is used for the analysis, find an effective way to combine them.
- Determine whether imputed values must be included in the analysis and, if so, how they should be processed.
- Specify in the analysis how total or partial non-response was treated, and consider the importance and types of missing data in the sources used.
- Determine what other methods can be used to correctly account for the effect of non-response in the analysis, if the imputed values are not used.
- Provide a cautionary note, when necessary, on how the methods used to treat the missing data could influence results.

- Avoid drawing conclusions about causality.
 - When analyzing short-term trends, remember to also consider medium- and long-term trends. Short-term trends are often minor fluctuations around a more significant medium- and/or long-term trend.
 - Avoid arbitrary time reference points, such as the change between the previous year and the current one.
 - Take account of more meaningful points of reference, such as the last turning point for economic data, intergenerational differences for demographic statistics, and legislative changes for social statistics.
 - Consult with experts both on the subject matter and statistical methods.
 - Explain rounding practices and procedures.
 - Carefully explain the difference between the measures of percentage change and change in percentage points when presenting details on rates.
 - Define the basis for calculating rates and identify the conceptual base of the measures (for example, if in constant dollars or an index, indicate the reference year).
- **Disclosure control**
 - Consult the organization's disclosure control directives, or refer to Statistics Canada's Guidelines on the Management of Statistical Microdata Files and Aggregate Statistics to determine the most appropriate disclosure control methods for the types of data to be processed.
 - Determine the type of data to be processed in order to choose the appropriate disclosure control methods. Each type of data has its own methods.
 - Do not reveal the parameters and rules used for disclosure control. Knowing these parameters can help to better clarify the value of some respondents.
 - Determine the degree of confidentiality of each cell in relation to the organization's rules.
 - Avoid releasing a table of quantitative data if it contains values linked to cells considered confidential.
 - Consider the residual disclosure risk, which is the possibility of identifying confidential data by cross-referencing published information with other accessible information, including previous releases by the organization.
 - Determine if zero-frequency cells pose problems. Zero-frequency cells can reveal confidential information in quantitative data tables.
 - Suppress confidential cells in tables.
 - Check whether the categories and hierarchies used in the tables overlap.
 - Where appropriate, use a technique to round cell values in order to protect confidentiality.
 - With respect to microdata dissemination:
 - Evaluate the risk of disclosure in the microdata files
 - Apply one of the two disclosure control methods generally used to control disclosure risk
 - Ensure that the population is large enough for certain identifiable groups
 - Expand the categories of variables
 - Combine the upper and lower extreme values
 - Suppress certain variables from certain respondents
 - Suppress respondents from the file, if applicable.
 - See subsection 17(2) of the *Statistics Act*, which stipulates that certain types of confidential information may be released at the discretion of the Chief Statistician, by order.

- Consult the organization’s confidentiality resources. Otherwise, contact the following resources at Statistics Canada:
 - Microdata Access Division, which provides advice and guidance on policies related to the confidentiality of collected information
 - The Confidentiality and Legislation Committee and its subcommittees, the Microdata Management Access Committee and the Microdata Release Committee, which provide disclosure control strategies and practices
 - The Disclosure Control Resource Centre in the International Cooperation and Methodology Innovation Centre (ICMIC), which provides technical assistance, as well as the Generalized Systems Support Team for G-Confid.
- **Presenting the results**
 - Emphasize the variables and important topics when presenting the results.
 - Arrange ideas in a logical order and in order of relevance or importance.
 - Use headings, sub-headings and boxes to make it easier for users to consult the results.
 - Keep the language as simple as the topic permits. Depending on the target audience, it is sometimes worthwhile to be less precise to make the text more understandable.
 - Insert graphs and tables into the text to support the message.
 - Whenever possible, try to favour titles that carry a message (e.g., “Women’s incomes remain lower than men’s”), rather than conventional chart titles (e.g., “Income by age and sex”).
 - Comment on the information in the tables and charts to help readers understand it better.
 - Ensure that the tables not only present data clearly, but also prevent misinterpretation. This includes spacing, wording, placement and appearance of titles, line and column titles, and other labels.
 - Explain rounding practices or procedures.
 - Meet the confidentiality requirements of the statistical process for which the data are being presented.
 - Include information about the analytical methods and tools used. Include a section on the methods or a message telling readers where to get further information.
 - Include information on the accuracy of the results. Standard errors, confidence intervals or coefficients of variation provide readers with important information about data quality. The choice of indicator may vary depending on where the article is published.
 - Ensure that all references are accurate, consistent and defined in the text.
 - Check the article for errors. Also check the details, including ensuring that the figures in the text, tables and charts are consistent, the external data are accurate, and the mathematical calculations simple.
 - Ensure the intentions in the introduction are fulfilled in the article. Ensure the conclusions are consistent with the results of the analysis.
 - Have the article reviewed by others to verify its relevance, accuracy and interpretability, regardless of where it will be disseminated.
 - Ensure the text complies with the organization’s current publishing standards, or refer to Statistics Canada Publishing Guidelines. These standards apply to graphs, tables and style, among other things.
 - Ensure that all components to be disseminated are of high quality in both official languages and the data and text in both versions match.

