South African Statistical Quality Assessment Framework (SASQAF)
Operational standards and guidelines

First edition

Statistics South Africa

Preferred supplier of quality statistics

REPUBLIC OF SOUTH AFRICA
STATISTICIAN-GENERAL
NSS National Statistics System
QUALITY APPROVED

South African Statistical Quality Assessment Framework (SASQAF)

Operational Standards and Guidelines

First edition

Statistics South Africa
2010

Pali Lehohla
Statistician-General

Published by
Statistics South Africa, Private Bag X44, Pretoria, 0001

Website address: www.statssa.gov.za
Email: info@statssa.gov.za

Statistics South Africa Library Cataloguing-in-Publication (CIP) Data

South African Statistical Quality Assessment Framework (SASQAF) Operational Standards and Guidelines. 1st ed. / Statistics South Africa – Pretoria: Statistics South Africa, 2010

v, p.87
ISBN 978-0-621-39104-6

1. Statistics - Data quality
2. Data Quality - Assessment.
3. Assessment-Standards
4. Standards - Guidelines

I. South Africa. Statistics South Africa.
II. Title.
(LCSH 24)

A complete set of Stats SA publications is available at the Stats SA Library and the following libraries:
National Library of South Africa, Pretoria Division
National Library of South Africa, Cape Town Division
Library of Parliament, Cape Town
Bloemfontein Public Library
Natal Society Library, Pietermaritzburg
Johannesburg Public Library
Eastern Cape Library Services, King William's Town
Central Regional Library, Polokwane
Central Reference Library, Nelspruit
Central Reference Collection, Kimberley
Central Reference Library, Mmabatho

Enquires regarding the content of this book:
Email: quality@statssa.gov.za

Preface

This publication, SASQAF Operational Standards and Guidelines, is a companion to the South African Statistical Quality Assessment Framework (SASQAF), the framework for assessment of quality of statistics intended for public consumption in South Africa. When SASQAF was first released in 2008, we noted an increased stakeholder interest in data quality, in data quality assessment tools, and in its use among users and producers of statistics. The interest has been stimulated by mounting pressure on heads of government departments to manage for results, especially delivery of services. Accordingly departments have established programmes for the effective monitoring and evaluation of principally their outputs and to a lesser extent outcomes and impacts.

A main concern of government and other stakeholders is the unknown and oftentimes poor quality of the data used to populate indicators that feed the planning and performance assessment processes in all three spheres of government. To counter the lack of small area statistics, municipalities often have to run surveys by hiring service providers to provide the required data, which also turn out to be of unknown quality and incomparable with existing data. It is thus imperative that data collection is designed and developed to meet certain quality standards specified and known to all users and producers of the data. Evaluation of these data must shed light on their quality. SASQAF provides clear criteria and transparent procedures for evaluation of official statistics and other data, but can also be used by data producers for self-assessment of their own products. In addition, the framework can be used in outlining the terms of reference for the production of statistics by those who rely on service providers for collection of their data.

The imperative for this companion to SASQAF emanates from experience gained from efforts to implement SASQAF. It was soon realised after publication of SASQAF that the framework merely provides a listing of indicators and levels of their satisfaction against which to judge the quality of a given dataset. SASQAF thus identifies and describes what needs to be done; it does not identify or describe how it should be done.

SASQAF Operational Standards and Guidelines describes how SASQAF should be implemented by developing generic *standards*, *benchmarks* and *guidelines* with respect to each indicator of SASQAF. A standard explicitly or formally specifies a set of requirements for a given SASQAF indicator. The standard includes metrics. A *benchmark* is in practice a reference point against which to measure performance of an indicator; benchmarks are often established in a comparative context. *Guidelines* specify the procedure to be followed when computing estimates for the indicator. Because the standards, benchmarks and guidelines are generic, they are expected to vary in some definite way among different data collections. Accordingly, it is expected that a similar publication will be produced specific to each of the various statistics series. Thus *SASQAF Operational Standards and Guidelines* constitutes a supplement to SASQAF. The quality indicators are complemented by a standard, or a set of standards, with accompanying benchmarks. The document also provides guidelines for the implementation of the framework and includes a detailed discussion of requirements needed to satisfy the many aspects relating to quality and performance of the standards expected to be achieved for each output.

It is intended to support users and producers of statistics in assessing the quality of statistical output as well as managing quality in the production of statistics.

During the development of this document, some improvements in the quality dimensions and indicators in SASQAF were deemed necessary. The definitions of some of the dimensions were expanded. Coherence was renamed “Comparability and Coherence” since there was no clear distinction previously made between the two concepts. Some indicators were dropped while new, and more appropriate, indicators were developed and others were moved to more suitable dimensions. These improvements have been taken into consideration during the revision of the first edition of SASQAF. Changes to these indicators can also be found in the Appendix C

While its primary audience is departments and agencies that produce data and statistical information, the document will be of interest to other stakeholders, for example those engaged in research, practitioners of monitoring and evaluation, members of the academic community, and service providers to government agencies. This book is the culmination of a comprehensive process of learning and consultation, and has benefited from the input of a number of individuals from the following divisions: National Statistics System and Strategy Division; Standards Division; Methodology and Evaluation Division; Survey Coordination, Monitoring and Evaluation Division; and National Accounts Division.

PJ Lehohla

Statistician-General

Contents

1. Introduction

A. Background

Statistics South Africa (Stats SA), as the agency responsible for collection and dissemination of official statistics, has a particularly central role in evaluation and improvement of data quality. To assist in this process of evaluation, ranking and certification, Stats SA has developed the South African Statistical Quality Assessment Framework (SASQAF). The first edition of this framework was issued in 2008.

The main purpose of SASQAF is to provide a flexible structure for the assessment of statistical products and to enable certification of statistics as official. SASQAF can be used for self-assessment by producers of statistics; reviews performed by a Data Quality Assessment Team (DQAT) for certification purposes; assessment by data users based on the producing agency's quality declaration; assessment by international organizations, again based on the quality declaration.

In the context of SASQAF, quality is defined in terms of prerequisites of quality, as well as the eight dimensions of quality, namely, relevance, accuracy, timeliness, accessibility, interpretability, coherence, methodological soundness, and integrity. Each dimension consists of a number of quality indicators, together with standards to be adhered to. Although these are in place, it is still not easy for producers of statistics to implement SASQAF as intended; it was therefore deemed necessary to develop a document that will facilitate the implementation of SASQAF.

B. Purpose of the document

This document should serve as a guide for data producers and assessors in determining the quality of statistics through the implementation of SASQAF. The document outlines standards that need to be adhered to in relation to a quality indicator, as well as providing benchmarks for the standards which determine the four different quality levels as prescribed in SASQAF edition 1. The document also provides guidelines on how data producers can best achieve good quality products by adhering to the standards specified. Although the guidelines are primarily aimed at producers of official statistics, they can be used by anyone wanting to improve on the quality of their statistical products.

This document attempts to facilitate assessment by specifying the standards that needs to be adhered to in order to satisfy the requirements of each quality indicator. Each chapter addresses a single quality dimension, giving the description of the quality dimension, the key components, which outline what the dimension entails, as well as the purpose of managing the particular dimension. It also aligns each quality indicator within the dimension with standards that need to be adhered to, followed by a set of guidelines that assist in managing the dimension.

C. Glossary or 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 recognised standard, or a reference point, that forms the basis for assessment or comparison.

catalogue: an ordered list of statistical products available in the organisation

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 unauthorised 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 the prerequisites and eight 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 *survey*s and administrative sources collecting information on a particular topic, including definitions, statistical units, 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

D. Definition of data quality

Stats SA defines data quality in terms of 'fitness for use'. Data quality is further defined in terms of prerequisites and the eight dimensions of quality, namely, relevance, accuracy, timeliness, accessibility, interpretability, coherence, methodological soundness and integrity. Five of the eight SASQAF quality dimensions are also covered in the Data Quality Assessment Framework of the International Monetary Fund (IMF).

Prerequisites of quality refer to the institutional and organisational conditions that have an impact on data quality. These include the institutional and legal environment, and availability of human, financial and technological resources.

The *relevance* of statistical information reflects the degree to which it meets the real needs of clients. It is concerned with whether the available information sheds light on the issues of most importance to users.

The *accuracy* of statistical information is the degree to which the output correctly describes the phenomena it was designed to measure. It relates to the closeness between the estimated and the true (unknown) values. Accuracy is measured by means of two major sources of error, namely, sampling error and non-sampling error.

The *timeliness* of statistical information refers to the delay between the reference points to which the information pertains, and the date on which the information becomes available. It also considers the frequency and punctuality of release. The timeliness of information will influence its relevance.

The *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.

The *interpretability* of statistical information refers to the ease with which users can understand statistical information through provision of metadata. This information normally includes the underlying concepts, definitions and classifications used the methodology of data collection and processing, and indicators or measures of the accuracy of the statistical information.

The *coherence* of statistical information reflects the degree to which it can be successfully brought together with other statistical information within a broad analytical framework and over time. The use of standard concepts, classifications and target populations promotes coherence, as does the use of common methodology across surveys.

Methodological soundness refers to the application of international, national, or peer-agreed standards, guidelines, and practices to produce statistical outputs. Application of such standards fosters national and international comparability.

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

These dimensions of quality are overlapping and interrelated. Achieving an acceptable level of quality is the result of addressing, managing and balancing these elements of quality over time with careful attention to programme objectives, costs, respondent burden and other factors that may affect information quality or user expectations. Each dimension has to be adequately managed if information is to be fit for use. Failure to comply with any one dimension will impair the usefulness of the information.

2. Quality dimensions

The chapters which follow specify the quality dimensions of SASQAF in considerable detail, setting out descriptions and key components, the related indicators for each dimension, standard(s) that need to be adhered to in relation to a quality indicator, as well as providing applicable guidelines for each indicator or a group of related indicator and standards .

The prerequisites and eight dimensions of quality

- Prerequisites of quality
- Relevance
- Accuracy
- Timeliness
- Accessibility
- Interpretability
- Comparability and Coherence
- Methodological soundness
- Integrity

Chapter 1: Prerequisites of quality

1.1 Description

The prerequisites of quality refer to the institutional and organisational 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 good quality statistics. It therefore serves as the foundation on which all other dimensions of data quality should be premised on.

1.2 Key components

- Legal and institutional environment (including Memoranda of Understanding (MoUs) or Service Level Agreements (SLAs).
- Privacy and confidentiality.
- Commensurability of resources
- Quality as the cornerstone of statistical work.

1.3 Guidelines

Prerequisites refer to the institutional arrangements which may have a significant influence on the effectiveness and credibility of the agency producing the statistics, and on the quality of statistics in general. Consideration of the institutional environment associated with a statistical product is important as it enables assessment of the surrounding context, which may influence the validity, reliability and appropriateness of the product and ultimately its quality. Institutional frameworks, such as Acts and other legal arrangements, are the superstructures forming the necessary basis for all other measures an agency producing statistics needs for improving quality at statistical output and product level. The issues to consider are the mandate for statistical production, data confidentiality, adequacy of resources, and commitment to quality.

Quality Indicator	1.1 The responsibility for producing statistics is clearly specified.
Standard	1.1.1 A legal arrangement exists that explicitly mandates the production of statistics.

Within the context of the framework, the legal issues surrounding data products include enhancing data product safety and minimizing product liability. Any statistics producer (owner) should be able to demonstrate a clear mandate in one or more of the following ways:

- An Act (e.g. statistics law or departmental act).

- Terms of Reference that gives the appropriate authority to mandate organisations with data collection.
- Memorandum of understanding.
- Service Level Agreement.

While the foregoing list may not be exhaustive, it is incumbent upon any data collecting agency or statistics producer to demonstrate that it has a legal basis for collecting information and that the basis for collection has been obtained from the necessary authority. The mandate should be clear on who should be producing which statistics or data and for what purposes, and thus there are no ambiguities on mandates and responsibility by each agency or department. The producer has a responsibility to act on the mandate given within a specified period of time by establishing arrangements that are consistent with this assignment of responsibility.

Quality Indicator	1.2 Standards and policies are in place to promote consistency of methods and results.
Standard	1.2.1 A set of policies must exist which covers all aspects of the statistical value chain.
	1.2.2 A set of standards related to appropriate policies must exist.

Generally, standards are agreed principles for doing things in a common way. A statistical standard, in particular, provides a comprehensive set of guidelines for surveys and administrative sources collecting information on a particular topic. Data producing agencies are also responsible for developing corporate policies which covers all aspects of the statistical value chain. These policies should be implemented and maintained to be consistent with other policies, especially other quality policies.

All data producing agencies are encouraged to establish an infrastructure or functions that will be responsible for the development and implementation of standards and related policies. Such infrastructure or function will coordinate the development, and promote the use of statistical standards, and liaise with other agencies and will support the rest of the programmes in the organisation by providing standards, as well as the framework for implementation of these standards. The implementation of statistical standards is fundamental to providing a high quality and dynamic statistical service. The availability of standards in the organisation will provide enhanced and consistent access to authoritative and reliable statistical information by users. The development of, and compliance to these standards should cover all aspects of the SVC.

Quality Indicator	1.3 Data sharing and coordination among data-producing agencies are clearly specified.
Standard	1.3.1 A legal arrangement must exist which allows for the timely and efficient sharing of data between the collecting agency and the secondary user.

Arrangements and procedures should be in place to ensure the efficient and timely flow of data between the agency with the primary responsibility for compiling the statistics and other data producing agencies. Data sharing and coordination should occur within a legal framework (e.g. Service Level Agreement, Memoranda of Understandings, or an Act) that directs data-producing agencies to share statistical information. As an example, the Statistics Act (Act no.6 of 1999) mandates Stats SA to coordinate the production

of statistics across data producing agencies in the country. Ideally a legal agreement may exist both within and between agencies. Data producing agency or statistics producer should provide sufficient metadata on applicable processes of the Statistical Value Chain (SVC).

Quality Indicator	1.3 Data sharing and coordination among data-producing agencies are clearly specified.
Standard	1.3.2 Regular contact must occur between the data-producing agencies and secondary users/agencies to resolve statistical issues.

It is advisable for the agencies that share data/depend on each other for data to have regular contact through meetings, workshops and other forums to ensure proper understanding of data requirements. These meetings can serve a variety of purposes and some of these are enumerated below. Firstly, this allows for interaction between the user and producer of the data to promote the idea that statistical needs have to be built into administrative collections since they may not have been designed with statistical production in mind. Secondly, these forums can serve to address statistical issues that prevent the duplication of effort and reduce respondent burden. Thirdly, user needs can be addressed in a structured and formalised process that may enhance the relevance of the statistical product. Lastly, these forums can also assist in aligning operational and statistical definitions and facilitate comparability and coherence with other information within the same domain.

Quality Indicator	1.4 Measures are in place to ensure that individual data are kept confidential, and used for statistical purposes only.
Standard	1.4.1 There must be a law or policy that ensures information collected are kept confidential and used for statistical or administrative purposes only.

In the production of statistics, whether through administrative collections or surveys, data collection agencies should put in place a process that ensures the respondent information remains confidential. The process should be backed up through a law, policy or any other formal provision that clearly states individual responses are to be treated as confidential. The provision should clearly state that responses shall not be disclosed or used for any purpose other than statistical purposes unless disclosure is agreed to in writing by the respondent. It is important that confidentiality be maintained throughout the statistical value chain, while preserving the usefulness of the data outputs to the greatest extent possible.

Data collection agencies should have rules and regulations (e.g. code of conduct, Acts) governing staff to prevent disclosure of information. These rules could include restricting access to unit record data to those who require the information in performing their duties. Steps should be taken to secure the premises of the data producing agency and its computer systems to prevent unauthorized access to individual data. Whether in electronic or paper format, confidentiality of data should be appropriately guarded during storage and during the process of the destruction of records. When the duration for the archiving of forms or questionnaires has lapsed, they should be disposed of in a manner that complies with the applicable policy.

In some cases, due to the uniqueness of certain individual respondents, individual records may be identified even though the data is presented at an aggregate level. For example, it is possible to identify companies with high turnover by sector. A way of avoiding this is to develop special aggregation / anonymisation rules to deal with such

situations. Ideally, before information is solicited from the respondent, officials of the agency should seek to reassure respondents that their information will be treated as confidential and used for statistical purposes only. In order to achieve this, a message to this effect can be stated on the forms/questionnaires, website of the data collecting agency, terms and conditions pamphlet or through a variety of media forming part of a publicity campaign.

Where data is to be exchanged with other agencies, procedures should be in place to adequately secure data while in transit and prevent the negligent loss of data. If the transfer is electronic, then data can be encrypted to ensure that if intercepted the data is of no value to the interceptor. In the same way, necessary precautions should be taken in cases of manual transfer of data. Amongst others, these include a combination of password protection, use of courier services, and encryption of data files.

Quality Indicator	1.5. Measures to oblige response are ensured through law.
Standard	1.5.1. 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.

In collecting data, respondents should be informed of their obligation under law with regard to the provision of information, as well as the consequences for not complying. At a minimum, when soliciting information from respondents they need to be provided with three pieces of information viz. the collecting agency's mandate to collect data; the purpose of collection; and who the data will be shared with. If possible, when collecting data, partially completed forms / questionnaires with information previously supplied and that have remained unchanged should be pre-populated. In this way, the data producing agency demonstrates that it carefully considers respondent burden by shortening the time required in completing the questionnaire or form.

