

REPUBLIC OF CAMEROON
Peace – Work – Fatherland

**Ministry of Economy, Planning and
Regional Development**

National Statistics Council



REPUBLIQUE DU CAMEROUN
Paix – Travail – Patrie

**Ministère de l’Economie, de la Planification
et de l’Aménagement du Territoiret**

Conseil National de la Statistique



NATIONAL QUALITY ASSURANCE FRAMEWORK

**GUIDELINE MANUAL FOR THE MANAGEMENT
AND DISSEMINATION OF DATA, MICRODATA
AND METADATA IN CAMEROON**

VOLUME 03
March 2022



National Statistical Council Secretariat

PO Box: 134 Yaounde Telephone: (237) 222 22 04 45 Fax: (237) 222 23 24 37
Website: www.ins-cameroon.cm

REPUBLIC OF CAMEROON
Peace-Work-Fatherland

**Ministry of Economy, Planning and
Regional Development**

National Statistical Council



REPUBLIC OF CAMEROON
Paix-Travail-Patrie

**Ministère de l'Economie, de la
Planification et de l'Aménagement du
Territoire**

Conseil National de la Statistique

National Quality Assurance Framework

Volume 03

Guideline manual for the management and dissemination of data, microdata and metadata in Cameroon



National Statistical Council Secretariat

PO Box: 134 Yaounde Telephone: (237) 222 22 04 45 Fax: (237) 222 23 24 37

Website: www.ins-cameroon.cm

March 2022

Contents

Acronyms and abbreviations -----	5
Foreword -----	7
Background -----	9
Goal and Scope of the Manual -----	10
Methodological Approach to the Development of the Manual -----	11
Definition of Concepts-----	12
Directives for the Dissemination of Data, Metadata and Microdata-----	13
I. Directives for Dissemination of Data and Metadata -----	13
I.1 Statement of directives on the Dissemination Data and Metadata-----	13
I.2 Expected Changes from Compliance with Data Dissemination directives -----	13
I.3 Entities responsible for Implementation within the Administrations -----	14
II. Directives for the Management and Dissemination of Microdata -----	14
II.1 Statement of the directives on the Management and Dissemination of Microdata-----	14
II.2 Expected Changes from Compliance with the directives on the Management and Dissemination of Microdata-----	14
II.3 Entities in Charge of the Implementation within the Administrations -----	14
Section I: Guidelines for the Management and Dissemination of Data and Metadata -----	15
I.1 Preparing the Publication Catalogue-----	15
I.2 Preparing the Advance Release Calendar -----	16
I.3 Managing Access to Statistical Data-----	16
I.4 Communicating Information -----	17
I.5 Disseminating Revised Information -----	17
Section II: Guidelines for the Management and Dissemination of Microdata -----	19
II.1 Main types of Dissemination Microdata files -----	19
II.1.1. “Public Use” Files -----	19
II.1.2. Scientific Use Files -----	19
II.1.3. Enclave Files-----	20
II.2 Compilation of Microdata Files-----	20
II.2.1. Preparing files -----	20
II.2.2. Assessing Disclosure Risk-----	21
II.2.3. Processing of Microdata Files -----	22
II.2.4. Assessing Information Loss -----	23
II.3 Retroceding Microdata-----	23
Bibliography -----	25
Editorial Team -----	26

Acronyms and abbreviations

ACS	African Charter on Statistics
ARC	Advance Release Calendar
BUCREP	Central Bureau of the Census and Population Studies
BUNEC	National Civil Status Registration Office
FPOS	Fundamental Principles of Official Statistics
GDDS	General Data Dissemination System
GSBPM	Generic Statistical Business Process Model
IMF	International Monetary Fund
NIS	National Institute of Statistics
NQAF	National Quality Assurance Framework
NSC	National Statistical Council
NSDS	National Strategy for the Development of Statistics
NSIS	National Statistical Information System
PUMF	Public Use Microdata File
SDDS	Special Data Dissemination Standard
SUF	Scientific Use Files

Foreword

By adopting the National Quality Assurance Framework (NQAF) in 2018, Cameroon's National Statistical Council (NSC), the main regulatory body of the National Statistical Information System (NSIS), committed to focus on the quality of its official statistics in order to support the effective management of public policies.

The implementation of this reference framework is supported by the development, provision and implementation of operational tools including the ***Guideline Manual for the Management and Dissemination of Data, Metadata and Microdata***.

This effort is a milestone in the migration, or even the adherence of the NSIS to the Special Data Dissemination Standard (SDDS), the normative framework for the production and dissemination of official statistics implemented by the International Monetary Fund (IMF). This ambition to migrate to the SDDS is enshrined in the vision of the National Strategy for the Development of Statistics (NSDS 30), which focuses on the effective use of statistics for the management and steering of the performance of development initiatives. This ultimately contributes to the improvement of public finance management in Cameroon; a country that has subscribed to the SDDS is committed to transparency in publishing its official statistics, and therefore of public finance statistics. This gives it easy and affordable access to financial markets.

