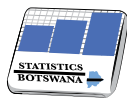




## Botswana Data Quality Assessment Framework



# STATISTICS BOTSWANA

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

Private Bag F193,  
City of Francistown Botswana  
**Tel.** 241 5848, **Fax.** 241 7540

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

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

**E-mail:** [info@statsbots.org.bw](mailto:info@statsbots.org.bw) **Website:** <http://www.statsbots.org.bw>

# **Botswana Data Quality Assessment Framework**

**Published by**

**Statistics Botswana  
Private Bag 0024,  
Gaborone**

**Contact: Quality Assurance Unit  
Tel: 3671300/ Ext 377/ 415  
Fax : 3952201  
E-mail: [info@statsbots.org.bw](mailto:info@statsbots.org.bw)**

## Foreword

Statistics Botswana and the broader National Statistical System (NSS) are required to produce and disseminate quality statistical products to inform national planning, policy and decision making. Therefore the quality of statistics is of high priority and must satisfy the users' needs. As a result, the Botswana Data Quality Assessment Framework (BDQAF) has been developed to provide criteria for evaluating and certifying statistics produced by Statistics Botswana and other stakeholders within the wider NSS. Statistics Botswana is geared towards promoting production of quality statistical products and services that meet quality standards.

The NSS must focus on the improvement and maintenance of quality statistics by ensuring that statistics are in compliance with all the required dimensions of quality in order for their products to be certified as official statistics, as per the Statistics Act of 2009. The dimensions of quality entail assurances of integrity, methodological soundness, accuracy and reliability, and serviceability including relevance, timeliness and accessibility.

The BDQAF was developed in accordance with UN Fundamental Principles of Official Statistics, which underpins the reliability and objectivity of official statistics. It is a major milestone in the improvement of quality statistics within the National Statistical System which is envisaged to inform all development agendas nationally, regionally and internationally in planning, policy formulation and evidence based decision making.

It is worth noting that the BDQAF was developed in consultation with producers and users of statistics. Therefore its implementation will significantly contribute to the production of quality official statistics and enhance its usage.



---

**B. S. Mguni (Dr)**  
**Statistician General**  
January 2020

# Acronyms

<b>BDQAF</b>	Botswana Data Quality Assessment Framework
<b>CI</b>	Confidence Interval
<b>COICOP</b>	Classification of Individual Consumption according to Purpose
<b>CPI</b>	Consumer Price Index
<b>CPC</b>	Central Product Classification
<b>CV</b>	Coefficient of Variation
<b>DBA</b>	Database Administrator
<b>DQAT</b>	Data Quality Assessment Team
<b>DQAF</b>	Data Quality Assessment Framework
<b>DMID</b>	Data Management and Information Delivery
<b>EA</b>	Enumerator Area
<b>Eurostats</b>	European Statistics
<b>GDDS</b>	General Data Dissemination Standard
<b>GDP</b>	Gross Domestic Product
<b>ICD-10</b>	International Classification of Diseases (10th Revision)
<b>ICT</b>	Information and Communication Technology
<b>ILO</b>	International Labour Organization
<b>IMF</b>	International Monetary Fund
<b>IT</b>	Information Technology
<b>ISIC</b>	International Standard Industrial Classification
<b>JSTOR</b>	Journal of Official Statistics
<b>LFS</b>	Labour Force Survey
<b>MDGs</b>	Millennium Development Goals
<b>MSE</b>	Mean Square Error
<b>NSO</b>	National Statistics Office
<b>NSS</b>	National Statistics System
<b>NGO</b>	Non-governmental Organization
<b>SVC</b>	Statistical Value Chain

Foreword.....	4.
ACRONYMS.....	5.
A. Introduction.....	7.
B. Objective of the Framework.....	7.
C. Scope.....	7.
D. Applicability of the Framework.....	7.
E. Exemptions.....	7.
F. Roles and Responsibilities.....	8.
G. Definition of Terms.....	8.
H. KEY ASPECTS OF THE BDQAF.....	9.
I. Botswana Data Quality Assessment Framework.....	12.
1. Prerequisites of quality.....	12.
2. Credibility.....	16.
3. Comparability and Coherence.....	18.
4. Methodological Soundness.....	21.
5. Relevance.....	25.
6. Accuracy.....	27.
7. Timeliness and Punctuality.....	36.
8. Accessibility.....	40.
9. Interpretability.....	43.
J. References.....	44.
<b>Annexure A:</b> Protocol Specifying the Procedure for the Statistician General to Designate Statistics as Official Statistics.....	45.
<b>Annexure B:</b> Fundamental Principles of Official Statistics Relative to Statistics Botswana.....	51.
<b>Annexure C:</b> The Process of Assessing Statistics and Designating Them as Official.....	53.
<b>Annexure D:</b> Statistical Value Chain (Svc).....	54.
<b>Annexure E:</b> Mapping Quality Dimensions to Activities in the Statistical Value Chain.....	70.

## A. Introduction

Statistics Botswana, like other statistical agencies, is mandated to be the pre-eminent national agency responsible for the development and management of official statistics. In addition, the organization is mandated to coordinate the National Statistical System (NSS). Thus, Statistics Botswana has the responsibility to inform users and producers of statistics on the concepts, definitions and methodologies used in collecting, processing and analyzing data, the accuracy of the data produced, and any other features that may affect data quality. The organization has developed a National Statistics Strategy in an effort to realize its objective on quality statistics. The development of the Data Quality Assessment Framework (DQAF) plays an important role in promoting the production of statistics that conform to the UN Fundamental Principles of Official Statistics.

The framework provides a rational, transparent and sustainable methodology for quality assessment of statistical products and services. This document also outlines procedures for designating statistics as 'official statistics'. According to Section 29 (2) of the Statistics Act of 2009, official statistics shall be in accordance with the standards and principles issued by the United Nations Statistical Commission, which include;

- a) Practical utility, relevance of, and equal access to, statistics produced;
- b) Credibility through the exercise of professional independence, accountability and transparency in data production and dissemination;
- c) Legitimacy through a clear mandate for data access, production and dissemination;
- d) data quality covering relevance, accuracy, reliability, coherence, comparability, timeliness and use of international standards;
- e) right of statistical agencies to comment on misleading interpretation and misuse of statistics;
- f) cost-effectiveness through the maximum use of data from various sources provided that quality, cost and burden on respondents are considered;
- g) confidentiality protection of individual data collected for statistical purposes;
- h) national coordination for consistency and efficiency of the System; and
- i) Co-operation with agencies of the System, bilateral and multilateral organizations in order to share best practices and improve standards.

## B. Objective of the Framework

To ascertain that all the processes during production of statistics are adhered to within NSS. These entail conforming to international standards, guidelines and policies. Thus, the availability of the BDQAF enables facilitation of the assessment to ensure quality of the statistical products.

## C. Scope

The BDQAF is applicable to all producers and users of statistics.

## D. Applicability of the Framework

The application of the BDQAF is based on the following criteria;

- **The producing agency must be a member of the NSS.**
- **The statistical product should be part of a sustainable series.**

## E. Exemptions

Researches and studies are not subjected to BDQAF. However, they could adopt and implement BDQAF for self-assessment as well as ensuring compliance to standards.

## F. Roles and Responsibilities

### QA lead

- Ensure proper usage of available standards and methods to gain the maximum benefit of the QA effort.
- Responsible for defining DQA framework during assessment.
- Responsible for leading and directing the DQA team.
- Participate in interviews, induction, training and performance evaluation of all QA team.
- Provide technical leadership and expertise within the field of Quality Assurance and assessment.
- Ensuring that all sectors adhere to the principles, guidelines and best practices of the QA framework as defined.
- Focus on continuous QA improvements including usage of appropriate tools and techniques during surveys and censuses.
- Manage training and continuous learning of sectors on the DQAF by means of workshops and meetings.

### Statistician General

- Monitor the quality of the data produced during surveys and censuses.
- Identify opportunities to drive improvement into existing systems and activities.
- Ensure effective linkages between the different activities within the statistics value chain and assure compliance with current standards.
- Ensure that changes for improvement remain in place over time.

### The Board

The Board of Directors for Statistics Botswana is responsible for assuring the quality framework is designed and implemented.

## G. Definition of Terms

**Administrative data:** the set of units and data derived from an administrative source.

**Administrative source:** a data holding containing information collected and maintained for the purpose of implementing one or more administrative regulations.

**Benchmark:** a recognized standard, or a reference point, that forms the basis for assessment or comparison.

**Catalogue:** an ordered list of statistical products available in the organization.

**Micro-data:** observation data collected on an individual object or statistical unit.

**Data confidentiality:** a property of data, usually resulting from legislative measures, which prevents it from unauthorized disclosure.

**Data quality:** fitness for use of statistical information, i.e. the degree to which a set of inherent characteristics in the statistical data fulfils user requirements; measured in terms of pre-requisites to quality and eight (8) dimensions of quality, namely: relevance, accuracy, timeliness, accessibility, interpretability, comparability and coherence, methodological soundness and integrity.



**Electronic media:** dissemination media that allow electronic exchange of data such that software, or a combination of individuals and software, can put the data in a compatible form at the receiving end.

**Estimate:** the particular value yielded by an estimator in a given set of circumstances.

**Estimator:** a rule or method of estimating a parameter of a population, usually expressed as a function of sample values.

**Guidelines:** directions or principles used in the development, maintenance and application of rules; they may or may not be necessarily mandatory, but are provided as an aid to interpretation and use of rules.

**Misclassification:** when a subject is falsely classified into a category in which the subject does not belong. It may result from misreporting by study subjects, from the use of less than optimal measurement devices, or from random error.

**Quality indicator:** an attribute of statistical information that is used to measure its quality

**Reference period:** the period of time relevant for a particular question.

**Respondent burden:** the effort, in terms of time and cost, required for respondents to provide satisfactory answers to a survey.

**Scope:** coverage or sphere of what is to be observed. It is the total membership or population of a defined set of people, objects or events.

**Standard:** a comprehensive set of guidelines for surveys and administrative sources collecting information on a particular topic, including definitions, statistical units, and classifications, coding processes, questionnaire modules, and output categories.

**Survey:** a process which collects, examines, and reports on data concerning variables of interest for a reference period.

**Statistical programme:** a programme for producing statistics in a particular socio-economic sphere.

## H. Key Aspects of the BDQAF

The BDQAF consists of a set of pre-requisites and eight dimensions of data quality being: credibility, comparability and coherence, methodological soundness, relevance, accuracy, timeliness and punctuality, accessibility and interpretability. For each dimension, the BDQAF identifies key components of good practice and for each key component there are standards and indicators. Below are dimensions and key components:

- 1. Pre-requisite of quality refers to the institutional and organizational conditions that have an impact on data quality. It defines the minimum set of necessary conditions that have to be met in order to produce quality statistics. It therefore serves as the foundation on which all other dimensions of data quality should be premised.**

### Key components

- Legal and institutional environment (Statistics Act and other Acts including Memoranda of Understanding (MoUs) or Service Level Agreements (SLAs)
  - Confidentiality
  - Commensurate resources
  - Quality as the cornerstone of statistical work
  - Risk Management
- 2. The credibility of statistical information refers to values and related practices that maintains users' confidence in the agency producing statistics and ultimately in the statistical product.**

### Key components

- Professionalism and ethical standards which guide policies and practices.
- Assurances that statistics are produced on an impartial basis
- Ethical standards are guided by policies and procedures.

### 3. The Comparability and Coherence

Comparability of statistical information is the ability to compare statistics on the same characteristic between different points in time, geographical areas or statistical domains; while coherence of statistical information reflects the degree to which it can be successfully brought together with other similar statistical information from different sources within a broad analytic framework and over time. It is the extent to which differences between two sets of statistics are attributable to differences between the estimates and the true value of the statistics.

### Key components

- The use of common concepts and definitions within and between series.
- The use of common variables and classifications within and between series.
- The use of common methodology and systems for data collection and processing within and between series.

### 4. Methodological soundness refers to the application of international, regional and national standards, guidelines, and good practices to produce statistical outputs. Application of such standards fosters national and international comparability.

### Key components

- Application of International, regional and national standards and methods.
- Data compilation methods employ acceptable procedures.
- Statistical procedures employ sound statistical techniques.
- Transparent revision policy and studies of revisions are done and made public.

### 5. The Relevance of statistical information reflects the degree to which the statistical product meets the needs of users.

### Key components

- Identify the need to conduct the survey or collect data
- Identify users of statistics and their needs
- Determine whether users are satisfied with statistics produced
- Monitor user needs and incorporate their feed back into the design process

- 6. The accuracy of statistical information is the degree to which the product correctly describes and or estimates the phenomena it was designed to measure. Accuracy also refers to the closeness of the values provided to the (unknown) true values.**

#### Key components

- Assessment of sampling errors where sampling was used
- Assessment of non-sampling errors including
- Frame coverage errors (coverage of data collection in comparison to the target population)
- Measurement errors (data collection errors)
- Processing errors (data capture, coding, editing, etc.)
- Non-response errors (response rates with a view to determine usability of the data)
- Assessment of the impact of the imputation
- Assessment of source data accuracy or consistency problems with register based statistics

#### 7. Timeliness and Punctuality

The timeliness of statistical information refers to the time lag between the reference point to which the information pertains and the date on which the information becomes available. Timeliness also addresses aspects of periodicity and punctuality of production activities within the statistical value chain. Punctuality of statistical product is the time difference between the date the data are released and the target date on which they were scheduled for release, as announced in an official release calendar and laid down by regulations or previously be agreed with users.

#### Key components

- Statistics production time
- Periodicity and punctuality of statistical release.

- 8. Accessibility of statistical information refers to the ease with which it can be obtained from the agency. This includes the ease with which the existence of information can be ascertained, as well as the suitability of the form or medium through which the information can be accessed. The cost of the information may also be an aspect of accessibility for some users.**

#### Key components

- Catalogue systems are available in the parastatals/government or statistical agency
- Delivery systems to access information
- Measure of release calendar and delivery systems performance

- 9. The Interpretability of statistical information refers to the ease with which users understand statistical information through the provision of supplementary information (metadata and relevant supporting documents).**

#### Key components

- Metadata on concepts and definitions, classifications and methodology used within the statistical value chain
- Key findings giving the summary of the results;
- Presentation of statistics in a meaningful way.

## Level of Assessment

### The framework certifies national statistics on one of the four levels:

**Level Four: Quality Statistics**- these are statistics that meet all the quality requirements as set out in the BQAF. They are designated as quality statistics to the extent that deductions can be made from them and are 'fit for use' for the purpose for which they were designed. Level four applies to highly-developed statistical activities with respect to the corresponding indicator.

**Level Three: Acceptable statistics** – These are statistics that meet most, but not all the quality requirements as stipulated in the BQAF. They are designated as acceptable to the extent that, despite the limitations, deductions can be made, and are 'fit for use' for the purpose for which they were designed. Level three refers to moderately well-developed activities with reference to a particular indicator.

**Level Two: Questionable Statistics** – these are the statistics that meet few of the quality requirements as provided in the BQAF. They are designated as questionable to the extent that very limited deductions can be made and they are therefore not 'fit for use' for the purpose for which they were designed. Level two refers to statistical activities that are developing but still have many deficiencies.

**Level One: Poor Statistics** – These are statistics that meet almost none of the quality requirements as provided in the BQAF. They are designated as poor statistics to the extent that no deductions can be made from them and are not 'fit for use' for the purpose for which they were designed. Level one refers to activities that are underdeveloped.

## I. Botswana Data Quality Assessment Framework

This part of the document discusses the prerequisites to quality and the eight (8) dimensions of quality. All the dimensions share a common three-part approach.

- **Description, defines a particular dimension of quality.**
- **Key Components introduces concepts to be discussed in that particular dimension.**

Standards are indicators of each dimension and accompanied by different quality levels.

The following chapters discuss the dimensions in this framework together with the indicators of such.

## 1. Prerequisites of quality

### a. Description

The prerequisites of quality refer to the institutional and organizational conditions that have an impact on data quality. It defines the minimum set of necessary conditions that have to be met in order to produce quality statistics. It therefore serves as the foundation on which all other dimensions of data quality should be premised.