While punitive measures may be applied to secure data from respondents it is advisable for collecting agencies to create goodwill with those from whom they seek to obtain information. This can be achieved in a variety of ways and some of these are outlined below. Collection agencies can establish a single point of contact through service centres to deal with requests from respondents on how to complete questionnaire or forms. These service centres may provide support to respondents who enter from the street or it could be through a call centre or both. Either way the collecting agency demonstrates it goodwill towards the respondent may making a service available that assist in the data collection process. As discussed in the guidelines for standard 0.3.2 goodwill can be created by having pre-populated form sent to respondents where this information has or is unlikely to change.

Quality Indicator	1.6 Resources are commensurate with the needs of statistical programmes. 1.7 Measures to ensure efficient use of resources in 1.6 are implemented • Staff
Standard	1.6.1 Collection programmes must be adequately staffed with skilled personnel.
	1.6.2 There must be a statistics unit or component responsible for compiling statistics.
	1.7.1 Staff of a statistical programme must be employed in positions that are aligned with their skills profile.

Data collecting agencies should build a culture that focus on quality and place emphasis on objectivity and professionalism. Adequate resources and skills should be made available for the purpose intended. A separate unit responsible for the compilation of statistics should exist. This requires that statistical outputs be planned and budgeted for as well as staffed with capable and competent personnel. In the public sector, this will require that agencies whose core business is other than production of statistics plan for this as part of the MTEF.

Human resources should be commensurate with the needs of statistical programmes. Good recruitment processes, such as interviews and competency tests, etc., should be in place with appropriate remuneration and rewards aligned to skills and competencies of the staff it desires to attract. Overall, the staff complement should be adequate to perform the required tasks with none being overloaded with work. Recruitment processes should be aligned with the skills need of the organisation with an initiative such as an internship programme attract young graduates. Job satisfaction should be measured, with areas of concern addressed. Staff should be encouraged to attend conferences and seminars to be keep abreast of new developments in their line of work. Staff should be provided with both formal and on-the-job training in all aspects of statistical value chain, for example in methodology, information technology (IT), analysis, type-setting methods, communication skills, and management etc to acquire and hone the skills necessary to perform their tasks. Sound management principles ensure the retention of core staff which includes methods such as career-pathing strategies, staff mentoring, and succession planning.

The organisation should regularly conduct a skills audit exercise to determine training, recruitment and staff re-assignment needs as well as develop a skills profile for each staff member. Staff, both skilled & semi-skilled, of a statistical programme should be employed in positions that are aligned with their skills profile with documented reasons for deviations from this profile.

Quality Indicator	1.6 Resources are commensurate with the needs of statistical programmes 1.7 Measures to ensure efficient use of resources in 1.6 are implemented • Facilities
Standard	1.6.3 Facilities must have the infrastructure to manage the needs of statistical programmes
	1.7.2 Asset management policies that prevent the abuse of facilities (e.g. vehicles, telephones, etc.) must be developed adopted and implemented

The buildings and other storage facilities used should be secure with restricted access control and have adequate infrastructure to manage the needs of the statistical programme.

Amongst others, this will include:

- a designated warehouse/storeroom to keep questionnaires, maps, listing books, and other relevant documents secure,
- adequate office space for staff to work in,
- facilities for handling forward and reverse logistics of questionnaires efficiently,
- a stores management system that monitors the allocation and return of questionnaires during processing,
- adequate working environment such as lighting, heating, clean air, adequate office space and the provision of office furniture for performing the required tasks,
- a printing facility for dissemination, and
- adequate logistics (vehicles, telephone, fax machines, photocopiers, etc.) to ensure that statistical production occurs in an efficient way.

Asset management policies should be in place to prevent abuse of telephones, printers and other facilities. Policies regarding security of the facility should be in place, especially for areas where sensitive information such as data is stored.

Computing resources should be commensurate with their usage on behalf of the statistical programme. A situation should not exist where management have the most up-to-date and powerful computers whereas data processing and analyses are hamstrung due to the lack of sufficient or powerful computers. All computing resources earmarked for statistical production should be used for such and not be diverted to accommodate systems and processes unrelated to statistical production. Research should take place in the usage of new more efficient technologies

ICT resources are crucial for any modern and efficient statistical programme where collection is through administrative sources, derived surveys, and census or sample surveys. An appropriate type of ICT should be used during data collection programme. This can be achieved by addressing issues related to hardware, software, network and usage of the internet. Failure of ICT systems can have a devastating impact on any statistical programme if many activities of the SVC are automated. It is crucial therefore that loss of data, systems failure, electrical outages, and natural disasters be anticipated and that contingency plans exist to deal with them resulting in minimum impact on the statistical programme.

Quality Indicator	1.7 Measures to ensure efficient use of resources in 1.6 are implemented
Standard	1.7.3 Policies and procedures governing the use of ICT resources must exist, so as to maximise the return on investment.

Policies governing the use and abuse of computing resources should be in place. Amongst others, these policies could address software usage and include:

- Internet;
- Email;
- data dissemination;

- data storage;
- network resources; and
- data access policies that prevent abuse of computing resources.

Quality Indicator	1.6 Resources are commensurate with the needs of statistical programmes. • Computing
Standard	1.6.5 A disaster recovery and business continuity plan must exist.

In view of the foregoing, all statistical programmes should have a disaster recovery and business continuity plan in place that ensures that work continues in the event of incidents listed above. In particular, business continuity plans should allow ICT related statistical activities to continue at another location if access to the organisation's facilities is unavailable. In most cases, the disaster recovery and business continuity plans will be drawn up for the organisation and not the statistical programme. It is incumbent on the owners of the statistical programme to ensure that the business continuity and disaster recovery plans adequately caters for their needs and does not compromise their work.

Network, hardware and software resources

As statistical programmes become more sophisticated, the need arises for data to be moved from disparate geographic locations to a central data repository. By necessity, the use of communications networks is introduced. The use of Local Area Networks, Wide Area Networks, Virtual Private Networks and their associated carrying capacity can become major bottlenecks in ensuring the seamless transfer of regional data to a central repository. Special attention need to be given to the maintenance and upgrading of communications infrastructure if they are used in collecting data from disparate geographic locations. Failure to do so will result in significant degradation of the network's performance.

This will include ensuring that there is sufficient bandwidth where electronic data transfer occurs between locations. Infrastructure such as network switches, and routers need to be maintained and upgraded to ensure the predictable transfer of data collected from various location to a central repository. This becomes crucial in an environment where administrative data collections are conducted on a daily basis and in a variety of locations across the country. This will require the implementation of scheduled and documented maintenance procedures.

Quality Indicator	1.6 Resources are commensurate with the needs of statistical programmes. • Computing 1.7 Measures to ensure efficient use of resources in 1.6 are implemented.
Standard	1.6.4 Computer hardware resources must be adequate in terms of data storage, data backup media, power supply (uninterrupted), memory, and other necessary equipment (notebooks, desktops etc.).
	1.7.3 Policies and procedures governing the use of ICT resources must exist, so as to maximise the return on investment

The demand for ICT resources is critical when they form part of a sustainable statistical programme. Hence, at a minimum, the agency should ensure that hardware for statistical programme has sufficient Random Access Memory (RAM) and data storage capacity that are commensurate with the needs of the programme. Additionally, ICT equipment such as file and database servers are sensitive to changes in electrical current as well as electrical outages. To mitigate the risks associated with the foregoing, it is advisable that servers have Uninterrupted Power Supply (UPS) units. The units will ensure the controlled shutdown of these machines when there is a power outage.
Regardless of the precautions taken above, even the best systems experience unanticipated problems that result in complete systems failure. When this happens the system will have to be restored to state at a given point in time. For this to proceed there has to be regular scheduled backup procedures of databases that will allow for restore points to be established. If the backup procedures are too infrequent then the restoration points may compromise the timeliness of the activity and ultimately impact negatively on the timeliness of the entire SVC.
In the event of hardware upgrades and additional hardware purchases it is imperative that none of the activities linked to the SVC becomes a bottleneck due to the lack of ICT hardware resources. This will require that all hardware resources are equitably distributed across all activities for the entire statistical programme to ensure that they proceed efficiently.

Quality Indicator	1.6 Resources are commensurate with the needs of statistical programmes. • Computing
Standard	1.6.6 Computer software resources must be adequate in terms of capturing systems, editing systems, coding systems, statistical software, up-to-date licences, virus protection, and appropriate access rights.

From the foregoing it is self-evident to have sufficient ICT hardware resources. However, having the appropriate software will allow for the maximum exploitation of hardware resources. The production of statistical products requires the use of specialised software. Amongst others, these may include software such as SAS, STATA and SPSS to perform statistical computations. Where these are used, the agency need to ensure that they have up-to-date software licences that will allow access to the entire suite of statistical functions it is capable of.

The SVC includes activities related to coding, editing, data capturing and analysis that may require additional specialist software. Even though many of these can be done manually, there has been a clear move in automating these by purchasing specialised application software. While clearly desirable, all purchases of specialist software have to be accompanied by a detailed specifications document and business case outlining the proposed benefits it will bring to the statistical programme. This will ensure that funds are not needlessly used to purchase expensive software that does not result in material benefits to the statistical programme such as improved efficiencies through economies of scale.

Even though statistical programmes require specialist statistical software there continues to be a demand for generic business software. These include database software provided major software vendors that allows for managing and administering databases. It is crucial that any data collection programme has sophisticated database application software that allows it to manage the data in the associated databases efficiently.

Quality Indicator	1.6 Resources are commensurate with the needs of statistical programmes. 1.7 Measures to ensure efficient use of resources in 1.6 are implemented. • Financing.
Standard	1.6.7 Budgets must be adequate.
	1.7.4 Budgets must be reviewed and audited to ensure that financial resources are used in the best possible way.

The budget allocated should be adequate for all production and related activities with forward plans put in place to allocate budgetary resources to future statistical development based upon identified statistical needs.

Periodic reviews of budgeting procedures should be undertaken to ensure that financial resources are best employed in addressing major data problems or meeting new data priorities.

Quality Indicator	1.8 Processes are in place to focus on, monitor and check quality.
Standard	1.8.1 The agency must have a quality management system in place.

There is a need for all data producers to develop and implement a quality management system. This system will include tools for quality declaration and assessment, auditing, policies around quality management, metadata management framework, and other interventions to improve the quality of products. The quality management system will provide a structure to ensure that process of managing quality during production of statistics is carried out in a formal and systematic way.

Data quality awareness programmes should be conducted to communicate to the employees the aim of the data quality management system; the advantage it offers to employees, customers and the organisation; how it will work; and their roles and responsibilities within the system. The awareness programme should emphasize the benefits that the organisation expects to realise through its data quality management system. Since data quality management systems affect many areas in the organisation, training

programmes should be devised for different categories of employees. The data quality implementation plan should make provision for this training. The training should cover the basic concepts of data quality management systems, the standard to be adhered to, and their overall impact on the strategic goals of the organisation, the changed processes, and the likely work culture implications of the system.

Quality Indicator	1.8 Processes are in place to focus on, monitor and check quality.
Standard	1.8.2 The data producing agency must have an independent data quality audit process.

Independent quality assessment in the NSS has as its goal to produce more official statistics to inform government on progress made in the implantation of its programmes and projects. Clearly then, data producers should have processes to monitor and assess the quality of processes for the entire SVC. Firstly there needs to be a clear separation of functions of those involved in the statistical production process and those who conduct the quality assessment as this obviates the outcome where assessment and production functions are vested in the same individual. Secondly the quality assessment process requires the existence of quality standards and benchmarks that forms part of a broader quality framework.

Given that assessment is the mapping of a process or system, it follows that assessment presupposes the existence of metadata that has been collected as part of monitoring the operations of the entire SVC. Poor monitoring of processes along with inadequate supporting documentation (i.e. metadata) makes it impossible to arrive at a favourable assessment as an outcome. The assessment process should be undertaken periodically to ensure that where quality standards have been met, they continue to be so in the future; and where this is not the case, incremental improvements are made in a given period of time. A favourable assessment can then lead to the process of certification recommending that the SG approve the product as official statistics and in this way grow the stock of official statistics. Where no such assessment and monitoring system exists chances are slim that statistical product will qualify as official statistics and that the agency only pays lip-service to quality management.

Quality Indicator	1.8 Processes are in place to focus on, monitor and check quality
Standard	1.8.3 Staff members in production areas must have a data quality management requirement as part of their performance agreements or job descriptions.

Managers and professionals should be made sensitive to issues of data quality. Also, data quality criteria should be built into the job description or performance agreements of employees, making them accountable for data quality. Sanctions should be applied for failure to comply with all known quality regulations. Each phase of the value chain should have built in quality assurance (QA) processes. In turn, each QA process should have a checklist of activities to perform. Each quality assurance process should be thoroughly documented and recorded and form part of the operational metadata. Data collection agencies should periodically undertake self-assessment on the quality of the statistics they produce.

Chapter 2: Relevance

2.1 Description

Relevance of statistical information reflects the degree to which the data meet the real needs of clients. It is concerned with whether the available information sheds light on the issues of most importance to users.

2.2 Key components

- Why do you need to conduct the survey or collect data?
- Who are the users of the statistics?
- What are their known needs?
- How well does the output meet these needs?
- Are user needs monitored and fed back into the design process?

2.3. Guidelines

The main purpose of measuring relevance of statistical products is to ensure that statistics are useful to a wide range of users. Maintaining relevance requires keeping abreast of the full range of current and potential user needs. User's requirements for statistical information emanate from their subject matter problems. These may concern issues of economic, demography, environment, and many more. The preferable choices of units, population, variables, etc. in the target characteristics depend on the subject matter problem. A National Strategy for the Development of Statistics (NSDS) serves as the foundation for the production of statistics. It should be based on the most important information needs of the country. The planning and production process of the producing agency should be aligned with the NSDS. To ensure relevance, three primary processes need to be in place, i.e. documenting and monitoring user needs (Classification and a description of the users/database), analysis of user need and implementations of findings of the user satisfaction assessment.

Quality Indicator	2.1 Have both the internal and external users of the data been identified?
Standard	2.1.1 An up-to-date user database must exist.

Measuring relevance requires the identification of user groups and their needs. Since needs evolve over time, a process for continuously reviewing programmes in the light of user needs and making necessary adjustments is essential. The data producing agency should ensure that processes are in place to ascertain user needs and what they use the data for. To ensure that the contents of a product reflect the needs of intended users, data producers should consider user needs early in the development process. This should occur ideally at the conceptual stage of the product development, since most statistical products are conceived due to the desire to answer a question or resolve a need.

Data producers should have strategies in place to collect user feedback on the utility of its products and solicit recommendations for making data more useful. This strategy should include the development of a database containing a list of users classified according to different categories, their contact details, specific usage, classification of users by usage, and their needs. This will allow the data producer to categorise statistical products by the different classes of users. It is recommended that this user database should be kept up-to-date.

The following guideline for the broad classification of users can be used.

- national, provincial and local government; Governments
- research and educational institutions;
- private sector;
- media;
- international agencies;
- non-governmental organisations (NGOs);
- trade unions and associations; and
- others not classified elsewhere.

Quality Indicator	2.2 Is there a process to identify user needs?
Standard	2.2.1 A process to identify user needs must exist.

The following are some mechanisms that may be used to identify user needs and obtain feedback from current users of products on their level of satisfaction, and to identify potential new markets for information:

- user feedback through User Information Service (UIS) and customer satisfaction surveys provide information on the strengths and weaknesses of existing products and gaps in the product line;
- bilateral and multilateral relations with major users,
- user feedback on existing products and services
- a statistics council can provide overall advice on policies and priorities for statistical programmes;
- advice provided by the National Statistics Office (NSO)
- consultative groups
- professional advisory committees in major subject areas may also be set up to regularly review statistical programmes and plans to amend them;
- participation of the head of the data producing agency in cluster meeting discussions with other heads of departments;
- participation of the head of the data producing agency in cabinet lekgotla with ministers and deputy ministers also keeps the producing agency's management aware of current and emerging issues within government;
- liaison through the NSS and consultation arrangements with provincial or local government officials;
- periodic liaison with business associations and labour unions that help to understand information needs and reporting preferences in the business sector;
- ad hoc consultations with interested groups on particular programmes, e.g. on census content, provide input to the design of these programs;
- bilateral liaison with foreign statistical agencies, and multilateral liaison through international organisations e.g., OECD, or other international statistical institutes helps to identify information needs and emerging reporting obligations.

A description of what the data is used for should aim at classifying and ranking the usage of the data. A guideline for the broad classification of data usage follows.

- business planning,
- monitoring and evaluation system (government wide monitoring and evaluation (GWM&E), integrated development plan (IDP) or other performance indicators),
- research,
- international reporting obligation,
- publicity, etc.

Quality Indicator	2.3 Are user needs and the usage of statistical information analysed?
Standard	2.3.1 A report containing the findings of user needs and the usage of statistical information must be available.

The analysis of user needs should be conducted with the intention of assessing them and translating them into statistical objectives. This will allow the producer of statistics to define the statistical parameters of the survey (indicators, precise definitions of the indicators, domains of study, etc.) An analysis of the usage of the data provides feedback on information gaps, the adequacy of the data and their limitations. An active programme of analysis is encouraged and could be nurtured through several mechanisms, viz.