This is how the NSIS players become “transformers” of public finance management. Hence the importance of strict application of the guidelines in this manual.

Background

Since 2016, the National Statistical Information System has been committed to a quality approach in order to provide public authorities, the private sector and other users with reliable, useful and timely official statistics. This quality approach is gradually being implemented through the development of operational tools such as guidelines for surveys and censuses, as well as for the production of statistics from administrative sources.

The dissemination and use of the products of statistical operations are critical in this quality approach; a dissemination that concerns statistical information as well as microdata and metadata. In its vision set out in the National Strategy for the Development of Statistics (NSDS) since 2015, Cameroon has shown its ambition to migrate to the Special Data Dissemination Standard, which is more restrictive with regard to compliance with the dissemination standards of the General Data Dissemination System (GDDS) and the enhanced GDDS (e_GDDS) to which Cameroon subscribed since the year 2000. This is in line with the Fundamental Principles of Official Statistics (FPOS) adopted by the African Charter on Statistics (ACS), which emphasize the responsibility and transparency of producers for the correct use of data, prevention of the misuse of statistics, and confidentiality of individual data collected in the production of statistics.

In 2015, the National Institute of Statistics (NIS) developed the national policy on communication and dissemination of statistical information, which sets out the guiding principles for dissemination and communication of statistical information. This policy needs to be implemented through guidelines.

Microdata are increasingly used by the academic and research community, other public administrations and even the private sector for analytical purposes. Dissemination of these microdata should be increased to meet the demands of a growing number of users for a variety of purposes. This will help to support research, strengthen the credibility of official statistics, improve the reliability and relevance of data, etc. The new statistical law of 20 July 2020 provides a framework for making these individual data available, emphasizing the obligation of statistical confidentiality for data producers.

Goal and Scope of the Manual

These guidelines are intended to enable the NSIS structures to disseminate the various statistical products effectively. This dissemination concerns not only statistical information (data) with its metadata, but also microdata to make wider use possible. The aim is to ensure that all users have easy and equitable access to data in a timely manner, in accordance with the recommendations of the United Nations' FPOS, taken up by the African Charter on Statistics, and the guidance contained in the national policy on communication and dissemination of statistical information in Cameroon.

The statistical production process already adopted in previous manuals includes a dissemination stage, which is part of the overall production process. This manual addresses the specific issue of dissemination in all its aspects, taking into account the products already available, and emphasizing the relationships that should exist between producers and users to ensure easy access, efficient and correct use of data.

Moreover, the manual sets out the principles, procedures and practices for the dissemination of microdata, data and metadata. It is not a methodological document that develops the technical tools for the dissemination of these different data attributes.

It should be noted that with regard to the dissemination of microdata, the obligation to protect the privacy of respondents should not be used as an excuse to restrict dissemination. The purpose of these guidelines is to maximize the dissemination of microdata while respecting the privacy of respondents.

Methodological Approach to the Development of the Manual

The preparation of this manual consisted in defining the guidelines for the implementation of the national policy on communication and dissemination of statistical information in Cameroon. The idea was precisely to make operational the guiding principles set out in the said policy, by first setting out the directives, and subsequently the guidance for their implementation. These guidelines were developed based on best practices identified in reference statistical systems, but also within the NSIS itself and in particular at the NIS.

Importance was placed on the management and dissemination of microdata, which are gradually requested by the research community, other public administrations and even the private sector.

The manual has three sections. The first section outlines directives for dissemination, and the second and third sections present guidelines for the implementation of the prescribed directives. However, it should not be expected that the guidelines are developed in a strict order in relation to the different directives, or that a directive matches exactly a set of guidelines. The headings for the guidelines may follow a different order from the directives, the important thing being that guidance is provided for the implementation of all the directives.

These directives will enter into force on the date of their adoption by the National Statistical Council (NSC).

Definition of Concepts

Confidentiality: Protection against the disclosure of personally identifiable information about an individual, company or organization. Confidentiality refers to the dissemination of information derived from the data collected.

Data source: The place where the data used originate. A data source can be the place where the data were created (point of observation of the raw data), or where the physical information was digitized.

Directive: Official instruction that requires producers of statistics to take (or avoid) a specific action. The directive is binding.

Dissemination: Making available to the public the statistical data produced by the National Statistical Information System in all forms permitted by the regulations in force and respecting the privacy of the natural and legal persons who may have provided the information used for their compilation.

Guideline: Guidance, advice, explanations to managers or functional specialists. The guideline is a recommendation that is strongly suggested to be followed.

Individual data: Any information relating to a natural or legal person who is identified or can be identified, directly or indirectly, by reference to an identification code or to one or more elements specific to that person.

Metadata: Data about data, additional information needed to interpret and use the data (information about variables, classifications and concepts, data collection and processing methods, etc.).

Microdata: Immediate results of observations of characteristics and statistical variables collected from an observation unit and not subject to statistical processing.

