REPUBLIC OF SOUTH AFRICA

STATISTICIAN-GENERAL

NSS
National Statistics System

QUALITY APPROVED

Statistics South Africa
Preferred supplier of quality statistics

South African Statistical Quality Assessment Framework (SASQAF)

First edition

South African Statistical Quality Assessment Framework (SASQAF)

First edition

Statistics South Africa
2008

Pali Lehohla
Statistician-General

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 manual:

Data Management and Information Delivery (DMID):   (012) 310 8602
National Statistics Systems Division (NSSD):   (012) 310 8635
Email: quality@statssa.gov.za

Preface

Democracy involves contestation, debate and disagreement between contending interests. This is both necessary and healthy. Statistics are often mobilised in support of, or in opposition to, several of these contending interests. This, too, is part of the ordinary lifecycle of any democratic society, especially where statistics are used as a basis for evaluating and measuring the impact of policies, estimating progress in meeting national priorities such as economic growth and job creation, and assessing the success of initiatives aimed at reducing scourges such as crime and poverty.

Evaluation of statistics collected in different ways, over time, and for different purposes, is an essential element in assessing their reliability and quality. However, this sort of comparison is no easy matter, unless the evaluation is based on common and standard criteria, which are broadly agreed upon by both the users and producers of statistical information.

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. This is not only because of its responsibility for the quality and reliability of the official statistics it produces. It is also because the Statistics Act (Act No. 6 of 1999) mandates the Statistician-General (SG) to put a framework in place to enable evaluation of statistics collected by organs of state. Section 14(7) of the Act empowers the SG to '*designate as official statistics any statistics or class of statistics*' produced by Stats SA or any other organ of state.

In addition, the Statistician-General may, in certain circumstances and, on request by a producer of statistics which is not an organ of state, comment on the statistics it produces, and evaluate and rank those statistics. Ranking of such statistics produced by, for example, the private sector, non-governmental organisations, or research institutes, must follow the identical procedures used for the ranking and certification of statistics produced by organs of state.

To assist in this process of evaluation, ranking and certification, Stats SA has developed the South African Statistical Quality Assessment Framework (SASQAF). The first draft of this framework, which was issued in 2006, drew extensively from the International Monetary Fund's Data Quality Assessment Framework (DQAF). The new version of the document is a more developed framework, incorporating comments and suggestions from a range of users and producers of statistics.

Four requirements have to be met for statistics to be certified as official. Before the data can be considered for certification, it needs to be established whether the statistics collected go beyond the needs of the producer; whether the series involved is sustainable in terms of human and financial resources; and whether the producing agency applying for certification has membership of the National Statistics System (NSS). Once these three preconditions have been met, the data is evaluated against prerequisites, and the eight dimensions of quality set out in SASQAF. These cover relevance, accuracy, timeliness, accessibility, interpretability, coherence, methodological soundness, and integrity. The framework details the key criteria to be met in each of these dimensions, together with their related measures or indicators.

Transparent procedures and criteria are essential if producers and users of statistics are to willingly accept and embrace official evaluation of data for quality. Certification is based on these procedures. The development of SASQAF by Stats SA enhances and extends transparency in data evaluation. It also effectively calls on other producers of statistics to be transparent in informing users of the concepts, definitions, classifications, methodologies, and frames used in collecting, processing and analysing their data, as well as informing them on the accuracy of the data, and any other features that may affect the quality of the data or their "fitness for use".

Within Stats SA, SASQAF is already being used to evaluate the quality of our statistics according to the eight dimensions specified. Through putting SASQAF into operation, we have found that not every measure or indicator applies equally to every statistical series or product. Extending SASQAF as a tool to evaluate statistics collected by other organs of state, especially where they seek to have their statistics declared official, will assist in establishing which measures are most applicable to the state's various statistical collections.

PJ Lehohla
Statistician-General

Contents

1. Introduction

South Africa's first democratically elected government was voted into power with a strong mandate to transform society. Provision of services to the historically dispossessed, education, housing, poverty alleviation, job creation, economic development, and more-equitable distribution of wealth were all identified as high priorities.

However, the new government inherited a statistics void, at least as far as reliable information fit to be used in benchmarking and monitoring progress in service delivery was concerned. Equally, statistics gathered at the time had little value as a basis for informed decision-making, the development of policies, or the planning required for a massive programme of social transformation.