### b. Key components

- Legal and institutional environment (Statistics Act and other Acts including Memoranda of Understanding (MoUs) or Service Level Agreements (SLAs))
- Confidentiality
- Commensurate resources
- Quality as the cornerstone of statistical work
- Risk Management

### c. Quality indicators, standards and benchmarks

## Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
			Level 3	Level 2	Level 1
<b>1.1</b> The responsibility for producing statistics is clearly specified.	<b>1.1.1</b> A legal arrangement exists that explicitly mandates the production of statistics.	A law or legal arrangement exists that explicitly provides the mandate for the production of statistics.	A law or legal arrangement exists that implies that statistical production is part of its mandate.	No law or legal arrangement exists, but an informal agreement exists for statistical production.	No arrangement exists.
<b>1.2</b> Standards and policies are in place to promote consistency of methods and results.	<b>1.2.1</b> A set of policies must exist which covers all aspects of the statistical value chain.	Policies exist which cover all aspects of the statistical value chain and are adhered to.	Policies exist which cover all aspects of the statistical value chain but are not always adhered to.	Policies exist, but do not cover all aspects of the statistical value chain.	No policies exist.
	<b>1.2.2</b> A set of standards related to appropriate policies must exist.	Standards exist which cover all aspects of the statistical value chain and are adhered to.	Standards exist which cover all aspects of the statistical value chain but are not adhered to OR existing standards do not cover all aspects of the statistical value chain but are all adhered to.	Standards exist, but do not cover all aspects of the statistical value chain OR Standards exist but are not adhered to.	No standards exist.
<b>1.3</b> Data sharing and coordination among data-producing agencies are clearly specified.	<b>1.3.1</b> A legal arrangement must exist which allows for the timely and efficient sharing of data between the collecting agency and the user.	A legal arrangement exists and is adhered to.	A legal arrangement exists but is not adhered to.	An informal arrangement exists	No arrangement exists.
	<b>1.3.2</b> Regular meetings must occur between the data-producing agencies and users/agencies to resolve statistical issues.	Regular meetings are held to address all statistical issues to be resolved. Commitment to resolving them is achieved.	Regular meetings are held to address all statistical issues to be resolved. Issues remain unresolved.	Ad hoc meetings occur.	No meetings.
<b>1.4</b> Measures are in place to ensure that individual data are kept confidential, and used for statistical and administrative purposes only.	<b>1.4.1</b> There must be a law or policy that ensures information collected is kept confidential and used for statistical or administrative purposes only.	A law/policy guaranteeing the confidentiality of respondents' information, and used for statistical purposes exists and is adhered to at all times.	A law/policy guaranteeing the confidentiality of respondents' information, and used for statistical purposes only exists but is not adhered to at all times.	Although no law/policy exists, efforts are made to keep respondent information confidential and ensure they are used for statistical purposes only.	No law or policy exists.
<b>1.5</b> Measures to oblige response are ensured through law.	<b>1.5.1</b> There must be a law or other formal measures that inform respondents of their obligation to provide information; and any sanctions which may apply if they fail to do so.	There is a law that requires respondents to provide information. Respondents are informed of sanctions which may occur if they fail to respond, and sanctions are enforced.	There is a law that requires respondents to provide information. Respondents are informed of any sanctions which may occur if they fail to respond, though sanctions are not enforced.	A law exists, but there are no efforts made to inform respondents of the consequences of non-response.	No law or other formal measures exists.

## Cont'd: Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
			Level 3	Level 2	Level 1
<b>1.6</b> Resources are commensurate with the needs of statistical programmes. <ul style="list-style-type: none"> <li>• Staff</li> <li>• Facilities</li> <li>• Computing resources</li> <li>• Financing.</li> </ul>	<b>1.6.1</b> Have adequately skilled personnel within the Statistical value chain.	The programme is adequately staffed with skilled personnel.	The staffing is inadequate with respect to the number of skilled personnel.	The staffing with respect to the number of personnel is adequate, but their skill levels are low.	The staffing is inadequate with respect to the number of personnel as well as the skills they possess.
	<b>1.6.2</b> There must be a Statistics Unit or component responsible for compiling statistics.	A statistics unit or component responsible for statistics compilation with qualified statisticians exists.	A statistics unit exists but there are no qualified statisticians.	No statistics unit exists, but there are qualified statisticians.	No statistics unit exists. There are no statisticians.
	<b>1.6.3</b> Facilities must have the infrastructure to manage the needs of statistical programmes. (facilities – office space, furniture)	Facilities are well equipped and have the necessary infrastructure for the production of statistics.	Facilities are well equipped but do not have necessary infrastructure for the production of Statistics.	N/A	Facilities are ill-equipped and do not have the necessary infrastructure for the production of statistics.
	<b>1.6.4</b> Computer hardware resources must be adequate in terms of <ul style="list-style-type: none"> <li>• data storage;</li> <li>• data backup media;</li> <li>• power supply (uninterrupted);</li> <li>• memory; and</li> <li>• Other necessary equipment (notebooks, desktop, etc.).</li> </ul>	The hardware computing resources is sufficient in terms of data storage, data backup media, power supply, memory, and computers.	The hardware computing resources is sufficient in terms of data backup media, power supply, memory, and computers.	The hardware computing resources is sufficient in terms of power supply, memory, and computers.	The hardware resources are entirely insufficient.
	<b>1.6.5</b> A disaster recovery and business continuity plan must exist.	A disaster recovery and business continuity plan exists.	N/A	N/A	A disaster recovery and business continuity plan does not exist.
	<b>1.6.6</b> Computer software resources must be adequate in terms of <ul style="list-style-type: none"> <li>• capturing systems;</li> <li>• editing systems;</li> <li>• coding systems;</li> <li>• statistical software;</li> <li>• up-to-date licenses;</li> <li>• virus protection; and</li> <li>• Appropriate access rights.</li> </ul>	Computer software resources are adequate in terms of statistical & other specialized software, up-to-date licenses, virus protection, and appropriate access rights.	Computer software resources are adequate in terms of appropriate access rights, virus protection, statistical & other specialized software.	Computer software resources are adequate in terms of appropriate access rights, statistical & other specialized software.	Computer software resources are entirely inadequate.
	<b>1.6.7</b> Budgets must be adequate.	Budgets are sufficient for the needs of statistical programmes.	N/A	N/A	Budgets are not sufficient.

Cont'd: Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
			Level 3	Level 2	Level 1
<p><b>1.7</b> Measures to ensure efficient use of resources in 1.6 are implemented.</p> <p>metrics</p> <p><math>a = \frac{\text{number of aligned of staff}}{\text{total aligned of staff}} \times 100</math></p>	<p><b>1.7.1</b> Staff of a statistical programme must be employed in positions that are aligned with their skills profile.</p>	90% a. 100%.	75% a. 89%.	40% a. 74%.	0% a. 39%.
	<p><b>1.7.2</b> Asset management policies that prevent the abuse of facilities (e.g. vehicles, telephones etc.) must be developed, adopted and implemented.</p>	Asset management policies have been developed, adopted and monitored for compliance.	Asset management policies have been developed and adopted. No compliance monitoring exists.	Asset management policies have been developed but are not adopted by the organization.	No asset management policies exist.
	<p><b>1.7.3</b> Policies, guidelines and procedures governing the use of ICT resources must exist, so as to maximize the return on investment</p>	All documented policies, guidelines and procedures on the usage of IT resources have been implemented.	There are well documented policies, guidelines and procedures on the usage of IT resources. Some have been implemented.	There are policies, guidelines and procedures on the usage of IT resources. None have been implemented.	No policies, guidelines or procedures on IT resource usage exist.
	<p><b>1.7.4</b> Budgets must be reviewed and financials audited to ensure that resources are used in the best possible way.</p>	Budgets are reviewed and financial audits onto ensure that resources are used in the best possible way.	Financial audits take place; however, budgets are not reviewed as per the policy.	Budgets are reviewed but there is no financial auditing procedure in place.	Budgets are neither reviewed nor audited.
<p><b>1.8</b> Processes are in place to focus on, monitor and check quality.</p>	<p><b>1.8.1</b> The agency must have a quality management system in place.</p>	The agency has a quality management system in place.	N/A.	N/A.	The agency has no quality management system in place.
	<p><b>1.8.2</b> The data producing agency must have an independent data quality audit process.</p>	The data producing agency has an independent data quality audit process.	The data producing agency has developed and implemented a data quality audit process, but it is not independent.	The data producing agency has developed a data quality audit process but not implemented,	The data producing agency has no audit process developed.
	<p><b>1.8.3</b> Staff members in production areas must have a data Quality Management requirement as part of their performance agreements or job descriptions.</p>	All staff members have a data quality management requirement as part of performance agreements or job descriptions, with clear sanctions for failure to comply.	All staff members have a data quality management requirement as part of performance agreements or job descriptions, without clear sanctions for failure to comply.	Staff members do not have a data quality management requirement as part of performance agreements or job descriptions.	There are no performance agreements or job descriptions.
<p><b>1.9</b> Policies and frameworks are in place to manage risk in the statistical value chain</p>	<p><b>1.9.1</b> A risk management framework, policies and registrar which cover all processes in the statistical value chain must exist.</p>	A risk management framework, policy and register exists which cover all processes in the statistical value chain.	A risk management framework and policy exist which cover all processes in the statistical value chain and risks are not registered	N/A	A risk management framework, policies and register which cover all processes in the statistical value chain do not exist



## 2. Credibility

### a. Description

The credibility of statistical information refers to values and related practices that maintain users' confidence in the agency producing statistics and ultimately in the statistical product.

### b. Key components

- Professionalism and ethical standards which guide policies and practices.
- Assurances that statistics are produced on an impartial basis
- Ethical standards are guided by policies and procedures.

### c. Quality indicators, standards and benchmarks

## Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
<b>2.1</b> The terms and conditions, including confidentiality, under which statistics are collected, processed and disseminated, are available to the public.	<b>2.1.1</b> A terms and conditions document must be available, accessible to the public and follow the UN fundamental principles of official statistics.	Terms and conditions documents is available, accessible to the public and follow the UN fundamental principles of official statistics.	Terms and conditions documents is available but not accessible to the public and follow the UN fundamental principles of official statistics.	A terms and conditions document is available but neither accessible to the public nor follow the UN fundamental principles of official statistics.	A terms and conditions document does not exist.
<b>2.2</b> The conditions under which users have access to data are described and published.	<b>2.2.1</b> A data dissemination policy detailing the conditions under which users have access to the data must be available.	A data dissemination policy detailing the conditions under which users have access to the data is published and adhered to.	A data dissemination policy detailing the conditions under which users have access to the data is not published but adhered to.	N/A	A data dissemination policy detailing the conditions under which users have access to the data does not exist.
<b>2.3</b> Advance notice is given of major changes in methodology and source data.	<b>2.3.1</b> Advance notice of at least 6 months must be given of major changes in methodology and source data.	Advance notice of at least 6 months is given of major changes in methodology and source data.	Advance notice of between 3 months to 6 months is given of major changes in methodology and source data.	Advance notice of less than 3 months is given of major changes in methodology and source data.	No advance notice is given of major changes in methodology and source data.
<b>2.4</b> Government commentary, when data are released, should be identified as such, and not be seen as part of the official statistics.	<b>2.4.1</b> Government commentary, when data are released, must be identified as such, and not be seen as part of the official statistics.	Government commentary, when data are released, is clearly identified as such, and not seen as part of the official statistics.	N/A	N/A	Government commentary is regarded as part of the official statistics.
<b>2.5</b> Choice of source data, techniques and dissemination decisions are informed solely by statistical considerations.	<b>2.5.1</b> The choice of source data, techniques and dissemination decisions must be informed solely by statistical considerations.	The choice of source data, techniques and dissemination decisions are solely informed by statistical considerations.	N/A	N/A	The choice of source data, techniques and dissemination decisions are not solely informed by statistical considerations.
<b>2.6</b> Ethical guidelines for staff behavior are in place and are well known to the staff.	<b>2.6.1</b> A professional code of conduct must be in place providing ethical guidelines for staff behavior.	A professional code of conduct is in place providing ethical guidelines for staff behavior. The code of conduct is well known and enforced.	A professional code of conduct is in place providing ethical guidelines for staff behavior. The code of conduct is well known, but is not enforced.	A professional code of conduct is in place providing ethical guidelines for staff behavior. The code of conduct is not well known and not enforced.	A professional code of conduct is not in place.

### 3. Comparability and Coherence

#### a. Description

Comparability of statistical information is the ability to compare statistics on the same characteristic between different points in time, geographical areas or statistical domains. The coherence of statistical information reflects the degree to which it can be successfully brought together with other similar statistical information from different sources within a broad analytic framework and over time. It is the extent to which differences between two sets of statistics are attributable to differences between the estimates and the true value of the statistics.

#### b. Key components

- The use of common concepts and definitions within and between series.
- The use of common variables and classifications within and between series.
- The use of common methodology and systems for data collection and processing within and between series.

#### c. Quality indicators, standards and benchmarks

## Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
3.1 Data within series and administrative systems are based on common concepts and definitions, classifications, and methodology, and departures from this are identified in the metadata.	3.1.1 All data (including source data, related frame data, and related survey data) within the same series must use the same concepts and definitions. Departures from common concepts and definitions must be identified, documented in the metadata and archived.	All statistical products within the same series use the same concepts and definitions. Departures must be identified, documented in the metadata and archived.	At least one statistical product within the same series uses different concepts and definitions. Departures from common concepts and definitions are identified, documented in the metadata and archived.	At least one statistical product within the same series uses different concepts and definitions. Departures from common concepts and definitions are identified in the metadata, but not archived.	At least one statistical product within the same series uses different concepts and definitions. Departures from common concepts and definitions are not identified in the metadata and are not archived.
	3.1.2 All data (including source data, related frame data, and related survey data) within the same series must use the same classifications. Departures from common classifications must be identified in the metadata, documented and archived.	All statistical products within the same series use the same classifications. Departures must be identified, documented in the metadata and archived.	At least one statistical product within the same series uses different classifications. Departures from common classifications are identified in the metadata, documented and archived.	At least one statistical product within the same series uses different classifications. Departures from common classifications are identified in the metadata but not documented and archived.	At least one statistical product within the same series uses different classifications. Departures from common classifications are not identified in the metadata and are not documented and archived.
	3.1.3 All data (including source data, related frame data, and related survey data) within the same series must use the same methodology. Departures from common methodology must be identified in the metadata, documented and archived.	All statistical products within the same series use the same methodology. Departures must be identified, documented in the metadata and archived.	At least one statistical product within the same series uses different methodology. Departures from common methodology are identified in the metadata, documented and archived.	At least one statistical product within the same series uses different methodology. Departures from common methodology are identified in the metadata, but not documented and archived.	At least one statistical product within the same series uses different methodology. Departures from common methodology are not identified in the metadata and are not archived.
3.2 Statistics are consistent or reconcilable over time.	3.2.1 Statistics must be consistent over time.	Statistics are consistent over time.	N/A	N/A	Statistics are not consistent over time.
	3.2.2 The statistics must follow an expected trend established over time. An inconsistencies in the key variables must be reconciled and included in the metadata.	The statistics follow an expected trend established over time. Inconsistencies in the key variables are reconciled and included in the metadata	The statistics follow an expected trend established over time. Inconsistencies in the key variables are reconciled and not included in the metadata	N/A	The statistics do not follow an expected trend established over time and not documented.

## Cont'd: Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
3.3 Data across comparable series, or source data are based on common frames, identifiers, concepts and definitions, and classifications, and departures from these are identified in the metadata.	3.3.1 Data across comparable series or source data must be based on common identifiers. Departures from common identifiers must be identified in the metadata and archived.	Data across comparable series or source data are based on common identifiers. Usage of different identifiers is identified in the metadata, and is archived.	Data across comparable series or source data are not based on common identifiers. Usage of different identifiers is identified in the metadata, and is archived.	Data across comparable series or source data are not based on common identifiers. Usage of different identifiers is identified in the metadata, but is not archived.	Data across comparable series or source data are not based on common identifiers. Usage of different identifiers is not identified in the metadata and is not archived.
	3.3.2 Data across comparable series or source data must use the same concepts and definitions. Departures from common concepts and definitions must be identified in the metadata and archived.	Data across comparable series or source data are based on common concepts and definitions. Usage of different concepts and definitions is identified in the metadata, documented and is archived.	Data across comparable series or source data are not based on common concepts and definitions. Usage of different concepts and definitions is identified in the metadata, documented and is archived.	Data across comparable series or source data are not based on common concepts and definitions. Usage of different concepts and definitions is identified in the metadata, but is not documented and archived.	Data across comparable series or source data are not based on the same concepts and definitions. Usage of different concepts and definitions is not identified in the metadata and is not archived.
	3.3.3 Data across comparable series or source data must use the same classifications. Departures from common classifications must be identified in the metadata and archived.	Data across comparable series or source data are based on common classifications. Usage of different classifications is identified in the metadata, and is archived.	Data across comparable series or source data are not based on common classifications. Usage of different classifications is identified in the metadata, and is archived.	Data across comparable series or source data are not based on common classifications. Usage of different classifications is identified in the metadata, but is not archived.	Data across comparable series or source data are not based on common classifications. Usage of different classifications is not identified in the metadata and is not archived.
3.4 Statistics are checked for consistency with those obtained through other data sources.	3.4.1 Statistics must be checked for consistency with a comparable dataset. Inconsistencies must be reconciled.	The data producer demonstrates that the statistics are consistent. Any inconsistencies in the key variables are reconciled.	The data producer does not demonstrate that the statistics are consistent. Any inconsistencies in the key variables are reconciled.	The data producer demonstrates that the statistics are consistent. Any inconsistencies in the key variables are not reconciled.	The data producer does not demonstrate that the statistics are consistent. Any inconsistencies in the key variables are not reconciled.
3.5 A common set of identifiers (for the purpose of record matching) exist and have been agreed upon by data producers.	3.5.1 A common identifier must be agreed upon by the data producers.	The identifier is the same across all datasets and is agreed upon by the data producers.	The identifier is the same across all datasets but is not agreed upon by the data producers.	There is a common identifier across any two datasets. However, there is not one identifier common to all datasets.	There is no common identifier across all datasets to be matched.
	3.5.2 The common identifier must be unique in every dataset. Rules and practices must be agreed upon to ensure uniqueness.	The common identifier is unique in every dataset. Rules and practices to ensure uniqueness are agreed upon.	The common identifier is unique in every dataset. Rules and practices to ensure uniqueness exist independently of each other.	The common identifier is unique in every dataset. Rules and practices to ensure uniqueness do not exist.	The common identifier is not unique in every dataset.