Statistical confidentiality: Provision to ensure that natural or legal persons who provide information to be used for official statistics respect the confidentiality of information relating to their personal and family life, or commercial secrecy for enterprises.

Statistical data: Numerical information relating to all areas of the life of the Nation and obtained by appropriate processing using statistical methods.

Directives for the Dissemination of Data, Metadata and Microdata

These directives deal with the dissemination of data and metadata, on the one hand, and the dissemination of microdata on the other. The directives on the dissemination of data and metadata aim to provide users with easy access to the statistical information produced by the National Statistical Information System, and relate to the provision of information and the support of users in accessing and using the information. Regarding microdata, the aim is to ensure that they are made available for popularization and to ensure greater use of NSIS databases, in compliance with the statistical law in force.

For each section, after the statement of the directives, the changes expected from the respect of these directives as well as those responsible for their implementation are presented.

I - Directives for Dissemination of Data and Metadata

I.1 Statement of directives on the Dissemination Data and Metadata

The directives cover six important elements for dissemination (i) coverage of statistical output, (ii) prior information to users, (iii) access to data, (iv) dissemination of revised information, (v) communication of data sources and (vi) communication of information.

D1. The NIS should inform users about all statistical products made available, by publishing annually a catalogue of statistical publications produced by the NSIS.

D2. The NSIS shall, through the NIS, inform users in advance of forthcoming publications by publishing a monthly Advance Release Calendar (ARC) in accordance with the requirements of the SDDS.

D3. Any producer of official NSIS data shall ensure that users have easy access to the published data. Such data shall be accompanied by metadata.

D4. Producers of official NSIS data shall disseminate revised data, clearly informing users of the reasons for the revision.

D5. Producers of official NSIS data should inform users of the origin of the data used in statistical production, with all data sources being accurately and appropriately reflected in the dissemination of statistical products.

D6. Each NSIS member administration should establish a framework for permanent consultation with other stakeholders (suppliers, users, partners, media), and develop communication materials appropriate to each category.

I.2 Expected Changes from Compliance with Data Dissemination directives

- Users are aware of all statistical products made public by the NSIS and the dates of the next release of information.
- Users have access, as appropriate, to data and metadata of statistical operations carried out in the system.
- Statistical data producers are transparent in the production of statistics, notably through communication on the fundamental revisions made and on the data.
- Users are satisfied with the communication of statistical products by producers.

I.3 Entities responsible for Implementation within the Administrations

(a) At the level of the NIS, they are:

- the Technical departments which are responsible for developing the expected statistical products and transmitting them to the department in charge of statistical dissemination;
- the Department in charge of coordination, which will be responsible for drawing up the publication schedule and the catalogue by using information from all NSIS member administrations;
- the General Management, which is responsible for all requests for access to statistical data produced by the Institute and for institutional communication.

(b) At the level of the other NSIS member administrations:

- BUCREP, BUNEC and the statistical units of the sectoral administrations of the NSIS, which must provide the NIS with the information needed to draw up the publication schedule and the catalogue.

II - Directives for the Management and Dissemination of Microdata

II.1 Statement of the directives on the Management and Dissemination of Microdata

The directives on the management and dissemination of microdata focus on making microdata available and managing them to meet user needs.

D1. Microdata collected as part of statistical operations should be made available to users, in accordance with the statistical law in force, and particularly with the provisions on the protection of personal data. Producers must ensure that the data they make available to users are free from all risks of identification of respondents.

D2. Producers should ensure adequate management of microdata to meet specific user requests.

II.2 Expected Changes from Compliance with the directives on the Management and Dissemination of Microdata

- Users have easy access to microdata and use them without risk of disclosure.
- Value of databases is increased through greater access to microdata files.
- Users are assisted by producers in accessing and using microdata.

II.3 Entities in Charge of the Implementation within the Administrations

(c) At the level of the NIS:

- the Department in charge of database management ensures the establishment and management of the Institute's statistical operation databases. It will also be responsible for all necessary processing of the databases in order to better address users' requests in strict compliance with the requirements of the statistical law;
- the General Management, to which all requests for access to statistical data produced by the Institute are addressed.

(d) At the level of the other NSIS member administrations:

- BUCREP, BUNEC and the statistical units of the NSIS sectoral administrations establish and manage the databases in accordance with the confidentiality principle.

Section I: Guidelines for the Management and Dissemination of Data and Metadata

This section presents guidelines for the implementation of the prescribed directives for the management and dissemination of data and metadata. For each directive, the idea is to specify the purpose before stating the guidelines. The scope and relevance for both producers and users are presented before best practices for implementation is stated.