Institutionally, the new government needed to integrate and rationalise the production of national statistics by amalgamating the homeland statistical offices with the then Central Statistical Service (CSS), the statutory agency responsible for national statistics. Then the CSS itself had to be transformed both by reviewing and re-engineering the statistical series it produced, and by transforming its human capacity and statistical infrastructure.

The CSS was transformed into Statistics South Africa (Stats SA), a national government department deriving its new mandate and role from the Statistics Act (Act No.6 of 1999). Although Stats SA was the only institution tasked with producing official statistics, there were – and continue to be – many other producers of statistical information: market research companies, parastatal bodies, government departments, universities and research institutions, and the private sector. In this way, the decentralised and fragmented system of statistical production inherited in 1994 has endured into the present.

As South Africa's first Statistician-General (SG), Pali Lehohla, has argued in the preface to this document that the existence of competing statistics which are mobilised in the interests of various contending parties is part of an ordinary and healthy democratic process. However, this does not mean that all statistics are of equal value or quality. Statistics can and should be assessed in terms of their quality and fitness for required purpose, and this requires common standards and criteria as a basis for evaluation. This is one of the statutory mandates allocated to Stats SA under the Statistics Act.

Although there has been some progress in implementing this mandate, the current state of national statistics is still characterised by:

- an information gap in terms of relevant statistics to meet the needs of users;
- a quality gap in terms of common standards, including concepts, definitions, classifications, methodologies and sampling frames; and
- a capacity gap in terms of both human resources and infrastructure.

Stats SA has used its statutory mandate for statistical leadership to develop various strategies aimed at closing these gaps. One of the most important of these involves implementation of a National Statistics System (NSS) to align the use and production of statistics, particularly those collected within the various institutions and organisations of national, provincial and local government, and other organs of state. Statistics are also produced by other institutions and organisations, in the private sector, research bodies and non-governmental organisations. Where these statistics are in the public domain, and where they might have an influence on the development of government policy, or on the measurement and monitoring of government programmes; they too can be evaluated within the framework of the NSS using the South African Statistical Quality Framework (SASQAF).

Alignment and evaluation of statistics requires a rational, transparent and sustainable framework for assessing the quality of those statistics. SASQAF has been developed for this purpose: it provides the framework and criteria used for evaluating and certifying statistics produced by government departments and other organs of state and, in some circumstances, by non-governmental institutions and organisations.

Within the NSS framework, SASQAF draws a distinction between 'official' and 'national' statistics. National statistics refer to those statistics used in the public domain but which the SG has not certified as being official. Official statistics are those statistics that have been certified by the SG as being official in terms of Section 14(7)(a) of the Statistics Act.

Certification of statistics produced by organs of state involves a standard assessment procedure undertaken by a Data Quality Assessment Team (DQAT), established by the SG.[1]

For assessment of data for quality to begin, the submitting organ of state and the statistics under review need to comply with three initial criteria:

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

On meeting these initial criteria, assessment of the data begins. The DQAT is required to report on the statistics, classifying them as one of the following:

- quality statistics;
- acceptable statistics;
- questionable statistics; or
- poor statistics.

If the statistics being evaluated are not classified as *quality statistics*, the DQAT is required to make recommendations indicating areas for improvement which might lead to this status. Once the statistics are classified as being quality statistics, in line with the quality dimensions set out in SASQAF, the SG will formally designate the data as *official statistics*, which become subject to periodic reviews determined by the SG in consultation with the head of the producing agency or department.

2. Purpose of the framework

The main purpose of SASQAF is to provide a flexible structure for the assessment of statistical products. SASQAF can be used for:

- self-assessment by producers of statistics;
- reviews performed by a DQAT in the context of the NSS work;
- assessment by data users (e.g. financial market participants) based on the producing agency's quality declaration;
- assessment by international agencies (e.g. the International Monetary Fund) based on the quality declaration.

3. 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.

---

[1] See Annexure A

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.

4. Structure of the framework

SASQAF covers the various quality aspects of the entire statistical value chain (i.e. need, design, build, collection, processing, analysis and dissemination), and certifies national statistics on one of four levels. Level 4, certification (quality statistics), indicates optimal conditions for statistical production, while Level 1 (poor statistics) indicates the least favourable conditions.

In outline, the four levels of certification are as follows:

Level Four: Quality Statistics – These are statistics that meet all the quality requirements as set out in SASQAF. They are designated as **quality statistics** to the extent that deductions can be made from them, and are 'fit for use' for the purpose for which they were designed. Level 4 applies to highly-developed statistical activities with respect to the corresponding indicator.