## 4. Methodological Soundness

### a. Description

It refers to the application of international, regional and national standards, guidelines, and good practices to produce statistical products. Application of such standards fosters national and international comparability.

### b. Key components

- Application of International, regional and national standards and methods.
- Data compilation methods employ acceptable procedures.
- Statistical procedures employ sound statistical techniques.
- Transparent revision policy and studies of revisions are done and made public.

### c. Quality indicators, standards and benchmarks

## Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
<p><b>4.1</b> Concepts, definitions, and classifications used follow accepted standards, guidelines or good practice (international, regional, national,).</p>	<p><b>4.1.1</b> The concepts and definitions must satisfy accepted standards, guidelines or good practice in line with international, regional, national standards; and must be documented. Deviations from the standard must be formally approved, and be fully documented.</p>	<p>The concepts and definitions satisfy accepted standards, guidelines or good practice in line with international, regional, national standards, and are documented. Deviations from the standard are formally approved, and fully documented.</p>	<p>The concepts and definitions satisfy accepted standards, guidelines or good practice in line with international, regional, national standards, and are documented. Deviations from the standard are not approved, and not fully documented.</p>	<p>The concepts and definitions are documented, but do not satisfy accepted standards, guidelines or good practice.</p>	<p>No documented concepts and definitions exist.</p>
	<p><b>4.1.2</b> The classifications must satisfy accepted standards, guidelines or good practice in line with international, regional, national standards; and must be documented. Deviations from the standard must be formally approved, and be fully documented.</p>	<p>Classifications satisfy accepted standards, guidelines or good practice in line with international, regional, national standards; and are documented. Deviations from the standard are formally approved, and fully documented.</p>	<p>Classifications satisfy accepted standards, guidelines or good practice in line with international, regional, national standards; and are documented. Deviations from the standard are not approved, and not fully documented.</p>	<p>Classifications are documented, but do not satisfy accepted standards, guidelines or good practice.</p>	<p>No documented classifications exist.</p>
<p><b>4.2</b> The scope of the study is consistent with accepted standards, guidelines or good practices.</p>	<p><b>4.2.1</b> The scope of the study must be appropriate for the intended topic and must be consistent with accepted standards, guidelines or good practices in line with the survey constraints</p>	<p>The scope of the study is appropriate for the intended topic and is consistent with accepted standards, guidelines or good practice in line with the survey constraints.</p>	<p>N/A</p>	<p>N/A</p>	<p>The scope of the study is inappropriate and inconsistent for the intended topic.</p>

## Cont'd: Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
<p><b>4.3</b> Methodologies used follow accepted standards, guidelines or good practice (international, regional, national standards); viz:</p> <ul style="list-style-type: none"> <li>questionnaire design</li> <li>sampling methods</li> <li>sampling frame development</li> <li>frame maintenance</li> <li>piloting</li> <li>data collection methods</li> <li>data editing, capturing, coding and imputation methods</li> <li>data analytical methods</li> <li>revision procedures</li> </ul>	<p><b>4.3.1</b> The designing of the questionnaire must follow accepted standard, sets of guidelines or good practice.</p>	The designing of the questionnaire follows accepted standard, sets of guidelines or good practice.	N/A	N/A	The designing of the questionnaire does not follow accepted standard, sets of guidelines or good practice.
	<p><b>4.3.2</b> The sampling methods must follow accepted standard, sets of guidelines or good practice.</p>	The sampling methods follow accepted standard, sets of guidelines or good practice.	N/A	N/A	The sampling methods do not follow accepted standard, sets of guidelines or good practice.
	<p><b>4.3.3</b> The frame maintenance methods must follow accepted standard, sets of guidelines or good practice.</p>	The frame maintenance methods follow accepted standard, sets of guidelines or good practice.	N/A	N/A	The frame maintenance methods do not follow accepted standard, sets of guidelines or good practice.
	<p><b>4.3.4</b> The piloting methods must follow accepted standard, sets of guidelines or good practice.</p>	The piloting methods follow accepted standard, sets of guidelines or good practice.	N/A	N/A	The piloting methods do not follow accepted standard, sets of guidelines or good practice.
	<p><b>4.3.5</b> Data collection methods must follow accepted standard, sets of guidelines or good practice.</p>	Data collection methods follow accepted standard, sets of guidelines or good practice.	N/A	N/A	Data collection methods do not follow accepted standard, sets of guidelines or good practice.
	<p><b>4.3.6</b> Editing, coding, capturing and imputation methods must follow accepted standard, sets of guidelines or good practice.</p>	Editing and imputation methods follow accepted standard, sets of guidelines or good practice.	N/A	N/A	Editing and imputation methods do not follow accepted standard, sets of guidelines or good practice.
	<p><b>4.3.7</b> The methods of analysis used must follow accepted standards, sets of guidelines or good practice.</p>	The methods of analysis used follow accepted standards, sets of guidelines or good practice.	N/A	N/A	The methods of analysis used do not follow accepted standards, sets of guidelines or good practice.
	<p><b>4.3.8</b> Revision methods used must follow accepted standards, sets of guidelines or good practice.</p>	Revision methods used follow accepted standards, sets of guidelines or good practice.	N/A	N/A	Revision methods used do not follow accepted standards, sets of guidelines or good practice.



Cont'd: Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
<b>4.4</b> Revisions schedules are followed, regular and transparent.	<b>4.4.1</b> A revision schedule must exist for statistical products, where applicable and must be applicable and must be publicly available , accessible and adhered to.	A revision schedule is publicly available, accessible and adhered to.	The revisions schedule is available and publicly accessible but not adhered to	A revision schedule is available but not publicly Accessible and not adhered to.	There is no revisions schedule.
<b>4.5</b> Preliminary and revised data are identified in the metadata.	<b>4.5.1</b> Preliminary and revised data must be identified in the metadata. Metadata must contain an explanation of the changes.	Preliminary and revised data are identified in the metadata. Metadata contains an explanation of the changes.	Preliminary and revised data are identified in the metadata. Metadata does not contain an explanation of the change.	N/A	No revisions are identified in the metadata.
<b>4.6</b> Studies of revisions and their findings are made public.	<b>4.6.1</b> Regular studies of revisions or upcoming revisions must be done and their findings must be made public.	Regular studies of revisions or upcoming revisions are done and their findings are made public.	Regular studies of revisions or upcoming revisions are done and their findings are not made public.	Studies of revisions are done on an ad hoc basis.	No revision studies are done.

## 5. Relevance

### a. Description

Relevance of statistical information reflects the degree to which the statistical product meets the needs of users.

### b. Key components

- Identify the need to conduct the survey or collect data
- Identify users of statistics and their needs
- Determine whether users are satisfied with statistics produced
- Monitor user needs and incorporate their feed back into the design process

### c. Quality indicators, standards and benchmarks

## Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
5.1 The internal and external users of the statistics have been identified.	5.1.1 An up-to-date user database must exist.	An up-to-date user database exists.	A user database exists but is not up-to-date.	Users are known but not recorded in a database.	Users have not been identified.
	5.1.2 A documented process to identify user needs must exist.	A documented process to identify user needs exists.	A process to identify user needs exists but not documented.	N/A	A process to identify user needs does not exist.
5.2 User needs and the usage of statistical information analysed	5.2.1 A report containing the findings of the usage of statistical information must be available.	A report containing the findings of the usage of statistical information is available	User needs and the usage of statistical are analysed but a report is not available.	One of user needs or usage of statistics information is analysed but a report is not made available.	User needs and usage of statistical information are not analysed
	5.2.2 Findings of user needs must be assessed and availed to users	Findings of user needs are assessed and availed to users	Findings of user needs are assessed but not availed to users	N/A	User needs are not assessed
5.3 Changes are made as a result of user needs assessments.	5.3.1 The results of the user need assessment must influence decisions on the statistical value chain of the survey or on administrative data collection systems, where feasible. Documented reasons for not implementing user needs must be provided as feedback to users.	The results of the user need assessment influence decisions on the statistical value chain of the survey or administrative data collection systems, where feasible. Documented reasons for unimplemented user needs are provided as feedback to users.	The results of the user need assessment influence decisions on the statistical value chain of the survey or administrative data collection systems, where feasible. Documented reasons for unimplemented user needs are not provided as feedback to users	The results of the user need assessment do not influence in any way decisions on the statistical value chain of the survey or administrative data collection systems. Documented reasons for not implementing user needs are provided as feedback to users.	The results of the user need assessment do not influence in any way decisions on the statistical value chain of the survey or on administrative data collection systems. Documented reasons for not implementing user needs are not provided as feedback to users.
5.4 There is a process to determine the satisfaction of users with the statistical information.	5.4.1 A formal process to determine the satisfaction of users with the statistical information exists.	A formal process to determine the satisfaction of users with the statistical information exists.	N/A	N/A	A formal process to determine the satisfaction of users with the statistical information does not exist.

## 6. Accuracy

### a. Description

The accuracy of statistical information is the degree to which the product correctly describes and or estimates the phenomena it was designed to measure. Accuracy refers to the closeness of the values provided to the (unknown) true values.

### b. Key components

- Assessment of sampling errors where sampling was used
- Assessment of non-sampling errors include:
  - Frame coverage errors ( Coverage of data collection in relation to target population)
  - Measurement errors ( Data collection errors)
  - Processing errors ( Data capture, coding, editing, etc)
  - Non-response errors ( Response rates with a view to determine usability of the data)
- Assessment of source data accuracy or inconsistency problems with register based statistics

### c. Quality indicators, standards and benchmarks

## Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
<p><b>6.1</b> Measures of sampling errors for key variables are calculated. Amongst others these are:</p> <ul style="list-style-type: none"> <li>• standard error (SE)</li> <li>• coefficient of variation (CV)</li> <li>• Confidence interval (CI)</li> <li>• Mean Square Error (MSE)</li> <li>• Design Effect (DEFF).</li> <li>• Intra-Cluster Correlation (ICC)</li> </ul>	<p><b>6.1.1</b> Measures of sampling errors must be calculated for the main variables. They must be available for the other variables on request.</p>	<p>Measures of sampling errors are published for the main variables. Measures for other variables are available on request. Measures of sampling errors are published for the main variables. Measures for other variables are available on request.</p>	<p>Measures of sampling errors are published for the main variables. Measures for other variables are not available on request</p>	<p>Measures of sampling errors are available on request for the main variables only.</p>	<p>Measures of sampling errors are not calculated.</p>
	<p><b>6.1.2</b> 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 this exist), is explained</p> <p>Metrics:</p> $SE = \sqrt{\text{Var}(\hat{\theta})}$ $CV = \frac{\sqrt{\text{Var}(\hat{\theta})}}{E(\hat{\theta})}$ $MSE(\hat{\theta}) = \text{Var}(\hat{\theta}) + B^2(\hat{\theta}),$ <p><math>B^2(\hat{\theta})</math> is as the result of Bias creeping in, compromising quality of data. MSE is a measure of the quality of an estimator—values closer</p>	<p>CV ≤ 5% and <math>B^2(\hat{\theta}) = 0</math> CV ≤ 0.05</p>	<p>CV ≤ 5% and <math>B^2(\hat{\theta}) = 0</math> CV ≤ 0.05 <math>B^2(\hat{\theta}) = 0</math></p>	<p>5% &lt; CV ≤ 15% and <math>B^2(\hat{\theta}) = 0</math> 0.05 &lt; cv ≤ 0.10 <math>B^2(\hat{\theta}) = 0</math></p>	<p>CV &gt; 15% and <math>B^2(\hat{\theta}) = 0</math> CV &gt; 0.10 <math>B^2(\hat{\theta}) = 0</math></p>

Cont'd: Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
	<p><b>6.1.3</b> Scientific Sampling techniques must be used.</p> <p>Metrics:</p> <p><math>\alpha = \text{Design effect}</math>  <math>\theta) = 1 + \delta(n-1)</math></p> <p><math>\delta(n-1)</math> is a positive integer which indicates the deviation from accuracy. <math>\delta(n-1)</math> when calculated, it depicts the effect of complexity of the survey design, the larger it becomes, the lesser the accuracy on the survey design. It measures how worse the design is from the SRS method.</p>	<p>Scientific sampling techniques are used.</p> <p><math>1.0 &lt; \text{Deff} \leq 1.5</math></p>	<p>Scientific Sampling techniques are used.</p> <p><math>1.5 &lt; \text{Deff} \leq 2.0</math></p>	<p>Scientific Sampling techniques are used.</p> <p><math>2.0 &lt; \text{Deff} \leq 2.5</math></p>	<p>Scientific Sampling techniques are used.</p> <p><math>&gt; 2.5</math></p>
<p><b>6.2</b> Measures of non-sampling errors are calculated, viz:</p> <ul style="list-style-type: none"> <li>• Frame coverage errors (e.g. duplication in the frame/register used to conduct a survey, number of statistical units out of scope (i.e. number of ineligible units)</li> <li>• Misclassification errors</li> <li>• Systematic errors to determine the extent of bias introduced for both administrative records and surveys</li> <li>• Measurement errors (e.g. questionnaire effects, data collection effects, interviewer effects and respondent effects)</li> <li>• Processing errors (e.g. data entry errors rates, coding errors, editing failure rates, imputation rates)</li> </ul>	<p><b>6.2.1</b> The extent of measures of non-sampling errors must be kept to an acceptable level.</p> <p>Metrics:</p> $a = \frac{\sum   \text{final weight} - \text{design weight}  }{\sum \text{design weight}}$ $b = \max \left[ \frac{  \text{final weight} - \text{design weight}  }{\text{design weight}} \right]$ <p>Design weight is the product of the inverse of probability of selection both at EA and Household level (<math>W1*W2</math>) &amp; final weight is the product of weights calculated at EA level (<math>W1</math>), household level (<math>W2</math>) &amp; result code (response) <math>W3</math> (i.e. Final Weight is (<math>W=W1*W2* W3</math>))</p>	<p><math>a &lt; 15\%</math> and <math>b &lt; 15\%</math>.</p>	<p><math>15\% \leq a &lt; 25\%</math>; or <math>15\% \leq b &lt; 25\%</math>.</p>	<p><math>a \leq 25\%</math> or <math>b \geq 25\%</math>.</p>	<p>a. and b. are not calculated.</p>

Cont'd: Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
	<b>6.2.2</b> Delays between newly-registered administrative units and the corresponding statistical unit births must be known. Update procedures are enforced to correct for under-coverage.	Delays between newly- registered administrative units and the corresponding statistical unit births are known. Corrective actions for under-coverage are undertaken.	Delays between newly- registered administrative units and the corresponding statistical unit births are known. Corrective actions for under-coverage are not undertaken.	N/A	Delays between newly- registered administrative units and the corresponding statistical unit births are unknown.
	<b>6.2.3</b> The measures of under-coverage fall within acceptable standards.  Where; x=under coverage rate	$x \leq 0.05$	$0.06 \leq x \leq 0.1$  $0.05 \leq x \leq 0.1$	$0.11 \leq x \leq 0.15$	$x \geq 0.15$  Or $x > 0.15$  under-coverage rate is not calculated
	<b>6.2.4</b> Delays between de-registering of administrative units and the corresponding statistical unit deaths must be known. Update procedures are enforced to correct for over-coverage.	Delays between de-registering of administrative units and the corresponding statistical unit deaths are known. Corrective actions for over-coverage are undertaken.	Delays between de-registering of administrative units and the corresponding statistical unit deaths are known. Corrective actions for over-coverage are not undertaken	N/A	Delays between de-registering of administrative units and the corresponding statistical unit deaths are unknown.
	<b>6.2.5</b> The duplication rate must be at an acceptable level.  Where; x= duplication rate	$x \leq 0.05$	$0.06 \leq x \leq 0.1$  $0.05 < x \leq 0.10$	$0.11 \leq x \leq 0.15$  $0.10 < x \leq 0.15$	$x \geq 0.15$  Or $> 0.15$  Duplication rate is not calculated
	<b>6.2.6</b> The proportion of units which are out of scope must be at an acceptable level.  Where; x= the proportion of units out of scope	$x \leq 0.05$	$0.06 \leq x \leq 0.1$  $0.05 < x \leq 0.1$	$0.11 \leq x \leq 0.15$  $0.1 < x \leq 0.15$	$x \geq 0.15$  Or $> 0.15$  The proportion of units out of scope is not calculated
	<b>6.2.7</b> The proportion of units which are misclassified must be at an acceptable level.  Where; x= the proportion of units misclassified	$x \leq 0.05$	$0.06 \leq x \leq 0.1$  $0.05 < x \leq 0.1$	$0.11 \leq x \leq 0.15$  $0.10 < x < 0.15$	$x \geq 0.15$  Or $0.15$  The proportion of units misclassified is not calculated