I.1. Preparing the Publication Catalogue

Purpose: The catalogue presents the NSIS publications over a defined reference period. It reports on everything that has been published, both at the NIS and in the other NSIS administrations. It informs users about the availability of publications and other useful information to enable them to have an overview of all statistical publications and to access them easily. The catalogue primarily reports on publications made available over a defined period (one year, two years, five years, etc.), although major publications to be released in the near future can be added. Statistical publications cover all products, including databases, maps, etc. It is produced and disseminated by the NIS, which is responsible for coordinating the NSIS.

Guidelines: The preparation of a catalogue of publications consists mainly in the following:

- Designing the catalogue, which means sketching its layout, defining elements such as the scope of publications, useful information for the presentation of publications to facilitate access and use, etc. The scope must be defined in such a way as to cover the field of NSIS productions, taking into account both annual publications and those with a longer periodicity.
- Identifying the producers over the defined reference period.
- Carrying out an inventory of publications, which consists in listing all the publications that fall within the scope defined for the catalogue with each producer.
- Organizing statistical publications according to the United Nations classification of statistical activities.
- Presenting the publications according to the defined information, which are mainly:
 - *title and date of publication;*
 - *periodicity of production;*
 - *structure of the NSIS responsible for the publication;*
 - *content description, including key indicators, data sources, etc.;*
 - *photograph of the front cover;*
 - *languages of publication (inform the user whether the publication is available in both official languages, or only one);*
 - *available format (paper and digital);*
 - *access terms and conditions (users should be informed about where to get the publications, possible costs for access, etc.);*
 - *useful contacts for information.*
- Producing the catalogue: preliminary version to be validated with all stakeholders.
- Publishing and disseminating the catalogue via the websites and various possible channels in order to reach the largest number of users. Events such as the African Statistics Day and other open days are good opportunities for the promotion and dissemination of the catalogue.

1.2. Preparing the Advance Release Calendar

Purpose: The Advance Release Calendar is a prospective tool of the NSIS providing users with prior notice of the precise release dates when specific statistical information is to be released to the public. Published for all users at the same time, it is a tool for ensuring transparency, announcing the date (and if possible the time) when users should expect the announced publication. Its purpose is to enhance equality of access by ensuring simultaneous dissemination of statistical data to all users, and to create a special interest among users in the statistics to be released.

Adherence to the release schedule contributes to increasing user confidence in the NIS. The ARC is published at a defined frequency (usually monthly), and users are regularly informed about possible changes in the previously announced publication dates.

Guidelines: The following actions are necessary for the preparation of the release calendar:

- Designing the ARC, which means sketching the layout and defining the elements for the presentation of the publications.
- Collecting information from NSIS producers on a regular basis through a sheet designed for this purpose. The regular collection of information makes it possible to foresee possible shifts in the release dates already announced and to alert those responsible to the risks associated with failure to meet the release dates. The sheets are sent to the various NSIS focal points electronically.
- Filling in the publications to be released during the period, in the defined format: the calendar covers publications to be released for the current month and the next three months. The relevant information are:
 - *title of the publication;*
 - *abstract/description (the abstract should outline the main theme of the publication and the main indicators);*
 - *languages (this is the language used in the publication, usually English and/or French);*
 - *format and distribution channels (this is the announcement to users of the channels through which the publication will be made available);*
 - *publication date, which is the expected date on which the publication will be available (day, month, year and possibly the time);*
 - *author (this is the name of the administration that produced the publication);*
 - *Useful contacts (these are the contact details—surnames, first names, telephone numbers and email addresses—of the person(s) to be contacted if additional information is required).*
- Informing users of any discrepancies (on the publication date), and explain, as a priority for statistics whose release dates are regulated (national accounts, etc.). The objective is to comply with the announced publication dates.
- Validating and translating the ARC.
- Publishing.

1.3. Managing Access to Statistical Data

Purpose: Once the data have been produced and are ready for dissemination, it is necessary to define how they will be made available to the various users. This means defining the terms and conditions, as well as the mechanisms through which users will take possession of the data, with the aim of ensuring easy access and proper use of the data by all. It also implies the availability of metadata, as well as the establishment of an adequate mechanism for user support.

Guidelines: Managing access to statistical data consists in:

- defining the target audience for dissemination (main users, other users);
- choosing the dissemination media according to the data to be disseminated and the target audience;
- informing users about the terms and conditions of access (e.g. what is free and what is not, other conditions, procedures, costs for additional processing);
- providing users with contact details (user helpdesks, handling of queries about published information, etc.);
- disseminating information to users simultaneously, to ensure equitable access to information;
- indicating privileged (early) access to information (according to the regulation), where appropriate.

1.4. Communicating Information

Purpose: Statistical communication is the set of elements to be provided to accompany the publication of statistical information. The aim of statistical communication is to ensure its popularization, which makes it possible to increase statistical culture, ensure good understanding of published information, guarantee an adequate interpretation of published statistics, encourage the use of statistical information for decision-making at all levels (government, private sector, technical and financial partners, general public, etc.). This communication mainly refers to the definition of communication tools adapted to the various categories of users, preparation and dissemination of metadata.

