Compliance Guidelines For Statistical Censuses And Surveys

STATISTICS BOTSWANA

E-mail: info@statsbots.org.bw
Website: http://www.statsbots.org.bw

Compliance Guidelines for Statistical Censuses and Surveys

Published by

Statistics Botswana
Private Bag 0024, Gaborone

March 2021

Contact: Quality Assurance Unit
Tel: 3671300/ Ext 377/ 415
Fax : 3952201
E-mail: info@statsbots.org.bw

Table of contents

Preface

Statistics Botswana and the broader National Statistical System (NSS), are required to produce and disseminate quality statistical products to inform national planning, policy development, monitoring and evaluation of programs. Therefore statistics quality is of high priority and must satisfy the requirements of users within Botswana, regionally and globally. As a result, this document has been developed to provide standards which must be adhered to in the production of quality statistics.

This Compliance Guidelines document contains professional principles and practices that statistical agencies are required to adhere, as well as the level of quality and effort expected in all statistical activities. Each standard has accompanying guidelines that present recommended best practices to fulfil its goals. Thus, the guidelines provide a means to ensure consistency among and within statistical value chain activities and hence the implementation of and the guidelines will ensure that users of statistical information and/or products are provided with details on the principles and methods employed in the development, collection, processing, analysis, dissemination, and preservation of statistical information.

The focus of the guidelines are thus in the processes and quality measurement for statistical outputs through surveys, censuses and administrative records.

Dr. Burton Mguni
Statistician General
March 2021

INTRODUCTION

This document is developed to establish common statistical standards within the National Statistical System. It also provides standards that apply to censuses and surveys whose statistical purposes include the description, estimation, or analysis of the characteristics of groups, segments, activities, or geographic areas in any biological, demographic, economic, environmental, natural resource, physical, social, or other sphere of interest. The development, implementation, or maintenance of methods, technical or administrative procedures, or information resources that support such purposes are also covered by these standards.

In addition, these standards apply to censuses and surveys that are used in research studies or program evaluations if the purpose of the survey meets any of the statistical purposes noted above.

The compliance guidelines apply to all data producers within the National Statistics System. They also provide a high-level review of the extent to which the statistics meet the highest standards of trustworthiness, quality and public value.

Compliance checks are undertaken with one of the following objectives in mind:

- **To identify whether any immediate enhancements are necessary for the producer body to be able to maintain National Statistics status, or whether National Statistics status should be withdrawn**
- **To help evaluate and decide whether a full re-assessment would be appropriate to explore the statistics more deeply**

These standards document the professional principles and practices that statistical agencies are required to adhere to and the level of quality and effort expected in all statistical activities. Agency implementation of standards and guidelines ensures that users of statistical information products are provided with details on the principles and methods employed in the development, collection, processing, analysis, dissemination, and preservation of statistical information.

STANDARDS AND GUIDELINES FOR STATISTICAL SURVEYS

SECTION 1: DEVELOPMENT OF CONCEPTS, METHODS AND DESIGN

1.1 Survey Planning:

The first task in planning a statistical activity is to specify the objectives. A clear statement of objectives guides all subsequent steps and could be revised many times during the survey development. The information needs as stated in the objectives must justify any response burden that will be generated. The relevance of project deliverables to the targeted user community must also be clearly stated.

Agencies initiating a new survey or major revision of an existing survey must develop a written plan that sets forth a justification, including:

- The mandate to collect statistics (statistics act 2009)
- goals and objectives as well as potential users;
- the decisions the survey is designed to inform;
- key survey estimates;
- the precision required of the estimates ,
- the tabulations and the level of detail needed in tabulations as well as analytic results that will inform decisions and other uses;
- related and previous surveys;
- steps taken to prevent unnecessary duplication with other sources of information;
- when and how frequently users need the data;
- Indication of confidential micro data and public-use data files.

The following guidelines represent best practices that must be used in fulfilling the goals of survey planning. All documentation and reports should be made available:

a. An initial step in planning a new survey or a revision of an existing survey should be to contact the sponsoring agency to ensure the survey work is done in compliance with the law and regulations.

b. A justification for the survey, including the rationale for the survey, relationship to previous surveys, survey goals and objectives (including priorities within these goals and objectives), hypotheses to be tested and definitions of key variables. Consultations with potential users to identify their requirements and expectations are also important at this stage of the planning process.

c. A review of related studies, surveys, and reports from other sources to ensure that part or all of the survey would not unnecessarily duplicate available data from an existing source, or could not be more appropriately obtained by adding questions to existing statistical surveys. The goal here is to spend funds effectively and minimize respondent burden. If a new survey is needed, efforts to minimize the burden on individual respondents are important in the development and selection of variables.

A review of confidentiality and privacy acts and all other relevant laws, regulations and guidance is necessary.

d. A review of all survey data items, the justification for each item, and how each item can best be measured, for example, through questionnaires, tests, or administrative records. Agencies should assemble reasonable evidence that these items are valid and can be measured accurately and reliably, or develop a plan for testing these items to assess their accuracy and reliability.

e. A plan for pretesting the survey or survey system.

f. A plan for quality assurance during each phase of the survey process to permit monitoring and assessing performance during implementation. The plan should include contingencies to modify the survey procedures if design parameters appear unlikely to meet expectations (for example, if low response rates are likely). The plan should also contain general specifications for an internal project management system that identifies critical activities and key milestones of the survey that will be monitored, for example, availability of quality control plan, risk management plan, use of Microsoft Project, etc.

g. A plan for evaluating survey procedures, results, and measurement errors. There is need to develop Standard Operating Procedures (SOPs) for the statistical value chain (SVC).

h. An analysis plan that identifies analysis issues, objectives, key variables and proposed statistical tests.

i. An estimate of resources and target completion dates needed for the survey cycle, for example, survey budget and schedule.

j. A dissemination plan that identifies target audiences, proposed major information products, and the timing of their release.

k. A data management plan for the preservation of survey data, documentation of information and products, as well as the authorized disposition of survey records.

l. Maintain a consistent data series over time by using consistent data collection procedures for on-going data collections. Continuous improvement efforts sometimes result in a trade-off between the desire for consistency and a need to improve data collection. If changes are needed in key variables or survey procedures for a data series, consider the justification or rationale for the changes in terms of their usefulness for policymakers, conducting analysis, and addressing information needs; for transition from Paper Assisted Personal Interview (PAPI) to Computer Assisted Personal Interview (CAPI).

m. A project document detailing all processes and Standard Operating Procedures (SOPs), key players and their roles.