Cont'd: Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
	<b>6.2.8</b> Systematic errors must be identified and reported.	Systematic errors are both identified and reported.	Systematic errors are identified, but not reported.	N/A	Systematic errors are not identified.
	<b>6.2.9</b> Every k <sup>th</sup> statistical unit is independently double collected. The two outputs must be compared and corrective action must be taken. Records must be kept.	Every k <sup>th</sup> statistical unit is independently double collected. The two outputs are compared and corrective action is taken. Records are available.	Fewer than one-in-k statistical units are independently double collected. The two outputs are compared and corrective action is taken. Records are available.	Fewer than one-in-k statistical units are independently double collected. The two outputs are compared, but no corrective action is taken. Records are available.	Double collection is not done.
	<b>6.2.10</b> Data collection error rates calculated from fieldwork records must be at an acceptable level.  Where; x= data collection error rate	$x \leq 0.05$	$0.06 \leq x \leq 0.1$ $0.05 < x \leq 0.10$	$0.11 \leq x \leq 0.15$ $0.10 < x \leq 0.15$	$x \geq 0.15$ Or $>0.15$  Data collection error rate is not calculated
	<b>6.2.11</b> The effects of data collection instruments must be determined, reported and corrective measures taken.	The effects of data collection instruments are reported, determined and corrective measures are taken.	The effects of data collection instruments are determined, reported and no corrective measures are taken.	The effects of data collection instruments are determined but not reported	The effects of data collection instruments are not determined.
	<b>6.2.12</b> The effects of the data collection mode must be determined, reported and corrective measures are taken.	The effects of data collection mode are determined, reported and corrective measures are taken.	The effects of data collection mode are determined, reported and no corrective measures are taken	The effects of data collection mode are determined but not reported	The effects of data collection mode are not determined.
	<b>6.2.13</b> The effects of the interviewers must be determined, reported and corrective measures are taken	The effects the interviewers are determined, reported and corrective measures are taken.	The effects the interviewers are determined, reported and no corrective measures are taken	The effects the interviewers are determined but not reported	The effects of the interviewers are not determined.
	<b>6.2.14</b> Respondent effects must be determined, reported and corrective measures are taken.	Respondent effects are determined, reported and corrective measures are taken.	Respondent effects are determined, reported and no corrective measures taken	Respondent effects are determined but not reported	Respondent effects are not determined



## Cont'd: Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
	<p><b>6.2.15</b> Proxy responses must be separately categorized (flagged). Proxy response rate must be at an acceptable level.</p> <p>Where; x= proxy response rate</p> <p><b>NB:</b> Do not encourage proxy responses in a survey because it could bring bias, distortion, perceptions and speculations in the data which may not be a good reflection of the survey results.</p> <p>.</p>	$x \leq 0.05$	$0.06 \leq x \leq 0.1$ $0.05 < x \leq 0.1$	$0.11 \leq x \leq 0.15$ $0.10 < x \leq 0.15$	$x \geq 0.15$ Or $> 0.15$ Proxy response rate is not calculated
	<p><b>6.2.16</b> Data entry error must average an acceptable accuracy rate</p> <p>Where; x= data entry error</p>	$x \leq 0.05$	$0.06 \leq x \leq 0.1$ $0.05 < x \leq 0.1$	$0.11 \leq x \leq 0.15$ $0.10 < x \leq 0.15$	$x \geq 0.15$ Or $x > 0.15$ Data entry error is not calculated
	<p><b>6.2.17</b> Coding error must average an acceptable accuracy rate.</p> <p>Where; x= coding error</p>	$x \leq 0.05$	$0.06 \leq x \leq 0.1$ $0.05 < x \leq 0.1$	$0.11 \leq x \leq 0.15$ $0.10 < x \leq 0.15$	$x \geq 0.15$ Or $x > 0.15$ Coding error is not calculated
	<p><b>6.2.18</b> Editing rate must average an acceptable level.</p> <p>Where; x= editing rate</p>	$x \leq 0.05$	$0.06 \leq x \leq 0.1$ $0.05 < x \leq 0.1$	$0.11 \leq x \leq 0.15$ $0.10 < x \leq 0.15$	$x \geq 0.15$ Or $x > 0.15$ Editing rate is not calculated
	<p><b>6.2.19</b> Editing failure rate must average an acceptable level.</p> <p>Where; x= editing failure rate</p>	$x \leq 0.05$	$0.06 \leq x \leq 0.1$ $0.05 < x \leq 0.1$	$0.11 \leq x \leq 0.15$ $0.10 < x \leq 0.15$	$x \geq 0.15$ Or $x > 0.15$ Editing failure rate is not calculated
	<p><b>6.2.20</b> The imputation rate for item non-response must average an acceptable level.</p> <p>Where; x= imputation rate for item non-response</p>	$x \leq 0.05$	$0.06 \leq x \leq 0.1$ $0.05 < x \leq 0.1$	$0.11 \leq x \leq 0.15$ $0.10 < x \leq 0.15$	$x \geq 0.15$ Or $x > 0.15$ imputation rate for item non-response is not calculated

Cont'd: Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
	<p><b>6.2.21</b> The imputation rate for unit non-response must average an acceptable level.</p> <p>Where; x= Imputation rate for unit non-response</p>	$x \leq 0.05$	$0.06 \leq x \leq 0.1$ $0.05 < x \leq 0.1$	$0.11 \leq x \leq 0.15$ $0.10 < x \leq 0.15$	$x \geq 0.15$ Or $x > 0.15$ Imputation rate for unit non-response is not calculated
	<p><b>6.2.22</b> The model assumptions must be stated. All models used in the estimation of statistics must be described.</p>	The model assumptions are stated. All models used in the estimation of statistics are described. The model assumptions are valid, and evidence is provided.	The model assumptions are stated. Not all models used in the estimation of statistics are described. The model assumptions are valid, and evidence is provided	The model assumptions are not stated. All models used in the estimation of statistics are described. The model assumptions are not valid, or no evidence is provided for validity.	The model assumptions are not stated. All models used in the estimation of statistics are not described. The model assumptions are not valid, or no evidence is provided for this.
	<p><b>6.2.23</b> Item non-response rate must be within acceptable levels.</p> <p>Where; x= Item non-response rate</p>	$x \leq 0.05$	$0.06 \leq x \leq 0.1$ $0.05 < x \leq 0.1$	$0.11 \leq x \leq 0.15$ $0.10 < x \leq 0.15$	$x \geq 0.15$ Or $x > 0.15$ Item non-response rate is not calculated
	<p><b>6.2.24</b> Unit non-response rate must be within acceptable levels.</p> <p>Where; x= Unit non-response rate</p>	$x \leq 0.05$	$0.06 \leq x \leq 0.1$ $0.05 < x \leq 0.1$	$0.11 \leq x \leq 0.15$ $0.10 < x \leq 0.15$	$x \geq 0.15$ Or $x > 0.15$ Unit non-response rate is not calculated

Cont'd: Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
<b>6.3</b> The extent to which the primary data is appropriate for the statistical product produced	<b>6.3.1</b> Source data must be consistent with the scope, definitions, and classifications of the statistical product produced.	Source data are consistent with the scope, definitions, and classifications of statistical product produced.	Source data is consistent with scope and definitions.	Source data is consistent with the scope of the product.	Source data is consistent with only definitions and/or classifications of the statistical product.
<b>6.4</b> Data from the primary source have been quality assessed.	<b>6.4.1</b> Source data must be accompanied by a quality assessment report.	Source data is accompanied by a quality assessment report.	N/A	N/A	Source data does not have a quality assessment report.
<b>6.5</b> Register/frame maintenance procedures are adequate. <ul style="list-style-type: none"> <li>• Updates.</li> <li>• Quality assurance.</li> <li>• Data audit.</li> </ul>	<b>6.5.1</b> Maintenance procedures of register/frame must be documented. Registers/frames must be updated on regular basis in line with what has been documented.	Maintenance procedures of register/frame must be documented. Registers/frames must be updated on regular basis in line with what has been documented	Maintenance procedures of register/frame must be documented. Registers/frames are updated on an ad hoc basis.	Maintenance procedures of register/frame are not documented. Registers/frames are not updated on an ad hoc basis.	Maintenance procedures of register/frame are not documented and are not performed on a regular basis.
	<b>6.5.2</b> The impact of frame maintenance must be measured, monitored, analyzed and reported on.	Maintenance impact is measured, monitored analyzed and reported on.	Maintenance impact is measured and monitored, reported on but not analyzed.	Maintenance impact is measured but not monitored or analyzed nor reported on.	Maintenance impact is not measured, monitored or analyzed.
<b>6.6</b> Data collection systems are sufficiently open and flexible to cater for new developments.	<b>6.6.1</b> The system must be flexible enough to cater for new developments.	The system is flexible enough to cater for new developments.	N/A	N/A	The system is inflexible.

## 7. Timeliness and Punctuality

### a. Description

Timeliness of statistical information refers to the time lag between the reference point to which the information pertains and the date on which the information becomes available. Timeliness also addresses aspects of periodicity and punctuality of production activities within the statistical value chain. Punctuality of statistical product is the time difference between the date the data are released and the target date on which they were scheduled for release, as announced in an official release calendar and laid down by regulations or previously been agreed with users.

### b. Key components

- Statistics production time
- Periodicity and punctuality of statistical release.

### c. Quality indicators, standards and benchmarks

## Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
<p><b>7.1</b> Average time between the end of reference period and the date of the preliminary results.</p> <p>(ratings are informed by the - GDDS and SDDS Document)</p>	<p><b>7.1.1</b> The preliminary results must be released according to the prescribed standard.</p> <p>Metrics:</p> $a = \frac{\text{date of preliminary results} - \text{end of reference period}}{\text{prescribed timeframe}}$ <p>a. &gt; 1 implies that the archiving phase exceeded the planned timeframes.</p> <p>a. ≤ 1 implies that the archiving phase was completed within the planned timeframes.</p>	<p>0.9 &lt; a. ≤ 1. Preliminary results are released in accordance with prescribed timeframes.</p>	<p>1 &lt; a. &lt; 1.25 times the prescribed lapse.</p>	<p>a &lt; 0.9 or a ≥ 1.25 times the prescribed lapse.</p>	<p>No preliminary results are released.</p>
<p><b>7.2</b> Average time between the end of reference period and the date of the final results.</p>	<p><b>7.2.1</b> The final results must be released according to the prescribed standard.</p> $a = \frac{\text{date of final results} - \text{end of reference period}}{\text{prescribed timeframe}}$ <p>a. &gt; 1 means the prescribed timeframe has been exceeded.</p>	<p>0.9 &lt; a. ≤ 1. Final results are released in accordance with prescribed timeframes.</p>	<p>1 &lt; a. &lt; 1.25 times the prescribed lapse.</p>	<p>a &lt; 0.9 or 1.25 &lt; a. &lt; 1.5 times the prescribed lapse.</p>	<p>a. ≥ 1.5 times the prescribed timeframe.</p>
<p><b>7.3</b> Production activities within the statistical value chain are within the planned timelines, viz.:</p> <p>data collection</p> <ul style="list-style-type: none"> <li>• data processing</li> <li>• data analysis</li> <li>• dissemination</li> <li>• Archiving</li> </ul>	<p><b>7.3.1</b> Project plan/schedule of key deadlines related to the statistical value chain must be compiled.</p> <p><b>7.3.2</b> Updates to registers must occur within clearly specified timeframes.</p>	<p>A schedule of key deadlines does exist.</p> <p>Updates to registers occur within clearly specified timeframes</p>	<p>N/A</p> <p>N/A</p>	<p>N/A</p> <p>N/A</p>	<p>A schedule of key deadlines does not exist.</p> <p>There is no time frame to update the register.</p>

Cont'd: Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
	<p><b>7.3.3</b> Data collection must follow the project plan/schedule</p> <p>Metrics</p> $a = \frac{\text{actual duration} - \text{planned duration}}{\text{planned duration}} + 1$ <p>a. &gt; 1 implies that the collection phase exceeded the planned duration.</p> <p>a. ≤ 1 implies that collections phase completed within the planned duration.</p>	a. ≤ 1 times the planned duration for data collection.	1 < a. < 1.25 times the planned duration for data collection.	1.25 ≤ a. < 1.5 times the planned duration for data collection.	a. ≥ 1.5 times the planned duration for data collection.
	<p><b>7.3.4</b> A protocol for the timely delivery of administrative data must exist and must be adhered to.</p>	A protocol for the timely delivery of administrative data exists and it is adhered to.	A protocol for the timely delivery of administrative data exists and it is not adhered to.	N/A	A protocol for the timely delivery of administrative data does not exist.
	<p><b>7.3.5</b> Data processing must follow the project plan/schedule.</p> <p>Metrics:</p> $a = \frac{\text{actual duration} - \text{planned duration}}{\text{planned duration}} + 1$ <p>a. &gt; 1 implies that the processing phase exceeded the planned duration</p> <p>a. ≤ 1 implies that processing phase completed within the planned duration.</p>	a. ≤ 1 times the planned duration for data processing.	1 < a. < 1.25 times the planned duration for data processing.	1.25 ≤ a. < 1.5 times the planned duration for data processing.	a. ≥ 1.5 times the planned duration for data processing.
	<p><b>7.3.6</b> Data analysis must follow the project plan/schedule.</p> <p>Metrics:</p> $a = \frac{\text{actual duration} - \text{planned duration}}{\text{planned duration}} + 1$ <p>a. &gt; 1 implies that the analysis phase has exceeded the planned timeframes.</p> <p>a. ≤ 1 implies that the analysis phase was completed within the planned timeframes.</p>	a. ≤ 1 times the planned duration for data analysis.	1 < a. < 1.25 times the planned duration for data analysis.	1.25 ≤ a. < 1.5 times the planned duration for data analysis.	a. ≥ 1.5 times the planned duration for data analysis.

Cont'd: Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
	<p><b>7.3.7</b> Dissemination must follow the project plan/schedule.</p> <p>Metrics:</p> $a = \frac{\text{actual duration} - \text{planned duration}}{\text{planned duration}}$ <p>a. &gt; 1 implies that the dissemination phase exceeded the planned timeframes.</p> <p>a. ≤ 1 implies that the dissemination phase was completed within the planned timeframes.</p>	a. ≤ 1 times the planned duration for dissemination.	1 < a. < 1.25 times the planned duration for dissemination.	1.25 ≤ a. < 1.5 times the planned duration for dissemination.	a. ≥ 1.5 times the planned duration for dissemination.
	<p><b>7.3.8</b> Archiving of products must follow the project plan/schedule</p> <p>Metrics:</p> $a = \frac{\text{actual duration} - \text{planned duration}}{\text{planned duration}}$ <p>a. &gt; 1 implies that the archiving phase exceeded the planned timeframes.</p> <p>a. ≤ 1 implies that the archiving phase was completed within the planned timeframes.</p>	a. ≤ 1 times the planned duration for archiving.	1 < a. < 1.25 times the planned duration for archiving.	1.25 ≤ a. < 1.5 times the planned duration for archiving.	a. ≥ 1.5 times the planned duration for archiving.
<b>7.4</b> Periodicity of release.	<b>7.4.1</b> The periodicity (e.g. monthly, quarterly, and annual) of release must conform to a data dissemination standard.	The periodicity of release conforms to a data dissemination standard.	N/A	N/A	The periodicity of release does not conform to a data dissemination standard.

## 8. Accessibility

### a. Description

The accessibility of statistical information and metadata refers to the ease with which it can be obtained from the agency. This includes the ease with which the existence of information can be ascertained, as well as the suitability of the form or medium through which the information can be accessed. The cost of the information may also be an aspect of accessibility for some users.