Guidelines: To ensure good communication of information, it is necessary to:

- Identify communication tools, giving priority to modern tools, including social media, newsletters, press briefings and other platforms. These tools should be adapted depending on the category of users targeted.
- Prepare and disseminate metadata: this should be consistent with the data to be published. Users must be aware of details such as definitions, data coverage, collection and processing methods, interpretation of statistical information, possible limitations, as well as all the useful elements to enable them to fully understand and take stock of the quality of the statistical data they are using. The metadata sheets must be harmonized and used by all NSIS producers.
- Communicate data sources: the data source is an essential element of the metadata. Users need to know the exact source of the data they are using. In general, it is necessary to specify whether the data are primary (basic information directly collected from the individual or statistical unit: administrative registers, survey data, etc.) or secondary (data that have already been analysed, synthesized, explained or evaluated on the basis of the primary sources: reports, summary data, etc.), to indicate the statistical operation that made it possible to collect the data, to indicate all the sources that were combined to produce the information (if applicable).
- Use the communication languages adapted to each target audience.

1.5 Disseminating Revised Information

Purpose: When a change is made to previously published information, users should be informed as soon as possible. The change may be as a result of an error being identified, a change in methodology, etc. Informing users of revisions to already published data is a guarantee of transparency, and this practice enhances users' credibility in the NSIS and prevents them from using incorrect information that may lead to incorrect analyses and/or decision-making. This information should include both the revised data and the reasons for the changes. This revision does not concern statistics where the publication of preliminary data is recognized (e.g. national accounts), as the user is aware of their preliminary nature when using them.

Guidelines: The process of disseminating revised information depends on several factors, such as the importance of the revision, nature of the publication, time elapsed since the publication of the erroneous information, etc. In all cases, it is urgent to correct the information and make it available to the user. This involves:

- Adding the word “*corrected*” to the relevant data in the next publication for regular publications, which often concern data series (yearbooks, national accounts, prices, etc.). This allows the user to understand the discrepancy with the data they have already consulted.
- Replacing the online publication with the amended version as soon as possible.
- Producing an information notice presenting all the data that have been corrected and the reasons for doing so: for publications for which a large number of errors have been found.
- Informing users by all possible means: press releases, social media, websites, etc.

Please note: The revision of a published data naturally implies the revision of the indicators produced from this data. Thus, the information notice on the corrected data must take this aspect into account, in order to inform the user of all the data concerned by the revision.

Section II: Guidelines for the Management and Dissemination of Microdata

The management of the statistical environment requires that, at the dissemination stage, the individual data sets or microdata obtained in the statistical production process are disseminated in accordance with confidentiality requirements. This sub-process (dissemination) of the statistical production process includes all the activities needed to make the different micro-data files accessible to different types of users; the challenge being to maintain the confidence of information providers and to ensure the quality of the data collected. The search for data access solutions involves addressing a number of challenges, particularly ensuring secure access to data, maximizing the usefulness of data by minimizing the risk of disclosure, protecting individual data by all means (privacy, commercial confidentiality). Generally, finding the optimal balance between dissemination and protection is difficult and requires relevant technical expertise. Methods of disseminating microdata can vary depending on the types of files to be disseminated.

II.1. Main types of Dissemination Microdata files

Raw microdata files obtained during census or survey operations are rarely ready for dissemination. Otherwise referred to as master files, they need to be prepared in a format suitable for making them available to users.

In order to preserve the confidentiality of individual data required by statistical law, making microdata anonymous is imperative before dissemination. Depending on the level of anonymisation or accessibility, three types of microdata files are generally considered: (i) public use files, (ii) scientific use files, and (iii) enclave files.

II.1.1. “Public Use” Files

Purpose: “Public use” Microdata Files (PUMFs) are super-anonymised files that can be easily accessed by any user without conditions or under minimal, non-binding conditions of use (e.g. no sale, etc.). The PUMF (Public Use Microdata File) is a subset (of variables or records) of a master file, for which the level of anonymisation is such that identification of a respondent is almost impossible even when cross-referenced with other files. They are part of the statistical products expected from a production process, and should generally be designed and made available at the same time as these products (reports, databases, maps, etc.). The public use files, although highly anonymised and of limited use, should make it possible to find at least the main indicators of the operation. These files are for example used by teachers for educational purposes; the objective being more to support practical work than to carry out in-depth analysis of the data.

Advantages: public files make it possible for a minimum of variables to be made available, relevant to the main objectives of the operation. Access is easy and open to the widest possible audience.

Limitations: scope of public use files in terms of analytical value is limited, due to the various confidentiality safeguards applied.

II.1.2. Scientific Use Files

Purpose: Scientific Use Files (SUFs) are files intended to facilitate research by allowing in-depth analyses of microdata. They are generally intended for researchers and research centres, with more rigid access conditions (access authorization granted after a justified request has been submitted, signature of an agreement governing the use of the data, etc.). Although anonymisation is applicable to

these types of files and measures are taken during their preparation to make identification impossible when they are used in isolation, they may contain identifiable data when compared with other types of files. For this reason, scientific files are usually subject to use contracts between producers and users, and their dissemination may require the approval of a microdata review board. Users of scientific files are, for example, researchers, who have specific themes and need real data for their analyses.