Level Three: Acceptable Statistics – These are statistics that meet most, but not all, the quality requirements as stipulated in SASQAF. They are designated as **acceptable** to the extent that, despite their limitations, deductions can be made, and are 'fit for use' for the purpose for which they were designed. Level 3 refers to moderately well-developed activities with reference to a particular indicator.

Level Two: Questionable Statistics – These are statistics that meet few of the quality requirements as stipulated in SASQAF. They are designated as **questionable** to the extent that very limited deductions can be made, and they are therefore not 'fit for use' for the purpose for which they were designed. Level 2 refers to activities that are developing but still have many deficiencies.

Level One: Poor Statistics – These are statistics that meet almost none of the quality requirements as stipulated in SASQAF. They are designated as **poor statistics** to the extent that no deductions can be made from them, and are not 'fit for use' for the purpose for which they were designed. Level 1 refers to activities that are underdeveloped.

5. Quality dimensions

The tables which follow specify the quality dimensions of SASQAF in considerable detail, setting out the key components and their related indicators for each dimension. Additionally, they provide a benchmark for the four different levels against which statistics will be measured.

The prerequisites and eight dimensions of quality

- Prerequisites of quality

- Relevance

- Accuracy

- Timeliness

- Accessibility

- Interpretability

- Coherence

- Methodological soundness

- Integrity

Table 1: Prerequisites of quality

Quality dimension		Description			
0. Prerequisites of quality		Refers to the institutional and organisational conditions that have an impact on data quality.			
		Assessment Levels			
		Quality Statistics Level 4	**Acceptable Statistics** Level 3	**Questionable Statistics** Level 2	**Poor Statistics** Level 1
Key components	**Indicator**				
• Legal and institutional environment (including Memoranda of Understanding (MoUs) or Service Level Agreements (SLAs)) • Privacy and confidentiality • Resources are commensurate with the needs of statistical programmes • Quality is the cornerstone of statistical work	0.1 The responsibility for producing statistics is clearly specified.	The responsibility for producing statistics is explicitly specified through a legal framework.	The responsibility for producing statistics is specified through a legal framework.	The responsibility for producing statistics is implied through a legal framework.	The responsibility for producing statistics is not specified.
	0.2 Standards and policies are in place to promote consistency of methods and results.	All standards and policies are in place to promote consistency of methods and results, and are adhered to.	The majority of standards are in place to promote consistency of methods and results.	Some standards are in place to promote consistency of methods and results.	No standards are in place to promote consistency of methods and results.
	0.3 Data sharing procedures and coordination among data-producing agencies are clearly specified and adhered to.	Data sharing procedures and coordination among data-producing agencies are explicitly specified through a legal framework.	Data sharing procedures and coordination among data-producing agencies are specified through a legal framework.	Data sharing procedures and coordination among data-producing agencies are implied through a legal framework.	Data sharing procedures and coordination among data-producing agencies are not specified.
		A data-sharing policy exists and is regularly updated and adhered to.	A data-sharing policy exists and for the most part is adhered to. It may not be up to date.	A data-sharing policy exists, but is rarely adhered to. It may not be up to date.	No data-sharing policy exists.
	0.4 Measures are in place to ensure that individual data are kept confidential, and used for statistical purposes only.	Measures (e.g. policies, documented procedures) exist and are fully enforced so that individual data are always kept confidential.	Measures exist and are partially enforced so that individual data are always kept confidential.	Measures exist, but are not enforced to always keep individual data confidential.	There are no measures that ensure confidentiality.

Table 1: Prerequisites of quality (concluded)