### b. Key components

- Catalogue systems are available in the parastatals/government or statistical agency
- Delivery systems to access information
- Measure of release calendar and delivery systems performance

### c. Quality indicators, standards and benchmarks



## Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
<b>8.1</b> Statistical products (e.g. data, metadata) are available to the public	<b>8.1.1</b> The statistical products must be disseminated to the public.	The statistical products are disseminated to the public.	N/A.	The statistical products are disseminated to selected users.	The statistical products are not disseminated.
<b>8.2</b> Rules governing the restricted availability of administrative records are well described and documented.	<b>8.2.1</b> A policy document having clear rules governing the restricted availability of administrative records must exist.	A policy document having clear rules governing the restricted availability of administrative records exists and is adhered to	N/A	A policy document not having clear rules governing the restricted availability of administrative records exists and is not adhered to.	There is no policy document at all.
<b>8.3</b> Types of media and/or channels used for sharing data amongst stakeholders are adequate and preserve confidentiality.	<b>8.3.1</b> Data must be accessible through various channels with mechanisms that ensure confidentiality.	Data are accessible through at least 3 channels with mechanisms that ensure confidentiality.	Data are accessible through at most 2 channels with mechanisms that ensure confidentiality	N/A	At most 2 channels exist for stakeholders to access data and no mechanisms exist to ensure confidentiality
<b>8.4</b> Data are accessible in a user friendly format.	<b>8.4.1</b> The data must be available in different file formats.	Data are available in different file formats.	N/A	N/A	Data are available in only one file format.
<b>8.5</b> Statistics are released according to Release calendar.	<b>8.5.1</b> Statistics must be released according to Release calendar.	Statistics are released according to Release Calendar and made available to users annually.	Statistics are released without adherence to the Release calendar which is availed annually to the users.	Statistics are released without adherence to the Release calendar which is not availed annually to the users.	No Release calendar exists.
<b>8.6</b> Statistical releases are made available to all users at the same time.	<b>8.6.1</b> Statistical release must be made available to all users at the same time	Statistical release has embargo date and time (a release date and time) and is made available to all users at the same time.	Statistical release does not have an embargo date and time, but are made available to all users at the same time.	Statistical release has an embargo date and time, but is not made available to all users at the same time.	Statistical release does not have an embargo date and time, and are not made available to all users at the same time.
<b>8.7</b> Statistics/ administrative records not routinely disseminated are made available upon request.	<b>8.7.1</b> Statistics/ administrative records not routinely disseminated must be made available, and the terms and conditions on which they are made available must be publicized.	Statistics/ administrative records not routinely disseminated are made available, and the terms and conditions are publicized.	Statistics/ administrative records not routinely disseminated are made available, but the terms and conditions are not publicized.	N/A	Statistics/ administrative records not routinely disseminated are not available to users.
	<b>8.7.2</b> Special requests are considered and met.	80-100% of special requests are met	60-79% of special request are met	40-59% of special requests are met	0-39 of special requests are met

## Cont'd Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
<b>8.8</b> User support services exist and are publicized.	<b>8.8.1</b> User support services must exist and widely publicized.	User support services exist and are widely publicized	User support services exist, but are not publicized.	N/A	User support services do not exist.
	<b>8.8.2</b> User support services are effective	75- 100% effective	60 -74 effective	40-59% effective	40-59% effective
<b>8.9</b> A data dissemination policy exists, and it is accessible.	<b>8.9.1</b> A data dissemination policy must exist and be accessible.	A data dissemination policy exists, accessible and is adhered to.	A data dissemination policy exists, accessible but is not adhered to.	A data dissemination policy exists but it is not accessible.	No dissemination policy is in place.
<b>8.10</b> A pricing policy exist, and it is accessible	<b>8.10.1</b> A pricing policy must exist and be accessible.	A pricing policy exists, accessible and is adhered to.	A pricing policy exists, accessible but is not adhered to.	A pricing policy exists but it is not accessible.	No pricing policy is in place.
<b>8.11</b> Catalogues of publications and other services are available to users of statistics.	<b>8.11.1</b> Catalogues of publications and other services must be freely accessible to users of statistics.	Catalogues of publications and other services exist and freely accessible to users of statistics.	N/A.	Catalogues of publications and other services exist, but is not freely accessible to users of statistics.	No catalogue is in place.
<b>8.12</b> Metadata are readily accessible to users.	<b>8.12.1</b> Minimum metadata required for interpreting the product must be accessible.	Minimum metadata required for interpreting the product is readily accessible to users.	Minimum metadata required for interpreting the product is available, but not readily accessible to users.	Incomplete metadata is available.	No metadata is available at all.

## 9. Interpretability

### a. Description

Interpretability of statistical information refers to the ease with which users understand statistical information through the provision of supplementary information (metadata and relevant supporting documents).

### b. Key components

- Metadata on concepts and definitions, classifications and methodology used within the statistical value chain
- Key findings giving the summary of the results.
- Presentation of statistics in a meaningful way.

### c. Quality indicators, standards and benchmarks

#### Quality indicators, standards and benchmarks

Indicator	Standards	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
<b>9.1</b> Documented metadata (definitional, operational, methodological, system and dataset) are sufficient to understand data.	<b>9.1.1</b> Metadata must be documented according to the accepted standards, guidelines or good practices.	A complete set of metadata is documented according to the standard(s).	N/A	A complete set of Metadata is not documented according to the standard(s).	No metadata available.
<b>9.2</b> Statistics are presented in a simple and understandable manner.	<b>9.2.1</b> The presentation of the statistics must be according to the standard	The presentation of the statistics is according to the standard.	N/A	The presentation of statistics is not according to the standard.	The statistics are not presented.
<b>9.3</b> Statistical releases contain a summary of the key findings as defined in the major objectives.	<b>9.3.1</b> Statistical releases must contain a summary of key findings.	Statistical releases contain a summary of key findings as defined by the major objectives	N/A	Statistical releases contain inadequate summary of key findings as defined by the major objectives	Statistical releases are published without summary of key findings

## J. References

South African Statistical Quality Assessment Framework (SASQAF), second edition, Statistics South Africa, 2008

Uganda Quality Assurance Framework

Ethiopian Data Quality Assessment Framework, September 2011

Namibia's Data Quality Assessment Framework, November 2014

Brackstone, G. (1999). Managing Data Quality in a Statistical Agency. *Survey Methodology*, 25, 139-149

Eurostat, (2004) FAO Statistical Data Quality Framework: A multi-layered approach to monitoring and assessment

IMF (DQAF), Source: [http://dsbb.imf.org/vgn/images/pdfs/dqrs\\_Genframework.pdf](http://dsbb.imf.org/vgn/images/pdfs/dqrs_Genframework.pdf) 13 February 2008

UNECE (2009), Generic Statistical Business Process Model, Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS) Version 3.1 – December 2008

OECD (2003/1), Quality Framework and Guidelines for OECD statistical activities, Version 2003/1, September 2003

Statistics Act, (2009), Statistics Botswana.

Botswana Strategy for the Development of Statistics (BSDS),( 2015-2020)

Statistics Canada (2000). Policy on Informing Users of Data Quality and Methodology. Statistics Canada

Statistics Canada (2003). Statistics Canada's Quality Guidelines, Fourth Edition Catalogue no. 12-539-XIE, October 2003

Statistics Netherlands, (2004), Conceptual metadata and process metadata: Key elements to improve the quality of the statistical system

Statistics Act, Act No 6 of 1999, Republic of South Africa, 1999

Sundgren, B. (2000) The Swedish Statistical metadata system. Eurostat conference, 2000

Svante Öberg, Director General (2005) Quality Issues in the European Statistical System, Statistics Sweden

World Bank, Development Data Group and UNESCO: Institute for Statistics: A Framework for Assessing the Quality of Education Statistics

## Annexure A: Protocol Specifying the Procedure for the Statistician General to Designate Statistics as Official Statistics

### 1. Purpose of the Protocol

The purpose of this protocol is to specify the requirements that must be met, and the procedures to be followed, before statistics can be certified as official statistics.

### 2. Mandate

According to the Statistics Act 2009; 'Statistics Botswana shall be Pre-eminent national agency responsible for the development and management of official statistics; and the official source and the custodian of official statistics in the country'.

In addition, the Act Mandates Statistics Botswana to coordinate, monitor and supervise the activities of the National Statistical System (NSS). It further empowers the Statistician General to designate statistics as official statistics 'if

- a. Such statistics was published by Statistics Botswana in its statistical reports; or
- b. Declared by Statistics Botswana to be official statistics and sealed by Statistics Botswana with its official seal for that purpose'.

### 3. Definitions

**Official Statistics:** Statistics produced in accordance with UN Fundamental Principles of Official Statistics which emphasizes adherence to the national and international standards on concepts and definitions, classifications and methodologies; as per section 26 of the Statistics Act.

**Agency:** means a component of National Statistical System that is carrying on statistical business

**National Statistics:** Statistics produced by an agency that are within the public domain, and have not been designated as official statistics.

**Member of National Statistics System (NSS)**

A statistical agency or other organisation that produces, supplies or uses statistics, and has signed a memorandum of agreement or any other documents with Statistician General committing to adhere to common statistical quality criteria, standards and procedures as set down by the Statistician General.

### 4. Scope of the Protocol

The protocol covers the role of the Statistician General, the function of the National Statistical System (NSS), the general principles that guide the evaluation of statistics, and the procedure of evaluating, and designating statistics as official by the Statistician General as well as guidelines on the application of the seal.

### 5. Principles of Official Statistics

The Statistics Act 2009 empowers the Statistician General to coordinate statistics and develop standards for both Statistics Botswana and other data producing agencies. The statistics act specifies that the principles of should be in accordance with the standards and principles issued by the United Nations Statistical Commission, which include the following;

- a) Practical utility, relevance of, and equal access to, statistics produced:
- b) Credibility through the exercise of professional independence, accountability and transparency in data production and dissemination:
- c) Legitimacy through a clear mandate for data access, production and dissemination.
- d) Data quality covering relevance, accuracy, reliability, coherence, comparability, timeliness and use of international standards.
- e) Right of statistical agencies to comment on misleading interpretation and misuse of statistics.
- f) Cost effectiveness through the maximum use of data from various sources provided that data quality, cost and burden on respondents are considered.
- g) Confidentiality protection of individual data collected for statistical purposes.
- h) National coordination for consistency and efficiency of the system: and
- i) Cooperation with agencies of the System, bilateral and multilateral organizations in order to share best practices and improve standards

## 6. Procedure for Declaring Statistics as Official Statistics

**6.1** The Statistician-General will regularly publish and update BQAF as a framework with the producing agencies who may apply to have their statistical product declared as official statistics. The Statistician-General, in consultation with the Head of the producing agency, determines the output of the producing agency to be declared as official statistics.

**6.2** The following are the steps that will be followed in accordance with the Statistics Act Section 28 (4b) for statistics of the requesting agency to be declared as official statistics;

**6.2.1** An agency will apply to the Statistician-General to have their statistics declared as official statistics.

**6.2.2** Application will be referred to the Data Quality Assessment Team (DQAT) constituted by the Statistician General from;

- Quality Assurance Team
- Chief Internal Auditor (observer status).
- Applicant (producing agency)
- Subject-matter expert (recommended by the producing agency and/or the Statistician General)

**6.2.3** Appointed DQAT members will sign Terms of Reference for the review and an Oath of Confidentiality document.

**6.2.4** For assessment to begin, the submitting agency and the statistics under review need to comply with three initial criteria:

- The producing agency must be a member of the NSS.
- The statistics are used to meet user needs beyond those specific and internal to the producing agency.
- The statistics produced should be part of a sustainable series, not a once off collection.

**6.2.5** DQAT will assess the quality of the product in terms of BDQAF requirements, assigning a BDQAF quality level to the product. The assessment process is as per Annexure C and described as follows:

**i.** The applicant will identify all the BDQAF indicators that are relevant to the product under evaluation, and motivates why the remaining indicators are not relevant;

**NOTE:** the selection is based on the requirements of the product or properties of the data. In principle, indicators that provide useful information to users should be selected. Not all indicators are relevant for all products.

**ii.** Once DQAT and the applicant reach agreement on which indicators are relevant and on the standard for each indicator, they sign an agreement to this effect.

**iii.** The applicant will then be asked to produce a quality declaration (self-assessment) for their product, for all agreed indicators; and

**iv.** DQAT will assess these quality statement against relevant standards, and based on the results, assign one of the four quality levels (quality, acceptable, questionable or poor), and will identify areas of improvement in the quality statements.

**v.** It should be noted that the individual marks assigned per indicator are averaged to get the overall score for a particular dimension. Finally the marks for all the dimensions are averaged to get the final score of the product.

#### **Assessment Process:**

- Every dimension has a set of questions to be asked.
- The marks obtained for each question are averaged to get an overall mark for a particular dimension.
- Thereafter, all marks for different dimensions are averaged for a product.

**6.2.6** DQAT will recommend the overall assessment level of the product.

**6.2.7** If the product submitted for evaluation is not classified as quality statistics in terms of the BDQAF levels of evaluation, DQAT will advise the applicant on areas of improvement.

**6.2.8** If the product satisfies the requirements of quality statistics set out in BDQAF, Statistician General will declare the product as official statistics.

**6.2.9** Once the product has been declared as official statistics, it will be published with the Statistics Botswana seal of approval (the official statistics mark), and archived for public access.

**6.2.10** The Statistician General will publish a notice to the effect that a product has been declared as official statistics.

**6.2.11** All other subsequent statistical products shall be subjected to quality assessment process.

## 7. Guidelines for the Usage of the Seal for DQAF Process

### 7.1 Purpose and Scope of Guidelines

The purpose of these guidelines is to provide the National Statistical System (NSS) with guidance on the proper use of the Statistics Botswana seal in an effort to manage expectation. The seal is provided for under Section 5 and Section 28(4) of the Statistics Act 2009. Section 5 prescribes the generic use as well as the management of the seal while Section 28(4) prescribes the use of the seal for authenticating official statistics. The application of the seal for confirmation of official statistics relates to all statistics that are compiled and published from the wider NSS including Statistics Botswana reports. Thus the seal is a distinguishing mark that indicates to users of statistics that the product has successfully gone through the quality assessment process, qualified and hence certified as fit for use.

Statistics Botswana as the coordinator of NSS, by affixing the seal, assumes responsibility and is answerable for the quality of the product presented therein. Proper use of the seal is essential, not only for complying with the Statistics Act, but also for assuring the public that the seal represents the NSS commitment to producing quality official statistics.

### 7.2 Use of Seal

According to section 5 of the Statistics Act of 2009,

***'the seal of Statistics Botswana shall be such a device as may be determined by the Board and shall be kept by the Secretary of the Board. The Affixing of the seal shall be authenticated by the Chairperson of the Board and the Secretary or any other person authorized by a resolution of the board'.***

*Section 28 (4) prescribes the use of the seal for authenticating official statistics. It states that*

***'...statistics in existence at the coming into operation of this Act shall be deemed to be official statistics if –  
Such statistics were published by the Central Statistics Office in its statistical reports; or  
Declared by Statistics Botswana to be official statistics and sealed by Statistics Botswana with its official seal for that purpose. '***

### 7.3 Application of the Seal

It is important to maintain a standard application of the Seal whenever it is used. The Seal is a brand of Statistics Botswana and its application has to be consistent so as to maintain its integrity. Anybody who applies the Seal without the necessary authority shall be committing an offence. The seal shall be applied as a soft copy prior to print production. It will be provided in .png format for placement. The soft copy shall then be placed at the bottom center of the preface page, below the Statistician General's official signature.

The following are guidelines relating to the application of the seal;

- The seal shall be applied in a clear, visible and legible manner.
- The seal shall be used on final products which have been prepared by the member of the NSS.



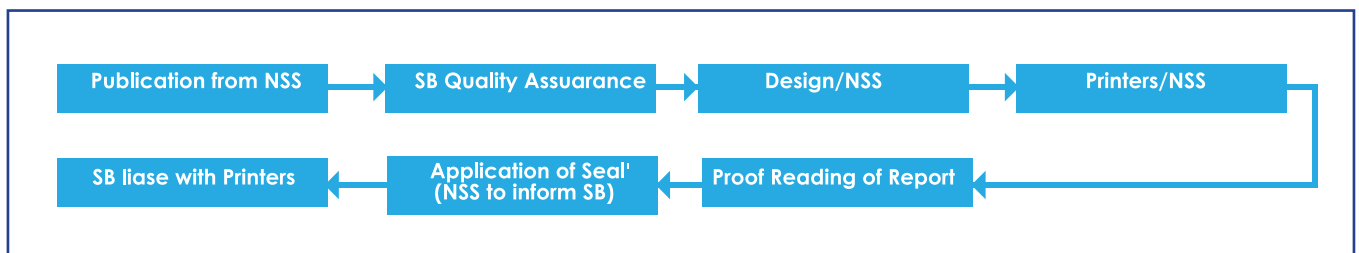
- The seal shall be applied on the preface page of the document at the bottom and shall be centered.
- All documents requiring the seal should undergo validation processes within Statistics Botswana before printing.
- Statistics Botswana shall keep the final copy of the 'sealed' document for reference purposes.
- Once the document is sealed, should there be any changes to the document, Statistics Botswana should be notified and be issued with final sealed copy of the said document.
- The Statistician General shall own up and sign for the quality of the products brought in for the seal.

#### 7.4 Professional Responsibility

The responsibility is always on the applicant to ensure that his/her application of the Seal is done in a legal, ethical and professional manner. It is the responsibility of the applicant to be aware of any legal or employer limitations or requirements on the use of the Seal. Statistics Botswana will assist by providing general guidelines as necessary

Below is a Process Map which cuts across all the necessary procedures up to the application of Seal by the Printers (in consultation with the owner of report, SB and Printers).