Main Dimension: Prerequisite of Quality

Quality Indicators

- The responsibility for producing statistics must be clearly specified.
- Standards and policies must be in place to promote consistency of methods and results.
- Measures must be in place to ensure that individual data are kept confidential, and used for statistical and administrative purposes only.
- Measures to oblige response must be ensured through law.
- Resources must be commensurate with the needs of statistical programmes (staff, facilities, computing resources, connectivity, financing…).
- Measures to ensure efficient use of resources must be implemented.
- Processes must be in place to focus on, monitor and check quality.
- Policies and frameworks must be in place to manage risk in the statistical value chain (business continuity plan, business registers…)

Relevance

Quality Indicators

- The internal and external users of the statistics must be identified
- Changes on the planned survey should be made on the basis of the results of user needs assessment(s)
- A process to determine the satisfaction of users with the statistical information must be in place
- User needs and the usage of statistical information must be available

1.2 Survey Design:

Sampling is a means of selecting a subset of units from a target population for the purpose of collecting information. This information is used to draw inferences about the population as a whole. The subset of units that are selected is called a sample. The sample design encompasses all aspects of how to group units on the frame, determine the sample size, allocate the sample to the various classifications of frame units, and finally, select the sample. Choices in sample design are influenced by many factors, including the desired level of precision and detail of the information to be produced, the availability of appropriate sampling frames, the availability of suitable auxiliary variables for stratification and sample selection, the estimation methods that will be used and the available budget in terms of time and resources.

Agencies must develop a survey design, including

- Defining the target population,
- Specifying the data collection instrument and methods,
- Developing a realistic timetable and cost estimate
- Designing the sampling plan,
- Selecting samples using generally accepted statistical methods; for example, probabilistic methods that can provide estimates of sampling error.
- Any use of non-probability sampling methods such as cut-off or model-based samples must be justified statistically and be able to measure estimation errors.
- The size and design of the sample must reflect the level of detail needed in tabulations and other data products, and the precision required of key estimates.
- Documentation of each of these activities and resulting decisions must be maintained in the project files.

The following guidelines represent best practices that must be used in fulfilling the goals of the survey design:

a. The survey design must include;

i. proposed target population
ii. response rate goals/targets
iii. frequency and timing of data collection, data collection methods
iv. sample design, sample size, precision requirements, and, where applicable include an effective sample size determination based on power analyses for key variables.

NB: Power analysis is the procedure that researchers use to determine if the test contains enough power to make a reasonable conclusion. From another perspective power analysis can also be used to calculate the number of samples required to achieve a specified level of power.

b. Ensure the sample design yield the data required to meet the objectives of the survey. Include the following in the sample design:

i. identification of the sampling frame and the adequacy of the frame;
ii. the sampling unit used ;
iii. sampling strata;
iv. power analyses to determine sample sizes and effective sample sizes for key variables by reporting domains (where appropriate);
v. criteria for stratifying or clustering, sample size by stratum, and the known probabilities of selection;
vi. response rate goals;
vii. estimation and weighting plan;
viii. variance estimation techniques appropriate to the survey design; and
ix. expected precision of estimates for key variables.

c. When a non-probabilistic sampling method is employed, include the following in the survey design documentation:

i. a discussion of options considered and why the final design was selected,
ii. an estimate of the potential bias in the estimates,
iii. the methodology to be used to measure estimation error. In addition, detail the selection process and demonstrate that units not in the sample are impartially excluded on objective grounds in the survey design documentation.

d. Include a pledge of confidentiality (if applicable), along with instructions required to complete the survey. A clear, logical, and easy-to-follow flow of questions from a respondent's point of view is a key element of a successful survey.

e. Include the following in the data collection plans:

i. frequency and timing of data collection;
ii. method of collection for achieving acceptable response rate;
iii. training of surey personnel such as enumerators, editors and coders) and cost estimates, including the costs of pre-tests, non-response follow-up, and evaluation studies.

f. Construct an estimate of total mean square error in approximate terms, and evaluate accuracy of survey estimates by comparing with other information sources. If probability sampling is used, estimate sampling error

g. Estimate the effects of potential non sampling errors including measurement errors due to interviewers, respondents, instruments, and mode; nonresponse error; coverage error; and processing error.

Main Dimension: Methodological Soundness

Quality Indicators

- The scope of the study must be consistent with accepted standards, guidelines or good practices.
- Methodologies used should follow acceptable standards, guidelines or best practices (international, national standards) with respect to questionnaire design, sampling methods, piloting, data collection, processing, imputation and analysis methods as well as revision procedures.
- Concepts, definitions and classifications used must follow accepted standards, guidelines or good practice (international, regional and national).
- Revision schedules should be followed, regular and transparent; and studies of revisions and their findings should be made public

Other: Timeliness and Punctuality

Quality Indicators

- Production activities within the statistical value chain must be within the planned timelines (data collection, processing, analysis, dissemination and archiving).
- Periodicity of release must be indicated.

1.3. Developing Sampling Frames:

A frame is any list, material or device that delimits, identifies, and allows access to the elements of the survey population. Frames are generally of two types: area frames and list frames. A list frame is a list of units in the survey population. Area frames are usually made up of a hierarchy of geographical units which in turn contain units in the survey population; that is, the frame units at one level can be subdivided to form the units at the next level. All of the elements included in the frame constitute the frame population. Frames are often much more than a simple list of units or a map with geographic units delineated. A frame usually includes other information (e.g., identification, contact, classification, address, size, maps in case of geographical units) to be used in carrying out the survey.

Agencies must ensure that the frames for the planned sample surveys (whether household or business based) or census are appropriate for the study design and evaluated against the target population for quality.

The following guidelines represent best practices that must be used in fulfilling the goals of **developing sampling frames**:

a. Describe target population and sampling frame.

b. Conduct periodic evaluation of coverage rates and coverage of the target population in survey frames that are used for recurring surveys. Coverage rates in excess of 95

percent overall and for each major stratum are desirable. If coverage rates fall below 95 percent, conduct an evaluation of the potential bias.

c. Consider using frame enhancements to increase coverage particularly in business surveys frame.