Quality dimensions and indicators at this phase

a. **Quality dimensions:** Relevance, accuracy, accessibility, interpretability, coherence.

b. **Indicators of quality:**

- Availability of a data validation report
- Availability of an analysis report
- Availability of a data dictionary
- Availability of quality indicators
- Availability of information on the methods used
- Measures in place for disclosure control
- Revision of the text and data to be disseminated.

8. Disseminate

Description of this phase and its sub-processes

The disseminate phase involves making the collected and processed data available to users by various means. It is important to effectively communicate the data to users and publicize that they are available. The dissemination and communication activities of any statistical process should be related, and should aim to optimize use of the data by meeting user needs and providing them with broad access to the information.

The disseminate phase includes five sub-processes, which are generally sequential, but can also occur in parallel, and can be iterative. The sub-processes are as follows:

8.1 Finalize dissemination systems	8.2 Produce dissemination components	8.3 Manage product dissemination	8.4 Promote dissemination products	8.5 Manage user support
--	--	--	--	-----------------------------------

- 8.1. Finalize dissemination systems:** This sub-process involves making final adjustments to output systems with the data and metadata to be published. This includes formatting data and metadata so they are ready to be loaded into the output module, and ensuring that both the data and relevant metadata are connected.
- 8.2. Produce dissemination components:** This sub-process involves finalizing data production as outlined in sub-process 2.1. It includes activities to prepare the product components, such as text, tables and graphs, and quality statements. It also includes assembling the prepared components and performing a final check to ensure the product meets initial objectives and complies with dissemination standards.
- 8.3. Manage product dissemination:** This sub-process involves putting all the elements in place for dissemination, including mechanisms to manage the period when it will take place. The activities in this sub-process include information sessions for the media and other stakeholders interested in the topic, presentations to certain government ministers who are required to answer questions regarding them, managing access to confidential data for authorized users, etc. Presentation activities that take place before the actual dissemination are always under embargo.
- 8.4. Promote dissemination products:** This sub-process promotes the products resulting from a particular statistical business process to give them more media visibility and help them reach the widest possible audience.
- 8.5. Manage user support:** This sub-process involves ensuring that requests related to the disseminated product are recorded and forwarded to the correct location and that responses are provided as quickly as possible. This applies to requests for information or services from clients, access to microdata, etc.

Guidelines for ensuring quality at this phase

► **All statistical processes**

- Consult the organization’s dissemination guidelines, or refer to Statistics Canada’s Policy on Official Release.

- Consult the organization’s microdata release guidelines, or refer to Statistics Canada’s Directive on the Release of Microdata Files.
- Consult the organization’s guidelines on publication highlights, or refer to Statistics Canada’s Policy on Highlights of Publications.
- Consult the organization’s publication guidelines, or refer to the Statistics Canada Publishing Guidelines.
- Avoid preparing products (preliminary drafts) when data processing is in progress.
- Ensure the data to be disseminated correspond to the source data obtained.
- Thoroughly review all data (including underlying products) prior to dissemination to ensure they are accurate, the analysis is thorough, the processing was appropriately executed, the publication is relevant for the organization, and that communication is effective.
- Use automated tools, such as Smart Publishing or a tool to compare texts, to reduce the risk of human error.
- Test electronic products (data tables and other links) before disseminating to ensure they perform as intended.
- Ensure that written products are reviewed by someone who was not involved in the statistical activity.
- Carefully check numbers, reference periods (e.g., in the last half of the year or the last quarter), and words used to describe trends (e.g., upward or downward) in articles and publications to ensure they are accurate.
- Validate figures cited in articles and publications against other tabular products.
- Ensure the text is of high quality in both official languages and that the data and text in both versions match.
- Ensure that data quality measures or, if possible, the tools required to calculate them (e.g., validation tables for coefficients of variation, sampling variance programs) are included with the product being disseminated.
- Ensure the documentation describing the quality and methodology used is ready to be included in the product that is being disseminated.
- Ensure the data dissemination includes the name of one or more resource persons, a telephone number, and an email address if users require additional information.
- Ensure the resource persons are prepared to respond to requests for official interviews from the media, to provide comments, and to interpret data.
- Assess the impact of any incorrect statements or misinterpretations conveyed by the media and determine the best way to act on it.