- Publication from NSS
- SB Quality Assurance
- Design/NSS (& Design Quality Assurance) • Printers/ NSS
- Proof Reading of Report
- Application of Seal (NSS to inform SB)
- SB liaise with Printers



#### 7.5 Presentation of Statistics Botswana Seal



**Below is an example of application of the seal**

## **PREFACE**

The aim of this report is to publish the latest consolidated labor statistics in Botswana drawn from quarterly publications on formal sector employment stats briefs. This report covers quarters ending March, June, September and December 2012.

Data on work-permit holders is also included in the report. The Department of Labor and Social Security issues work-permits and copies of approved work-permit applications are then forward to Statistics Botswana for processing. Also included in the report, is data on Minimum Hourly Wage Rates set for the Private Sector by the Department of Labor and Social Security.

This report is mainly the work of Labor Statistics Unit of Statistics Botswana. The Office wishes to acknowledge the co-operation of many people in the Private sector, Parastatal Organizations and Government departments who assisted in providing the information used in this report, not forgetting officers who made it possible at all costs for its existence.

---

**B.S Mguni**  
**Statistician General**  
**June 2018**



## **Annexure B: Fundamental Principles of Official Statistics Relative to Statistics Botswana**

Statistics Botswana has adopted the following United Nations fundamental principles of statistics developed by the Economics and Social Council Statistical Commission of the United Nations;

### **Principle 1: Relevance, Impartiality and equal access**

Official statistics provide an indispensable element in the information system of a democratic society, serving the Government, the economy and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis to honour citizen's entitlement to public information.

### **Principle 2: Professional standards and ethics**

To retain trust in official statistics, Statistics Botswana will decide, according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and presentation of statistical data.

### **Principle 3: Accountability and transparency**

To facilitate the correct interpretation of data, Statistics Botswana will present information according to scientific standards on the sources, methods and procedures of statistics.

### **Principle 4: Prevention of misuse**

Statistics Botswana is entitled to comment on erroneous interpretation and misuse of statistics.

### **Principle 5: Cost effectiveness**

Data for statistical purposes may be drawn from all types of sources, be it statistical surveys or administrative records. Statistics Botswana will choose the source with regard to quality, timeliness, costs and the burden on respondents.

### **Principle 6: Confidentiality**

Individual data collected by Statistics Botswana for statistical compilation, whether they refer to natural or legal persons, will be strictly confidential and used exclusively for statistical purposes.

### **Principle 7: Legislation**

The laws, regulations and measures under which the statistical systems operate will be made public.

### **Principle 8: National Coordination**

Statistics Botswana will promote coordination among statistical producers within Botswana in order to advance consistency and efficiency in the statistical system.

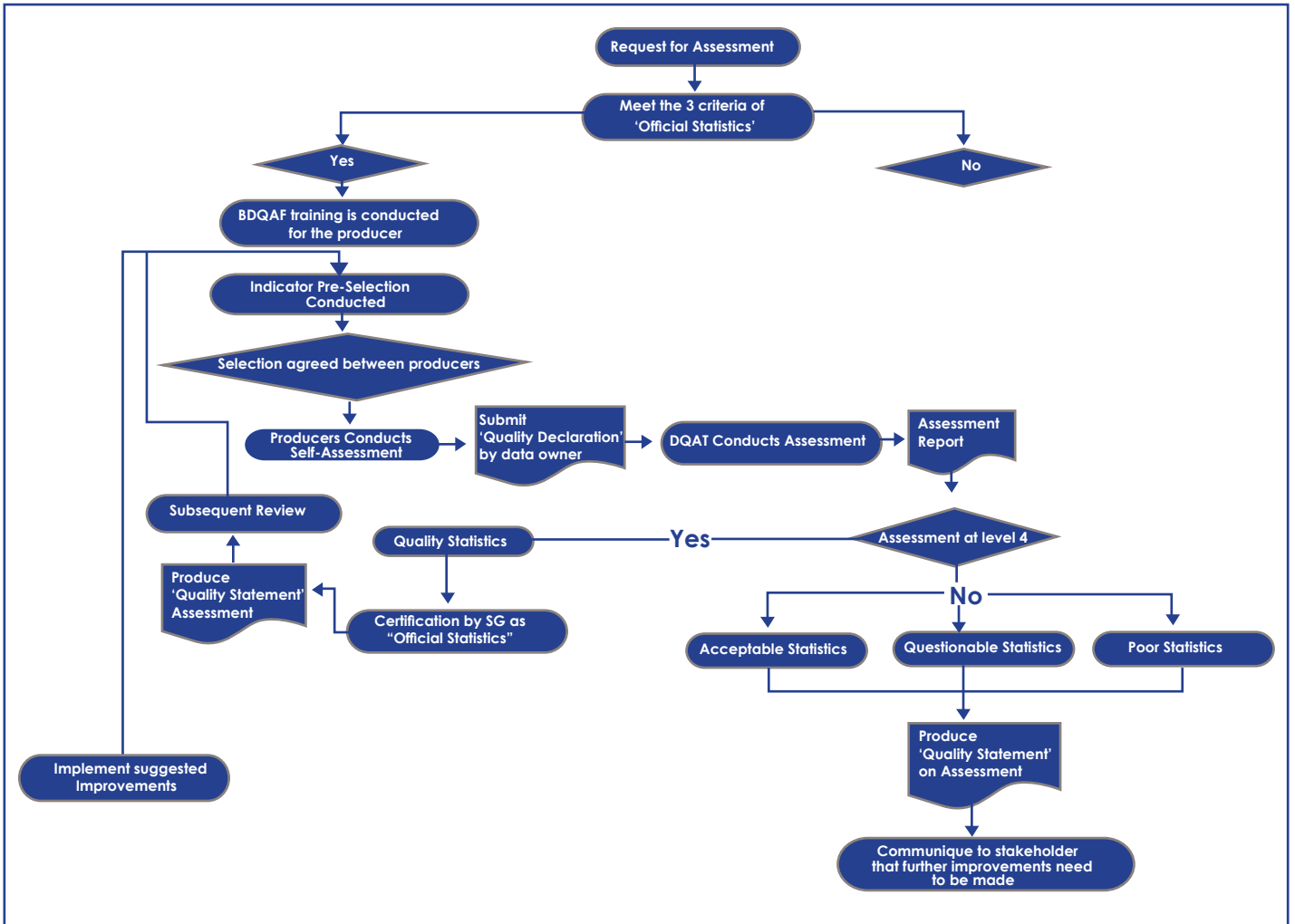
### **Principle 9: International Standards**

Statistics Botswana will use international concepts, classifications and methods, where possible, to promote the consistency and efficiency of statistical systems between countries.

## **Principle 10: International cooperation**

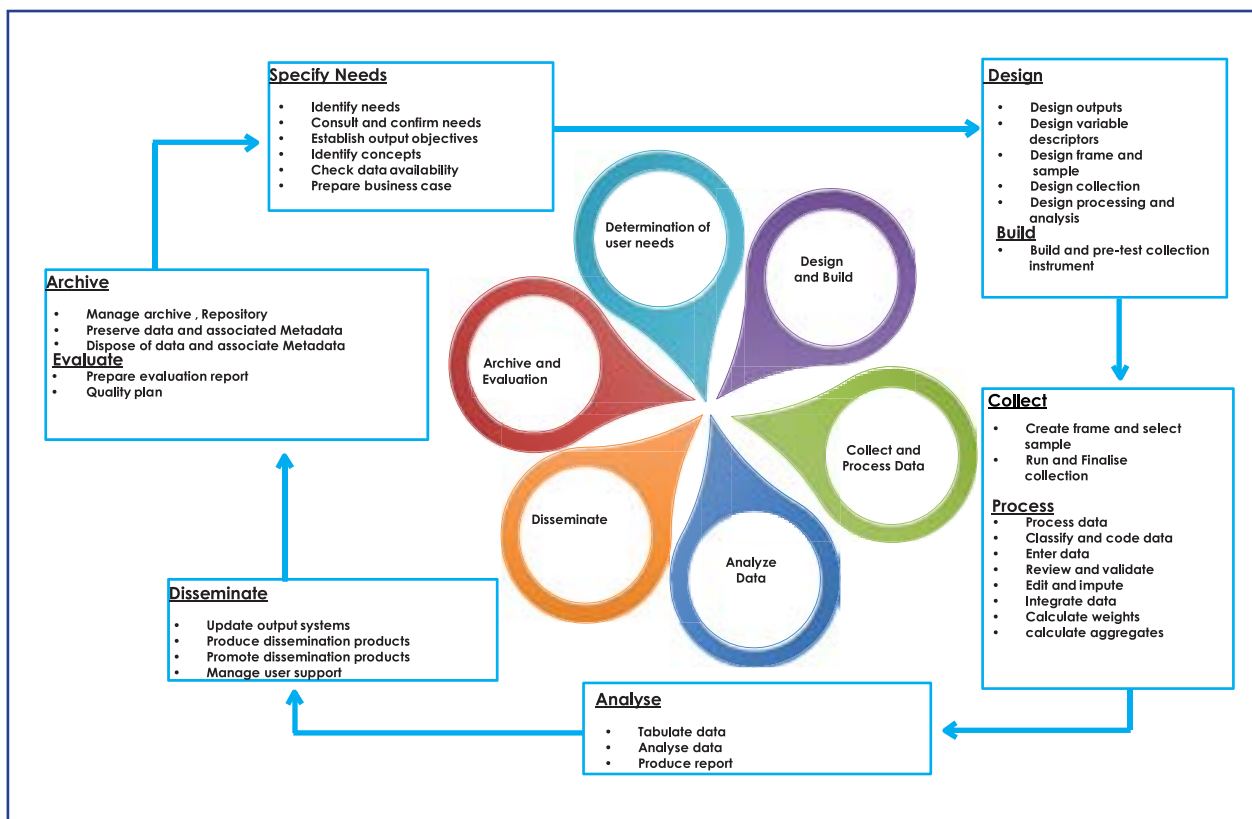
Bilateral and multilateral cooperation in statistics contributes to the improvement of systems of official statistics in all countries.

## Annexure C: The Process of Assessing Statistics and Designating Them as Official



## Annexure D: Statistical Value Chain (Svc)

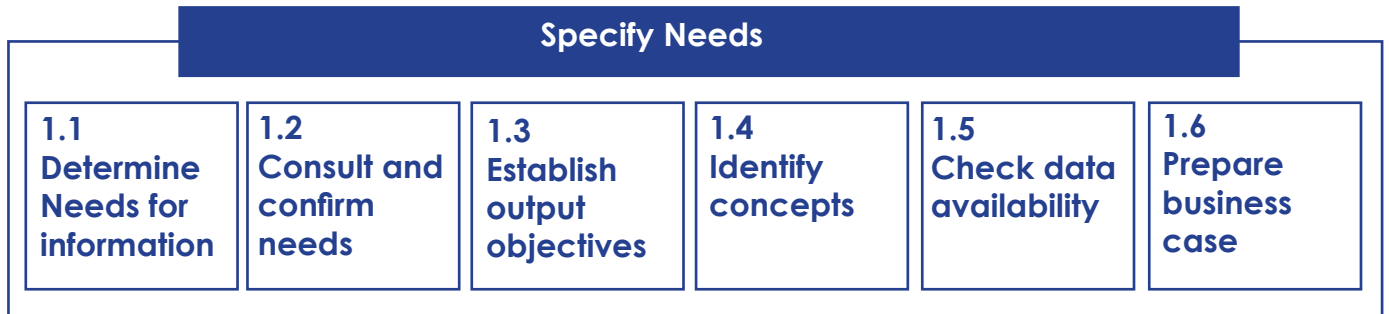
The statistical value chain depicts the primary or core processes that convert inputs into suitable products and services. This SVC was adopted from the Botswana Strategy for the Development of Statistics (BNSDS).



The SVC provides a standardised process approach for statistics production, categorised under different phases, with specific sub processes and components, and ensures these processes flow logically and interacts systematically to deliver high quality outputs and services

As indicated above, phases 1 to 9 which includes determination of user needs, design, build, collect, process, analyse, disseminate, archive and evaluate constitute the statistical value chain. The model guides systems development in an organisation, quality management, metadata development and resource & performance management at all levels.

The summary below describes each of the nine (9) phases, their related processes and sub-processes as adapted from the United Nations Economic Commission for European Countries (UNECE, 2009), Generic Statistical Business Process Model



## Phase 1 – Specify Needs

This phase is triggered when a need for new statistics is identified, or feedback about current statistics initiates a review. It determines whether there is a presently unmet demand, externally and/or internally, for the identified statistics, and whether the statistical organisation can produce them. In this phase the organisation:

- Determines the need for statistics
- Confirms in more detail the statistical needs of the stakeholders
- Establishes the high level objectives of the statistical outputs
- Identifies the relevant concepts and variables for which data are required
- Checks if current collections and /or methodologies can meet these needs
- Prepares the business case to get approval to produce the statistics

This phase is broken down into 6 sub processes. These are generally sequential, from left to right but can also occur in parallel and can be iterative.

The sub-processes are:

### 1.1 Determine need for information

This sub-process includes:

- Initial investigation and identification of what statistics are needed and what is needed of the statistics.
- Consideration of practice amongst other (national and international) statistical organisations, and methods used by those organisations.

### 1.2 Consult and confirm needs

This sub-process focuses on:

- Consulting with the stakeholders and confirming in detail the need for the statistics. Statistical organisations should know what is expected to deliver, when, how, and, perhaps most importantly, why.
- Determining whether previously identified needs have changed. This detailed understanding of user needs is the critical part of this sub-process.

### 1.3 Establish output objectives

This sub-process includes:

- Identifying the statistical output that is required to meet the user needs identified in sub-process 1.2 (consults and confirm need).
- Agreeing the suitability of the proposed outputs and their quality measures.

### 1.4 Identify concepts

This sub-process clarifies the required concepts to be measured by the business process from the point of view of the user. At this stage the concepts identified may not align with existing statistical standards. This alignment, and the choice or definition of the statistical concept and variables to be used, takes place in sub-process 2.2

### 1.5 Check data availability

This sub-process includes

- Checking whether current data sources could meet user requirements, and the conditions under which they would be available, including any restrictions on their use.
- Researching into potential administrative data sources and their methodologies, to determine whether they would be suitable for use for statistical purposes.
- Preparing a strategy for filling any remaining gaps in the data requirement.

**NOTE:** This sub-process also includes a more general assessment of the legal framework in which data would be collected and used, and may therefore identify proposals for changes to existing legislation or the introduction of a new legal framework.

### 1.6 Prepare business case

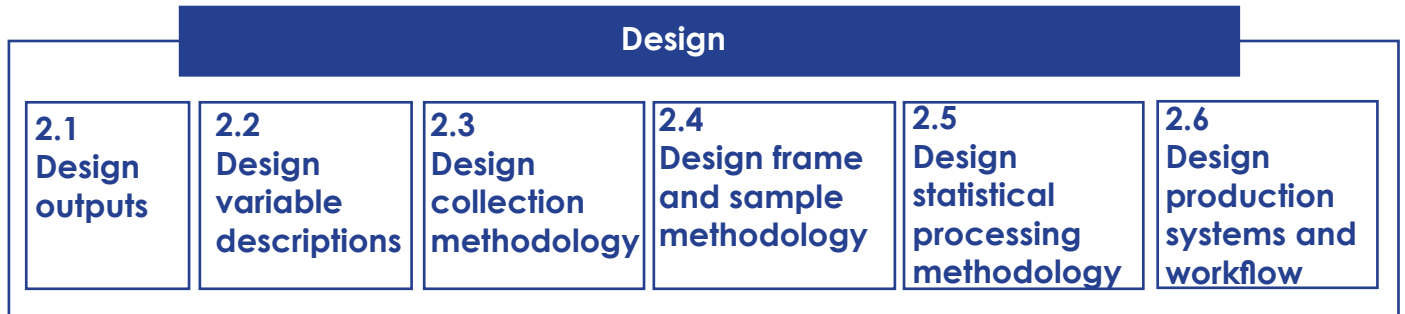
This sub-process documents the findings of the other sub-processes in this phase in the form of a business case to get approval to implement the new or modified statistical business process.

Such a business case would typically include;

- A description of the “As-Is” business process (if it already exists), with information on how the current statistics are produced highlighting and inefficiencies and issues to be addressed;
- The proposed “To-Be” solution, detailing how the statistical business process will be developed to produce the new or revised statistics;
- An assessment of costs and benefits as well as any external constraints.



## Phase 2- Design



The phase describes the development and design activities, and any associated practical research work needed to define the statistical outputs, concepts, methodologies, collection instruments and operational processes. For statistical outputs produced on a regular basis, this phase usually occurs for the first iteration, and whenever improvement actions are identified in phase 9 (Evaluate).