Main Dimension: Methodological Soundness

Quality Indicators

- Methodologies used for sampling frame development and maintenance should follow acceptable standards, guidelines or best practices (international, national standards).

1.4. Survey Response Rates:

Response here refers to all data obtained either directly from respondents or from administrative data. This broad definition of response is necessary to reflect the increased use of different collection strategies in the same survey. As with survey data, administrative data is not exempted from non-response, whether it be partial or total. Non-response is sometimes the result of lateness in obtaining all the administrative data.

Agencies must design the survey to achieve the highest response rates which commensurate with the importance of survey uses, respondent burden, and data collection costs. This will ensure that survey results are representative of the target population so that they can be used with confidence to inform decisions. Non-response bias analysis must be conducted when unit or item response rates or other factors suggest the potential for bias to occur.

The following guidelines represent best practices that may be useful in fulfilling the goals of achieving the **highest response rates**:

a. When designing the survey, use previous experience with similar surveys, the total budget and allocation of the budget to various operations for guidance;

b. Design data collection tools/methods that will be used for sample frames for other surveys to meet a target unit response rate of at least 75 percent for business surveys and at least 95 percent for household based surveys;

c. Prior to data collection, identify expected unit response rates at each stage of data collection, based on content, use, mode, and type of survey;

d. Plan for a non-response bias analysis if the expected unit response rate is below 75 or 95 percent or depending on the base survey;

e. Plan for a non-response bias analysis if the expected item response rate is below 95 percent for any items used in a report;

f. Identify the survey frame's quality (in terms of population coverage and the facility of establishing contact with respondents), the population observed and the sampling method;

g. Specify the data collection method (for example, by mail, personal interview or

computer-assisted telephone interview, by electronic data reporting (EDR), the Internet or a combination of methods), the time of year and the length of the collection period;

h. Develop the communication strategy to be used to inform respondents of the importance of the survey and to maintain a relationship with respondents;

i. Identify the response burden imposed (length of interview, difficulty of subject matter, timing and interview periodicity); the subject's nature and sensitivity, questionnaire length and complexity; questionnaire language and respondents' cultural backgrounds;

j. Collect staff's prior experience and skills in interpersonal relationships; their workload; factors related to the interviewers themselves, such as training; and potential staff turnover;

k. Specify the effectiveness and scope of follow-up methodology and expected difficulties in tracing respondents who have moved.

Main Dimension: Methodological Soundness

Quality Indicators

- The scope of the study must be consistent with accepted standards, guidelines or good practices.
- Methodologies used with respect to questionnaire design, sampling methods, frame maintenance, data collection, as well as revision procedures should follow acceptable standards, guidelines or best practices (international, national standards)
- Revision methods used should follow accepted standards, sets of guidelines or good practice. Similarly, revision schedules must be followed, regular and transparent.

Others: Timeliness and Punctuality

Quality Indicators

- Production activities within the statistical value chain must be within the planned timelines (data collection).

Accuracy

Quality Indicators

- Measures of non-sampling errors with respect to frame coverage errors and misclassifications must be calculated.
- Register/frame maintenance procedures must be adequate (updates, quality assurance and data audits).

Credibility

Quality Indicators

- The terms and conditions including confidentiality, under which statistics are collected, processed and disseminated, must be availed to the public.

1.5. Pre-testing Survey Systems:

Agencies must ensure that all components of a survey function as intended when implemented in the full-scale survey and measurement error are controlled by conducting a pre-test of the survey components.

The following guidelines represent best practices that must be used in fulfilling the goals of the Pre testing Survey system.

a. Test new components of a survey using methods such as cognitive testing i.e. measuring trust in official statistics, focus groups, and usability testing, where applicable. Prior to a field test of the survey system, incorporate the results from these tests into the final design.

b. Use field tests prior to implementation of the full-scale survey when some or all components of a survey system cannot be successfully demonstrated through previous work. The design of a field test should reflect realistic conditions, including those likely to pose difficulties for the survey. Elements to be tested include, for example, frame completeness, sample selection, questionnaire design, survey personnel capabilities, data collection, item feasibility, electronic data collection capabilities, edit specifications, data processing and estimation, file creation, and tabulations. A complete test of all components (sometimes referred to as a dress rehearsal) is highly desirable for all surveys.

Main Dimension: Methodological Soundness

Quality Indicators

- Methodologies used should follow acceptable standards, guidelines or best practices (international, national standards) with respect to piloting or pretesting (sampling design, questionnaire design, sampling methods, as well as data collection, processing, imputation and analysis methods).

2.1 Required Notifications to Potential Survey Respondents:

Agencies must ensure that each collection of information instrument clearly states the reasons the information is planned to be collected;

- the way such information is planned to be used to further the proper performance of the functions of the agency;
- whether responses to the collection of information are voluntary or mandatory (citing authority);
- the nature and extent of confidentiality to be provided, if any, citing authority;

The following guideline represents best practices that must be used in fulfilling the goal of notifying potential survey respondents

a. Provide appropriate informational materials to respondents, addressing respondent burden as well as the scope and nature of the questions to be asked. The materials may include a pre-notification letter, brochure, a set of questions and answers as well as media coverage.

Main Dimension: Credibility

Quality Indicators

- Advance notice should be given of major changes in methodology and source data

2.2 Data Collection Process

Agencies must design and administer data collection instruments and methods in a manner that achieves the best balance between maximizing data quality and controlling measurement error while minimizing respondent burden and cost.