Advantages: scientific use files are more useful for analysis (compared to the public use file) because of less stringent anonymisation settings. They are thus more likely to meet specific analytical needs.

Limitation: access to scientific files is strict, and risk of disclosure is higher.

II.1.3. Enclave Files

Purpose: Enclave files are files that can only be accessed in secure centres established by the producing organization. They allow access to raw microdata files (master files), while controlling the use of the data. Authorized users do not obtain the microdata, but run their programs on the microdata in real time through a secure device. This access can be within the producing organization or remotely, and disclosure control is automated. This data access solution requires adequate infrastructure. As a rule, staff at the organization supervise access and use of the data, computers must not be connected to an external network, and researchers' results must be checked by an analyst to ensure data confidentiality.

Advantages: users have access to master files, which are of great analytical value.

Limitations: users are subject to very strict access conditions; implementation of this practice requires adequate technological infrastructure and significant human deployment.

II.2. Compilation of Microdata Files

Compilation of raw microdata files includes all the procedures for adjusting and/or anonymising their content (variables and/or number of records) in order to make them suitable for dissemination. This involves modifying the file content by disrupting/modifying all direct or indirect identification variables that could lead to the identification of a respondent. The challenge in preparing microdata files is to ensure the trade-off between the usefulness of the data for analyses and protection of individual data. In other words, the usefulness of the data must be maximized while minimizing the risk of identification of respondents, which is a precondition for their trust in the producers.

This compilation of microdata files includes 4 critical stages: (i) preparation of the files, (ii) assessment of the risks of disclosure, (iii) processing of the files and (iv) assessment of the loss of information.

The process is almost the same for public use files and scientific use files, with the only difference that anonymisation methods are used sparingly for SUFs.

II.2.1. Preparing files

Purpose: In the microdata file preparation phase, the expected products are pre-release microdata files, which will subsequently be tested for non-disclosure. This involves examining in detail the raw file obtained directly after data capture and processing/removing any information or combination of information that could identify respondents directly, or indirectly through matching or cross-referencing with other data files. In the production process, this phase is triggered automatically after the completion of data collection, and the complexity of the task requires dedicated experts.

The nature of the operations (census, household survey, or enterprise survey) may guide processing in this process; in particular, the micro-data files of large enterprises are easily identifiable due to their generally skewed distribution.

Expected output: At the end of this sub-process, the main expectation is the draft of the anonymised micro-data file (pre-file), as well as a summary report presenting the different operations completed in the file.

Guidelines: Preparing a microdata file usually means:

- Identifying/reporting the nature of the statistical units on which the statistical data collection has taken place. This information guides the identification of sensitive variables and information, which need to be addressed during the anonymisation process.
- Identifying direct identification variables: these are those that unambiguously identify respondents (names, addresses, telephone numbers, personnel numbers, company names, etc.).
- Removing all direct identification variables.
- Identifying indirect identification variables: these are variables which, by cross-referencing or reconciliation with other variables or other data sources, can reveal the identity of individuals (sex, age, income, turnover, company size, etc.).
- Identifying quasi-identifiers, which are variables that identify respondents with some degree of ambiguity (e.g. geographic coordinates). The combination of quasi-identifiers allows for unambiguous identifiers.
- Identifying confidential outcome variables (sensitive variables with or without identifiers), which contain sensitive information about respondents (salaries, turnover, political affiliation, ethnicity, illness, etc.).
- Checking that the remaining variables (non-confidential variables) are cleared.
- Documenting the entire process.

II.2.2. Assessing Disclosure Risk

Purpose: Disclosure risk assessment refers to the set of analyses to be carried out on prepared microdata files to determine their disclosure risks. The aim here is to identify all possible disclosure risks from the different categories of identification variables identified. It is assumed at this stage that the direct identification variables have been removed from the files in the previous stage. At this stage, the main task is to identify the possible scenarios leading to the disclosure of individual data and to choose the method(s) that can lead to the measurement of the disclosure risk. It is usually assumed at this level (strong assumption) that an external user has information on the quasi-identifiers that can be matched with the microdata file to be disseminated. The intuitive idea to consider in risk assessment is that an individual is at risk of identification if a user is able to isolate him or her in the microdata file from the quasi-identifiers. The risk can be assessed by the probability of identification of an individual, or the number of individuals at risk of identification, for example.

Expected output: a report presenting the different disclosure scenarios and estimates of the different disclosure risks.

Guidelines: In practice, risk assessment consists in:

- Determining the risks of identity disclosure: this is the recognition of an individual in the database. In addition to direct identification variables (already removed), this can occur when, with information about an individual, the latter can be formally recognized in the database.
- Assessing the risk of attribute disclosure: this is the act of accurately determining an individual's response on a sensitive variable from quasi-identifiers. An individual cannot formally be recognized; however, from some information and data on a sensitive variable, their response may be deduced.