### 2.1 Design Output

This sub-process covers:

- A detailed design of the statistical outputs to be produced, the related development work and preparation of the systems and tools used in the Disseminate phase.
- Output should be designed, whenever possible, to follow existing standards, so inputs to this process may include metadata from similar or previous collections, international standards, and information about practices in other statistical organisations from sub-process 1.1 Determine need for information.

### 2.2 Design Variable descriptions

The sub- process:

- defines the statistical variables to be collected via the data collection instrument, as well as any other variables that will be derived from them in sub-process 5.5 (Derive new variables and statistical units), and any classifications that will be used. It is expected that existing national and international standards will be followed wherever possible.
- may need to run in parallel with sub-process 2.3 (Design data collection methodology), as the definition of the variables to be collected, and the choice of data collection instrument may be inter-dependent to some degree. Preparation of metadata descriptions of collected and derived variables and classifications is a necessary precondition for subsequent phases.

### 2.3 Design data collection methodology

This sub-process includes:

- determining the most appropriate data collection method(s) and instrument(s). The actual activities in this sub-process will vary according to the type of collection instruments required, which can include computer assisted interviewing, paper questionnaires, administrative data interfaces and data integration techniques
- the design of questions and response templates (in conjunction with the variables and classifications designed in sub-process 2.2 (Design variable descriptions))
- the design of any formal agreements relating to data supply, such as memoranda of understanding, and confirmation of the legal basis for the data collection
- enabling tools such as question libraries (to facilitate the reuse of questions and related attributes), questionnaire tools (to enable the quick and easy compilation of questions into formats suitable for cognitive testing) and agreement templates (to help standardize terms and conditions)
- the design of process-specific provider management systems.

### 2.4 Design frame and sample methodology

This sub-process:

- identifies and specifies the population of interest, defines a sampling frame (and, where necessary, the register from which it is derived), and determines the most appropriate sampling criteria and methodology (which could include complete enumeration). Common sources are administrative and statistical registers, censuses and sample surveys;
- describes how these sources can be combined if needed. Analysis of whether the frame covers the target population should be performed. A sampling plan should be made: The actual sample is created sub-process 4.1 (Select sample), using the methodology, specified in this sub-process.

### 2.5 Design statistical processing methodology

This sub-process designs the statistical processing methodology to be applied during phase 5 (Process), and Phase 6 (Analyse). This can include specification of routines for coding, editing, imputing, estimating, integrating, validating and finalising data sets.

### 2.6 Design production systems and workflow

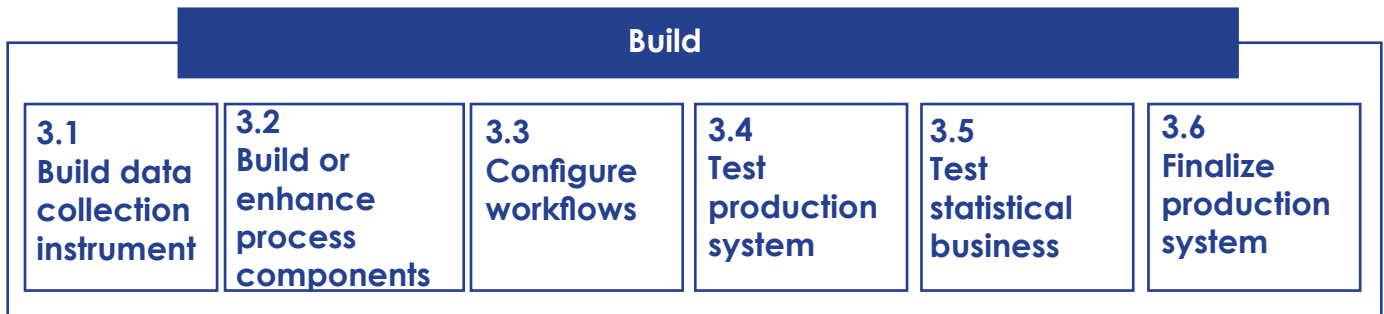
This sub-process includes:

- determining the workflow from data collection to archiving, taking an overview of all the processes required within the whole statistical production process, and ensuring that they fit together efficiently with no gaps or redundancies
- Various systems and databases that are needed throughout the process. A general principle is to reuse processes and technology across many statistical business

processes, so existing systems and databases should be examined first, to determine whether they are fit for purpose for this specific process, then, if any gaps are identified, new solutions should be designed

- considering how staff will interact with systems, and who will be responsible for what and when.

## Phase 3 – Build



This phase builds and tests the production system to the view point where they are ready for use in the “live” environment. For statistical outputs produced on a regular basis, this phase usually occurs for the first iteration, and following a review or a change in methodology, rather than for iteration. It is broken down into five sub-processes, which are generally sequential, from left or right, but can also occur in parallel, and can be iterative.

### 3.1 Build data collection instrument

This sub-process includes:

- Describing the activities to build the collection instruments to be used during the Phase 4 (collect). The collection instrument is generated or built based on the design specifications created during Phase 2 (Design). A collection may use one or more collection modes to receive the data, e.g. personal or telephone interviews; paper, electronic or web questionnaires. Collection instruments may also be data extraction routines used to gather data from existing statistical or administrative data sets
- Preparing and testing the contents and functioning of that instrument (e.g. testing the questions in a questionnaire). It is recommended to consider the direct connections of the collection instruments to the statistical metadata system, so that metadata can be easily captured in the collection phase. Connection of metadata and data at the point of capture can save work in later phases.

### 3.2 Build or enhance process components

This sub-process describes the activities to build new and enhance existing software components needed for the business process, as designed in Phase 2 (Design). Components may include dashboard functions and features, data repositories, transformation tools, workflow framework components, provider and metadata management tools.

### 3.3 Configure workflows

This sub-process configures the workflow, systems and transformations used within the statistical business process and transformations used within the statistical business processes, from data collection, right through archiving the final statistical outputs. It ensures that the workflow specified in sub-process 2.7 (Processing system and workflow) works in practice.

### **3.4 Test Production System**

This sub-process includes:

- testing of computer systems and tools
- technical testing and sign-off of new programmes and routines, as well as confirmation that existing routines from other statistical business processes are suitable for use in this case. Whilst part of this activity concerning the testing of individual components could logically be linked with sub-process 3.2 (Build or enhance process components)
- testing of interactions between components, and ensuring that the production system works as a coherent set of components.

### **3.5 Test statistical business process**

This sub-process includes:

- describing the activities to manage a field test or pilot of the statistical business process
- a small-scale data collection, to test collection instruments, followed by processing and analysis of the collected data, to ensure the statistical business process performs as expected. Following the pilot, it may be necessary to go back to a previous step and make adjustments to instruments, systems or components. For a major statistical business process, e.g. a population census, there may be several iterations until the process is working satisfactorily.

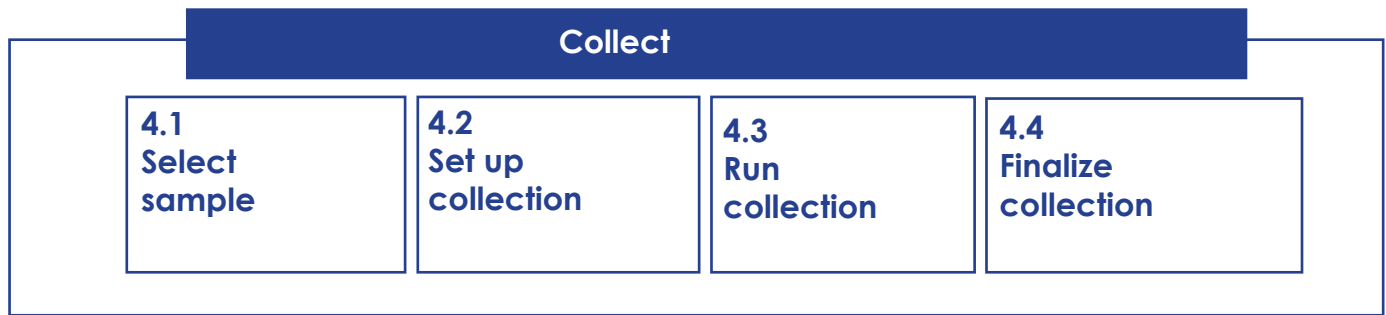
### **3.6 Finalise productions systems**

This sub-process includes activities to put the process, including workflow systems, modified and newly-built components into production ready for by business areas.

The activities include;

- Producing documentation about the process components, including technical documentation and user manuals
- Training the business users on how to operate the process;
- Moving the process components into the production environment and ensuring they as expected in that environment.

## Phase 4 - Collect



This phase collects all necessary data, using different collection modes (including extractions from administrative and statistical registers and databases), and loads them into the appropriate data environment. It does not include any transformations of collected data, as these are all done in phase 5 (Process). For statistical outputs produced regularly, this phase occurs in each iteration.

The Collect phase is broken down into four sub-processes, which are generally sequential, from left to right, but can also occur in parallel, and can be iterative. These sub-processes are:

### 4.1 Select sample

The sub-process includes:

- Establishing the frame and selecting the sample for this iteration of the collection, as specified in sub-process 2.2 (Frame and sample methodology)
- the coordination of samples between instances of the same statistical business process (for example to manage overlap or rotation), and between different process using a common frame or register (for example to manage overlap or to spread response burden). Quality assurance, approval and maintenance of the frame and selected sample are also undertaken in the sub-process, though maintenance of underlying registers, from which frames for several statistical business process are drawn, is treated as a separated business process
- The sampling aspect of this sub-process is not usually relevant for the processes based entirely on the use of pre-existing data sources (e.g. administrative data) as such processes generally creates frames from the available data and then follow a census approach.

### 4.2 Set up collection

This sub-process ensures that the people, processes and technology are ready to collect data, in all modes as designed. It takes place over a period of time, as it includes the strategy, planning and training activities in preparation for the specific instance of the statistical business process. Where the process is repeated regularly, some (or all) of these activities may not be explicitly required for each iteration. For one-off and new processes, these activities can be lengthy.

This sub-process includes:

- preparing a collection strategy
- training collection staff
- ensuring collection resources are available e.g. laptops
- configuring collection systems to request and receive the data
- ensuring the security of data to be collected
- preparing collection instruments (e.g. printing questionnaires, pre-filling them with existing data, loading questionnaires and data onto interviewers' computers etc.).

### **4.3 Run Collection**

This sub-process includes:

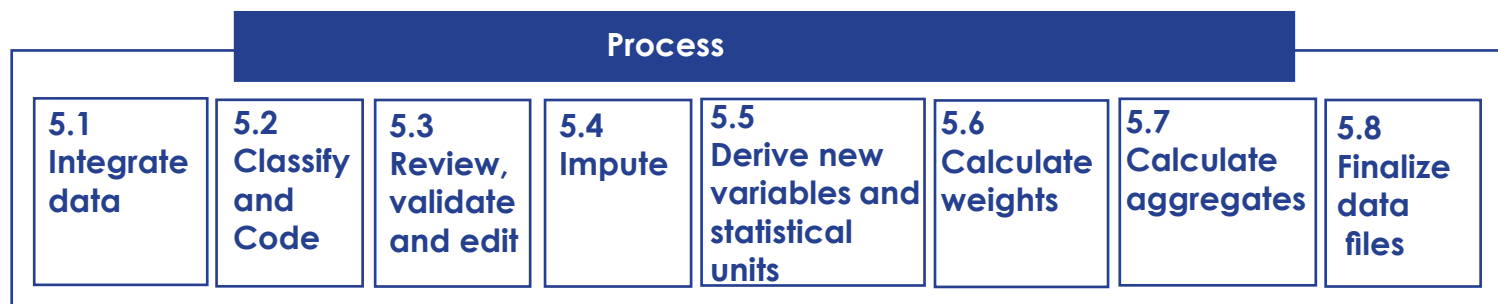
- collection implementation, with the different collection instruments being used to collect the data
- the initial contact with providers and any subsequent follow-up or reminder actions. It records when and how providers were contacted, and whether they have responded
- the management of the providers involved in the current collection, ensuring that the relationship between the statistical organization and data providers remains positive, and recording and responding to comments, queries and complaints. For administrative data, this process is brief: the provider is either contacted to send the data, or sends it as scheduled. When the collection meets its targets (usually based on response rates) the collection is closed and a report on the collection is produced.

### **4.4 Finalize Collection**

This sub-process includes:

- loading the collected data and metadata into a suitable electronic environment for further processing in phase 5 (Process).
- automatic data take-on, for example using optical character recognition tools to extract data from paper questionnaires, or converting the formats of data files received from other organizations.
- cases where there is a physical data collection instrument, such as a paper questionnaire, which is not needed for further processing, this sub-process manages the archiving of that material in conformance with the principles established in phase 8 (Archive).

## Phase 5 –Process



This phase describes the cleaning of data records and their preparation for analysis. It is made up of sub-processes that check, clean, and transform the collected data, and may be repeated several times. For statistical outputs produced regularly, this phase occurs in each iteration. The sub-processes in this phase can apply to data from both statistical and non-statistical sources (with the possible exception of sub-process 5.6 (Calculate weights), which is usually specific to survey data).

The “**Process**” and “**Analyse**” phases can be iterative and parallel. Analysis can reveal a broader understanding of the data, which might make it apparent that additional processing is needed. Activities within the “**Process**” and “**Analyse**” phases may commence before the “Collect” phase is completed. This enables the compilation of provisional results where timeliness is an important concern for users, and increases the time available for analysis. The key difference between these phases is that “Process” concerns transformations of microdata, whereas “Analyse” concerns the further treatment of statistical aggregates.

This phase is broken down into eight sub-processes, which may be sequential, from left to right, but can also occur in parallel, and can be iterative. These sub-processes are:

### 5.1 Integrate Data

This sub-process integrates data from one or more sources. The input data can be from a mixture of external or internal data sources, and a variety of collection modes, including extracts of administrative data. The result is a harmonized data set.

Data integration typically includes:

- matching / record linkage routines, with the aim of linking data from different sources, where those data refer to the same unit
- prioritising, when two or more sources contain data for the same variable (with potentially different values).

Data integration may take place at any point in this phase, before or after any of the other sub-processes. There may also be several instances of data integration in any statistical business process. Following integration, depending on data protection requirements, data may be anonymized, that is stripped of identifiers such as name and address, to help to protect confidentiality.

### 5.2 Classify and code

This sub-process classifies and codes the input data. For example, automatic (or clerical) coding routines may assign numeric codes to text responses according to a pre-determined classification scheme.

### 5.3 Review, validate and edit

This sub-process applies to collected micro-data, and looks at each record to try to identify (and where necessary correct) potential problems, errors and discrepancies such as outliers, item non-response and miscoding. It can also be referred to as input data validation. It may be run iteratively, validating data against predefined edit rules, usually in a set order. It may apply automatic edits, or raise alerts for manual inspection and correction of the data. Reviewing, validating and editing can apply to unit records both from surveys and administrative sources, before and after integration. In certain cases, imputation (sub-process 5.4) may be used as a form of editing.

### 5.4 Impute

Where data are missing or unreliable, estimates may be imputed, often using a rule-based approach. Specific steps typically include:

- the identification of potential errors and gaps
- the selection of data to include or exclude from imputation routines
- imputation using one or more pre-defined methods e.g. “hot-deck” or “cold-deck”
- writing the imputed data back to the data set, and flagging them as imputed;
- the production of metadata on the imputation process;

### 5.5 Derive new variables and statistical units

This sub-process derives:

- variables and statistical units that are not explicitly provided in the collection, but are needed to deliver the required outputs.
- new variables by applying arithmetic formulae to one or more of the variables that are already present in the dataset. This may need to be iterative, as some derived variables may themselves be based on other derived variables. It is therefore important to ensure that variables are derived in the correct order;
- New statistical units by aggregating or splitting data for collection units, or by various other estimation methods. Examples include deriving households where the collection units are persons or enterprises, where the collection units are legal units.

### 5.6 Calculate Weights

This sub process creates weights for unit data records according to the methodology created in sub-process 2.5 (Design statistical processing methodology). These weights can be used to “gross-up” sample survey results to make them representative of the target population, or to adjust for non-response in total enumerations.

### 5.7 Calculate aggregates

This sub process creates aggregate data and population totals from micro-data. It includes summing data for records sharing certain characteristics, determining measures of average and dispersion, and applying weights from sub-process 5.6 to sample survey data to derive population totals.