- direct analyses of user needs through customer satisfaction surveys, focus groups;
- number of queries, requests, complaints etc. received;
- through published articles with external authors, often academics and through the peer review processes for these articles and;
- through feedback in reaction to and commentary on analytical results;
- decentralised analysis in subject-matter divisions such as demographic, economic or poverty analyses; and
- contracts with external analysts to produce analytical reports for the data producing agency.

This information provides information for management to decide on revising the data production programme. After gathering user needs, they need to be translated information statistical requirements. The results of user needs analysis will provide information regarding estimates, scope, period for data collection, geographical detail, data items to be covered, target population and other parameters.

Quality Indicator	2.4 Changes are made as a result of user needs assessments.
Standard	2.4.1 The results of the user needs 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.

After using a variety of mechanisms to keep abreast of users' information requirements, the data producing agency should identify areas of weakness and gaps; and provide room for change. The results of the assessments should be built into the data production processes, and influence decisions on the design of the survey or series. However, costs, respondent burden, public sensitivities, and the agency's capacity and expertise need to be taken into account when changes are considered. Some of the features of the process that are particularly important to managing this part of relevance are the following:

- an annual strategic planning meeting to identify major corporate priorities to be addressed in the coming long term planning round;
- an invitation to programme managers to submit new initiatives that would respond to user needs, especially in areas identified as corporate priorities;
- a review and screening of proposals from key users.

Users should be informed about the implementation of their recommendations.

Quality Indicator	2.5 Is there a process to determine the satisfaction of users with the statistical information?
Standard	2.5.1 A formal process must exist to determine the satisfaction of users with the statistical information.

The evaluation of whether users' needs have been satisfied should be carried out using all efficient available means. User satisfaction surveys is the best scenario, but failing in this, proxy measures and substitute indicators of user satisfaction should be produced using auxiliary means such as publication sales, number of questions received, complaints, etc. The results of the assessments should always be built into the corporate processes and influence decisions on the design of the survey. The data producing agency is required to put in place processes that monitor the relevance of its existing programs that identify new or emerging information gaps that the current program is not filling, and that lead to decisions on program change or new funding initiatives aimed at strengthening the relevance of the agency's overall program.

Chapter 3: Accuracy

3.1 Description

The accuracy of statistical information is the degree to which the output correctly describes the phenomena it was designed to measure. Source data available provide an adequate basis to compile statistics.

3.2 Key components

- Assessment of sampling errors where sampling was used.
- Assessment of coverage of data collection in comparison to the target population.
- Assessment of response rates and estimates of the impact of imputation.
- Assessment of non-sampling errors and any other serious accuracy or consistency problems with the survey results or register based statistics.
- Data capture, data coding and data processing errors.
- Source data available provide an adequate basis to compile statistics (e.g. administrative records).
- Source data reasonably approximate the definitions, scope, classifications, valuation, and time of recording required.
- Source data are timely.

3.3 Guidelines

Accuracy of statistics denotes the closeness of estimates to the exact or true values. It is usually characterized in terms of error in statistical estimates and is traditionally decomposed into bias (systematic error) and variance (random error) components. The errors may also be classified in terms of their sources, namely sampling error and non-sampling error. Sampling errors affect only sample surveys; they are simply due to the fact that only a subset of the population, ideally randomly selected, is enumerated, and non-sampling errors may arise from many different sources such as defects in the frame, faulty demarcation of sample units, defects in the selection of sample units, mistakes in the collection of data due to personal variations or misunderstanding or bias or negligence or dishonesty on the part of the investigator or of the interviewer, mistakes at the stage of the processing of the data, etc. This type of error affect sample surveys as well as censuses or statistics derived from administrative registers .Management of accuracy requires attention especially during the design and implementation of survey processes. Assessment of accuracy is also an important part in the management of accuracy, because it will tell how well the designs have been implemented. The results of assessment inform necessary corrections and future considerations.

Design

It is crucial to start off with a very good design in order to ensure a good quality end product. The following are some of the aspects that should be given attention during survey design: objectives of survey, target population, non-response, data collection methods, sample design,

- In designing a survey, data producer should use the appropriate survey frame. The sampling frame should be assessed for coverage of the population targeted by the survey. Sampling frames should be well maintained and updated. Maintenance procedures include, but are not limited to: procedures to eliminate duplication, out-of-scope units, and to update for births and deaths of units as well as changes in characteristics.
- The choice of the sampling method has a direct impact on accuracy. It is influenced by many factors, including the desired level of precision (sample size) of the information to be produced, the availability of appropriate frames, the availability of suitable stratification variables, the estimation methods that will be used and the available budgets.
- The following should be included in data collection plans viz.: frequency and timing of data collections; methods of collection for achieving acceptable response rates; training of enumerators and persons coding and editing the data; and cost estimates, including the costs of pre-tests, non-response follow-up, and evaluation studies.
- Questionnaire design considerations should adequately address the objectives of the survey.
- Trade-offs between accuracy, cost, timeliness and respondent burden should explicitly be considered during the design stage. In addition, evidence that proper consideration was given to these trade-offs should be available in the documentation.
- Alternative sources of data, including the availability of existing data in the organization, or use of administrative records, should be considered; this could help minimize costs associated with starting a new data collection. This issue focuses on minimising respondent burden and the avoidance of unnecessary collection as respondent fatigue is one of the causes of inaccurate responses.
- Quality assurance processes, for monitoring accuracy, should be in place for all stages of the SVC. There should be appropriate internal and external consistency checking of data with corresponding correction or adjustment strategies.

All of the above and more design considerations will improve accuracy of statistics, only if they are implemented according to plan; (a good design can be negated in implementation).

Implementation

While a very good design will contain built-in protection against implementation errors (through quality assurance processes, for example), things can always go wrong. From the management perspective, two types of information are needed at the implementation stage. The first is information enabling managers to monitor and correct any problems arising during implementation as they occur. This requires timely information to enable adjustments or correction of problems while the survey is in progress. The second need

is for the availability of metadata to assess, after the event, whether the design was carried out as planned, whether some aspects of the design were problematic during implementation, and what lessons were learned from implementing the design.

Assessment of errors

The assessment of accuracy entails determining what level of accuracy has actually been achieved. It needs to be a consideration at the design stage since the measurement of accuracy often requires information to be recorded as the survey is taking place. Depending on size of programme; level of accuracy may be assessed for amongst others, the following key estimates: Measures of sampling errors and measures of non-sampling errors, including assessment of coverage, non-response rate, imputation rate, systematic, measurement, processing and model assumptions errors

Quality Indicator	3.1 Measures of sampling errors for key variables are calculated. Amongst others these are: • standard error • coefficient of variation (CV) • confidence interval (CI) • mean square error (MSE) • design effect (DEFF)	
Standard	3.1.1	Measures of sampling errors must be calculated for the main variables. They must be available for the other variables on request.
	3.1.2	Measures of sampling errors must fall within acceptable standards. At a minimum the following must be calculated: standard error, coefficient of variation, confidence interval, mean square error. The low accuracy of variables (if these exist), are explained.
	3.1.3	Scientific Sampling techniques must be used.

Sampling is the selection of a set of units from a survey frame. This set of units is referred to as the sample. An appropriate sampling technique should be identified and used. A distinction in this context is that of probability samples (yielding control of sample inclusion probabilities) and non-probability samples. A probability sample is a safeguard against bias. Given that most sample designs have one or more of the following three characteristics: unequal probabilities of selection, stratification, and clustering, it is important to ensure that appropriate techniques for the estimation of variance in sample surveys are identified, implemented and documented. Variance estimates should be derived for all reported point estimates whether reported as a single, descriptive statistic, or used in an analysis to infer or draw a conclusion. The vast number of different study variables or population characteristics and the different domains of interest in a survey make it impractical and almost impossible to calculate and publish standard errors for each statistic (estimated value of a population variable or characteristic) and for each domain individually. Best practice dictates to publish S.E, CV, CI and MSE for the key study variables on selected domains. It should be remembered that although the standard states that measures of sampling error must fall within acceptable standards; the

extent of the errors is survey specific and has to be established during planning of the survey. However, existing surveys may have well-established standards for measures of sampling error. It is also advisable to caution users on use of impractical domains and disaggregation that result in low accuracy of variables. Whenever possible, construct an estimate of total mean square error in approximate terms, and evaluate accuracy of survey estimates by comparing with other information sources. Methods for deriving approximate standard errors should be indicated for estimates not provided with explicit standard errors.

Quality Indicator	3.2 Measures of non-sampling errors are calculated
Standard	3.2.1 The extent of non-sampling errors must be kept to an acceptable level.

Non-sampling errors should be controlled and reduced to a level at which their presence does not obliterate the usefulness of the final sample results. Unlike in the control of sampling error this error may increase with increases in sample size. If not properly managed non-sampling error can be more damaging than sampling error for large-scale household surveys. When possible, estimate the effects of potential non-sampling errors which include amongst other coverage error, measurement errors due to interviewers, respondents, instruments, and mode; non-response error; and processing error, and model assumption error. The extent of non-sampling error can be determined by formulae below.

$$a = \frac{\Sigma\left|\ \text{final weights} - \text{design weights}\ \right|}{\Sigma\ \text{design weights}}.$$

$$b = \max\left[\frac{\left|\text{final weights} - \text{design weights}\right|}{\text{design weights}}\right].$$

Quality Indicator	3.2 Measures of non-sampling errors are calculated • Frame coverage	
Standard	3.2.2	Delays between newly-registered administrative units and the corresponding statistical unit births must be known. Update procedures are enforced to correct for under-coverage.
	3.2.3	The measures of under-coverage fall within acceptable standards
	3.2.4	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.
	3.2.5	The duplication rate must be at an acceptable level.
	3.2.6	The proportion of units which are out of scope must be at an acceptable level.
	3.2.7	The proportion of units which are misclassified must be at an acceptable level.
	3.2.8	Systematic errors must be identified and reported

This section ensure that necessary steps are taken to develop and maintain data collections that are used as sampling frames, and that coverage of sampling frames is evaluated and documented. There is therefore a need to assess the coverage of the survey in comparison to a target population, for the population as a whole and for significant sub-populations. This may mean assessing the coverage of a sampling frame (e.g. a business register by industry), the coverage of a census that seeks to create a list of a population (e.g. the coverage of a census of population by province or by age and sex), or the coverage of an area sample survey in comparison to independent estimates of the target population (e.g. the difference between sample based population estimates from a household survey and official population estimates). Frames are constructed and maintained through the development of a statistical units register. The statistical units register is a list or equivalent procedure of identifying population units that are part of the survey population. The statistical units register is then maintained through updates from source administrative register (introduction of new units into the survey population and removal of dead units from the survey population based on updates of the administrative register).

The frame often gives auxiliary information about the units, for example industrial classification and economic size of businesses. Incorrect auxiliary information ascribed to units leads to misclassification error. If the variable that is incorrectly classified is also used as a stratification variable for sampling, this has an effect on the accuracy of the estimates. The first time a survey is conducted, background design and coverage work should be done before choosing the frame. Alternative frames, if applicable, should be considered and compared against frames found inside and outside of the production agency, considering total list count comparisons, matching operations, and dual-frame estimation procedures using capture/recapture procedures to estimate non-coverage, and providing an estimation of missing units.

Coverage errors such as over- and under-coverage, non-contacts, classification, temporal errors, and other listing errors should be minimized before the use of a frame. Techniques such as list supplements, multiplicity estimation, half-open intervals, and un-duplication can be used to reduce these errors and improve coverage of the frame.

Any possible changes to frame variables identified by sample survey staff should be reported to the staff responsible for the data collection being used as the frame. Coverage errors are due to divergences between the target population and the frame population. The target population is the population which we would like to describe. *Under-coverage* occurs when some units which are supposed to be included in the frame are excluded, thus giving them no chance of being selected. Under-coverage may happen due to delays in update procedures, lost registration applications for businesses or residential units, units which are live marked improperly as dead on the main frame, and thus not appearing on the main sampling frame. The characteristics of the excluded units introduce a bias in the estimates; for example, if the frame is based on all units which have an email address, it is biased towards units which are technologically enabled in this way. The effect of under-coverage may also underestimate the variance. This is so, if the excluded units differ from the rest of the sampled population.

Over-coverage occurs when units are present in the frame and in fact do not belong to the target population. This may be due to time lags in taking death of a business or person into account, units which are dead marked as live improperly on the main frame, and thus incorrectly appearing on the main sampling frame. The effect of over-coverage may be to underestimate the variance; for example, if the frame is based on all units which have an email address, then email addresses on the frame for units who have died in a random fashion has the effect of introducing randomness to the number of target population units in sample. The frame may contain multiple listings. That is, target population units may be present more than once in the frame. For example, if someone has more than one email address. This gives rise to duplication error. Since a unit which appears more than once on the frame has a greater chance of being selected by probability sampling, the sample is biased towards these duplicate units. If the variable which is used as a sampling unit contains the same information, the effect will be to underestimate the sampling variance. The most effective way to reduce coverage error is to improve the frame by excluding erroneous units and duplicates and updating the frame through filed work to identify units missing from the frame.

Quality Indicator	3.2 Measures of non-sampling errors are calculated • Measurement errors
Standard	3.2.9 Every 10th statistical unit is independently double collected. The two outputs must be compared and corrective action must be taken. Records must be kept.

Data collecting agency should develop protocols to monitor data collection activities, with strategies to correct identified problems. This will include good practice to minimize interviewer falsification, such as protocols for monitoring interviewers and re-interviewing respondents; Rules and guidelines to be followed by data collectors (field staff) should be in place. Inadequate instructions to field staff are another source of error. For some surveys instructions are vague and unclear leaving enumerators to use their own judgment in carrying out fieldwork. Proper training for field staff on these rules and guidelines should also be considered. An internal reporting system that provides timely reporting of response rates and the reasons for non-response throughout the data should be used collection. These systems should be flexible enough to identify important subgroups with low response rates for more intensive follow-ups.

Quality Indicator	3.2 Measures of non-sampling errors are calculated • Measurement errors
Standard	3.3.10 Data collection error rates calculated from fieldwork records must be at an acceptable level.
	3.2.11 The effects of data collection instruments must be determined and reported.

Data producing agency should design the data collection instrument in a manner that minimizes respondent burden, while maximizing data quality.
These are errors that occur during data collection and cause the recorded values to be different than the true ones. The causes are categorised as questionnaire effects, data collection mode effects, interviewer effects and respondent effects. Questionnaire effects (or collection instrument) may be directly due to the layout of the questionnaire that is being filled or due to the type of questions that have been asked.

- Categories in questions may not be mutually exclusive, or they may not be exhaustive.
- Leading or presumptuous questions will cause errors.
- Absence of a correct scale/unit of measurement leads to errors.
- Sensitive questions can lead to false responses or non-responses.
- Close ended questions can limit the respondent.
- Language of questionnaire also can lead to problems, especially misunderstanding of questions.
- Long questionnaires can cause respondent fatigue and thus to incorrect information being provided.

Quality Indicator	3.2. Measures of non-sampling errors are calculated • Measurement errors
Standard	3.2.12 The effects of the data collection mode must be determined and reported.

The choice of physical mechanism to collect data can lead to errors. For example, data can be collected via postal, face-to-face interviews, telephone interviews, leaving the questionnaire for self-enumeration, diary system, administrative record system, direct observation or via the internet.

- Telephonic interviews exclude people without access to a phone.
- Face-to-face interviews may not occur due to remoteness of some locations.
- Problems can occur through failure to understand handwriting, and/or the type of writing tool used.
- Postal questionnaires are often not collected, and also may not be completed by the intended participant.

It is recommended that the effect of data collection mode should be determined and reported

Quality Indicator	3.2 Measures of non-sampling errors are calculated
Standard	3.2.13 The effects of the interviewers must be determined and reported.

Interviewer effects occur when interviewers

- lead respondents in the way they phrase the question, and guess instead of asking the respondent;
- deliberately cheat by completing the forms “sitting under a tree”;
- do not treat questionnaires with proper care, for example, allowing the dog to eat it;
- and interviewees encounter a language barrier; or
- are not allowed to enter gated communities, for example, by not having suitable identification.

Quality Indicator	3.2 Measures of non-sampling errors are calculated
Standard	3.2.14 Respondent effects must be determined and reported.

Encourage respondents to participate to maximize response rates and improve data quality. The following data collection strategies can also be used to achieve high response rates:

- Ensure that the data collection period is of adequate and reasonable length;
- Send materials describing the data collection to respondents in advance, when possible;
- Plan an adequate number of contact attempts;

If applicable, train interviewers and other staff who may have contact with respondents in techniques for obtaining respondent cooperation and building rapport with respondents.

Quality Indicator	3.2 Measures of non-sampling errors are calculated
Standard	3.2.15 Proxy responses must be separately categorised (flagged). Proxy response rate must be at an acceptable level.

Proxy response is a response made on behalf of the sampled unit by someone other than the unit. This is the rate of complete interviews by proxy. It is an indicator of accuracy as information given by a proxy may be less accurate than information given by the desired respondent. The rate of complete interviews given partly by the desired respondent(s) and partly by proxy is an indicator of accuracy as information given by a proxy may be less accurate than information given by the desired respondent.