The following guidelines represent best practices that must be used in fulfilling the goals of the data collection process:

a. Design the data collection instrument in a manner that minimizes respondent burden, while maximizing data quality. The following strategies may be used to achieve these goals;

i. the questions are clearly written and skip patterns are easily followed;
ii. the questionnaire is of reasonable length;
iii. the questionnaire is pretested to identify problems with interpretability and ease in navigation;
iv. Methods to reduce item and unit non-responses are adopted.

b. Encourage respondents to participate in order to maximize response rates and improve data quality. The following data collection strategies must be used to achieve high response rates;

ii. ensure that the data collection period is of adequate and reasonable length;
iii. send materials describing the data collection to respondents in advance, where possible;
iv. Plan an adequate number of contact attempts (call backs); train trainers (accredited with further training), enumerators, supervisors, quality controllers and other staff who have contact with respondents in techniques for obtaining respondent cooperation and building rapport with respondents. Techniques for building rapport include respect for respondents' rights, follow-up skills, knowledge of the goals and objectives of the data collection, and knowledge of the uses of the data,

c. The way data collection is designed and administered also contributes to data quality. The following issues are important to consider: The agency should;

i. give the characteristics of the target population, the objectives of the data collection, resources available, and time constraints to determine the appropriateness of the method of data collection;
ii. collect data at the most appropriate time of year, when relevant;
iii. establish the data collection Standard Operating Procedures (SOP) to be followed by the survey staff;
iv. provide training for field staff on new protocols, with refresher training on a routine basis or recurring cycle;
v. establish best practice mechanisms to minimize interviewer falsification, such as protocols for monitoring interviewers and re-interviewing respondents;
vi. conduct response analysis surveys; i.e. ensure that response rates have reached the internationally accepted levels
vii. establish protocols that minimize measurement error, such as conducting response analysis surveys to ensure records exist for data elements requested, establishing recall periods (reference points)that are reasonable for surveys, and developing computer systems to ensure Internet data collections function properly; and quantify non-sampling errors to the extent possible.

d. Develop protocols to monitor data collection activities, with strategies to correct identified problems. The following must be adhered to;

i. Implement quality and performance measurement and process control systems to monitor data collection activities and integrate them into the data collection process. Thus, managers should be able to identify and resolve problems and ensure that the data collection is successfully completed. Additionally, these measurements should provide survey designers and data users with indicators of survey performance and resultant data quality;
ii. Use internal reporting systems that provide timely reporting of response rates and the reasons for non-response throughout the data collection.

Main Dimension: Pre Requisite of Quality

Quality Indicators

- The responsibility for producing statistics must be clearly specified
- Measures to oblige response must be ensured through law
- Measures should be in place to ensure that individual data are kept confidential and used for statistical and administrative purposes only
- Standards must be in place to promote consistency of methods and results
- Resources in terms of trained and skilled personnel, facilities, computing resources and financing should commensurate with the needs of statistical programmes
- Measures to ensure efficient use of resources should be implemented
- Policies and frameworks should be in place to manage risk in statistical value chain (data collection)
- Processes must be in place to focus on, monitor and check quality

Others: Credibility

Quality Indicators

- Choice of source data, techniques and dissemination decisions should be informed solely by statistical considerations
- Data sharing and coordination among data-producing agencies must be clearly specified.
- Ethical guidelines for staff behaviour should be put in place and well known to the staff

Accuracy

Quality Indicators

- Data collection systems must be sufficiently open and flexible to cater for new developments.
- Measures of non-sampling errors in terms of systematic and measurement errors must be kept to acceptable levels

2.3 Administrative data

Administrative data is generated during the process of executing administrative functions to inform policy and decision-making. For it to serve a statistical purpose, there is need to compile administrative data following a systematic, objective and standardized approach. Statistical use of administrative data include creation and maintenance of frames, the complete or partial (via record linkage) replacement of statistical collection, editing, imputation and weighting of data from statistical collection and evaluation of statistical outputs.

The following guidelines should be followed in administrative data production;

a. The agency or organization should describe the primary purpose of data collection through administrative sources;
b. Actively investigate and assess all potential sources of administrative data;
c. Describe the format in which the administrative data are available;
d. Describe and document the concepts, definitions and procedures underlying the collection of administrative data;
e. Each administrative dataset should be accompanied by metadata about its contents so that users assess their suitability for their purposes;
f. Describe the main uses of administrative data, including where applicable the statistical processes and/or outputs that require data from administrative sources;
g. Describe the extent of coverage of the administrative data and any known coverage problems;
h. Describe the known sources of error in administrative data;
i. Describe the timescale since the last update of data from administrative sources;
j. Describe the common identifiers of population units in administrative data;
k. Set up an edit and imputation procedure or a weight adjustment procedure to deal with non-response;
Describe the extent to which the data from administrative sources meets statistical requirements;
m. Describe any changes in the legislative environment through which the administrative data are provided and the effects on the statistical product;
n. The confidentiality implications of the publication of information from administrative records should be considered;
o. Maintain continuing collaboration with the providers and suppliers of administrative data; and

p. Implement continuous or periodic quality assessments for administrative data production.

Dimension: Pre Requisite of Quality

Quality Indicators

- The responsibility for producing statistics must be clearly specified
- Measures to oblige response must be ensured through law
- Measures should be in place to ensure that individual data are kept confidential and used for statistical and administrative purposes only
- Data sharing and coordination among data-producing agencies should be clearly specified.
- Standards and policies must be in place to promote consistency of methods and results
- Processes must be in place to focus on, monitor and check quality

Others: Accuracy

Quality Indicators

- Measures of non-sampling errors with respect to frame coverage errors and misclassifications must be calculated.
- The extent to which the primary data is appropriate for the statistical product produced must be assessed
- Data from the primary source must be quality assessed.
- Describe record-matching methods and techniques used on the administrative data sources.
- Quality report should accompany administrative data
- Register/frame maintenance procedures must be adequate (updates, quality assurance and data audits).

Credibility

Quality Indicators

- Choice of source data, techniques and dissemination decisions should be informed solely by statistical considerations
- Data sharing and coordination among data-producing agencies must be clearly specified.
- The terms and conditions, including confidentiality, under which statistics are collected, processed and disseminated should be availed to the public.

Accessibility

Quality Indicator

- Rules governing the restricted availability of administrative records are well described and documented

SECTION 3: **DATA PROCESSING**

3.1 Data editing:

Agencies must edit data based on editing manuals to mitigate or correct detectable errors.