Quality dimension		Description			
0. Prerequisites of quality		Refers to the institutional and organisational conditions that have an impact on data quality.			
		Assessment Levels			
		Quality Statistics Level 4	**Acceptable Statistics** Level 3	**Questionable Statistics** Level 2	**Poor Statistics** Level 1
Key components	**Indicator**	**Level 4**	**Level 3**	**Level 2**	**Level 1**
• Legal and institutional environment (including Memoranda of Understanding (MoUs) or Service Level Agreements (SLAs)) • Privacy and confidentiality • Resources are commensurate with the needs of statistical programmes • Quality is the cornerstone of statistical work	0.5 Resources are commensurate with the needs of statistical programmes • Staff • Facilities • Computing resources • Financing.	All resources are completely commensurate with statistical programmes.	Resources are partially commensurate with statistical programmes.	Resources are inadequately commensurate with statistical programmes.	Resources are not commensurate with statistical programmes.
	0.6 Measures to ensure efficient use of the above resources in 0.5 are implemented.	Measures (e.g. project plans and sign-off documentation) to ensure efficient use of resources are systematically implemented.	Measures to ensure efficient use of resources are often implemented.	Measures to ensure efficient use of resources are seldom implemented.	Measures to ensure efficient use of resources are not implemented.
	0.7 Processes are in place to focus on, monitor and check quality.	Processes are consistently in place to focus on, monitor and check quality.	Processes are to some extent in place to focus on, monitor and check quality.	Processes are seldom in place to focus on, monitor and check quality.	Processes are not in place to focus on, monitor and check quality.

Table 2: Relevance

Quality dimension		Description			
1. Relevance		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.			
		Assessment Levels			
		Quality Statistics	**Acceptable Statistics**	**Questionable Statistics**	**Poor Statistics**
Key components	**Indicator**	**Level 4**	**Level 3**	**Level 2**	**Level 1**
• Why do you need to conduct a 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?	1.1 Have both the internal and external users of the data been identified?	All users of the data have been identified with their most recent contact details.	All users of the data have been identified with some of the contact details not up to date.	Attempts have been made to create a user list.	No attempt has been made to create a user list.
	1.2 Is there a process to identify user needs?	User needs are identified as a matter of course.	User needs are usually, but not always, identified.	User needs are identified on an ad hoc basis.	No attempt is made to identify user needs.
	1.3 Are user needs and the usage of statistical information analysed?	User needs and the usage of statistical information are always analysed.	User needs and the usage of statistical information are often analysed.	User needs and the usage of statistical information are seldom analysed.	There is no effective interaction with users.
	1.4 Changes made as a result of user needs assessments.	The results of the assessment are always built into the corporate processes and influence decisions on the design of the survey/series.	The results of the assessment are often built into the corporate processes and influence decisions on the design of the survey/series.	The results of the assessment are seldom built into the corporate processes and influence decisions on the design of the survey/series.	No action taken to incorporate the results of assessments.
	1.5 Is there a process to determine the satisfaction of users?	User satisfaction is measured and to a large extent has made an impact on the output.	User satisfaction is measured and to some extent has made an impact on the output.	User satisfaction is measured but has made no impact on the output.	User satisfaction is not measured.
	1.6 To what extent are the primary data (e.g. administrative data and other data) appropriate for the statistical product produced?	The primary data are fully aligned to the statistical product released.	The primary data are mostly aligned to the statistical product released.	The primary data have limited relevance to the statistical product released.	The primary data are not at all relevant to the statistical product released.
	1.7 Were special requests for estimates of statistical characteristics met?	All special requests were met.	Some special requests were met.	The majority of special requests were not considered.	No special requests were met.

Table 3: Accuracy

Quality dimension		Description			
2. Accuracy		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.			
		Assessment Levels			
Key components	Indicator	Quality Statistics Level 4	Acceptable Statistics Level 3	Questionable Statistics Level 2	Poor Statistics Level 1
• 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 • Data capture 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	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).	Sampling errors are calculated for the main variables and are available for the other variables on request, and fall within acceptable standards.	Sampling errors are calculated and made available for the main variables, and fall within acceptable standards.	Sampling errors are calculated but not made available, and fall outside the acceptable standards.	No sampling errors are calculated.

	2.2 Measures of non-sampling errors are calculated, viz.:
	• Frame coverage errors
	  - under-coverage errors
	  - 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.
		Non-sampling errors are extensively described and analysed, and the measures fall within acceptable standards.	Non-sampling errors are described and analysed, and the measures are not far off from acceptable standards.	Non-sampling errors are described and analysed, and the measures are far off from acceptable standards.	Non-sampling errors are not described.

Table 3: Accuracy (continued)

Quality dimension		Description			
2. Accuracy		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.			