Quality Indicator	3.2 Measures of non-sampling errors are calculated
Standard	3.2.16 Data entry error must average an acceptable accuracy rate

Data entry errors occur when the data capturer erroneously enters items that are not in the questionnaire, captures items at the wrong location, or omits some of the items recorded in the questionnaire. A questionnaire which is not designed for scanning may give faulty results if it is scanned. These errors should be determined and follow within an acceptable level.

Quality Indicator	3.2 Measures of non-sampling errors are calculated • Processing errors
Standard	3.2.17 Coding must average an acceptable accuracy rate.
	3.2.18 Editing rate must average an acceptable level.
	3.2.19 Editing failure rate must average an acceptable level.
	3.2.20 The imputation rate for item non-response must average an acceptable level.
	3.2.21 The imputation rate for unit non-response must average an acceptable level

Processing error are errors that occur during data processing (coding, editing and imputation), and cause the recorded values to be different than the true ones. Data entry errors occur when the data capturer erroneously enters items that are not in the questionnaire, enters items at the wrong location, or omits some of the items recorded in the questionnaire. A questionnaire which is not designed for scanning may give faulty results if it is scanned. Coding errors occur when survey items are given the wrong code. Data set should be coded to indicate any actions taken during editing, and/or retain the unedited data along with the edited data. For example, codes that clearly identify missing data and cases when entry is not expected (e.g., skipped over by skip pattern) should be inserted into the data set. Standardized codes should be used, if they exist, when converting text data to codes. When setting up a manual coding process to convert text to codes, create a quality assurance process that verifies at least a sample of the coding to determine if a specific level of accuracy coding is being maintained

Prior to imputation the data must be edited. Data editing is an iterative and interactive process that includes procedures for detecting and correcting errors in the data. Data editing must be repeated after the data are imputed, and again after the data are altered during disclosure risk analysis. At each stage, the data must be checked for credibility based on range checks to determine if all responses fall within a pre-specified reasonable range. Consistency based on checks across variables within individual records for non-contradictory responses and for correct flow through prescribed skip patterns; and completeness based on the amount of nonresponse and involves efforts to fill in missing data directly from other portions of an individual's record. Editing errors arise as an effect of checking for inconsistencies and outliers. Editing failure rates express the extent of

distortion occurring between the raw and the edited data. Imputation errors arise from correcting missing and identified-as-incorrect values. Data editing is an iterative and interactive process that includes procedures for detecting and correcting errors in the data.

As appropriate, check data for the following and edit if errors are detected:

- Responses that fall outside a pre-specified range
- Consistency, such as the sum of categories matches the reported total, or responses to different questions are logical;
- Contradictory responses and incorrect flow through prescribed skip patterns;
- Missing data that can be directly filled from other portions of the same record
- The omission and duplication of records
- Inconsistency between estimates and outside sources.

The imputation rate is the percentage of values which have been imputed. There are many techniques or models available to impute for missing values. Depending on the problem, good estimates can be obtained through multiple imputations which adopt a simulation approach, or weighted estimation numerical algorithms like the Expectation Maximisation algorithm. This is further discussed below under the sections covering model assumption and non response error.

Quality Indicator	3.2 Measures of non-sampling errors are calculated • Model assumption errors
Standard	3.2.22 The model assumptions must be stated. All models used in the estimation of statistics must be described.

The estimation process of parameters in a survey process will often incorporate various models. These include generalised regression estimators, seasonal adjustment, cut-off sampling, benchmarking, treatment of missing data (cf. imputation errors above), etc. The use of a model is concomitant with various assumptions of feasibility of the model. Using the wrong models might produce results that are counter intuitive or wrong and thus the assumptions made need to be revisited; resulting in the model being changed or amended. Model assumption errors will most likely lead to bias in the final statistics. The variability of their parameter's estimators will lead to increased variance of the statistics output at the end of the process. Prior to producing estimates, establish criteria for determining when the error (both sampling and non-sampling) associated with a direct survey estimate, model-based estimate, or projection is too large to publicly release the estimate/projection.

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

Document methods and models used to generate estimates and projections to help ensure objectivity, utility, transparency, and reproducibility of the estimates and projections.

Quality Indicator	3.2 Measures of non-sampling errors are calculated • Non-response
Standard	3.2.23 Item non-response rate must be within acceptable levels
	3.2.24 Unit non-response rate must be within acceptable levels

Nonresponse is the failure of a survey to collect the data on all survey variables, from all the units designated for data collection in a sample and/or a complete enumeration. There are three types of non-response:

- unit non-response which occurs when no data is collected about a designated population unit; and
- item non-response which occurs when data on only some but not all the survey variables are collected about a designated population unit.
- overall response rate: is the ratio of the number of units which have responded in some way over the total number of units designated for data collection.

The contribution of non-responding units may be estimated in two ways. Weighted response rates are computed by summing the sample weights of the units. Value weighted response rates are computed by summing auxiliary variables for the non-responding units.

Non-response results in both variability and bias. Variability is introduced since the number of available responses (for example, the sample size) becomes random. Bias may be introduced if respondents do not respond in a systematic way. For example, sensitive data typically results in non-response for that data.

Examples of acceptance level of non-response are as follows:

- For longitudinal sample surveys, the target unit response rate should be at least 70 percent. In the base year and each follow-up, the target unit response rates at each additional stage should be at least 90 percent.
- For cross-sectional samples, the target unit response rate should be at least 85 percent at each stage of data collection.
- For random-digit dial sample surveys, the target unit response rate should be at least 70 percent for the screener and at least 90 percent for each survey component.
- For household sample surveys, the target response rates should be at least 90 percent for the screener and at least 85 percent for the respondents.

Quality Indicator	3.3 To what extent is the primary data appropriate for the statistical product produced? 3.4 Data from primary source have been quality assessed • Accuracy • Timeliness • Comparability and coherence
Standard	3.3.1 Data from another source must be consistent with the scope, definitions, and classifications of the statistical product produced
	3.4.1 Data from another source must be assessed using SASQAF for accuracy, timeliness, and comparability and coherence.

The primary source of data is data that has not been collected by the user. These data may be sample survey, census or administrative record data. Ideally, the user and producer of these data should produce a quality declaration for the data used in producing the statistics. The user should then expect that data received are accompanied by a quality declaration from the primary producer. In the absence of the foregoing, it is incumbent upon the user to produce a quality declaration of its own. Statistical requirements of the output should be outlined and the extent to which the administrative source meets these requirements stated. Gaps between the administrative data and statistical requirements can have an effect on the relevance to the user. Any gaps and reasons for the lack of completeness should be described, for example if certain areas of the target population are missed or if certain variables that would be useful are not collected. Any methods used to fill the gaps should be stated

Quality Indicator	3.1 Register/frame maintenance procedures are adequate. • updates. • quality assurance. • data audit.
Standard	3.5.1 Maintenance procedures of register/frame must be documented and must be performed on a regular basis in line with what has been documented.
	3.5.2 The impact of frame maintenance must be measured, monitored, analysed and reported on.
	3.5.3 A survey for the assessment of quality must be conducted for every release, based on a sample drawn from the administrative records. A register improvement survey must be undertaken periodically to improve the quality of an identified deficiency. Feedback from both surveys must be regularly incorporated into update procedures.

To improve and/or maintain the level of quality of the register or frame, maintenance procedures should be incorporated to eliminate duplications and to update for births, deaths, out of scope units and changes in characteristics.

These procedures must ideally be continuous or be implemented as close as possible to the survey period. Various methods can be used for this purpose: for example, the use of alternative sources of data to complete the existing list and/or feedback from the collection procedures.

- Maintenance based on the most reliable stable source of sufficient coverage must be identified.
- Maintenance must consist of both automatic updates, and investigation cases.
- Profiling and delineation of large and complex businesses in the case of business statistics must be done according to a sound statistical units model. Profiling of standard units is aligned with the statistical units required in social statistics.
- Maintenance takes place to an acceptable level of improvement. Maintenance performance is measured by a quality management framework.
- Register improvement surveys are undertaken to correct particular problem areas. Survey area feedback is regularly incorporated into update procedures.

Quality Indicator	3.6 Are data collection systems sufficiently open and flexible to cater for new developments?
Standard	3.6.1 The system must be flexible in all its components. It must be designed to allow for new developments.

From time to time, changes in legislation, definitions, classifications, etc. can cause a change in some main variables. For example new demarcations which change provincial boundaries, new definitions such as classification of marital status, etc. Therefore the data collection system needs to be made sufficiently open and flexible enough to remain easy to use in the case of possible future developments that may or may not be foreseen at the development stage of the survey. Any current system needs to be periodically reviewed to determine whether it satisfies current survey requirements, and not necessarily the requirements that were expected at the development stage. The practice of adapting core business processes of statistical collection to suit the demands of a data collection system must be avoided. The number of ad hoc methods of using the system must be kept to a minimum. If this number gets too large or results in the risk of poor quality statistics, the system will require a complete overhaul.

Quality Indicator	3.7 Description of record-matching methods and techniques used on the administrative data sources. • match rate as a percentage of total records • measure of false negative matches (missed matches) • measure of false positive matches (mismatches)
Standard	3.7.1 Procedures and/or algorithms must be fully described. The description of the procedures and/or algorithms must be independently verified by replicating the output
	3.7.2 Match rate must be at an acceptable level.
	3.7.3 False negative matches (missed matches) created as a result of missed matches must be at an acceptable level.
	3.7.4 False positive matches (mismatches) created as a result of missed matches must be at an acceptable level.

Record matching is when different administrative records for the same unit are matched using a common, unique identifier or main variables common to both databases, to create a statistical register. This may be through passport numbers, identification numbers, medical aid numbers and enterprise numbers.

There are many different methods for carrying out this procedure:

1. The method (e.g. automatic, clerical matching) used to match administrative sources must be fully described to the extent that a skilled programming practitioner (e.g. database administrator (DBA)) unfamiliar with the techniques is able to reproduce any of the desired results to a comparable quality previously obtained.
2. The matching procedure needs to be periodically tested to see if it conforms to part 1 above.
3. The documentation of the system, algorithms and code must be complete and kept up-to-date. Any additions, deletions or changes in method need to be fully tested before implementation. Any changes must then be fully documented and conform to the same standards as outlined directly above in part. 1 and part. 2 above.
4. Performance statistics for the matching process are produced and monitored. It is expected that these include the following metrics.
    - Match rate;
    - Measure of false negative matches;
    - Measure of false positive matches.

A false negative match is when two records relating to the same entity are not matched, or the match is missed. A false positive match is when two records are matched although they relate to two different entities. False positives and negatives can only be estimated if double matching is carried out. Double matching is where the matching is done twice, and then any discrepancies between the two versions can then be investigated.

Chapter 4: Timeliness

4.1 Description

Timeliness of statistical information refers to the delay 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.

4.2 Key components

- Statistics production time
- Timely receipt of administrative records.
- Periodicity of statistical release.
- Punctuality of statistical release.

4.3 Guidelines

Key issues to be addressed in this dimension are issues around periodicity and punctuality. Periodicity refers to the frequency of compilation and publication (monthly, quarterly, annual etc.) of the data. Among other things, periodicity should consider cost, ease of obtaining information, and user needs. The main goal is for a statistical series to be of appropriate periodicity for user needs. Punctuality refers to whether any phase of the SVC occurred within the planned timeframe and according to schedule. The foregoing holds true regardless whether statistics are compiled from censuses, sample surveys or administrative collections. Two internationally accepted standards for punctuality of dissemination and periodicity are given by the Special Data Dissemination Standard (SDDS) and the General Data Dissemination System (GDDS). The SDDS is applicable to economic and financial statistics while the GDDS is a lesser standard followed by developing countries. Unlike the SDDS, the GDDS also prescribes punctuality and periodicity standards for social statistics.

It must be acknowledged that different quality dimensions have an impact on one another. For information to remain relevant, for example, it needs to be made available in a timely manner since there is an important trade-off between accuracy and timeliness. These considerations are of particular importance in the field of economic statistics since the period for which the data remains useful varies with the rate of change of the phenomena being measured, with the frequency of measurement, and with the immediacy of response that users might make to the latest data. To illustrate, the Monetary Policy Committee of the South African Reserve Bank requires CPI information prior to setting the official bank repurchase (repo) rate. In this instance, the period for which the data remains useful is less than a month. In turn, users take financial positions based on indices

like this as well as the Reserve Bank's repo rate. In a policy environment, it becomes self-evident that failure to provide statistical information on time impacts on the statistical releases' quality since it is no longer fit for purpose. At the other end of the spectrum statistical information derived from population censuses remain useful on the order of years, not months. This demonstrates that timeliness considerations are product specific and are largely determined by the needs of users.

Quality Indicator	4.1 Average time between the end of reference period and the date of the preliminary results. 4.2 Average time between the end of reference period and the date of the final results.
Standard	4.1.1 The Preliminary results must be released according to the prescribed standard.
	4.2.1 The final results must be released according to the prescribed standard.

Reliable data, fit for the purpose for which they were compiled, should be ready for dissemination as soon as feasible after the period to which they relate. For some programmes, the release of preliminary statistical information followed by revised and final figures is used as a strategy for the timely release of statistical information. In such cases, there is a trade-off between timeliness and accuracy. The earlier data is released, the less complete and accurate it is. Nevertheless preliminary data can be used to inform interim decision-making. The tracking of the size and direction of revisions can serve to assess the appropriateness of the chosen timeliness-accuracy trade-off. Where it is applicable to publish preliminary results, these should be communicated to users well in advance of the release date. It is desirable that the release of preliminary results be communicated when the survey manager makes public his/her schedule of key deadlines. All preliminary and final results are to be released in accordance with prescribed standards. Where no standard(s) exist, the decision is informed by user requirements balanced with the feasibility of such releases. This also applies to situations where only one set of results is applicable. Any deviations should be explained through an accessible medium and should clearly state that the results are preliminary findings.

Where no prescribed timeframes for final results exist, the following guidelines may be followed:

Periodicity	**Release date**
Monthly	6 weeks after the end of the reference period.
Quarterly	Within a quarter of the end of the reference period
Annual	Within 12 months of the end of the reference period.
Periodic	Within 12 months of the end of the reference period
Population census:	Within 2 years

Quality Indicator	4.3 Production activities within the statistical value chain are within the planned timelines, viz.: • data collection • data processing • data analysis • dissemination.
Standard	4.3.1 Project plan / schedule of key deadlines related to the statistical value chain must be compiled**.**

Before proceeding with statistical production activities it is desirable to compile a project plan for the statistical programme. This plan should say what process follows which and how long each process will take. The project plan should be of sufficient detail to compile a schedule of key deadlines within the SVC and compliance with this project plan and schedule should be closely monitored. An overrun of one process could adversely affect the quality of output from another process, thereby affecting the quality of the overall statistical product. Anticipated delays across the SVC should be noted and planned for with lag time built into the project timeframe. This implies that overruns for each of the phases in the statistical value chain are anticipated and planned for.

Quality Indicator	4.3 Production activities within the statistical value chain are within the planned timelines, viz.: • data collection • data processing • data analysis • dissemination
Standard	4.3.2 to 4.3.6 All processes (register update, data collection/ delivery of administrative data, processing, analysis, and dissemination) must follow the project plan / schedule.

It is important to monitor the timely implementation of the project plan, and not only concentrate on the dissemination date as this will ensure that justice is done to all production processes and contribute toward quality statistical outputs. Survey managers are therefore expected to be meticulous insofar as updating project plans and schedules of key activities since the burden of proof resides with production staff to demonstrate that activities have occurred and were completed according to plan. The documentation should be part of the operational and system metadata.

Where administrative records are used in compiling statistics and to ensure their timely receipt, it would be beneficial to have a formal arrangement between the statistics producing agency and the data collection agency. Formalised arrangements through SLAs will protect both parties since it will unambiguously indicate what each agency is responsible for to ensure that statistical activities are not negatively affected due to a delay in receiving data.

An indication of the lapse in time since the last update of administrative sources will provide the user with an indication of whether the statistical product is timely enough to meet their needs. This should be updated frequently, according to a prescribed time table and standard. Those who have the responsibility for maintaining the registers should make sure that this time has been adhered to and made known to all users.

Quality Indicator	4.4 Periodicity of release.
Standard	4.4.1 The periodicity (e.g. monthly, quarterly, and annual) of release must conform to a data dissemination standard.

The periodicity (annual, bi-annual, quarterly, monthly, etc.) should be appropriate for the series and should satisfy the needs of users. Where there is a compelling argument and sufficient demand for an increase in the periodicity of a series, this should be considered. There may be additional cost implications, but a detailed cost-benefit analysis should be undertaken to decide whether the series' periodicity should be increased. When disseminating the results, the periodicity of release should be clearly identified and reported as part of the metadata with the periodicity displayed on the front cover of the publication; and in an advance release calendar, when the product is advertised. It is important to follow the organisational standard in this regard, but users also need to be consulted when considering a review of the standard.

Chapter 5: Accessibility

5.1 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.

5.2 Key components

- Catalogue systems are available in the organ of state or statistical agency
- Delivery systems to access information
- Information and metadata coverage is adequate
- Measure of catalogue and delivery systems performance
- Means of sharing data between stakeholders

5.3 Guidelines

Accessibility has to do with the ease with which data is obtained. Thus maximising accessibility is directly related to maximising the number of users who get the data they require. This would include considerations of whether the data is affordable to users, the number and ways of how users are able to get the data (e.g. channels and media), and whether potential users know about the data. Output catalogues; delivery systems, distribution channels and media, and strategies for engagement with users and the data dissemination policy are all important considerations relating to this quality dimension. Statistical information that users don't know about, can't locate, or, haven't located but can't access or afford, is of no value to them.