The following guidelines represent best practice that must be used in fulfilling the goals of Data Editing;

a. Check and edit data to mitigate errors. Data editing is an iterative and interactive process that includes procedures for detecting and correcting errors in the data. Editing uses available information and some assumptions to derive substitute values for inconsistent entries in a data file. When electronic data collection methods are used, data are usually edited both during and after data collection. Include results from analysis of data and input from subject matter specialists in the development of edit rules and edit parameters.

b. As appropriate, check data for the following and edit if errors are detected;

- responses that fall outside a pre-specified range (e.g., based on expert judgment or previous responses) or, for categorical responses that are not equal to specified categories;
- Consistency;
- Contradictory responses and incorrect flow through prescribed skip patterns;
- Missing data that can be directly filled from other portions of the same record (including the sample frame);
- The omission and duplication of records; and
- Inconsistency between estimates with outside sources.

c. Possible actions for failed edits include the following:

- Automated correction within specified criteria and data verified by respondent.
- Corrected data provided by respondents;
- Corrected data available from other sources;
- If unable to contact respondent and after review by survey staff, an imputed value may be substituted for a failed edit.

d. Code the data set to indicate any actions taken during editing for reference purposes, and/or retain the unedited data along with the edited data.

e. Prepare electronic edit programs for further editing after data entry. Normally the programs correct inconsistencies within and between sections of the questionnaire (e.g. a male having children). In this regard the program should be based on the edit specification prepared.

3.2 Coding:

Agencies must add codes to collected data to identify aspects of data quality from the collection (e.g., missing data) in order to allow users to appropriately analyse the data. Codes added to convert information collected as text into a form that permits immediate analysis must use standardized codes(ISCED , ISCO & ISIC etc.), when available, to enhance comparability.

The following guidelines represent best practices that must be used in fulfilling the goals for **Coding**:

a. Insert codes into the data set that clearly identify missing data and cases where an entry is not expected (e.g. skipped over by skip pattern). Do not use blanks and zeros as codes to identify missing data, as they tend to be confused with actual data.

b. When converting text data to codes to facilitate easier analysis, use standardized codes, if they exist, otherwise use relevant coding schemes that suit your purpose. Develop other types of codes using existing agency practice (for example locally developed codes, like Institutional types and employment type codes in SBR) or standard codes from industry or international organizations, if they exist.

c. When setting up a manual or electronic coding process to convert text to codes, create a quality assurance process that verifies at least a sample of the coding to determine if a specific level of coding accuracy is maintained. For instance, double checking of codes by Quality Controllers and Headquarters personnel after field supervisors have coded.

3.3 Data Entry

Agencies must enter/capture data where electronic data collection is not implemented. The following guidelines represent best practices that must be used in fulfilling the goals of **Data Entry**;

a. Design the data capture process in order to reduce capture cost and maximise timeliness and accuracy of the data. Data items could be captured during survey collection by the respondents through the use of Web Based Interviews (WABI) or interviewers through the use of Computer Assisted Personal Interviews (CAPI). This obviously reduces the cost of data capture, increases the timeliness and has the potential of improving accuracy through edit rules being integrated into the computer application. When it is not feasible to integrate data capture with collection, the capture is performed either by operators (manual key entry) or in an automated fashion (scanning followed by Intelligent Character Recognition). The latter is preferred as it reduces cost and often enhances accuracy of the data.

b. For CAPI interviewers who often perform data capture and coding during collection;

- use standard collection tools and process (e.g. standard screens and standardized questions) to ease interviewer work and limit the risk of introducing capture errors;
- Integrate edit rules in the collection system to validate the entry of data items and allow for potential corrections of errors (i.e. keying error, response error, missing items) at the time of collection.

c. Data capture operators are critical to the success of the capture operations. Agencies should;

- Ensure that data capture operators have appropriate training and tools.
- Prepare training material and procedures for the operators and deliver training sessions. This will enhance the skills of the staff and thus ensure accurate capture of data collected.
- Use quality control methods to verify whether the accuracy of capture performed by operators meets the pre-established levels and
- Provide them with feedback for improvement.

d. Manual data capture from paper questionnaires or scanned images is subject to entry errors, therefore;

- Incorporate online edits for error conditions that the data capture operator can correct (i.e. edits that will identify keying errors).
- Record these cases for later review and analysis.
- Test the manual operation prior to conducting the survey.

e. For automated data capture, ensure that the questionnaire is designed to ease the scanning and the intelligent character recognition. This refers to a hand writing recognition system that allows fonts and handwriting to be learned by a computer during processing to improve accuracy and recognition levels

f. When automated capture is used, some questionnaires cannot be scanned and others can be scanned but characters cannot be recognised.

g. Data capture operators should also be used to conduct a sample study assessment of the accuracy of automated capture. The results of such a study can be used to improve the process.

h. Institute effective control of systems to ensure the security of data capture, transmission and handling, especially with new technologies such as cell phone and Internet data collection. This would prevent loss of information and the resulting decline in quality, and potential credibility due to system failures or human errors.

i. Develop procedures for destroying the data when no longer needed.

3.4 Data Protection:

Agencies must implement safeguards throughout the production process to ensure that survey data are handled to avoid disclosure.

The following guidelines represent best practices that must be used in fulfilling the goals of Data Protection;

a. For purposes of surveys, establish procedures and mechanisms to ensure confidentiality and that the information is protected during the production, use, storage, transmittal, and disposition of the survey data in any format.

b. Ensure that individually-identifiable survey data are protected; data systems and electronic products are protected from unauthorized intervention (persons); and data files, network segments, servers, and desktops or PCs are electronically secure from malicious software and intrusion using best available information resource security practices that are periodically monitored and updated.

c. Ensure controlled access to data sets so that only specific, named individuals working on a particular data set can have read only, or write only, or both read and write access to that data set. Data set access rights are to be periodically reviewed by the project manager responsible for that data set in order to guard against unauthorized release or alteration. For instance, in the QMTHS, only Enumerators had access rights to make corrections whereas the Supervisors had access rights to code only as well as Quality Controllers; while Headquarters personnel had access rights to 'view the data only'.

3.5 Evaluation:

Agencies must evaluate the quality of the data and make the evaluation public (through technical notes and documentation included in reports of results or through a separate report) to allow users to interpret results of analyses and to help designers of recurring surveys focus on improvement efforts.