		Assessment Levels			
		Quality Statistics	**Acceptable Statistics**	**Questionable Statistics**	**Poor Statistics**
Key components	**Indicator**	**Level 4**	**Level 3**	**Level 2**	**Level 1**
• 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 • Data capture 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	• Measurement errors - questionnaire effects - data collection mode effects - interviewer effects - respondent effects • Processing errors - data entry errors rates - coding errors - editing failure rates - imputation rates • Model assumption errors • Non-response errors -overall response rate - item response rates - unit non-response (e.g. weighted and unweighted response rates) 2.3 Data from the primary source have been quality assessed • Coverage • Timeliness • Coherence	Quality declaration is attached and shows that data comply with acceptable standards.	Quality declaration is attached and shows that the deficiencies in the data do not invalidate use of the data.	Quality declaration is attached and shows that data deviate significantly from acceptable standards.	Quality declaration is not attached.
	2.4 Does an agreement for relevant deadlines for transfer of data from the primary source exist and are they adhered to?	Measures (agreements, documented procedures) exist to ensure that agreed deadlines are adhered to.	Measures exist to ensure adherence to agreed deadlines but there are minor discrepancies regarding adherence.	Deadlines for reporting exist with no follow-up procedures to ensure the timely receipt of data.	No deadlines for reporting and no procedures to ensure timely receipt of data exist.

Table 3: Accuracy (continued)

Quality dimension		Description			
2. Accuracy		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.			
		Assessment Levels			
		Quality Statistics	**Acceptable Statistics**	**Questionable Statistics**	**Poor Statistics**
Key components	**Indicator**	**Level 4**	**Level 3**	**Level 2**	**Level 1**
• 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 • Data capture 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	2.5 Register / frame maintenance procedures are adequate.	Maintenance and update procedures of register/frame are adequate, thoroughly documented and performed on a regular basis.	Maintenance and update procedures are adequate and performed on a regular basis, but are not thoroughly documented.	Maintenance and update procedures are inadequate and are performed on an ad hoc basis. Some documentation exists.	No maintenance and update procedures exist.
	• Updates	Updates are typically live and are registered on the occurrence of the event.	Updates are typically after the event, but occur at regular intervals.	Updates are typically after the event, but occur on an ad hoc basis.	No maintenance and update procedures exist.
	• Quality assurance	A regular follow-up survey is conducted based on a sample drawn from the administrative records and matches the frequency of the release.	A follow-up survey is conducted but is inadequate given the frequency of the release.	The follow-up survey is conducted on an ad hoc basis	No follow-up survey is conducted.
	• Data audit	An analysis of alternate data source/s is conducted to determine the • cause, • extent and • type of errors in the administrative record system / frame and matches the frequency of the release.	An analysis of alternate data sources is done on a regular basis but is inadequate given the frequency of the release.	An analysis of alternate data sources is done on an ad hoc basis	No analysis of alternate data sources is done.

Table 3: Accuracy (concluded)

Quality dimension		Description			
2. Accuracy		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.			
		Assessment Levels			
		Quality Statistics	**Acceptable Statistics**	**Questionable Statistics**	**Poor Statistics**
Key components	**Indicator**	**Level 4**	**Level 3**	**Level 2**	**Level 1**

• 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
 	
• Data capture 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
	2.6 Are data collection systems sufficiently open and flexible to cater for new developments (e.g., changes in definitions, classifications, etc.)?
	
	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)
		Data collection programmes are sufficiently robust, with changes causing minimal impact on systems.
		
		Manual and electronic techniques used for matching records are thoroughly documented.
			Data collection programmes are sufficiently robust, with changes causing significant impact on systems.
			
			Although incomplete, a high degree of documentation exists on manual and electronic record-matching techniques used.
				Although the data collection programmes are weak, the changes would result in significant system changes (not a major overhaul).
				
				Some documentation exists on manual and electronic record-matching techniques used.
					Data collection programmes are weak, with changes requiring an overhaul of the entire system.
					
					Manual and electronic techniques used for matching records are not documented at all.