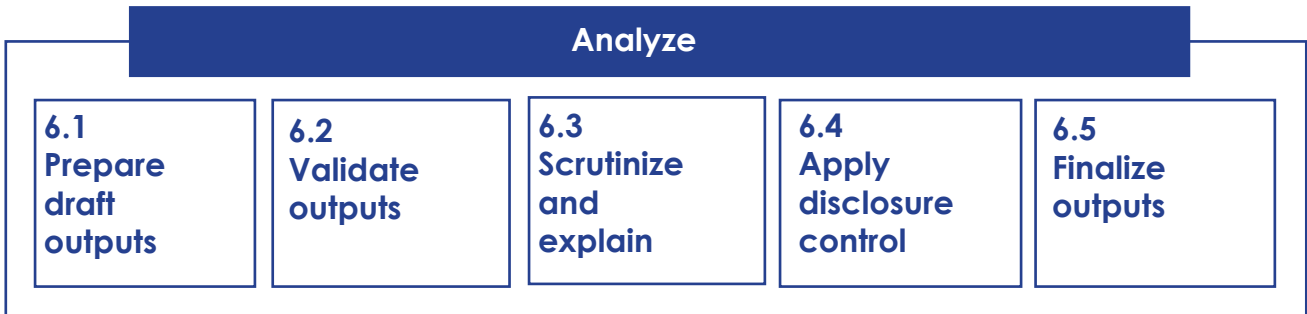
### 5.8 Finalize data files

This sub-process brings together the results of the other sub-processes in this phase and results in a data file (usually of macro-data), which is used as the input to phase 6 (**Analyse**).



Sometimes this may be an intermediate rather than a final file, particularly for business processes where there are strong time pressures, and a requirement to produce both preliminary and final estimates.

## Phase 6 – Analyse



In this phase, statistics are produced, examined in detail and made ready for dissemination. This phase includes the sub-processes and activities that enable statistical analysts to understand the statistics produced. For statistical outputs produced regularly, this phase occurs in every iteration. The Analyse phase and sub-processes are generic for all statistical outputs, regardless of how the data were sourced.

The Analyse phase is broken down into five sub-processes, which are generally sequential, from left to right, but can also occur in parallel, and can be iterative. The sub-processes are:

### 6.1 Prepare draft outputs

This sub-process is where the data collected are transformed into statistical outputs. It includes the production of additional measurements such as indices, trends or seasonally adjusted series, as well as the recording of quality characteristics.

### 6.2 Validate outputs

- This sub-process is where statisticians validate the quality of the outputs produced, in accordance with a general quality framework and with expectations;

This sub-process also includes activities involved with the gathering of intelligence, with the cumulative effect of building up a body of knowledge about a specific statistical domain. This knowledge is then applied to the current collection, in the current environment to identify any divergence from expectations and to allow informed analyses.

Validation activities can include:

- ✓ checking that the population coverage and response rates are as required
- ✓ comparing the statistics with previous cycles (if applicable)
- ✓ confronting the statistics against other relevant data (both internal and external)
- ✓ investigating inconsistencies in the statistics
- ✓ performing macro editing and
- ✓ validating the statistics against expectations and domain intelligence.

### 6.3 Scrutinize and explain

This sub-process is where the in-depth understanding of the outputs is gained by statisticians. They use that understanding to scrutinize and explain the statistics produced for this cycle by assessing how well the statistics reflect their initial expectations, viewing the statistics from all perspectives using different tools and media, and carrying out in-depth statistical analyses.

### 6.4 Apply disclosure control

This sub-process includes:

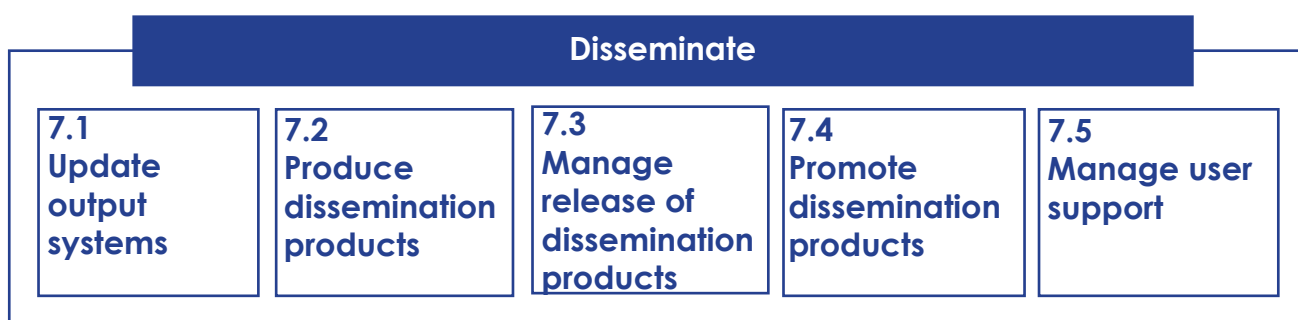
- ensuring that the data (and metadata) to be disseminated do not breach the appropriate rules on confidentiality.
- checking for primary and secondary disclosure, as well as the application of data suppression or perturbation techniques.

### 6.5 Finalize outputs

This sub-process ensures the statistics and associated information are fit for purpose and reach the required quality level, and are thus ready for use. It includes:

- completing consistency checks;
- determining the level of release, and applying caveats;
- collating supporting information, including interpretation, briefings, measures of uncertainty and any other necessary metadata;
- producing the supporting internal documents;
- pre-release discussion with appropriate internal subject matter experts;
- approving the statistical content for release.

## Phase 7 – Disseminate



This phase manages the release of the statistical products to customers. For statistical outputs produced regularly, this phase occurs in each iteration. It is made up of five sub-processes, which are generally sequential, from left to right, but can also occur in parallel, and can be iterative. These sub-processes are;

## 7.1 Update output systems

This sub-process manages the update of systems where data and metadata are stored for dissemination purposes, including:

- formatting data and metadata ready to be put into output databases
- loading data and metadata into output databases
- ensuring data are linked to the relevant metadata.

**NOTE:** formatting, loading and linking of metadata should preferably mostly take place in earlier phases, but this sub-process includes a check that all of the necessary metadata are in place ready for dissemination.

## 7.2 Produce dissemination products

This sub-process produces the products, as previously designed (in sub-process 2.1), to meet user needs. The products can take many forms including printed publications, press releases and web sites. Typical steps include:

- preparing the product components (explanatory text, tables, charts etc.)
- assembling the components into products
- editing the products and checking that they meet publication standards.

## 7.3 Manage Release of Dissemination products

This sub process includes:

- ensuring that all elements for the release are in place including managing the timing of the release
- briefings for specific groups such as the press or ministers, as well as the arrangements for any pre-release embargoes. It also includes the provision of products to subscribers.

## 7.4 Promote dissemination products

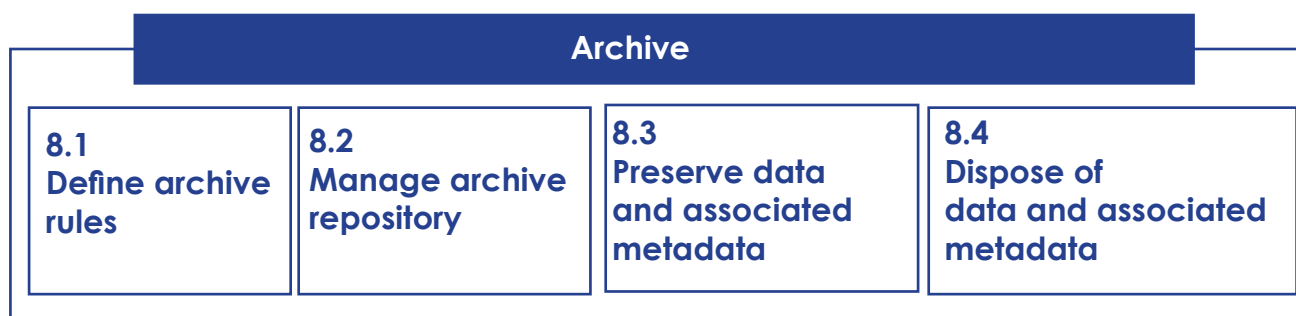
This sub-process includes

- the active promotion of the statistical products produced in a specific statistical business process, to help them reach the widest possible audience
- the use of customer relationship management tools, to better target potential users of the products, as well as the use of tools including web sites, wikis and blogs to facilitate the process of communicating statistical information to users.

## 7.5 Manage user support

This sub-process ensures that customer queries are recorded, and that responses are provided within agreed deadlines. These queries should be regularly reviewed to provide an input to the over-arching quality management process, as they can indicate new or changing user needs.

## Phase 8 – Archive



This phase manages the archiving and disposal of statistical data and metadata. Given the reduced costs of data storage, it is possible that the archiving strategy adopted by a statistical organization does not include provision for disposal, so the final sub-process may not be relevant for all statistical business processes. In other cases, disposal may be limited to intermediate files from previous iterations, rather than disseminated data.

For statistical outputs produced regularly, archiving occurs in each iteration. However, defining the archiving rules is likely to occur less regularly. This phase is made up of four sub-processes, which are generally sequential, from left to right, but can also occur in parallel, and can be iterative. These sub-processes are:

- Checking that the population coverage and response rates are as required
- Comparing the statistics with previous cycles (if applicable)
- Confronting the statistics against other relevant data (both internal and external)
- Investigating inconsistencies in the statistics
- Verify the statistics against expectations and domain intelligence.

### 8.1 Define archive rules

This sub-process is where the archiving rules for the statistical data and metadata resulting from a statistical business process are determined. The requirement to archive intermediate outputs such as the sample file, the raw data from the collect phase, and the results of the various stages of the process and analyse phases should also be considered. The archive rules for a specific statistical business process may be fully or partly dependent on the more general archiving policy of the statistical organization, or, for national organizations, on standards applied across the government sector. The rules should include consideration of the medium and location of the archive, as well as the requirement for keeping duplicate copies. They should also consider the conditions (if any) under which data and metadata should be disposed of.

**NOTE:** This sub-process is logically strongly linked to Phase 2 (Design), at least for the first iteration of a statistical business process).

### 8.2 Manage archive repository

This sub-process covers:

- the management of one or more archive repositories. These may be databases, or may be physical locations where copies of data or metadata are stored. It includes:
- maintaining catalogues of data and metadata archives, with sufficient information to

- ensure that individual data or metadata sets can be easily retrieved;
- testing retrieval processes;
- periodic checking of the integrity of archived data and metadata;
- upgrading software-specific archive formats when software changes.
- a specific statistical business process or a group of processes, depending on the degree of standardization within the organization. Ultimately it may even be considered to be an over-arching process if organization-wide standards are put in place.

### 8.3 Preserve data and associated metadata

This sub-process is where the data and metadata from a specific statistical business process are archived. It includes:

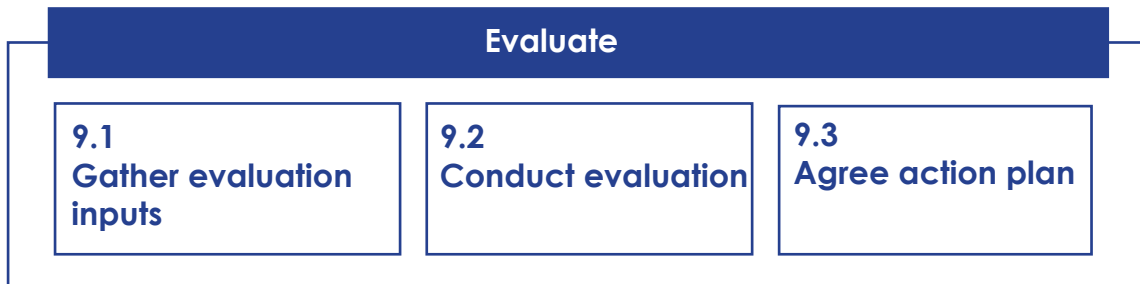
- identifying data and metadata for archiving in line with the rules defined in 8.1;
- formatting those data and metadata for the repository;
- loading or transferring data and metadata to the repository;
- cataloguing the archived data and metadata;
- verifying that the data and metadata have been successfully archived.

### 8.4 Dispose of data and associated metadata

This sub-process is where the data and metadata from a specific statistical business process are disposed of. It includes:

- identifying data and metadata for disposal, in line with the rules defined in 8.1;
- disposal of those data and metadata;
- recording that those data and metadata have been disposed of.

## Phase 9- Evaluate



This phase manages the evaluation of a specific instance of a statistical business process, as opposed to the more general over-arching process of statistical quality management described in Section VI. It logically takes place at the end of the instance of the process, but relies on inputs gathered throughout the different phases. For statistical outputs produced regularly, evaluation should, at least in theory occur for each iteration, determining whether future iterations should take place, and if so, whether any improvements should be implemented. However, in some cases, particularly for regular and well established statistical business processes, evaluation may not be formally carried out for each iteration. In such cases, this phase can be seen as providing the decision as to whether the next iteration should start from phase 1 (Specify needs) or from some later phase (often phase 4 (Collect)).

This phase is made up of three sub-processes, which are generally sequential, from left to right, but which can overlap to some extent in practice. These sub-processes are:

## 9.1 Gather evaluation inputs

Evaluation material can be produced in any other phase or sub-process. It may take many forms, including feedback from users, process metadata, system metrics and staff suggestions. Reports of progress against an action plan agreed during a previous iteration may also form an input to evaluations of subsequent iterations. This sub-process gathers all of these inputs, and makes them available for the person or team producing the evaluation.

## 9.2 Conduct Evaluation

This sub-process analyses the evaluation inputs and synthesizes them into an evaluation report. The resulting report should note any quality issues specific to this iteration of the statistical business process, and should make recommendations for changes if appropriate. These recommendations can cover changes to any phase or sub-process for future iterations of the process, or can suggest that the process is not repeated.

## 9.3 Agree on an action plan

This sub-process includes:

- bringing together the necessary decision-making power to form and agree on an action plan based on the evaluation report.
- consideration of a mechanism for monitoring the impact of those actions, which may, in turn, provide an input to evaluations of future iterations of the process.

## Quality Management

Quality management will be exercised throughout the statistical value chain processes. It is closely linked to phase 9 (evaluate), which has the specific role of evaluating individual stages of a statistical business process. It is also necessary to evaluate separate phases and sub-processes each time they are applied as well as evaluating iterations of a process as per the agreed schedule. These evaluations can apply within a specific process, or across several processes that use common components.

## Metadata Management

Good metadata management are crucial for the efficient operation of statistical business process. Metadata are present in every phase, either created or carried forward from a previous phase. The key challenge is to ensure that they are captured as early as possible, stored and transferred from phase to phase alongside the data they refer to. A metadata management strategy and system(s) are therefore crucial to the operation of this model

## Annexure E: Mapping Quality Dimensions to Activities in the Statistical Value Chain

Activities in the Statistical Value Chain		Quality Dimension
Phases	Sub processes	Quality Dimension
<b>Need</b>	Determine need for information	Pre-requisites of Quality
		Relevance
	Consult and confirm needs	Relevance
	Establish output objectives	Relevance
	Identify concepts	Methodological soundness
	Check data availability	Relevance
<b>Design</b>	Outputs	Comparability and coherence
	Variable descriptions	Comparability and coherence
		Methodological soundness
		Relevance
	Data collection methodology	Accuracy
	Frame and sample methodology	Methodological soundness
		Accuracy
Statistical Processing methodology	Comparability and coherence	
Production systems and workflow	Prerequisite to quality	
<b>Build</b>	Data collection instrument	Accuracy
	Process components	Relevance
	Configure workflow	Pre-requisite of quality
	Test production system	Accuracy
	Test Statistical business process	Accuracy
	Finalize production systems	Accuracy
<b>Collect</b>	Select sample	Accuracy
	Set up collection	Methodological soundness/Accuracy
	Run collection	Methodological soundness/Accuracy
	Finalize collection	Methodological soundness/Accuracy
<b>Process</b>	Integrate data	Comparability and coherence
		Accuracy
	Classify and code	Comparability and coherence
		Methodological soundness
	Review, validate and edit	Accuracy
	Impute	Accuracy
	Derive new variables and statistical units	Comparability and coherence
	Calculate weights	Accuracy
Finalize data files	Accessibility	

Activities in the Statistical Value Chain		Quality Dimension
Phases	Sub processes	Quality Dimension
<b>Analyse</b>	Prepare draft outputs	Comparability and coherence Accessibility
	Validate outputs	Accuracy
	Scrutinize and explain	Interpretability
	Apply disclosure control	Credibility
	Finalize outputs	Accessibility
<b>Disseminate</b>	Update output systems	Accessibility
	Produce dissemination products	Interpretability
	Manage release of dissemination products	Credibility /methodological soundness
	Promote dissemination products	Accessibility
	Manage user support	Accessibility
<b>Archive</b>	Define archive rules	Comparability and coherence Accessibility
	Manage archive repository	Accessibility
	Preserve data and associated metadata	Accessibility
	Dispose of data and associated metadata	Accessibility
<b>Evaluate</b>	Gather evaluation inputs	Prerequisite to quality
	Conduct evaluation	Prerequisite to quality
	Agree on action plan	Accuracy





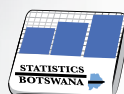
Private Bag 0024, Gaborone  
**Tel:** 3671300 **Fax:** 3952201  
**Toll Free:** 0800 600 200

Private Bag F193,  
City of Francistown Botswana  
**Tel.** 241 5848, **Fax.** 241 7540

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

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

**E-mail:** [info@statsbots.org.bw](mailto:info@statsbots.org.bw) **Website:** <http://www.statsbots.org.bw>



**STATISTICS BOTSWANA**