The following guidelines represent best practices that must be used to fulfil the goal of Evaluation;

a. Include an evaluation component in the survey plan that evaluates survey procedures, results, and measurement error. Review past surveys similar to one being planned to determine likely sources of error, appropriate evaluation methods, and problems that are likely to be encountered.

b. Where appropriate, develop and implement methods for estimating the non-sampling error from each source identified in the evaluation plan.

Dimensions: Accuracy

Quality Indicators

- Measures of non-sampling errors are calculated, vitz: terms of frame **coverage, misclassifications, systematic, measurement and processing errors.**

Methodological Soundness

Quality Indicators

- Concepts, definitions, and classifications used must follow accepted standards, guidelines or good practice (international, regional, national,).
- Methodologies used must follow accepted standards, guidelines or good practice (international, regional, national standards) - **in terms of editing, coding, capture and imputation.**

Comparability and Coherence

Quality Indicators

- Ethical data within series and administrative systems are based on common concepts, definitions, classification, and methodology. Departures from this can be identified in the metadata.
- Data across comparable series, or source data must be based on common frames, identifiers, concepts and definitions, and classifications, and departures from these are identified in the metadata.

Timeliness and Punctuality

Quality Indicators

- Production activities within the statistical value chain are within the planned timelines in terms of data processing (editing, coding, capture and imputation).

SECTION 4: **PRODUCTION OF ESTIMATES AND PROJECTIONS**

A typical survey objective is to estimate descriptive population parameters, as well as analytical parameters, on the basis of a sample selected from a population of interest. Examples of parameters include simple descriptive statistics such as totals, means, ratios and percentiles. Examples of analytical parameters include regression coefficients, correlation coefficients and measures of income inequality.

In a probability-based survey, a design weight is associated with each sampled unit. The design weight can be interpreted as the number of typical units in the survey population that each sampled units represents. Estimates can be calculated using the design weights or estimation weights obtained by adjusting the design weights. Common adjustments include those that account for non-response and that incorporate auxiliary information.

4.1 Developing Estimates and Projections:

Agencies must use accepted theory and methods when deriving direct survey based estimates, as well as model-based estimates and projections that use survey data.

- Error estimates must be calculated and disseminated to support assessment of the appropriateness of the uses of the estimates or projections.
- Agencies must plan and implement evaluations to assess the quality of the estimates and projections.

The following guidelines represent best practices that must be used in fulfilling the goals of **Developing Estimates and Projections:**

a. Develop direct survey estimates according to the following practices:

i. Employ weights appropriate for the sample design to calculate population estimates. However, an agency may employ an alternative method (e.g., ratio estimators) to calculate population estimates if the agency has evaluated the alternative method and determined that it leads to acceptable results.

ii. Calculate variance estimates by a method appropriate to a survey's sample design taking into account probabilities of selection, stratification, clustering, and the effects of non-response, post-stratification, and raking. The estimates must reflect any design effect resulting from a complex design.

b. Develop model-based estimates according to accepted theory and practices (e.g. assumptions, mathematical specifications).

c. Develop projections in accordance with accepted theory and practices (e.g. assumptions, mathematical specifications).

d. Subject any model used for developing estimates or projections to the following:

i. Sensitivity analysis to determine if changes in key model inputs cause key model outputs to respond in a sensible fashion; (recalculating outcomes under alternative model assumptions to determine the impact of the variables)

ii. Model validation to analyse a model's performance by comparing the results to available independent information sources; and

iii. Demonstration of reproducibility to show that, given the same inputs, the model produces similar results.

e. Prior to producing estimates, establish criteria for determining when the error (both sampling and non-sampling) associated with a direct survey estimate, model-based estimate, or projection is too large to publicly release the estimate/projection.

f. Document methods and models used to generate estimates and projections to help ensure objectivity, utility, transparency, and reproducibility of the estimates and projections. Also, archive data and models so the estimates/projections can be reproduced.

Dimension: Accuracy

Quality Indicators

- Measures of non-sampling errors must be calculated and be within acceptable levels in terms of;
  - imputation methods,
  - model assumptions,
  - response rates,
  - frame coverage errors,
  - misclassification errors,
  - systematics errors,
  - measurement errors and
  - data processing errors.

- Measures of sampling errors for key variables must be calculated, in terms of
  - standard error (SE),
  - Coefficient of variation (CV),
  - Mean Square Error (MSE),
  - Design Effect (DEFF) and
  - Intra Cluster Correlation (ICC).

SECTION 5: **DATA ANALYSIS**

Data analysis is the process of developing answers to questions through the examination and interpretation of data. The basic steps in the analytic process consist of identifying issues, determining the availability of suitable data, deciding on which methods are appropriate for answering the questions of interest, applying the methods and evaluating, summarizing and communicating the results.

Data analysis is essential for understanding results from surveys, administrative sources and pilot studies; for providing information on data gaps; for designing and redesigning surveys; for planning new statistical activities; and for formulating quality objectives.

5.1 Analysis and Report Writing:

Agencies must develop a plan for the analysis of survey data prior to the start of a specific analysis to ensure that statistical tests are used appropriately and that adequate resources are available to complete the analysis.

The following guidelines represent best practices that must be used in fulfilling the goals of **Analysis and Report Writing**:

a. Include the following in the analysis plan:

i. An introduction that describes the purpose (objectives of the survey), the research question, relevant literature, data sources (including a brief description of the survey data and any limitations of the data), key variables to be used in the analysis, type of analysis, and significance level to be used; and

ii. A framework for technical notes including, as appropriate, the history of the survey program, data collection methods and procedures, sample design, response rates and the treatment of missing data, weighting methods, computation of standard errors, instructions for constructed variables, limitations of the data, and sources of error in the data.

5.2 Inference and Comparisons:

Agencies must base statements of comparisons and other statistical conclusions derived from survey data on acceptable statistical practice.