- Assessing the risk of inferential disclosure: this is the act of predicting an individual's characteristics with high degree of accuracy. Data without identity and attributes can still be used to predict with a high probability what an individual's response is.
- Assessing the risks of residual disclosure, which can occur when two or more secure data sets are put together to create the risk of disclosure.
- Assessing the risks associated with hierarchical files: For example, in a file where individuals are grouped into households, if one household member is re-identified, then all household members are re-identified.
- Estimating the risk of disclosure on the basis of the quasi-identifier information. As several estimation methods exist, documentation of the risk measurement process is necessary.
- Identifying appropriate processing methods for individual variables (recoding, rounding, anonymisation, aggregations, local deletion, etc.).
- Identifying appropriate processing methods for combinations of variables or the whole file (perturbation, synthetic data, etc.).
- Documenting.

In practice, several methods exist to assess risks (probabilistic or not), and their application requires adequate expertise.

II.2.3. Processing of Microdata Files

Purpose: Once the overall risk has been identified for each individual (or for the whole microdata file), the risk of disclosure should be eliminated or reduced to the set threshold. Risk reduction is therefore the set of techniques and tools used to reduce the risk of identification of individuals to the desired level. This step is delicate because it depends on the perception of potential users. Risk reduction can be achieved without introducing false information into the files to be protected (non-disruptive method), with disruptive methods or by generating synthetic or hybrid data. Several risk reduction methods can be used consecutively.

Expected product: The main expected product is microdata files protected at a certain threshold of the risk of disclosure of individual data, depending on the type of file (public use or scientific file) and a data protection report.

Guidelines: processing of microdata files has to do with:

- Checking whether direct identifiers have actually been removed from the microdata file. These identifiers are assumed to have been identified and processed in the preparation phase of the microdata files.
- Identifying the types of files to be disseminated (public use, scientific, etc.). As a general rule, the definition of the appropriate level of data protection or disclosure risk reduction depends on the type of file to be disseminated.
- For each data disclosure risk, defining the attack scenarios. This involves making a number of assumptions about the users who might attack the microdata file. It is usually assumed that they have all the information on the quasi-identification variables.
- Applying appropriate data protection methods (perturbation, anonymisation, rounding, recoding, synthetic data, etc.).
- Assessing the degree of confidentiality conferred on the file.
- Documenting the process.

II.2.4. Assessing Information Loss

Purpose: After the application of protection techniques, there may be a distortion of the information contained in the microdata file. The issue here is to define the analytical value or utility of the dissemination microdata file obtained after the data protection process, compared to the original file. The value of this sub-process also lies in the fact that it may allow for a comparison of different anonymisation methods and thus the choice of the one that best preserves the utility of the files. Elements to be considered in measuring the usefulness of the microdata file may be the number of missing values, number of modified values, ability to reproduce basic statistics (mean, standard deviation, frequencies, etc.), indicators, levels of information details, etc.

This stage also requires assumptions to be made about the use that will be made of the data in order to make an ideal choice of the metric for measuring the information (quantitative or qualitative) lost.

The file processing and information loss assessment procedures are iterative until the balance between protection and utility is achieved.

Expected output: a final microdata file for dissemination and a report on the usefulness of the file.

Guidelines: assessment of the usefulness of a microdata file consists mainly in:

- Identifying the metrics or methods by which information loss should be assessed. The choice of assessment methods depends on the nature of the variables that have been modified (deletion, recoding, perturbation, etc.).
- Evaluating the information in the base file (not anonymised).
- Evaluating the information in the file after risk reduction.
- Choosing an optimal method for the final protection of the microdata file with regard to information loss.
- Preparing a detailed internal report for replication and supervision.
- Preparing a report for users: this does not involve giving users details of the anonymisation procedures and techniques used, but rather informing them of the usefulness of the data made available to them (types of analysis, level of disaggregation for indicators, etc.).

II.3. Retroceding Microdata

Purpose: Once the microdata files have been finalized and validated for dissemination, they should be made available to users. To do this, the terms and conditions of access to these files should be defined, as well as all the conditions of use of the files.

All stages resulting in the dissemination of microdata (preparation, protection, assessment) should be completed before the release of the results of the operation so as to minimize the time lag between the date of release of the results of a statistical operation and the availability of the microdata to users.

Expected output: microdata files accessible to various users

Guidelines: transfer of microdata involves:

- Formalizing the obligations and responsibilities of the user for scientific use files: this involves setting up a protocol for access to microdata files, which establishes the rules for using the files. These rules may be the prohibition for the user to transfer the file to other persons without the agreement of the producing structure, the obligation to submit a copy of the publications or research results made on the data, the obligation to refer to the source for the products obtained, etc.