Table 4: Timeliness

Quality dimension		Description			
3. Timeliness		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 as well as across the activities within the statistical value chain. Frequency of data collection and release.			
		Assessment Levels			
Key components	**Indicator**	**Quality Statistics** Level 4	**Acceptable Statistics** Level 3	**Questionable Statistics** Level 2	**Poor Statistics** Level 1
• Production time (for the entire survey) • Frequency of release • Punctuality of release	3.1 Average time between the end of reference period and the date of the first results.	Preliminary results are released within the recommended timeframes as specified in the relevant standards and good practices.	Preliminary results released approach the relevant standards and good practices.	Preliminary results released lag behind relevant standards and good practices.	Preliminary results released lag far behind the relevant standards and good practices.
	3.2 Average time between the end of reference period and the date of the final results.	Final results are released within the recommended timeframes as specified in the relevant standards and good practices.	Final results released approach the relevant standards and good practices.	Final results released lag behind relevant standards and good practices.	Final results released lag far behind the relevant standards and good practices.
	3.3 Production activities within the statistical value chain are within the planned timelines, viz.: • Data collection • Data processing • Data analysis • Dissemination.	All elements within the statistical value chain are within the planned timelines.	Some elements within the statistical value chain are within the planned timelines.	Few elements within the statistical value chain are within the planned timelines.	All elements within the statistical value chain are not within the planned timelines.
	3.4 Report on the frequency of release.	The standards and guidelines for the frequency of release exist and are adhered to.	The standards and guidelines for the frequency of release exist, but only some are adhered to.	The standards and guidelines for the frequency of release exist, but are not adhered to.	No standards and guidelines exist for the frequency of release.
	3.5 Punctuality of time schedule for publication.	Statistical outputs released are always within the relevant standards and good practices, e.g. see GDDS and SDDS as a standard.	Statistical outputs released are most of the time within the relevant standards and good practices.	Statistical outputs lag behind the relevant standards and good practices.	Statistical outputs lag far behind the relevant standards and good practices.

Table 5: Accessibility

Quality dimension		Description			
4. Accessibility		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.			
		Assessment Levels			
Key components	**Indicator**	**Quality Statistics** Level 4	**Acceptable Statistics** Level 3	**Questionable Statistics** Level 2	**Poor Statistics** Level 1
• 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 • Presentation of statistics in a meaningful way • Means of sharing data between stakeholders	4.1 Are data and information available to the public?	All statistics disseminated are available from a publicly accessible medium.	Most of the statistics disseminated are available from a publicly accessible medium.	Few statistics disseminated are available from a publicly accessible medium.	Statistics disseminated are not available from a publicly accessible medium.
	4.2 Rules governing the restricted availability of administrative records are well described and documented.	All rules governing the restricted availability of administrative are well described and documented.	Some of the rules are defined and documented	Some of the rules are defined and documented	No rules are defined or documented
	4.3 Legal arrangements are in place to access administrative records via manual/automated/ electronic systems.	Only those with whom legal arrangements are in place are able to access administrative data via manual/automated/ electronic systems	Administrative records are made accessible to those with whom legal arrangements are not officially in place, but are pending via manual/automated/ electronic systems.	Administrative records are made accessible to those without any legal arrangements in place, but discussions have been entered into.	Administrative records are made accessible without any legal arrangements in place.
	4.4 Types of media/channels used for sharing data amongst stakeholders are adequate and preserve confidentiality.	Data are accessible through a variety of channels with mechanisms that ensure confidentiality.	Data are accessible through a variety of channels though loopholes exist that may compromise confidentiality.	Limited channels exist for stakeholders to access data and no mechanisms exist to ensure confidentiality.	No channels exist for stakeholders to access data.
	4.5 Data is accessible in a format beyond the producing agency.	Data is accessible in a variety of formats that satisfies the requirements of all users.	Data is accessible in a variety of formats that satisfies the requirements most users.	Data is accessible in a variety of formats that satisfies the requirements of some users.	Data is accessible in a format that only meets the needs of the producing agency.

Table 5: Accessibility (continued)

Quality dimension	Description
4. Accessibility	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.

		Assessment Levels			
Key components	**Indicator**	**Quality Statistics** **Level 4**	**Acceptable Statistics** **Level 3**	**Questionable Statistics** **Level 2**	**Poor Statistics** **Level 1**
• Catalogue systems are available in organ or statistical agency • Delivery systems to access information • Information and metadata coverage is adequate • Measure of catalogue and delivery systems performance • Presentation of statistics in a meaningful way • Means of sharing data between stakeholders	4.6 Statistics are released on a pre-announced schedule.	Statistics are always released according to an advance release calendar.	Statistics are most of the time released according to an advance release calendar.	Statistics are sometimes released according to an advance release calendar.	There are no advance release calendars.
	4.7 Statistics are made available to all users at the same time.	Statistics are always made available to all users at the same time.	Statistics are often made available to all users at the same time.	Statistics are seldom available to all users at the same time.	Statistics are never released simultaneously to all interested parties.
	4.8 Statistics/administrative records not routinely disseminated are made available upon request.	Statistics not routinely disseminated are always available on request; or Administrative records not routinely shared are always available on request (where a legal framework is in place).	Statistics not routinely disseminated are usually available on request; or Administrative records not routinely shared are usually available on request (where a legal framework is in place).	Statistics not routinely disseminated are occasionally available on request; or Administrative records not routinely shared are occasionally available on request (where a legal framework is in place).	Statistics/ administrative records not routinely disseminated are not available on request.
	4.9 User support services are widely publicised.	User support services are well known and widely utilized.	User support services are well known and utilized by some users.	User support services are known but they are not used.	User support services do not exist.
	4.10 Does a data dissemination policy exist, and is it maintained and accessible?	A data dissemination policy exists, and is available and up to date.	A data dissemination policy exists but is outdated.	A data dissemination policy is under development.	No data dissemination policy exists.