The following guidelines represent best practices that must be used in fulfilling the goals of Inference and Comparisons:

a. Specify the criterion for judging statistical significance for tests of hypotheses (Type I error) before conducting the testing.

b. Before including statements in information products that two characteristics being estimated differ in the actual population, make comparison tests between the two estimates, if either is constructed from a sample. Use methods for comparisons appropriate for the nature of the estimates.

c. When performing multiple comparisons with the same data between subgroups, include a note with the test results indicating whether or not the significance criterion (Type I error) was adjusted and, if adjusted, by what method.

d. When performing comparison tests, test and report only the differences that are substantively meaningful (i.e., don't necessarily run a comparison between every pair of estimates; run only those that are meaningful within the context of the data, and report only differences that are large enough to be substantively meaningful, even if other differences are also statistically significant).

e. Given a comparison that does not have a statistically significant difference, conclude that the data do not support a statement that they are different. If the estimates have apparent differences, but have large standard errors making the difference statistically insignificant, note this in the text or as a note with tables or graphs.

f. Support statements about monotonic trends (strictly increasing or decreasing) in time series using appropriate tests. If extensive seasonality, irregularities, known special causes, or variation in trends are present in the data, take those into account in the trend analysis.

g. If part of a historical series is revised, data for both the old and the new series should be published for a suitable overlap period for the use of analysts.

Main Dimension: Interpretability

Quality Indicators

- Documented metadata (definitional, operational, methodological, system and dataset) must be sufficient to understand data.
- Statistics must be presented in a simple and understandable manner.
- Statistical releases must contain a summary of the key findings as defined in the major objectives.

SECTION 6: **REVIEW PROCEDURES**

6.1 Review of Information Products (Editorial process):

Agencies are responsible for the quality of information that they disseminate and must institute appropriate content/subject matter, statistical and methodological review procedures to comply with agency Information Quality Guidelines.

The following guidelines represent best practices that must be used in fulfilling 7 the goals of Reviewing of Information Products:

a. Conduct a content/subject-matter review of all information products that present a description or interpretation of results from the survey, such as analytical reports or "briefs." Select reviewers with appropriate expertise in the subject matter, operation, or statistical program discussed in the document.

b. Conduct a statistical and methodological review of all information products. Select reviewers with appropriate expertise in the methodology described in the document (auditing officers).

c. Ensure that any product that is to be disseminated via special software is tested for accessibility and interpretability prior to dissemination.

Main Dimension: Interpretability

Quality Indicator

- Statistics must be presented in a simple and understandable manner.

SECTION 7: **DISSEMINATION OF INFORMATION PRODUCTS**

7.1 Releasing Information

Agencies must release information intended for the general public according to a dissemination plan that provides for equivalent, timely access to all users and provides information to the public about the agencies' dissemination policies and procedures including those related to any planned or unanticipated data revisions.

The following guidelines represent best practices that must be used in fulfilling the goals of Releasing Information:

a. Dissemination procedures for major information products include the following:

i. Develop schedule (Release Calendar) and mode for the release of information products;
ii. Inform targeted audiences; and
iii. Ensure equivalent, timely access to all users.

b. Protect information against any unauthorized pre-release, and release information only according to established release procedures and protocols.

c. If revisions to estimates are planned, establish a schedule for anticipated revisions, make it available to users, and identify initial releases as preliminary.

d. Establish a policy for handling unscheduled corrections due to previously unrecognized errors. The policy may include threshold criteria (e.g., the correction will change a national level total value by more than one percent or a regional value by more than five percent) identifying conditions under which data will be corrected and re-disseminated.

e. When information products are disseminated, provide users access to the following information:

i. Definitions of key variables;
ii. source information, such as a survey form number and description of methodology used to produce the information or links to the methodology;
iii. Quality-related documentation such as conceptual limitations and non-sampling errors;
iv. Variance estimation documentation;
v. Time period covered by the information and units of measurement;
vi. Data taken from alternative sources;
vii. Point of contact to whom further questions can be directed;
viii. Software or links to software needed to read/access the information and installation/operating instructions, if applicable;
ix. Date the product was last updated; and
x. Standard dissemination policies and procedures.

7.2 Data Protection and Disclosure Avoidance for Dissemination

When releasing information products, agencies must ensure strict compliance with any confidentiality pledge to the respondents and all applicable legislation and regulations.

The following guidelines represent best practices that must be used in fulfilling the goals of **Data Protection and Disclosure:**

a. For survey information collected under a pledge of confidentiality, employ sufficient procedures and mechanisms to protect any individually-identifiable data from unauthorized disclosure.

b. Do not publicly reveal parameters associated with disclosure limitation rules.

7.3 Survey Documentation

Agencies must produce survey documentation that includes those materials necessary to understand how to properly analyse data from each survey, as well as the information necessary to replicate and evaluate each survey's results. Survey documentation must be readily accessible to users, unless it is necessary to restrict access to protect confidentiality.

The following guidelines represent best practices that must be used in fulfilling the goals of **Survey Documentation**:

a. Survey system documentation includes all information necessary to analyse the data properly

b. Retain all survey documentation according to appropriate records disposition and archival policy.

7.4 Documentation and Release of Public-Use Micro data:

Agencies that release micro data to the public must include documentation clearly describing how the information is constructed and provide the metadata necessary for users to access and manipulate the data. Public-use micro data documentation and metadata must be readily accessible to users.

The following guidelines represent best practices that must be used in fulfilling the goals of Documenting and Releasing of Public-Use Micro data:

a. Provide complete documentation for all data files.
b. Provide a file description and record layout for each file. All variables must be clearly identified and described.
c. Make all micro data products and documentation accessible to users with generally available software.
d. Clearly identify all imputed values on the data file.
e. Release public-use micro data as soon as practicable to ensure timely availability for data users.
f. Retain all micro data products and documentation according to appropriate records disposition

---

Main Dimension: Accessibility Timeliness and Punctuality

Quality Indicators

- A data dissemination policy must exists and be accessible
- A pricing policy must exists and be accessible
- Metadata must be readily available
- Catalogues of publications and other services must be made available to users of statistics
- Statistics be released according to the release calendar

- Data must be accessible in a user friendly format.
- Statistical products must be made available to all users at the same time
- Statistical products (e.g. data, metadata) must be made available to the public.
- Types of media and/or channels used for sharing data amongst stakeholders must be adequate and preserve confidentiality.
- User support services must exist and be publicized
- Statistics/administrative records not routinely disseminated must be made available upon request

Others: Timeliness and Punctuality

Quality Indicators

- Production activities within the statistical value chain must be within the planned timelines in terms of data dissemination
- Average time between the end of the reference period and the date of preliminary and final results must be in accordance with prescribed standards
- Periodicity of releases must be clearly specified as per the dissemination best practices

Interpretability

Quality Indicator

- Documented metadata (definitional, operational, methodological, system and dataset) must be sufficient to understand data

SECTION 8: **EVALUATION OF THE SURVEY**

Agencies must evaluate data appropriately, based on available information, evaluation procedures, to determine the appropriateness of the estimates and measures to correct detectable errors.