Quality Indicator	5.1 Are statistical products (e.g. data, metadata) available to the public?
Standard	5.1.1. The statistical products must be disseminated to the public.

To ensure that the contents of a product reflect the needs of intended users, authors should consider user needs early in the publication development process. A data dissemination plan should be available as early as in the planning of the product. Once the product has been approved for release the data producer should organise a

meeting to review proposed dissemination strategies including press releases, targeted mailings, libraries, the number of copies to be printed, web release, the use of print on demand, and the use of both print and electronic announcements.
Where possible, the strategy should be to make information of broad interest available free of charge through several media (including the press, the internet, research data centres).

Quality Indicators	5.2 Rules governing the restricted availability of administrative records are well described and documented.
Standard	5.2.1 A policy document having clear rules governing the restricted availability of administrative records must exist.

Administrative records are usually collected for purely administrative purposes and help the collecting agency conduct their day-to-day business. Personal/business details are often captured and this makes the data very sensitive. This makes the use of the data for statistical purposes difficult unless a legal framework is in place. Procedures to request access to administrative records should be clearly described. This legal arrangement is between the administrative record data collector and the data users. This arrangement could be via formal MoUs or less formal such as confidentiality clauses.

There are delivery systems available to disseminate data, this include putting data/statistics on the website, the traditional delivery system of printed publications which is still valued by many users, while electronic products on diskette or CD-ROM meet some needs.

Quality Indicator	5.3 Types of media and/or channels used for sharing data amongst stakeholders are adequate and preserve confidentiality.
Standard	5.3.1 Data must be accessible through a variety of channels with mechanisms that ensure confidentiality.

One of the key components necessary for the maximum accessibility is the number of ways data is available to the user. The more the options for the user obtaining the data, the more accessible it is. The following media and channels are commonly used to share data with other data producers: FTP, https, CDs/DVDs, forms, cassettes, external drives depending on the size and nature of the data. Security of the data is maintained by use of secure data transfer, password protection, courier services, locking offices and cabinets used for storing data.

Confidentiality of data may limit accessibility of data. Thus data disseminating agencies should always take into consideration complying with disclosure of private information while providing as much data as possible to users. Thus a clear description of conditions of access to data must be provided outlining possible restrictions, existing MoU and SLA.

Quality Indicator	5.4. Data is accessible in a format beyond the producing agency.
Standard	5.4.1 The data must be available in a variety of file formats.

Producing agencies should make sure that data is available in formats that satisfy the requirements of users. This should be agreed upon during planning of the product to be disseminated in consultation with users. Data should be released using formats that are required by the majority of users. For example, publishing data with SAS, SPSS, and Excel or ASCII format. The data production agency should also consider providing additional data sets in alternative formats that may be helpful to users. A variety of dissemination techniques should be used to accommodate the majority of users. Innovative ways to disseminate data/product should be explored.

Quality Indicator	5.5 Statistics are released on a pre-announced schedule.
Standard	5.5.1 Statistics must be released on a pre-announced schedule

Data producing agency should have advance release calendar which are published on a monthly basis or yearly on the internet. It will contain information on which statistics is being released at a particular time. These schedules also contain the dates and time of release of such product.

Quality Indicator	5.6 Statistical release is made available to all users at the same time.
Standard	5.6.1 Statistical release must be made available to all users at the same time.

All statistical releases or products should have an embargo date and time. All users get the information at the same time although the press and selected users, for example government departments may have limited access to key statistical releases prior to the embargoed time so that they are able to prepare their public responses. In such cases, measures should be in place to prevent misuse of data. A strict lock-up procedure is put in place that allows them to get access to the data one hour before the embargo.

Quality Indicator	5.7 Statistics/administrative records not routinely disseminated are made available upon request.
Standard	5.7.1 Statistics/administrative records not routinely disseminated must be made available, and the terms and conditions on which they are made available must be publicised.
	5.7.2 Special requests must be considered and be met.

Statistics or information that is not disseminated routinely is referred to as special requests. Data producing agency are encouraged to have a policy in place on handling these special request. These requests should be answered promptly (typically within 24 hours). Procedures to request access to confidentialised micro-data should be clearly described in a format easily understood by users, and made readily available. Non-published (but non-confidential) specialized tabulations (e.g., sub-aggregates of units of analysis) are made available upon request. Non-confidential micro-data files (e.g., with information permitting the identification of individual respondents removed) are available to permit analytical use by researchers and other users. The availability of non-published statistics and data, and the terms and conditions on which they are made available are publicised. This may include the publication of users' details and uses of data.

Quality Indicator	5.8 User support services exist and are widely publicised.
Standard	5.8.1 User support services must exist and be widely publicised.

User Support Services (USS) provides a single point of access to the organisation's information. This service will promote the increased effective use of the organisation's data products and services. In response to contacts, staff:

- help define the information requirements of the client;
- provide data and information on the data producer's products and services;
- develop customized, cost-efficient data solutions; and facilitate and serve as the direct link to the rest of the organisation's researchers, analysts, consultants and other technical experts.
- Users are advised to make their enquiries by telephone, fax, email, internet and post.

Any statistical release should have people assigned to key user response roles; e.g. analysis, queries, etc. The contact details for these people must be communicated to the central USS for every product.

Quality Indicator	5.9 Does a data dissemination policy exist, and is it accessible?
Standard	5.9.1 A data dissemination policy must be accessible.

A data dissemination policy is a document that guides the organisation in dealing with data dissemination issues such as:

- the nature of the data which is released (e.g. full data versus a 10% sample or unit record versus aggregated data);
- terms of condition for data use (i.e. copyright and acknowledgment of the producer of the data);
- confidentiality information;
- cost of data;

- periodicity of release;
- metadata availability and limitations on published documents;
- choice of media.

The organisation responsible for producing and disseminating the data has the right to amend this document from time-to-time without prior notice. The policy should be available to users.

Quality Indicator	5.10 Does the pricing policy governing dissemination exist, and is it accessible?
Standard	5.10.1 A pricing policy must be accessible.

Data collection, compilation, processing and dissemination are costly. A data producer may decide that users of the data must share the cost of the data. The amount of money charged may differ by the type of users e.g. student and academic, private and public sector organisations. In some instances, data are supplied free of charge to certain users such as government departments or government financed research institutes, international organisation such as the World Bank or UN agencies. Where data is supplied free of charge, the user may be asked to provide the media (e.g. disk or CD) or pay a nominal amount that covers the cost of the media and shipment. This pricing policy must be part of the data dissemination policy and should be made available to users.

Quality Indicator	5.11 Catalogues of publications and other services are available to users of statistics.
Standard	5.11.1 Catalogues of publications and other services must be accessible to users of statistics

A major component of ensuring accessibility is providing efficient search mechanisms to help users find what they need. The organisation should provide services and tools to help its users in this respect. The Internet site should offer an array of search and navigation tools and features that permit users to discover their information:

- data and product browsers and search, by theme and subject;
- catalogue search;
- key word search (supported by a terminology thesaurus);
- search of the organisational library;
- guides to data, to search tools and to methods;

Given the current rate of technology change, the nature of both catalogue and delivery systems is evolving fast. The traditional printed catalogue that was almost always out of date has given way to on-line catalogues of statistical products, whether printed or electronic, linked to metadata bases in which characteristics of the information can be found. Users are the main judge of accessibility systematic user feedback on catalogue and delivery systems. This feedback may be derived from (a) automated usage

statistics for the various components of these systems, (b) surveys of user satisfaction with particular products, services, or delivery systems, and (c) voluntary user feedback in the form of comments, suggestions, complaints, or plaudits.

Quality Indicator	5.12 Metadata are readily accessible to users.
Standard	5.12.1 Minimum metadata required for interpreting the product must be accessible.

To ensure the usefulness and usability of data files created, all data files or any statistical product should be accompanied by a readily accessible document containing metadata that clearly describe and explain the data. This is the minimum metadata required by users to know, locate, access or use data/statistics appropriately. The type of metadata needed will depend on the product to be disseminated. More details about the content of such metadata can be found in chapter 6.

Chapter 6: Interpretability

6.1 Description

Interpretability of statistical information is refers to the ease with which users understand statistical information through the provision of metadata.

6.2 Key components

- Concepts and definitions, and classifications that underlie the data;
- Metadata on the methodology used to collect and compile the data;
- Key findings, giving the summary of the results;
- Presentation of statistics in a meaningful way.

6.3 Guidelines

Providing sufficient information to allow users to properly interpret statistical information is the responsibility of any data producing agency. Managing interpretability is primarily concerned with the provision of metadata. Users have to know and understand the properties of the data. Metadata deals with the what, when, who and how of data. Metadata needs to be collected at different levels of detail to satisfy different purposes. Thus the metadata can be used to identify what data exists, to describe its content and geographic extent, to enable potential users to assess the suitability of the data for various purposes, and to indicate where more information about the data can be obtained.

It is important for the data producer to have a policy in place which informs users of the basic information they need to interpret data. This policy would prescribe what information should be provided with every release of data, and in what form it might be provided. It will also deal with the way in which metadata are stored, i.e. to ensure that the facilities for loading, maintaining and accessing metadata in the metadata store are available and fully used and that metadata is documented and reported in a standardised format; and that all metadata necessary to access and interpret the statistical outputs be held permanently in the metadata store. There is an element of user education in the provision of metadata. Spreading the message that all data should be used carefully, and providing the information needed to use data with care.

Definition of metadata

Metadata is defined as “data about data, and refers to the definitions, descriptions of procedures, system parameters, and operational results which characterise and summarises statistical programs. It can further be categorised into the following five categories;

- Definitional metadata: metadata that describes the concepts in the data, e.g. classifications, question and question module for collection instruments, statistical terms, etc
- Operational metadata: metadata arising from and summarising the results of implementing the procedures, e.g. response rates, edit failure rates, costs and other quality and performance indicators, etc.
- System metadata: active metadata used to drive automated operations, e.g. publication or dataset identifiers date of last update, file size, access methods to databases, mapping between logical names and physical names of files, etc
- Dataset metadata: metadata used to describe access and update dataset, data structures, e.g. textual description, data cell annotations, dataset title, keyword, distribution information, etc
- Procedural/Methodological metadata: metadata relate to the procedures by which data are collected and processed, e.g. sampling, collection methods, editing processes, etc.

Quality Indicator	6.1 Documented metadata (definitional, operational, methodological, system and dataset) are sufficient to understand data.
Standard	6.1.1 Metadata must be documented according to the standard; deviations from any standards must also be annotated in the metadata, together with reasons for deviation.

To make sure that information is interpretable, data producers are required to give descriptions of the underlying concepts, variables and classifications that have been used, the methods of collection, processing and estimation used in production of information and its own assessment of the quality of the information. Statistics released should be accompanied with complete metadata information, which is documented and complies with a metadata standard template. Completeness of metadata requires information on the extent to what metadata are available for the users and the extent to which it covers the topic. It refers to the completeness of the metadata published for a statistical domain/ theme etc.

In the case of public-use micro-data files, information regarding the record layout and the coding/classification system used to code the data on the file is an essential tool to allow users to understand and use the data files. Different types of metadata as indicated above should be documented according to standard(s), or sourced from the standard; for example:

- Concepts and definitions should be sourced from the relevant standard concepts and definition manual and made available to users;

- Classifications should be sourced from the standard classifications /coding system or document and made available to users;
- Variables should follow the standard variable naming convention;
- There should be a standard template for documenting a record layout which provides proper descriptions for the data set (flat file);
- A standard metadata capturing template should be provided that provides an overview of minimum metadata required to explain the data.

It is recommended that, where a deviation from the standard is required, such deviation should be documented, including reasons for deviating and approval thereof.

Quality Indicator	6.2 Statistics are presented in a clear and understandable manner.
Standard	6.2.1 The presentation of the statistical series must be commensurate with users' needs.

Statistics should be presented in a way that facilitates proper interpretation and meaningful comparisons (layout and clarity of text, tables, and charts). There is a need to have a standard for producing a report, statistical releases including a data tabulation standard. Data should be published in a clear manner; charts and tables are disseminated with the data to facilitate the analysis. It should offer adequate details and time series. Analysis of current period estimates should be available. Depending on the intended audience and purposes, data of different degree of aggregation, sub-components and additional data should be made available.

Quality Indicator	6.3 Statistical releases contain a summary of the key findings.
Standard	6.3.1 Statistical releases must contain a summary of key findings.

The statistical releases should contain a primary message that clarifies the interpretation of the data. Directed particularly at the media, such commentary increases the chance that at least the first level of interpretation to the public will be clear and correct.

Chapter 7: Comparability and Coherence

7.1 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.

7.2 Key components

- The use of common concepts and definitions within and between series.
- The use of common variables and classifications within and between statistical series.
- The use of common methodology and systems for data collection and processing within series.
- The use of common methodology for various processing steps of a survey such as editing and imputations within series.

7.3 Guidelines

The coherence of data products reflects the degree to which they are logically connected and mutually consistent. Managing comparability and coherence is enabled by ensuring that the concepts and definitions, classifications and methodology are consistently similar across programmes; and by using common frames and identifiers within and between statistical products. Maintaining consistency within a survey series promotes comparability in time of the series itself; that is, internal consistency. Two series which are both internally consistent are better suited for comparing with each other. Inconsistencies between two different sets of similar statistical information are reconciled to provide coherence. Statistics which are comparable are better suited for being brought together. Coherence has four important sub-dimensions: within a dataset, across datasets, over time and across countries.

Quality Indicator	7.1..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.
Standard	7.1.1 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 in the metadata and archived.
	7.1.2 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 and archived.
	7.1.3 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 and archived.

Coherence within a dataset implies that the elementary data items are based on common concepts and definitions, classifications and methodology; and can be meaningfully combined. This requires the development and use of standard frameworks, concepts, variables, classifications and methodology for all the subject-matter topics that are measured within the same series or administrative systems. This aims to ensure that the target of measurement is consistent within series; e.g. consistent concepts and definitions are used within series so that concepts such as “household” has the same meaning from year to year and quarter to quarter. Differences in concepts and definitions within the same series and administrative systems over time should be described and the reasons for and effects of these differences should be clearly explained and that the quantities being estimated bear known relationships to each other.

It is also recommended that standardised classification systems should be used for all categorical variables used in a survey or administrative system. Classifications not only help with the better management of data but also aid understanding and comparability of data. The realization of this element is normally through the adoption and use of frameworks such as the System of National Accounts and standard classification systems for all major variables. Some examples of such classifications follow.

- The Standard Industrial Classification (SIC) code which uses a standard numbering system identifying businesses and business units by type of economic activity. The system uses from a one-digit to an n-digit classification depending on how narrowly the business unit is defined.
- The Standard Occupational Classification (SOC) code which is used to classify workers into occupational categories. All workers are classified into over 820 occupations according to their occupational definition. To facilitate classification, occupations are combined into 23 major groups, 96 minor groups, and 449 broad occupations. Each broad occupation includes detailed occupation(s) requiring similar job duties, skills, education, or experience.
- The Standard International Trade Classification (SITC) code used to classify commodities by product types. Classifications can also be related to class sizes, gender, population group / race / ethnicity or geographical domains.
- The International Classification of Disease and related health problems (ICD-10) coding / classification used to code and classify causes of death data within the death register.

The use of the same classifications aims to ensure that the target of measurement is consistently classified in the same way within series or administrative systems. Differences in classification within the same series or administrative system should be described and the reasons for and effects of these differences should be clearly explained.

The data producing agency should develop and adopt common methodology for a series. These include the development and use of the common

- frames and source data,
- sampling techniques,
- frameworks such as the System of National Accounts used for the compilation of National Accounts Statistics,
- data collection and processing methodology,
- editing,
- seasonal adjustment methodology, etc.

This can be achieved by establishing centres of expertise in certain methodology and technologies, exchanging experiences, identifying good practices, developing standards, and training. Also the use of a data quality framework enhances consistency of methodology used. The use of the same methodology aims to ensure that the series consistently uses the same methodology all the time so that comparability over time for the series is feasible and any deviation from the trend is attributable to a large extent to statistical causes.

Changes in the legislative environment through which the administrative data are collected can cause a break in the administrative system which in turn can affect the comparability over time of the statistical product. Any changes of this kind should be described and the reasons for and effects of these changes should be clearly explained. In some cases, there might be a need for adjustment between the concepts and definitions used in the source data and the statistical product. A description of why the adjustment needed to be made and how the adjustment was made should be provided in the metadata to facilitate the assessment of the statistical product against other data. The issue of national or international comparability is addressed by considering the adherence of the standards adapted to national or international standards where these exist. Policies are required to define program responsibilities for ensuring that data are produced according to the standards adopted.

Quality Indicator	7.2 Statistics are consistent or reconcilable over time.
Standard	7.2.1 Statistics must be consistent over time.

This implies that the data is based on common concepts, definitions, and methodology over time, or that any differences are explained and accounted for. Incoherence over time refers to breaks in series resulting from changes in concepts, definitions, or methodology. The second element aims to ensure that the process of measurement does not introduce inconsistency between data sources even when the quantities being measured are defined in a consistent way. The development and use of common frames,

methodologies and systems for data collection and processing contribute to this aim. For example, the use of a common business register across all business surveys ensures that differences in frame coverage do not introduce inconsistencies in data the use of commonly formulated questions when the same variables are being collected in different surveys serves to minimize differences due to response error; the use of common methodology and systems for the various processing steps of a survey, especially edit and imputation, helps to ensure that these operations do not introduce spurious differences in data. All of these arguments apply across occasions of a particular survey, as well as across surveys.