Quality dimensions and indicators at this phase

a. **Quality dimensions at this phase:** Relevance, accessibility, timeliness, accuracy, coherence, interpretability (all six dimensions of the survey production process).

b. **Indicators of quality at this phase:**

- Availability of a dissemination schedule and a follow-up strategy
- Documented communication strategy in place
- Reasonable amount of time between the reference period or date, and the product dissemination date
- Reasonable amount of time between the planned and actual dissemination dates
- Availability of documentation of errors detected before and after dissemination, and evaluation of repercussions

- Availability of the frequency at which users access the information product over time, and assessment of the product’s utility.

9. Evaluate

Description of this phase and its sub-processes

The evaluate phase involves organizing one or more brainstorming sessions on the steps of the statistical business process. This phase, the last of any statistical production activity, is usually performed by the entire team that was involved in it. It is based on the analysis of all quantitative and qualitative inputs collected throughout the previous eight phases. It allows the team to identify strengths and weaknesses of the process, and to make recommendations on potential improvements for the future.

This phase includes three sub-processes, which are generally sequential. The sub-processes are:

9.1 Gather evaluation inputs	9.2 Conduct evaluation	9.3 Agree on an action plan
--	----------------------------------	---------------------------------------

- 9.1. Gather evaluation inputs:** This sub-process consists of gathering all the documentation produced during the various phases of the statistical business process and making it available to the evaluation team. This documentation comes in many forms, including user comments, process metadata and paradata, system metrics, and staff suggestions. Follow-up reports for the project action plan should also be used as input data for the evaluation.
- 9.2. Conduct evaluation:** This sub-process involves analyzing data collected in the previous sub-process and synthesizing them in an evaluation report. This report must describe all the problems encountered during the process, how they were managed, the results generated, and it should recommend changes as needed. These recommendations may include changes to any phase or sub-process, or a proposal that the entire statistical process not be repeated. It is also important to note the processes that worked well in the evaluation.
- 9.3. Agree on an action plan:** This sub-process involves seeking consensus from senior management on the need to develop an action plan based on the recommendations in the evaluation report and on how to implement them. This plan must include mechanisms for monitoring the impact of the planned measures, which can in turn be used as a basis for future evaluations.

Guidelines for ensuring quality at this phase

- Choose the person or team to evaluate the program (recurring production process) or the project (special production process).
- Ensure the person responsible for each phase in the process collects all inputs and makes them available to the person or team responsible for the evaluation. Inputs include:
 - Comments from users and members of the committees that participated in the consultations
 - Metadata and paradata on the phases and sub-processes
 - System metrics
 - Suggestions from staff
 - Follow-up reports for an action plan.
- Ensure the evaluation report is a synthesis of input data analysis, identifies all quality issues identified during the statistical business process, and recommends changes as needed.
- Bring together the necessary decision-making powers to establish an action plan based on the recommendations of the evaluation report and agree on this plan.

Quality dimensions and indicators at this phase

- Quality dimensions:** Interpretability, relevance.
- Indicators of quality:** Evaluation report for the statistical business process.