Table 5: Accessibility (concluded)

Quality dimension		Description			
4. Accessibility		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.			
		Assessment Levels			
Key components	**Indicator**	**Quality Statistics** Level 4	**Acceptable Statistics** Level 3	**Questionable Statistics** Level 2	**Poor Statistics** Level 1
• Catalogue systems are available in organ or statistical agency • Delivery systems to access information • Information and metadata coverage is adequate • Measure of catalogue and delivery systems performance • Presentation of statistics in a meaningful way • Means of sharing data between stakeholders	4.11 Does the pricing policy governing dissemination exist, and is it available to users?	Pricing policy exists, and is available and up to date.	Pricing policy exists but is outdated.	Pricing policy is under development.	Pricing policy does not exist.
	4.12 Catalogue systems (for survey, administrative records and other services) to identify information are available to users and are updated regularly.	Catalogue systems to identify information are available and updated regularly.	Catalogue systems to identify information are partially available and updated regularly.	Catalogue systems are not readily available and are not updated regularly.	Information is not catalogued.
	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	Metadata are always documented, available, and readily accessible.	Metadata are available and accessible to some users	Metadata are available but not readily accessible.	Metadata is not documented

Table 6: Interpretability

Quality dimension		Description			
5. Interpretability		The interpretability of statistical information refers to the ease with which users understand statistical information through the provision of metadata.			
		Assessment levels			
Key components	**Indicator**	**Quality Statistics** **Level 4**	**Acceptable Statistics** **Level 3**	**Questionable Statistics** **Level 2**	**Poor Statistics** **Level 1**
• Concepts, definitions and classifications underlying the data • Metadata on the methodology used to collect and compile the data	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.	Concepts, definitions and classifications underlying the data are available, and any deviations from acceptable standards are annotated. All concepts used in administrative records are well defined and documented.	Some deviations from acceptable standards are annotated. The vast majority of the concepts, definitions and classifications used in administrative records are well defined and documented.	Few deviations from acceptable standards are annotated. Some of the concepts, definitions and classifications used in administrative records are well defined and documented.	Deviations from acceptable standards are not annotated. None of the concepts, definitions and classifications used in administrative records are defined or documented.
	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.	Adequate documentation on scope, basis of recording, data sources, and statistical techniques used is available and deviations from accepted standards, guidelines or good practices are annotated. The accepted standard is the metadata template.	Partial documentation on scope, basis of recording, data sources, and statistical techniques used is available and deviations from accepted standards, guidelines or good practices are annotated.	Inadequate documentation on scope, basis of recording, data sources, and statistical techniques used is available, and deviations from accepted standards, guidelines or good practices are annotated.	Scope, basis of recording, data sources, and statistical techniques used are not documented.
	5.3 All the statistical releases produced are accompanied by primary messages clarifying the key findings.	Primary messages clarifying all key findings on each statistical release are available in detail.	Primary messages clarifying some key findings on each statistical release are available in detail.	Primary messages clarifying a few key findings on each statistical release are available but not in detail.	No primary messages clarifying key findings on each statistical release.