Comparability over time of statistics is expected in the absence of error. Inconsistency over time occurs if these characteristics collected for a specific period are not comparable with the data for the following period. This is considered a break in the time series. Thus the length of the time series is the number of years or months that it remains unbroken. The longer the time series, the greater is the confidence that the true trend is described. An unbroken series is more likely to be comparable with similar statistics in another unbroken series. Any break needs to be analysed and understood before the next instance of the statistics. The data producing agency should reconcile the inconsistencies in key variables by adjusting the estimates. These adjustments should be clearly explained in the metadata.

Quality Indicator	7.2 Statistics are consistent or reconcilable over time.
Standard	7.2.2 The statistics must follow an expected trend established over time. Any inconsistencies in the key variables must be reconciled.

The statistics being produced should be compared against trends established by results of previous years or data derived from other systems. This is sometimes referred as output editing. This is where errors that could not be detected at earlier phases of the value chain can be detected. These statistics must be consistent with accepted trends or must be compatible with the theory stemming from the subject matter. For example, changes in values for the key variable from the previous period (or the same period in the previous year) appear impossible or are inconsistent with the changes in other related variables. Such inconsistencies must be avoided or must be resolved. Correction of the statistics may require going back to the source data or looking into the editing and analysis processes. If the inconsistencies cannot be explained, it is best practice to remove them.

Quality Indicator	7.3 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.
Standard	7.3.1 – 7.3.4 Data across comparable series or source data must be based on common frame, identifiers; or use the same concepts and definitions, classifications. Departures from common frame, identifiers, concepts and definitions and classifications must be identified in the metadata and archived.

The section focuses on the comparison and integration of data from different sources. Coherence across datasets implies that the data is based on common concepts, definitions and classifications, or that any differences are explained and can be account for. An example of incoherency across datasets would be if exports and imports in the national accounts could not be reconciled with exports and imports in the balance of payments. Data can be enriched by integrating data from multiple comparable sources, or data from comparable sources can be used to compare and improve the precision of estimates. These sources can include both survey data and data from administrative systems. Some integration activities are regular and routine, e.g. the integration of data in the national accounts, benchmarking or calibration of estimates to more reliable control totals, or seasonal adjustment of data to facilitate temporal comparisons. This indicator encourages and promotes the collection of comparable data based on common frames, survey instruments, rules and methodology; and advance the notion of an integrated approach to data collection. The use of common frames, identifiers, concepts and definitions, and classifications also allows for the integration of these various sources in a consistent fashion.
Departures from common practices, procedures, and rules (including the use of common concepts and definitions, classifications, frames and identifiers) across all comparable data sources should be communicated to stakeholders to allow for reconciliation of disparities due to said departures.

Quality Indicator	7.4 Statistics are checked for consistency with those obtained through other data sources.
Standard	7.4.1 Statistics must be checked for consistency with a comparable dataset. Inconsistencies must be reconciled.

It is good practice to verify statistical estimates using data from alternative sources. The alternative sources may in some instances serve as a benchmark. Examples of such exercises can be the reconciliation of survey data with administrative sources. The confrontations of data from different sources, and the subsequent reconciliation or explanation of differences, are necessary activities of the pre-release review of the data or certification process of the data. The differences in statistics should be quantified and the reasons should be described. Typically, some of the discrepancies may be attributed to differences in data collection, processes or differences in reporting units.

Quality Indicator	7.5 A common set of identifiers (for the purpose of record matching) exist and have been agreed upon by data producers.
Standard	7.5.1 A common identifier must be agreed upon by the data producers.
	7.5.2 The common identifier must be unique in every dataset. Rules and practices must be agreed upon to ensure uniqueness.

An identifier should be able to uniquely and unambiguously identify a statistical unit. A common identifier can be a variable or a combined set of variables.

This identifier can be a code with some meaning, i.e. a so-called smart number, for example:

- Identity number is composed of the date of birth and gender of the individual; and
- Primary sampling unit (PSU) that is a number that can identify a unit geographically.

Databases may have multiple identifiers. However there should be one unique identifier that is common to most data sources. The relevant stakeholders need to get together and decide on what the identifier will be, its future use, and what attributes would then be required for the identifier. Ownership of the identifier must be explicit and known to all stakeholders. This identifier should be used by various data producing agencies according to the agreed standards and related policy. If we can ensure that the data sources are using the same identifier, we know that comparability and coherence will be significantly enhanced.

This unique identifier is indispensable for record-matching between different databases to create statistical registers. The use of a unique identifier also allows for the confrontation of data from various sources that use the same identifier. Different administrative sources often have different population unit identifiers. The user can utilise this information to match records from two or more sources. Where there is a common identifier matching (integration) is generally more successful.

Chapter 8: Methodological soundness

8.1 Description

It refers to the application of international, national, or peer-agreed standards, guidelines, and practices to produce statistical outputs. Application of such standards fosters national and international comparability.

8.2 Key components

- International norms and standards on methods.
- Data compilation methods employ acceptable procedures.
- Other statistical procedures employ sound statistical techniques.
- Transparent revision policy and studies of revisions are done and made public.

8.3 Guidelines

Methodological soundness refers to the application of international standards, guidelines, and agreed practices to produce statistical outputs. It covers the idea that the methodological basis for the production of statistics should be sound, and that this can be attained by following these standards or accepted practices. Application of such standards fosters national and international comparability. It is necessarily dataset-specific, since datasets may be collected using a variety of different methodologies. This dimension covers soundness of the methodology used and adherence to international norms and standards on methods. This dimension has four elements, namely:

(i) concepts, definitions, and classification

(ii) scope of the study;

(iii) basis for recording

(iv) revision studies.

Quality Indicator	8.1 Concepts, definitions, and classifications used follow accepted standards, guidelines or good practices (national, international, peer-agreed).
Standard	8.1.1 The concepts and definitions must satisfy accepted standards, guidelines or good practice in line with national, international, peer-agreed norms; and must be documented. Deviations from the standard must be formally approved, and be fully documented.
	8.1.2 The classifications must satisfy accepted standards, guidelines or good practice in line with national, international, peer-agreed norms; and must be documented. Deviations from the standard must be formally approved, and be fully documented.

The conceptual basis for the statistics should follow international, national or peer-agreed norms such as standards, guidelines, and agreed practices. Concepts are the subjects of inquiry and analysis that are of interest to users. They refer to general characteristics or attributes of a statistical unit or of a population. Concepts are usually based on a theoretical or statistical frame of reference and are used to define a subject, the statistical units to be described and/or the population under study. Classifications of items allow better management of records. Classifications schemes are usually hierarchical and allow systematic grouping of records together. These are usually done by transcribing text descriptions of records into numbers or alphabetical ordering. Concepts, definitions and classifications should follow internationally acceptable practices. It is important to have concordances between organizational/national and international concepts, definitions and classifications.

The UN proposes some international classifications on: activity (ISIC); product (CPC, SITC, BEC); expenditure according to purpose (COFOG, COICOP, COPNI, COPP). Classifications that are used must reflect both the most detailed and the collapsed levels. For example, Stats SA has adopted some system of classifications of occupation, industry, products, geography etc.

Data producing agencies should source their concepts, definitions and classifications from international agencies, these practices will enhance comparability and integration of data by users. Not all concepts, classifications sourced elsewhere should be adopted as such. Some international definitions might need to be adapted to local needs, e.g. the $1 a day money metric definition of poverty has not been adopted in South Africa as a definition of poverty. It is recommended that agreed standards should be used for the compilation of such statistics. Since all categorical statistical data need to be classified for analysis, the classification criteria chosen to group data systematically need to be suitable for these analytical purposes. In case where there is a need to deviate from the standards, reasons for such deviations should be documented in the metadata.

Quality Indicator	8.2 The scope of the study is consistent with accepted standards, guidelines or good practices.
Standard	8.2.1 The scope of the study must be appropriate for the intended topic. The scope of the study must be consistent with accepted standards, guidelines or good practices in line with the survey constraints.

Statistics are sufficiently comprehensive in scope and in terms of conceptual development of concepts to adequately describe the subject area in question. Hence, the scope of the study should be consistent with peer- or nationally- or internationally accepted standards, guidelines or good practices. The study measures its intended topic, for example the Labour Force Survey (LFS) measures the level and pattern of unemployment. The scope of the study should be aligned with the project constraints. For example, a limited budget will result into a relatively small sample household LFS which may not give very accurate result for the level of employment nationally

Important considerations for the usefulness of statistics are that the methodological basis for the production of statistics is sound; the presented rules are clear and practical, and international comparability is possible. It is recommended that methods used should follow accepted standards, guidelines or good practices. Each phase of the SVC requires the development or adoption of different methodologies:

Quality Indicator	8.3 Methodologies used follow accepted standards, guidelines or good practices (national, international, peer-agreed). • Questionnaire design
Standard	8.3.1 The designing of the questionnaire must follow accepted standard, sets of guidelines or good practices.

Collection instruments (questionnaires) play a central role in the data collection process and have a major impact on respondent behaviour, interviewer performance, and respondent relations, all of which have a major impact on the quality of information collected. Questionnaire design should therefore follow international good practice and at the same time take into account user requirements, administrative requirements of the organization, processing requirements, as well as the nature and characteristics of the respondent population. All questionnaires and individual questions must undergo testing

Quality Indicator	8.3. Methodologies used follow accepted standards, guidelines or good practices (national, international, peer-agreed), • Sampling methods
Standard	8.3.2 The sampling methods must follow accepted standard, sets of guidelines or good practices.

The choice of sampling methodology has direct impact on data quality. This choice is influenced by factors like desired level of precision, availability of appropriate sampling frame, estimation methods used, and budget. All these factors should be taken into account when deciding on the sampling method; only scientific and statistically sound methods should be used.

Quality Indicator	8.3. Methodologies used follow accepted standards, guidelines or good practices (national, international, peer-agreed), • Frame maintenance
Standard	8.3.3 The frame maintenance methods must follow accepted standard, sets of guidelines or good practices.

Monitor the frame quality by periodically assessing its coverage. The statistical units register is then maintained through regular updates from source administrative register

Quality Indicator	8.3. Methodologies used follow accepted standards, guidelines or good practices (national, international, peer-agreed), • Piloting
Standard	8.3.4 The piloting methods must follow accepted standard, sets of guidelines or good practices.

For data collection instrument to measure what they were intended for, administering and testing such instrument should be done prior the actual collection period. The outcome can thus improve the quality of the instrument for future process of survey

Quality Indicator	8.3. Methodologies used follow accepted standards, guidelines or good practices (national, international, peer-agreed), • Data collection
Standard	8.3.5 Data collection methods must follow accepted standard, sets of guidelines or good practices.

Data collection methods should take into consideration the size of the survey, nature of respondent population, and the type of information needed, for example, business-based surveys may choose method of collection whereby responded complete the questionnaire themselves because they would have facilities and a required level of literacy, while household-based surveys may choose to have face to face interview whereby an enumerator asks the questions and complete the questionnaire for the respondent.

Quality Indicator	8.3. Methodologies used follow accepted standards, guidelines or good practices (national, international, peer-agreed), • Editing and imputation of data
Standard	8.3.6 Editing and imputation methods must follow accepted standard, sets of guidelines or good practices.

Data editing is done in order to take care of invalid/inconsistent entries so as to maintain credibility and to facilitate further automated data processing and analysis. Caution should be exercised against overuse of query edit as these may create bias through the addition of data which may or may not be correct

Quality Indicator	8.3. Methodologies used follow accepted standards, guidelines or good practices (national, international, peer-agreed), • Analysis of data
Standard	8.3.7 The methods of analysis used must follow accepted standards, sets of guidelines or good practices.

Procedures and guidelines followed in carry out data analysis must exist and well documented. This is a process of gathering, modeling, and transforming data with the goal of highlighting useful information. For example, to be consistent with international standard of reporting seasonal adjusted estimates must be reported.

Quality Indicator	8.3. Methodologies used follow accepted standards, guidelines or good practices (national, international, peer-agreed), • Revision of data
Standard	8.3.8 Revision methods used must follow accepted standards, sets of guidelines or good practices.

A clear and concise data revision policy for all published data must be produced for users to describe the main reasons why particular statistics are subject to revision. These policies will explain how frequently and to what extent revisions will be made. All revision should have taken place by a specified date in the survey with results clearly explained and analysed in the metadata. For example, In case where new and improved information is received, revision of estimates should be carried out. This will help on the improvement of the quality and reliability of the estimates

Data producing agency should identify a set of standards and guidelines or accepted practices either from international or national communities. These usually stem from research conducted on methods used in specific fields of study. Recommended accepted practice ensures consistent, cost-effective and repeatable methodologies. Thus for an agency to commit to using accepted practice in the area of statistics, is for them to commit to using all knowledge, technology and resources available in the field to ensure success.

The UN for example, has made available various documents on good practice in various topics of statistics e.g. civil registration and vital statistics; crime statistics, etc. (http://unstats.un.org/unsd/progwork/pwabout.asp). Hence the crucial part in the selection of good practice is the identification of the best ones from a pool of available ones. International organisations such as the OECD, WTO, ILO, IMF, Eurostat, UN etc. and countries with relatively advanced statistics systems can serve as models of such practices.

Peer agreed standards should have transparent and meaningful linkages to international standards. This should be developed by a community of expert in consultations with relevant stakeholders. The designing of the questionnaire, methods of sampling, sampling frame design, frame maintenance, piloting, data collection, editing and imputation of data, data analysis, and revisions of data methods should follow peer-, or nationally-, or internationally-accepted standard, sets of guidelines or good practices. Data producing agencies should describe the major methodological changes that have taken place during the reference period and how they affect the data quality.

Quality Indicator	8.4 Are revisions schedule followed? Are they regular and transparent? 8.5 Preliminary and revised data are identified in the metadata. 8.6 Studies of revisions and their findings are made public.
Standard	8.4.1 A revisions schedule must exist for surveys which conduct revisions. The revisions schedule must be publicly available and accessible. The revisions must take place as per the schedule.
	8.5.1 Preliminary and revised data must be identified in the metadata. Metadata must contain an explanation of the changes.
	8.6.1 Regular studies of revisions or upcoming revisions must be done and their findings must be made public

A revision schedule that represent the optimal revision cycle should exist for surveys which conduct revisions and this should be followed. A pattern of revisions should be regular in order not to introduce bias. It is important that all users of data are at all times made aware of the revision policy relating to these data. Thus measures should be taken to align patterns of availability of new data/information with the intended revisions patterns.

When publishing data that is likely to be subsequently revised (e.g. preliminary estimates) always indicate this to the users. Therefore, studies should assess preliminary estimates against final estimates and identify all changes. These studies of revision should be documented in detail. Adequate documentation on revisions should be maintained and includes descriptions of causes of revisions, methods used to incorporate new data sources, the way data are adjusted, the frequency of revisions and the magnitude of the revisions. Both preliminary and revised data are kept and made available to users and the metadata clearly identifies such revised data.

Findings from revision studies are usually used to refine preliminary data and data collection programs for the subsequent periods. These studies also investigate the source of errors and fluctuations in the data and are used to make appropriate adjustments to the data. Although some of these findings may routinely be analysed, they may mostly be used for internal quality control purposes. However, good practice suggests that these studies must also be published as they give insight in methods used and add to transparency in the production of statistics.

Chapter 9: Integrity

9.1 Description

The integrity of statistical information refers to values and related practices that maintain users' confidence in the agency producing statistics and ultimately in the statistical product. This includes, among others, the need for the statistical system to be based on the United Nations (UN) principles of official statistics and includes principles of objectivity in collection, compilation and dissemination of data to ensure unbiased statistics which are not subject to confidentiality breaches or premature releases.

9.2 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.

9.3 Guidelines

Data producing agencies should build a culture that focuses on quality, and emphasise objectivity and professionalism. The credibility of data producers refers to the confidence that users place in those products based simply on their image of the data producer, i.e. the brand image. One important aspect is trust in the objectivity of the data. This implies that the data is perceived to be produced professionally in accordance with appropriate statistical standards, and that policies and practices are transparent. For example, data is not manipulated, nor their release timed in response to political pressure.

Quality Indicator	9.1 The terms and conditions, including confidentiality, under which statistics are collected, processed and disseminated are available to the public and follow the UN principles of official statistics.
Standard	9.1.1 A terms and conditions document must be available and accessible to the public.

The United Nation's principles of official statistics provide the terms and conditions, including confidentiality, under which statistics collected, processed and disseminated are available to the public. For example Principle 2 of the UN Fundamental Principles of Official Statistics (1994) states: "to retain trust in official statistics, the statistical agencies need to 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". Data producing agencies are encouraged to follow the UN Fundamental Principles of Official Statistics. Alternatively,

they can develop their own terms and conditions under which statistics are collected, processed and disseminated building on the Fundamental Principles of Official Statistics. It is recommended that this document should cover at least the following items:

Purpose of Collection
This will include the name of an agency producing data, and the purpose of collecting such data.