The following guidelines represent best practices that must be used in fulfilling the goals of evaluating the survey:

a. For recurring surveys, produce a periodic evaluation report, such as a methodology report, that itemizes all sources of identified errors.

b. Provide estimates or bounds on the magnitudes of these errors; discuss the total error model for the survey; and assess the survey in terms of this model.

c. Evaluate all survey processes including planning, data collection, processing, production of estimates, analysis, dissemination, and archiving, to determine likely sources of error, appropriate evaluation methods, and problems that are likely to be encountered in the preceding surveys.

Main Dimension: Accuracy

Quality Indicators

- Measures of sampling errors must be calculated for the main variables. They must be available for the other variables on request.
- Measures of sampling errors must fall within acceptable standards. At a minimum the following must be calculated: standard error, coefficient of variation, confidence interval, mean square error. The low accuracy of variables (if these exist), are explained.
- The extent of measures of non-sampling errors must be kept to an acceptable level.

SECTION 9: **ARCHIVING OF SURVEY INFORMATION**

Agencies must verify data appropriately, based on available information, verifying manuals to mitigate or correct detectable errors as per archiving rules.

The following guidelines represent best practices that must be used in fulfilling the goals of archiving of survey information:

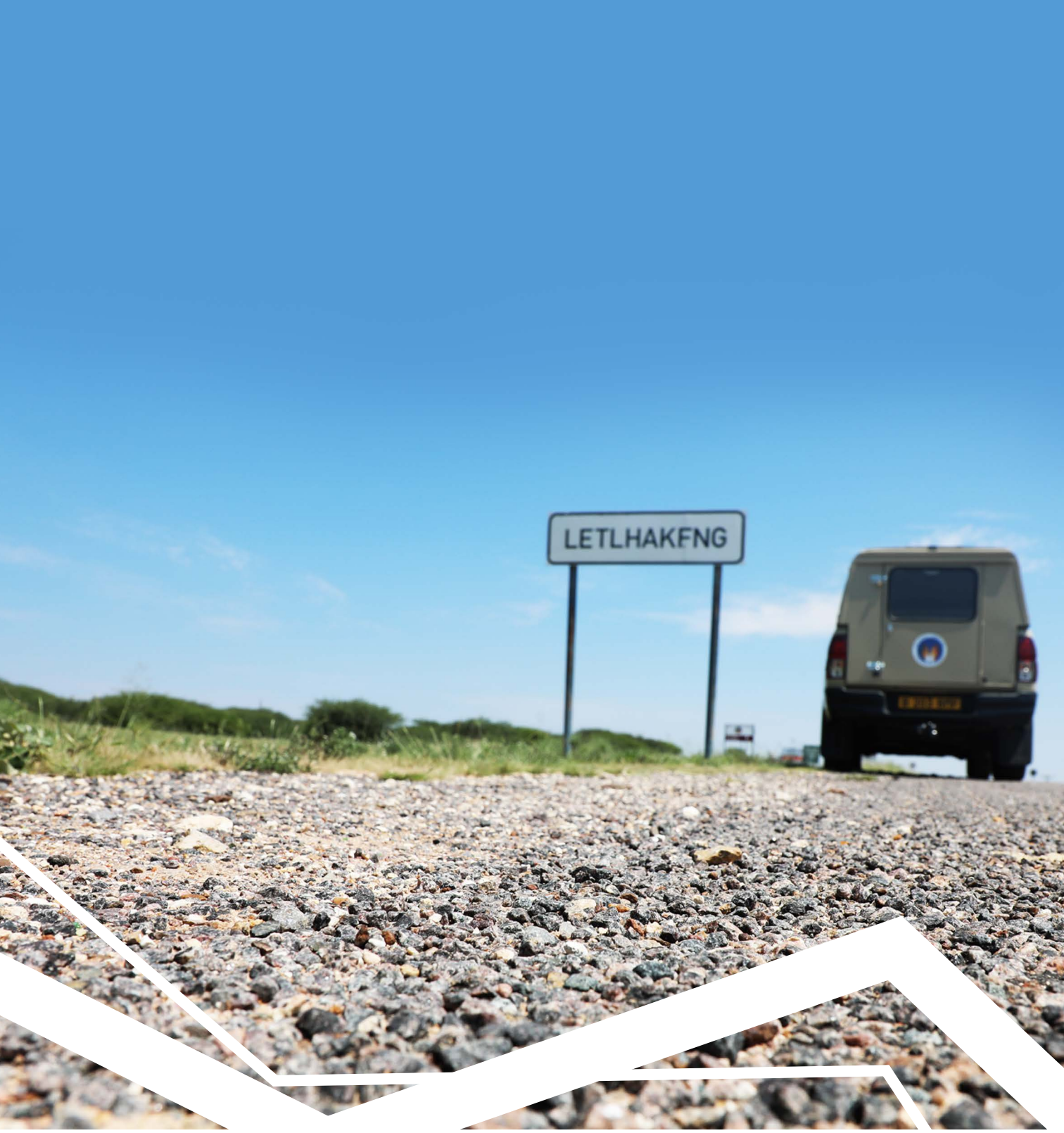
a. Archive data as appropriate, so that data are available for historical research in future years.

b. Retain all micro data products according to appropriate archival policy.

Main Dimension: Timeliness and Punctuality

Quality Indicator

- Production activities within the statistical value chain are within the planned timelines with respect to archiving



Private Bag 47
Maun
Tel: 371 5716
Fax: 686 4327

Private Bag 32
Ghanzi
Tel: 371 5723
Fax: 659 7506

Private Bag F193,
City of Francistown
Tel: 241 5848,
Fax: 241 7540

Private Bag 0024, Gaborone
Tel: 3671300
Fax: 3952201
Toll Free: 0800 600 200

STATISTICS BOTSWANA

E-mail: info@statsbots.org.bw
Website: http://www.statsbots.org.bw