- Preparing a form for access to the files: this form contains information such as the identity of the user, research project on the data, etc.
- Making the enclave microdata files and related metadata available to secure dissemination centres.
- For public use files, making them available online or by any other means, with access conditional on adherence to the terms and conditions of use.
- Making the report on the processing and usefulness of microdata (public and scientific use files) available to users.
- Establishing a user support mechanism for the centralization and processing of requests. This also makes it possible to monitor the use made of the files, and to ensure, for example, that the user's manipulations do not create identification risks (cross-checking with other data in the structure, or with data from other sources).
- Making users aware of the consequences of non-compliance with the law on the protection of personal data.
- Documenting.

Bibliography

Cameroon, National Institute of Statistics (2018): National Quality Assurance Framework of the National Statistical Information System

Cameroon, National Institute of Statistics (2019): Guideline manual for surveys and censuses, first edition

Cameroon, National Institute of Statistics (2021): Guideline manual for the production of statistics from administrative sources

Cameroon, National Institute of Statistics (2015): National information dissemination and communication policy in Cameroon

International Monetary Fund (2007): The Special Data Dissemination Standard – A guide for subscribers and users

Statistics Canada (2019): Quality guidelines, sixth edition

Olivier Dupriez and Ernie Boyko, IHSN Working Document (2010): Dissemination of Microdata Files. Principles, Procedures and Practices

Julien NICOLAS (2014): Confidentiality Management in Statistical Tables.

Editorial Team

Supervision

1. Tedou Joseph, *Director General, NIS*
2. She Etoundi Joseph, *Deputy Director General, NIS*

Technical coordination

1. Okouda Barnabé, *Head of the Department of Statistical Coordination, Cooperation and Research, NIS*
2. Kana Kenfack Christophe, *Head of the Division of Cooperation, Revision and Applied Research, NIS*
3. Abanda Ambroise, *Head of the Division of Statistical Coordination and Dissemination, NIS*
4. Kamgaing Yougbissi Léonie, *Assistant Research Officer, NIS*

Technical editorial team

- | | |
|---|--------------------------------------|
| 1. Abanda Ambroise, <i>NIS</i> | 12. Fotio Alain Hermann, <i>NIS</i> |
| 2. Tchomthe Séverin, <i>NIS</i> | 13. Modou Sanda, <i>NIS</i> |
| 3. Djenouassi Sébastien, <i>MINEPIA</i> | 14. Gabfoube Victor, <i>MINESEC</i> |
| 4. Essambe Bome Vincent Ledoux, <i>NIS</i> | 15. Mayap Fotso Nina, <i>NIS</i> |
| 5. Tchakoute Ngoho Romain, <i>NIS</i> | 16. Talla Jacques, <i>NIS</i> |
| 6. Tatsinkou Christophe, <i>NIS</i> | 17. Nguimdo Voufo Nadine, <i>NIS</i> |
| 7. Tchamago Olivier, <i>NIS</i> | 18. Ndzomo Nkama Alex, <i>BUCREP</i> |
| 8. Foundjo Eméline, <i>MINTSS</i> | 19. Olinga Martin Paul, <i>NIS</i> |
| 9. Waffo Jean Christophe, <i>MINADER</i> | 20. Djeukwi Vicky Laure, <i>NIS</i> |
| 10. Tiobo'o Papou Sédric Edmond, <i>NIS</i> | 21. Ngah Adèle Zoriphie, <i>NIS</i> |
| 11. Kamgaing Yougbissi Léonie, <i>NIS</i> | 22. Mbenty Jacqueline, <i>NIS</i> |

Administrations that attended the finalization and validation workshop

- | | |
|---|---|
| 1. Mbole Essiane Annie, <i>MINJUSTICE</i> | 10. Ngoube Essama Léon, <i>MINEE</i> |
| 2. Mache Patrice, <i>MINSANTE</i> | 11. Kimoun Temfemo Fatima, <i>MINADER</i> |
| 3. Foko Kamdem Mirabeau, <i>MINCOMMERCE</i> | 12. Ndzomo Nkama Félix, <i>BUCREP</i> |
| 4. Eloundou Adolphe Freddy, <i>MINTP</i> | 13. Bikoula Ondoua Franck, <i>MINDDEVEL</i> |
| 5. Nyam Ahola Djibring, <i>GICAM</i> | 14. Tchatchou Yannick, <i>MINAS</i> |
| 6. Ngwa Edielle T., <i>MINDDEVEL</i> | 15. Apeus Koume, <i>MINPOSTEL</i> |
| 7. Kamgaing Josiane, <i>MINHDU</i> | 16. Ndebi Ntamack Donatien, <i>MINEFOP</i> |
| 8. Djomo Sime Lyse, <i>MINTOUL</i> | 17. Yemene Samuel, <i>CCIMA</i> |
| 9. Dongmo Djiofack Miguel, <i>MINFOF</i> | 18. Ndzomo Nkama Alex, <i>BUCREP</i> |