Confidentiality
This item is already covered in detail in Chapter 1: Prerequisites of quality, and should include issues surrounding confidentiality and anonymization of data. For example, individual data collected by Stats SA for statistical compilation, whether they refer to natural or legal persons, will be strictly confidential and used exclusively for statistical purposes.

Dissemination
Data should be disseminated according to a data dissemination policy. If this policy exists, it should be in accordance with international best practice such as the GDDS, SDDS or national and peer agreed standards. There should be a policy or procedure to ensure confidentiality during data dissemination. These include data embargo and adopting access control systems. Data should be disseminated in an impartial manner:

- The timing of the data release should not be in response to political pressure.
- Released data should not be withdrawn in response to political pressure.
- The design of the survey should be described in the metadata and should accompany the published results.

Certain general criteria can bring a level of trust in the quality of the data that are being produced. These criteria can be defined as terms and conditions, could include the law regulating the production of the statistics and the impartiality of the statistics that are produced. The statistics should also be produced using adequate scientific standards which will permit agencies to defend them. In addition, the data that are produced should be readily available to all users as part of the citizen's entitlement to public information.

Quality Indicator	9.2 Describe the conditions under which policy-makers, specifically government, may have access to data before release. Are the conditions published?
Standard	9.2.1 A data dissemination policy detailing the conditions under which policy-makers have access to the data must be available.

Policy-makers, specifically government, may have access to embargoed data before release, but these conditions, list of users and their level of access to the data should be documented and be published as part of the agency's data dissemination policy. The producing agency should have a strictly controlled (lock-up) session where the public can read the results and formulate questions before release.

Quality Indicator	9.3 Advance notice is given of major changes in methodology and source data.
Standard	9.3.1 Advance notice of at least 6 months must be given of major changes in methodology and source data

Information about revisions and major changes in methodology should be communicated well in advance of the statistical release and before the data is made public. For example changes can be gazetted, be made in public announcements, on the internet (website) etc. Details of major changes and/or revisions to published data should be described in the explanatory notes of the relevant publication and metadata. Information about statistical standards, frameworks, concepts, definitions, sources and methods can also be released in a range of information papers and other publications to ensure that the public are informed about changes to statistical processes. It is also worth publishing discussion or technical documents well in advance where the nature of any changes is discussed.

Quality Indicator	9.4 Government commentary, when data are released, should be identified as such, and not be seen as part of the official statistics. 9.5 Choice of source data, techniques and dissemination decisions are informed solely by statistical considerations.
Standard	9.4.1 Advance notice of at least 6 months must be given of major changes in methodology and source data
	9.5.1 The choice of source data, techniques and dissemination decisions must be informed solely by statistical considerations.

Statistics should be compiled objectively, scientifically and impartially. The choice of data sources for statistical products as well as activities in the statistical value chain should be informed purely by strict professional considerations, including scientific principles and professional ethics. Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies should choose the source with regard to quality, timeliness, costs and the burden on respondents. The decision on the choice of techniques, methods, definitions and source data that are to be used must be left to the data producers. In doing this, data producers should neither favour any particular group in society nor be influenced by those in power to act against the principles of the prevailing professional ethics i.e. Code of Conduct. Politicians should not be able to suppress reports which reflect poorly on them.

The independence of the statistical authority from political and other external interference in producing and disseminating official statistics should be specified in law. The head of the data producing agency has the responsibility for ensuring that statistics are produced and disseminated in an independent manner. Statistical releases should be clearly distinguished and issued separately from political/policy statements. Data producing agencies, politicians, researchers, when appropriate, comments publicly on statistical issues, including criticisms and misuses of statistics. The data producing agency has a right to comment on the misuse of the statistics it produces, courtesy of the UN Fundamental Principles of Official Statistics. Such comments should be seen as such, not part of the statistics.

Quality Indicator	9.6 Ethical guidelines for staff behaviour are in place and are well known to the staff.
Standard	9.6.1 Professional code of conduct must be in place providing ethical guidelines for staff behaviour.

The data producer should have a code of conduct in place which guides the behaviour of staff with access to data. The code of conduct should address conflict of interest situations (e.g. payment of facilitation fees). There should be clear rules that make the connection between ethics and professional work (e.g., drawing a sample from Bishopscourt and Sandton and claiming that it represents South Africa; and over-reporting of learners by school principals for Education surveys). Management should acknowledge its status as role model and should be vigilant in following the code of conduct. New staff members should be made aware of the code of conduct when they join the organisation. Staff members should be reminded periodically of the code of conduct.

References

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 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

Statistics South Africa, (2008), South African Statistical Quality Assessment Framework (SASQAF), edition 1, Statistics South Africa, 2008

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

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: Statistical Value Chain (SVC)

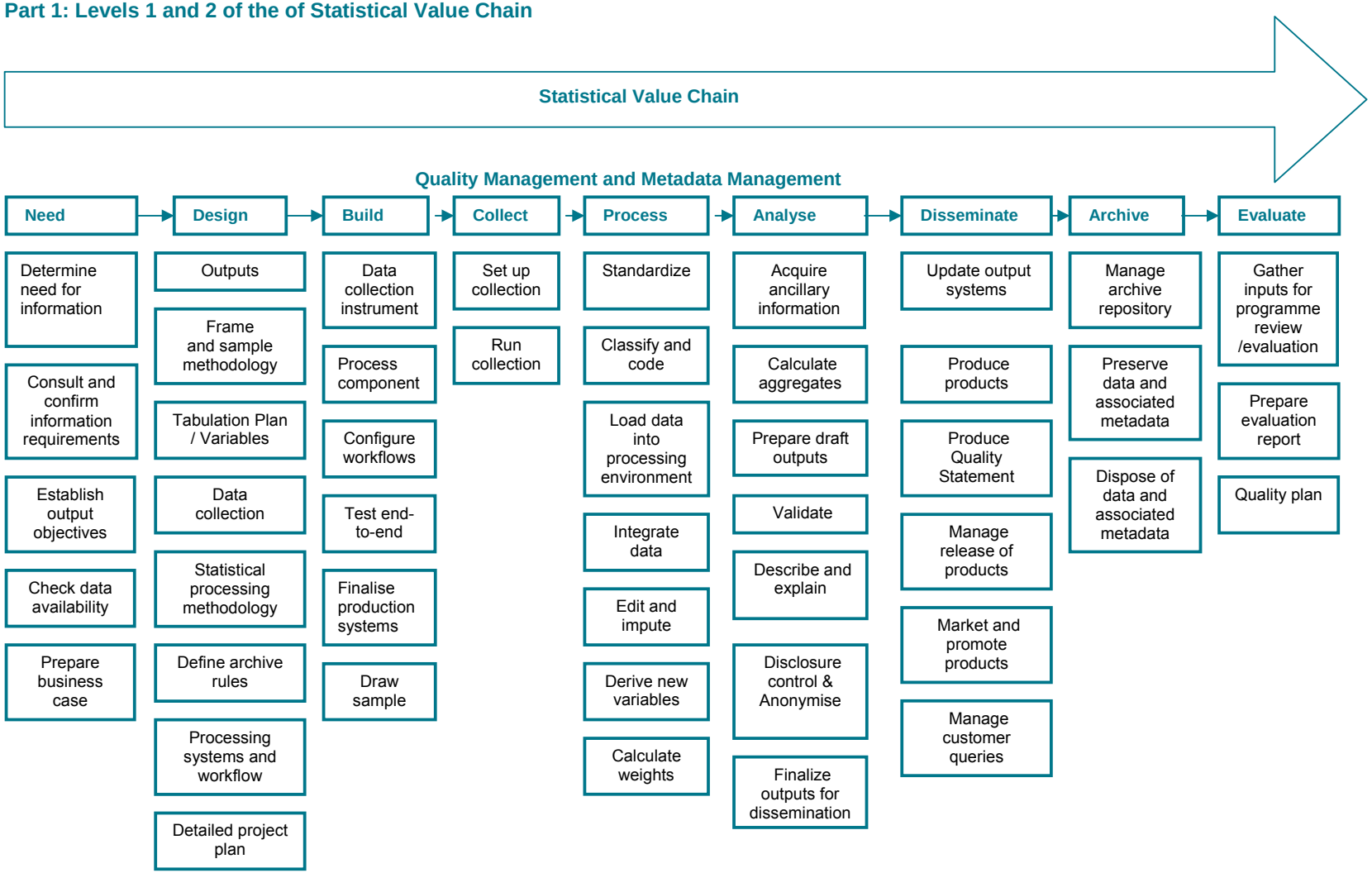
Statistics South Africa has adapted the Statistical Value Chain as developed by the Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS) in 2008. The statistical process involves a range of statistical operations, which are enabled by various support functions.

The Statistical Value Chain can be divided into four levels:

- Level 0: the statistical business process;
- Level 1: the nine phases of the statistical business process;
- Level 2: the sub-processes within each phase;
- Level 3: a description of those sub-processes.

A diagram showing the phases (Level 1) and sub-processes (Level 2) is included as Part 1. The sub-processes are described in detail in Part 2. The SVC also contains two over-arching processes that apply throughout the nine phases, and across statistical business processes, which are Quality Management and Metadata Management.

Part 1: Levels 1 and 2 of the of Statistical Value Chain



Part 2: Levels 2 and 3 of the Statistical Value Chain

This part of annexure considers each phase in turn, identifying the various sub-processes within that phase, and describing their contents. It therefore covers Levels 2 and 3 of the SVC.

1. Need

This is the first phase of the SVC, which involves all the necessary planning for a new survey. 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. The following are activities that need to be carried out during the Need phase:

1.1 Determine need for information

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

1.2 Consult and confirm need

- Consulting with the stakeholders and confirming in detail the need for the statistics. Statistical organisations should know what it 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

- Identifies the statistical outputs that are required to meet the user needs identified in sub-process 1.2 (Consult and confirm need).
- Agreeing the suitability of the proposed outputs and their quality measures with users.

1.4 Check data availability

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

1.5 Prepare business case

- 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 any 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.

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) of a previous iteration.

2.1 Outputs

- Contains the detailed design of the statistical outputs to be produced, including the related development work and preparation of the systems and tools used in the Disseminate phase.
- Outputs should be designed, wherever 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 Determine need for information.

2.2 Frame and sample methodology

- 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 in sub-process 3.6 (Draw sample), using the methodology specified in this sub-process.

2.3 Tabulation plan/Variables

- Defines the variables to be collected via the data collection instrument, as well as any other variables that will be derived from them in sub-process 5.6 (Derive new variables), and any classifications that will be used. It is expected that existing national and international standards will be followed wherever possible. This may need to run in parallel with sub-process 2.4 (Data collection), 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.

2.4 Data collection

- Determines the most appropriate data collection method(s) and instrument(s). The actual activities in this sub-process vary according to the type of collection instruments required, which can include computer assisted interviewing, paper questionnaires, administrative data interfaces and data integration techniques.
- Design of questions and response templates (in conjunction with the variables and classifications designed in sub-process 2.3 (Tabulation plan/Variables)).
- Design of any formal agreements relating to data supply, such as memoranda of understanding, and confirmation of the legal basis for the data collection.
- This sub-process is enabled by 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 standardise terms and conditions).
- Design of process-specific provider management systems.

2.5 Statistical processing methodology

- Designs the statistical processing methodology to be applied during Phase 5 (Process), and Phase 6 (Analyse).
- This can include developing and testing routines for coding, editing, imputing, estimating integrating, verifying and finalising data sets.

2.6 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 the 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 organisation, or, for national organisations, 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).

2.7 Processing systems and workflow

- Determines 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 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.

2.8 Detailed project plan

- Develop a project plan giving details on activities to be carried out, start date and duration of each activity; and human resources allocated to each activity.

3. Build

This phase builds and tests the production systems to the 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 every iteration. It 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.

3.1 Data collection instrument

- Describes 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 connection of collection instruments to the statistical metadata system, so that metadata can be more easily captured in the collection phase. Connection of metadata and data at the point of capture can save work in later phases.

3.2 Process components

- 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

- Configures the workflow, systems and transformations used within the statistical business processes, from data collection, right through to 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 end-to-end

- Describes the activities to manage a field test or pilot of the statistical business process. Typically it includes a small-scale data collection, to test collection instruments, followed by processing and analysis of the collected data, to ensure the statistical business process performs as expected.

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.

3.5 Finalise production systems

- Include activities to put the process, including workflow systems, modified and newly-built components into production ready for use 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 work as expected in that environment

3.6 Draw sample

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

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. For statistical outputs produced regularly, this phase occurs in each iteration.

4.1 Set up collection

- 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.).

4.2 Run collection

- This is where the collection is implemented, with the different collection instruments being used to collect the data. It includes 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.
- This includes 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 collection is produced.

4.3 Load data into processing environment

- Initial data validation, as well as loading the collected data and metadata into a suitable electronic environment for further processing in Phase 5 (Process). It may include 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 organisations. In 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).

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 5.7 (Calculate weights), which is usually specific to survey data).

5.1 Standardise and anonymise

- This is where statistical units are derived or standardised, and where data are anonymised.
- Depending on the type of source data, this sub-process may not always be needed.
- Standardisation includes converting administrative or collection units into the statistical units required for further processing.
- Anonymisation strips data of identifiers such as name and address, to help to protect confidentiality. Standardisation and anonymisation may take place before or after sub-process 5.2 (Integrate data), depending on the requirements for units and identifiers in that sub-process.

5.2 Integrate data

- Integrates one or more data sources. The input data can be from a mixture of external or internal data sources, and a variety of collection modes. The result is a harmonised data set. Data integration typically includes:
    - matching / record linkage routines, with the aim of linking data from different sources referring 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 subprocesses. There may also be several instances of data integration in any statistical business process.

5.3 Classify and code

- 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.

5.4 Edit and impute

- This applies to collected micro-data, and looks at each record to try to identify (and where necessary correct) missing data, errors and discrepancies. 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. Where data are missing or unreliable, estimates are imputed, often using a rule-based approach. Specific steps include:
    - the identification of potential errors and gaps;
    - the selection of data to include or exclude from editing and imputation routines;
    - editing and imputation using one or more pre-defined methods e.g. “hot-deck” or “cold-deck”
    - imputation;
    - writing the edited / imputed data back to the data set, and flagging them as edited or imputed;
    - the production of metadata on the editing and imputation process;
- Editing and imputation can apply to unit records both from surveys and administrative sources, before and after integration.

5.5 Derive new variables

- This sub-process creates variables that are not explicitly provided in the collection and are needed to deliver the required outputs. It derives these new variables by applying arithmetic formulae to one or more of the variables that are already present in the dataset. It 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.

5.6 Calculate weights

- This sub process creates weights for unit data records according to the methodology created in sub-process 2.5: 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.

6. Analyse

In this phase, statistics are produced, examined in detail, interpreted, 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.

6.1 Acquire ancillary information

- This sub-process includes many ongoing 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 allow informed analyses. Acquiring a high level of domain intelligence will allow a statistical analyst to understand the data better, and to identify where results might differ from expected values. This allows better explanations of these results in sub-process 6.5 (Describe and explain).

6.2 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.

6.3 Prepare draft outputs

- This sub-process is where domain intelligence is applied to the data collected to produce statistical outputs. It includes the production of additional measurements such as indices or seasonally adjusted series, as well as the recording of quality characteristics.

6.4 Validate

- This sub-process is where statisticians verify the quality of the outputs produced, in accordance with a general quality framework. Verification 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;
  - Verifying the statistics against expectations and domain intelligence.

6.5 Describe and explain

- This sub-process is where the in-depth understanding of the outputs is gained by statisticians. They use that understanding to interpret 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.

6.6 Disclosure control and anonymise

- This sub-process ensures that the data (and metadata) to be disseminated do not breach the appropriate rules on confidentiality. This may include checks for primary and secondary disclosure, as well as the application of data suppression or perturbation techniques.

6.7 Finalize outputs for dissemination

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

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.

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 and ready for dissemination.

7.2 Produce products

- This sub-process produces the products, as previously designed, 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 (text, tables, charts etc.);
    - assembling the components into products;
    - editing the products and checking that they meet publication standards.

7.3 Produce ‘quality statement’

This sub-process produces a quality report. This is a metadata required to declare the quality of a statistical product. This type of metadata should be documented according to the standard template for quality declaration.

7.4 Manage release of products

- This sub-process ensures that all elements for the release are in place including managing the timing of the release. It includes 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.

7.5 Market and promote products

- Whilst marketing in general can be considered to be an over-arching process, this sub-process concerns the active promotion and marketing of the statistical products produced in a specific statistical business process, to help them reach the widest possible audience. It includes 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, to facilitate the process of communicating statistical information to users.

7.6 Manage customer queries

- 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.

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 organisation 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.

8.1 Manage archive repository

- This sub-process concerns 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.

- This sub-process may cover a specific statistical business process or a group of processes, depending on the degree of standardisation within the organisation. Ultimately it may even be considered to be an over-arching process if organisation-wide standards are put in place.

8.2 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.

8.3 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.

9. Evaluate

This phase manages the evaluation of a specific instance of a statistical business process. 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)).