Bibliography

- Aitken, Alexis, Jan Hörngren, Nia Jones, Daniel Lewis and Maria João Zilhão. 2007. *Handbook on improving quality by analysis of process variables*. European Commission, Eurostat. 156 p.
- Bergdahl, Mats, Manfred Ehling, Eva Elvers, Erika Földesi, Thomas Körner, Andrea Kron, Peter Lohauß, Kornelia Mag, Vera Morais, Anja Nimmergut, Hans Viggo Sæbø, Ulrike Timm and Maria João Zilhão. 2007. *Handbook on Data Quality Assessment Methods and Tools*. European Commission, Eurostat. 141 p.
- Cai, Li and Yangyoung Zhu. 2015. “The Challenges of Data Quality and Data Quality Assessment in the Big Data Era.” *Data Science Journal*. Vol. 14, no. 2. p. 1–10.
- European Statistical System. 2015. *Quality Assurance Framework of the European Statistical System*. Version 1.2. European Statistical System. 52 p.
- Eurostat. 2012. *Guide to Statistics in European Commission Development Co-operation*. Version 2.0. Luxembourg. Methodologies & Working papers. European Commission, Eurostat. 460 p.
- Instituto Nacional de Estadística. 2015. *Directrices de calidad del INE*. Unidad de Calidad. 63 p.
- Lohr, Sharon L. and Trivellore E. Raghunathan. 2017. “Combining survey data with other data sources.” *Statistical Science*. Vol. 32, no. 2. p. 293–312.
- Naumann, Felix. 2013. “An Introduction to Data Profiling.” *Profiling & Cleansing*. Hasso Plattner Institut, IT Systems Engineering, Universität Potsdam. April 6. 58 slides.
- Naumann, Felix. “Data Profiling Revisited.” *SIGMOD Record*. Vol 42, no. 4. p. 40–49.
- Schomm, Fabian. 2016. “Data Profiling as a Process, Bridging the Gap Between Academia and Practitioners”. *GVD*.
- Statistics Canada. 2000. [Policy on Informing Users of Data Quality and Methodology](https://www.statcan.gc.ca/eng/about/policy/info-user).
https://www.statcan.gc.ca/eng/about/policy/info-user.
- Statistics Canada. 2003. *Survey Methods and Practices*. Statistics Canada Catalogue no. 12-587-XPE. Ottawa, Ontario.
- Statistics Canada. 2004. [Policy on Standards](https://www.statcan.gc.ca/eng/about/policy/standards). https://www.statcan.gc.ca/eng/about/policy/standards.
- Statistics Canada. 2009. *Statistics Canada Quality Guidelines*. Fifth Edition. Statistics Canada Catalogue no. 12-539-X. Ottawa, Ontario.
- Statistics Canada. 2012. *Policy on Microdata Access*.
- Statistics Canada. 2012. *Policy on Privacy and Confidentiality*.
- Statistics Canada. 2015. *Directive on the Validation of Statistical Outputs*. Version 1.4.
- Statistics Canada. 2015. *Directive on Obtaining Administrative Data under the Statistics Act*.
- Statistics Canada. 2015. [Policy on the Use of Administrative Data Obtained under the Statistics Act](https://www.statcan.gc.ca/eng/about/policy/admin_data).
https://www.statcan.gc.ca/eng/about/policy/admin_data.
- Statistics Canada. 2016. *Directive on the Transmission of Protected Information*.
- Statistics Canada. 2016. *Policy on Official Release*. Last updated September 27

Statistics Canada. 2017. *Directive on Corrections to Daily Releases and Statistical Products*. Last updated November 1st.

Statistics Canada. 2017. *Directive on Microdata Linkage*.

Statistics Canada. 2017. *Statistics Canada's Quality Assurance Framework*. Third Edition. Statistics Canada Catalogue no. 12-586-X. Ottawa, Ontario.

Statistics Canada. 2018. *Directive on the Management of Aggregate Statistics*. Last updated September 30.

Statistics Canada. 2018. *Directive on the Management of Statistical Microdata Files*. Last updated September 30.

Statistics Canada. 2018. *Directive on the Release of Microdata Files*.

UNECE. 2013. [Generic Statistical Business Process Model \(GSBPM\)](http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0), Version 5.0, December 2013, <http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0>

UNECE Big Data Quality Task Team. 2014. *A Suggested Framework for the Quality of Big Data*, version 4.1, UNECE, 32 p.

Yu, Frank. 2003. "Making Data Quality Visible in Practice." [Statistics Canada International Methodology Symposium 2003](https://www150.statcan.gc.ca/n1/en/pub/11-522-x/2003001/session14/7715-eng.pdf?st=6u72j3LU). Ottawa, Ontario. Date. 9 p. <https://www150.statcan.gc.ca/n1/en/pub/11-522-x/2003001/session14/7715-eng.pdf?st=6u72j3LU>.