9.1 Gather inputs for programme review/ evaluation

- Evaluation material can be produced in any other phase or sub-process. It may take many forms, including feedback from users, process metadata, and 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 Prepare evaluation report

- This sub-process analyses the evaluation inputs and synthesises 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 Quality plan

- This sub-process brings together the necessary decision-making power to form and agree an action plan based on the evaluation report. It should also include 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

This process is present throughout the model. It is closely linked to Phase 9 (Evaluate), which has the specific role of evaluating individual instances of a statistical business process. The overarching quality management process, however, has both a deeper and broader scope. As well as evaluating iterations of a process, it is also necessary to evaluate separate phases and sub-processes, ideally each time they are applied, but at least according to an agreed schedule. These evaluations can apply within a specific process, or across several processes that use common components.

Metadata management

Good metadata management is essential for the efficient operation of statistical business processes. 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, and stored and transferred from phase to phase alongside the data they refer to. A metadata management strategy and system(s) are therefore vital to the operation of this model.

Annexure B: Mapping quality indicators to activities in the Statistical Value Chain (SVC)

Activities of the statistical value chain		Quality dimensions and indicators	
Phases	**Sub-processes**	**Quality dimension**	**Quality indicator**
Need	Determine need for information	Prerequisites of quality	1.1 The responsibility for producing statistics is clearly specified.
Need	Determine need for information	Prerequisites of quality	1.2 Standards and policies are in place to promote consistency of methods and results.
Need	Establish output objectives	Prerequisites of quality	1.3 Data sharing and coordination among data-producing agencies are clearly specified.
Need	Establish output objectives	Prerequisites of quality	1.4 Measures are in place to ensure that individual data are kept confidential, and used for statistical purposes only.
Need	Determine need for information	Relevance	2.1 Have both the internal and external users of the data been identified?
Need	Determine need for information	Relevance	2.2 Is there a process to identify user needs?
Need	Establish output objectives	Relevance	2.3 Are user needs and the usage of statistical information analysed?
Need	Establish output objectives	Relevance	2.4 Changes are made as a result of user needs assessments.
Need	Check data availability	Relevance	2.5 To what extent is the primary data appropriate for the statistical product produced?
Need	Prepare business case	Timeliness	4.4 Periodicity of release.
Need	Check data availability	Accessibility	5.1 Legal arrangements are in place to allow access to administrative records via manual, automated or electronic processes.
Need	Frame and sample methodology, Data collection	Methodological soundness	8.1 The scope of the study is consistent with accepted standards, guidelines or good practices.
Need	Check data availability	Integrity	9.5 Choice of source data, techniques and dissemination decisions are informed solely by statistical considerations.
Design	Detailed project plan	Prerequisites of quality	Resources are commensurate with the needs of statistical programmes. • Staff • Facilities

Activities of the statistical value chain		Quality dimensions and indicators	
Phases	**Sub-processes**	**Quality dimension**	**Quality indicator**
			• Computing resources • Financing
Design	Detailed project plan	Prerequisites of quality	1.7 Measures to ensure efficient use of resources in 1.6 are implemented.
Design	Frame and sample methodology	Accuracy	3.5 Register/frame maintenance procedures are adequate. • updates. • quality assurance ▪ data audit.
Design	Data collection	Accuracy	3.6 Are data collection systems sufficiently open and flexible to cater
Design	Data collection	Accuracy	3.7 Description of record-matching methods and techniques used on the administrative data sources. • match rate as a percentage of total records • measure of false negative matches (missed matches) • measure of false positive matches (mismatches)
Design	Outputs, Frame and sample methodology, Tabulation plan/variables, Data collection Statistical processing methodology, Define archive rules,	Comparability and coherence	7.3 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.
Design	Tabulation plan/variables; Identify concepts	Methodological soundness	8.1 Concepts, definitions, and classifications used follow accepted standards, guidelines or good practices (national, international, peer-agreed).
Design	Statistical processing methodology	Methodological soundness	8.2 Methodologies used follow accepted standards, guidelines or good practices (national, international, peer-agreed), viz.: • questionnaire design • sampling • sampling frame design • frame maintenance

Activities of the statistical value chain		Quality dimensions and indicators	
Phases	**Sub-processes**	**Quality dimension**	**Quality indicator**
			• piloting • data collection • editing and imputation of data • analysis of data • revision data
Design	Tabulation plan/variables Identify concepts	Methodological soundness	8.1 Concepts, definitions, and classifications used follow accepted standards, guidelines or good practices (national, international, peer-agreed)
Design	Statistical processing methodology	Methodological soundness	8.3 Methodologies used follow accepted standards, guidelines or good practices (national, international, peer-agreed), viz.: • questionnaire design • sampling • sampling frame design • frame maintenance • piloting
Process	Integrate data	Comparability and coherence	7.4 A common set of identifiers (for the purpose of record matching) exist and have been agreed upon by data producers.
Analyse	Validate	Accuracy	3.1 Measures of sampling errors for key variables are calculated. Amongst others these are: • standard error • coefficient of variation (CV) • confidence interval (CI) • mean square error (MSE) • design effect (DEFF)

Activities of the statistical value chain		Quality dimensions and indicators	
Phases	**Sub-processes**	**Quality dimension**	**Quality indicator**
Analyse	Validate	Accuracy	3.2 Measures of non-sampling errors are calculated, viz.: • Frame coverage errors • Systematic errors • Measurement errors • Processing errors • Model assumption errors • Non-response errors
Analyse	Define archive rules, Processing systems and workflow Validate	Comparability and coherence	7.2 Statistics are consistent or reconcilable over time
Analyse	Validate	Comparability and coherence	7.3 Statistics are checked for consistency with those obtained through other data sources.
Disseminate	Produce quality statement	Accuracy	3.3 Data from the primary source have been quality assessed..
Disseminate	Manage release of products	Timeliness	4.1 Average time between the end of reference period and the date of the preliminary results.
Disseminate	Manage release of products	Timeliness	4.2 Average time between the end of reference period and the date of the final results.
Disseminate	Manage release of products	Accessibility	5.1 Are statistical products (e.g. data, metadata) available to the public?
Disseminate	Manage release of products	Accessibility	5.2 Rules governing the restricted availability of administrative records are well described and documented.
Disseminate	Produce dissemination products	Accessibility	5.3 Types of media and/or channels used for sharing data amongst stakeholders are adequate and preserve confidentiality.
Disseminate	Produce dissemination products	Accessibility	5.4 Data is accessible in a format beyond the producing agency.
Disseminate	Manage release of products	Accessibility	5.5 Statistics are released on a pre-announced schedule.
Disseminate	Manage release of products	Accessibility	5.6 Statistical products are made available to all users at the same time.

Activities of the statistical value chain		Quality dimensions and indicators	
Phases	**Sub-processes**	**Quality dimension**	**Quality indicator**
Disseminate	Manage customer queries	Accessibility	5.7 Statistics/administrative records not routinely disseminated are made available upon request.
Disseminate	Market and promote products	Accessibility	5.8 User support services exist and are widely publicised.
Disseminate	Manage release of products	Accessibility	5.9 Does a data dissemination policy exist, and is it accessible?
Disseminate	Manage release of products	Accessibility	5.10 Does the pricing policy governing dissemination exist, and is it accessible?
Disseminate	Market and promote products	Accessibility	5.11 Catalogues of publications and other services are available to users of statistics.
Disseminate	Manage release of products	Accessibility	5.12 Metadata are readily accessible to users.
Disseminate	Produce dissemination products	Interpretability	6.1 Documented metadata (definitional, operational, methodological, system and dataset) are sufficient to understand data.
Disseminate	Produce dissemination products	Interpretability	6.2 Statistics are presented in a clear and understandable manner. Statistical releases contain a summary of the key findings.
Disseminate	Produce dissemination products	Interpretability	6.3 Statistical releases contain a summary of the key findings.
Disseminate	Processing systems and workflow	Methodological soundness	8.4 Are revisions schedule followed? Are they regular and transparent?
Disseminate	Manage release of products	Methodological soundness	8.5 Preliminary and revised data are identified in the metadata.
Disseminate	Manage release of products	Methodological soundness	8.6 Studies of revisions and their findings are made public.
Disseminate	Manage release of products	Integrity	9.1 The terms and conditions, including confidentiality, under which statistics are collected, processed and disseminated are available to the public and follow the UN principles of official statistics.
Disseminate	Produce quality statement	Integrity	9.2 Describe the conditions under which policy-makers, specifically government, may have access to data before release. Are the conditions published?
Disseminate	Produce quality statement	Integrity	9.3 Advance notice is given of major changes in methodology and source data.

Activities of the statistical value chain		Quality dimensions and indicators	
Phases	**Sub-processes**	**Quality dimension**	**Quality indicator**
	Check data availability		
Disseminate	Outputs, Frame and sample methodology, Tabulation plan, Data collection Statistical processing methodology, Define archive rules	Integrity	9.4 Government commentary, when data are released, should be identified as such, and not be seen as part of the official statistics.
Evaluation	Gather inputs for programme review/evaluation	Prerequisites of quality	1.5 Measures to oblige response are ensured through law.
Evaluation	Gather inputs for programme review/evaluation	Relevance	2.6 Is there a process to determine the satisfaction of users with the statistical information?
All phases	All phases	Timeliness	4.3 Production activities within the statistical value chain are within the planned timelines, viz.: • data collection • data processing • data analysis • dissemination
All phase	All phases	Integrity	9.5 Ethical guidelines for staff behaviour are in place and are well known to the staff.

Annexure C: Differences between SASQAF Edition 1 and SASQAF operational standards and guidelines

The following table outlines the adjustments and additions made to the quality indicators in the first edition of SASQAF:

Old Indicator number	Indicator (SASQAF Edition 1) – September 2008	New Indicator number	Indicator (SASQAF operational standards and guidelines)
0.3	Data sharing procedures and coordination among data-producing agencies are clearly specified and adhered to.	1.3	Data sharing and coordination among data-producing agencies are clearly specified
	New indicator	1.5	Measures to oblige response are ensured through law.
0.6	Measures to ensure efficient use of above resources in 0.5 are implemented.	1.7	Measures to ensure efficient use of resources in 0.6 are implemented.
1.4	Changes made as a result of user needs assessments.	2.4	Changes are made as a result of user needs assessments.
1.5	Is there a process to determine the satisfaction of users?	2.5	Is there a process to determine the satisfaction of users with the statistical information?
1.6	To what extents are the primary data (e.g. administrative data and other data) appropriate for the statistical product produced?	3.3	Moved to accuracy
1.7	Were special requests for estimates of statistical characteristics met?		Deleted
2.1	Measures of sampling errors for key variables are calculated. Amongst others these are: • Standard error • Coefficient of variation (CV) • Confidence interval (CI) • Mean square error (MSE).	3.1	Measures of sampling errors for key variables are calculated. Amongst others these are: • standard error • coefficient of variation (CV) • confidence interval (CI) • mean square error (MSE) • design effect (DEFF)

Old Indicator number	Indicator (SASQAF Edition 1) – September 2008	New Indicator number	Indicator (SASQAF operational standards and guidelines)
2.2	Measures of non-sampling errors are calculated, viz.: • frame coverage errors o under-coverage errors o over-coverage errors • duplication in the frame/register used to conduct a survey • the number of statistical units out of scope(i.e. number of ineligible units) • misclassification errors • systematic errors to determine the extent of bias introduced for both administrative records and surveys • measurement errors o questionnaire effects o data collection mode effects o interviewer effects o respondent effects • processing errors o data entry errors rates o coding errors o editing failure rates o imputation rates • model assumption errors • non-response errors o overall response rate o item response rates o unit non-response (e.g. weighted and un-weighted response rates).	3.2	Measures of non-sampling errors are calculated, viz.: • 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) • Misclassification errors • Systematic errors to determine the extent of bias introduced for both administrative records and surveys • Measurement errors (e.g. questionnaire effects, data collection effects, interviewer effects and respondent effects) • Processing errors (e.g. data entry errors rates, coding errors, editing failure rates, imputation rates) • Model assumption errors • Non-response errors (e.g. item non-response rates, unit non-response rates and overall response rates).
1.6	Was moved from Relevance	3.3	To what extent is the primary data appropriate for the statistical product produced?

Old Indicator number	Indicator (SASQAF Edition 1) – September 2008	New Indicator number	Indicator (SASQAF operational standards and guidelines)
2.3	Data from the primary source have been quality assessed • accuracy. • timeliness. • coherence.	3.4	Data from the primary source have been quality assessed.
2.4	Does an agreement for relevant deadlines for transfer of data from the primary source exist and are they adhered to?		Deleted
2.6	Are data collection systems sufficiently open and flexible to cater for new developments (e.g., changes in definitions, classifications, etc.)?	3.6	Are data collection systems sufficiently open and flexible to cater for new developments?
2.7	Description of record-matching methods and techniques used on the administrative data sources. • match rate as a percentage of total records • measure of false negative matches (same unit but match was missed) • measure of false positive matches (record matched but relate to separate entities)	3.7	Description of record-matching methods and techniques used on the administrative data sources. • match rate as a percentage of total records • measure of false negative matches (missed matches) • measure of false positive matches (mismatches)
3.1	Average time between the end of reference period and the date of the first results.	4.1	Average time between the end of reference period and the date of the preliminary results.
3.4	Report on the frequency of release.	4.4	Periodicity of release
3.5	Punctuality of time schedule for publication.		Deleted
4.1	Are data and information available to the public?	5.1	Are statistical products (e.g. data, metadata) available to the public?
4.3	Legal arrangements are in place to access administrative records via manual/automated/electronic systems.		Deleted

Old Indicator number	Indicator (SASQAF Edition 1) – September 2008	New Indicator number	Indicator (SASQAF operational standards and guidelines)
4.4	Types of media/channels used for sharing data amongst stakeholders are adequate and preserve confidentiality.	5.3	Types of media and/or channels used for sharing data amongst stakeholders are adequate and preserve confidentiality
4.7	Statistics are made available to all users at the same time.	5.6	Statistical products are made available to all users at the same time.
4.9	User support services are widely publicised.	5.8	User support services exist and are widely publicised.
4.10	Does a data dissemination policy exist, and is it maintained and accessible?	5.9	Does a data dissemination policy exist, and is it accessible?
4.11	Does the pricing policy governing dissemination exist, and is it available to users?	5.10	Does the pricing policy governing dissemination exist, and is it accessible?
4.12	Catalogue systems (for survey, administrative records and other services) to identify information are available to users and are updated regularly.	5.11	Catalogues of publications and other services are available to users of statistics.
4.13	Metadata (a full range of information on underlying concepts, definitions, classifications, methodology, data sources, accuracy, etc.) are documented, available and readily accessible to users	5.12	Metadata are readily accessible to users.
5.1	Availability of concepts and definitions, classifications underlying the data (survey and administrative records). Differences from accepted standards, guidelines or good practices are annotated.	6.1	Documented metadata (definitional, operational, methodological, system and dataset) are sufficient to understand data.
5.2	Documents on scope, basis of recording, data sources and statistical techniques (methodology) used are available. Differences from accepted standards, guidelines or good practices are annotated.		5.1 and 5.2 merged
	New indicator	6.2	Statistics are presented in a clear and understandable manner
5.3	All the statistical releases produced are accompanied by primary messages clarifying the key findings.	6.3	Statistical releases contain a summary of the key findings.

Old Indicator number	Indicator (SASQAF Edition 1) – September 2008	New Indicator number	Indicator (SASQAF operational standards and guidelines)
6.1	Data within series and administrative systems are based on common frameworks, such as concepts, definitions, classifications, and methodologies, and departures from these are identified in the metadata.	7.1	Data within series and administrative systems are based on common concepts and definitions, classifications, and methodology, and departures from these are identified in the metadata.
6.2	Statistics are consistent and reconcilable over time.	7.2	Statistics are consistent or reconcilable over time.
6.3	Data across comparable series, or source data are based on common frames, common identifiers, concepts, definitions, and classifications, and departures from these are identified in the metadata.	7.3	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.
6.4	Statistics are checked for consistency with those obtained through other data sources (identify comparable datasets and incomparable ones).	7.4	Statistics are checked for consistency with those obtained through other data sources.
7.3	Methodologies used follow accepted standards, guidelines or good practices (national, international, peer-agreed), viz.: • Questionnaire design • Sampling methods • Sample frame design • Frame maintenance • Piloting • Standard collection methods • Standard editing and imputation methods • Standard analytical methods.	8.3	Methodologies used follow accepted standards, guidelines or good practices (national, international, peer-agreed), viz.: • questionnaire design • sampling methods • sampling frame design • frame maintenance • piloting • standard collection methods • standard editing and imputation methods • standard analytical methods • revision methods
7.4	Revisions schedule followed (explain the extent to which it is regular and transparent).	8.4	Are revisions schedule followed? Are they regular and transparent?

Old Indicator number	Indicator (SASQAF Edition 1) – September 2008	New Indicator number	Indicator (SASQAF operational standards and guidelines)
8.3	Advance notice is given of major changes in methodology, source data and statistical techniques.	9.3	Advance notice is given of major changes in methodology and source data.
8.4	Ministerial commentary, when data are released, should be identified as such, and not be seen as part of the official statistics.	9.4	Government commentary, when data are released, should be identified as such, and not be seen as part of the official statistics
8.5	Choice of source data, techniques and dissemination decisions are informed solely by statistical considerations (without political interference).	9.5	Choice of source data, techniques and dissemination decisions are informed solely by statistical considerations.
8.6	Ethical guidelines for staff behaviour are in place and are well known to the staff (professional code of conduct).	8.6	Ethical guidelines for staff behaviour are in place and are well known to the staff