Table 7: Coherence

Quality dimension		Description			
6. Coherence		The coherence of statistical information reflects the degree to which it can be successfully brought together with other statistical information within a broad analytic framework and over time.			
		Assessment levels			
Key components	**Indicator**	**Quality Statistics** **Level 4**	**Acceptable Statistics** **Level 3**	**Questionable Statistics** **Level 2**	**Poor Statistics** **Level 1**
• The use of common concepts within and between series • Common definitions within and between series • Common variables and classifications within and between statistical series • The use of common methodologies and systems for data collection and processing within series • Use of common methodology for various processing steps of a survey such as edits and imputations within series	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.	All data within series are based on common frameworks, concepts, definitions, classifications, and methodologies and departures from this are identified in the metadata.	Most of the data within series are based on common frameworks, concepts, definitions, classifications, and methodologies and departures from this are identified in the metadata.	Limited data within series are based on common frameworks, concepts, definitions, classifications, and methodologies and departures from this are identified in the metadata.	Data within series are not based on common frameworks, concepts, definitions, classifications, and methodologies.
	6.2 Statistics are consistent and reconcilable over time.	Statistics are always consistent and reconcilable over time.	Statistics are sometimes consistent and reconcilable over time.	Statistics are seldom consistent and reconcilable over time.	Statistics are neither consistent nor 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.	All data across comparable series, or primary source data are based on common frames, common identifiers, concepts, definitions, and classifications, and any differences are identified and can be allowed for in the interpretation.	Most data across comparable series, or primary source data are based on common frames, common identifiers, concepts, definitions, and classifications, and any differences are identified and can be allowed for in the interpretation.	Limited data across comparable series, or primary source data are based on common frames, common identifiers, concepts, definitions, and classifications, and any differences are identified and can be allowed for in the interpretation.	No data across comparable series or primary source data are based on common frames, common identifiers, concepts, definitions, and classifications.

Table 7: Coherence (concluded)

Quality dimension		Description			
6. Coherence		The coherence of statistical information reflects the degree to which it can be successfully brought together with other statistical information within a broad analytic framework and over time.			

		Assessment levels			
Key components	**Indicator**	**Quality Statistics** Level 4	**Acceptable Statistics** Level 3	**Questionable Statistics** Level 2	**Poor Statistics** Level 1
• The use of common concepts within and between series • Common definitions within and between series • Common variables and classifications within and between statistical series • The use of common methodologies and systems for data collection and processing within series • Use of common methodology for various processing steps of a survey such as edits and imputations within series	6.4 Statistics are checked for consistency with those obtained through other data sources (identify comparable datasets and incomparable ones). 6.5 A common set of identifiers (for the purpose of record matching) exist and have been agreed upon by the data producers.	Statistics are always checked for consistency with those obtained through other data sources. A common set of identifiers (for the purpose of record matching) exist and have been agreed upon by the data producers.	Statistics are sometimes checked for consistency with those obtained through other data sources. Some identifiers exist, facilitating record matching, but have not been agreed upon.	Statistics are rarely checked for consistency with those obtained through other data sources. Some identifiers exist, but are insufficient for accurate record matching	Statistics are not checked for consistency with those obtained through other data sources. No common identifiers exist

Table 8: Methodological soundness

Quality dimension	Description			
7. 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.			

		Assessment levels			
Key components	**Indicator**	**Quality Statistics** **Level 4**	**Acceptable Statistics** **Level 3**	**Questionable Statistics** **Level 2**	**Poor Statistics** **Level 1**
• International norms and standards on methods • Data compilation methods employ acceptable procedures • Other statistical procedures employ sound statistical techniques • Revision policy, transparent, and those studies of revisions are done and made public	7.1 Concepts, definitions, and classifications used follow accepted standards, guidelines or good practices (national, international, peer-agreed).	All concepts, definitions, and classifications follow accepted standards, guidelines or good practices (national, international, peer-agreed).	Most concepts, definitions, and classifications follow accepted standards, guidelines or good practices (national, international, peer-agreed).	Few concepts, definitions, and classifications follow accepted standards, guidelines or good practices (national, international, peer-agreed).	Concepts, definitions, and classifications do not follow any standards, guidelines or good practices (national, international, peer-agreed).
	7.2 The scope of the study is consistent with accepted standards, guidelines or good practices.	The scope of the study is completely consistent with accepted standards, guidelines or good practices.	The scope of the study is partially consistent with accepted standards, guidelines or good practices.	The scope of the study is inadequately consistent with accepted standards, guidelines or good practices.	The scope of the study is inconsistent with accepted standards, guidelines or good practices.
	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.	Methodologies used in all processes always follow accepted standards, guidelines or good practices.	Methodologies used in all processes sometimes follow accepted standards, guidelines or good practices.	Methodologies used in all processes seldom follow accepted standards, guidelines or good practices.	Non-standard methods used.

Table 8: Methodological soundness (concluded)

Quality dimension		Description			
7. 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.			
		Assessment levels			
		Quality Statistics	**Acceptable Statistics**	**Questionable Statistics**	**Poor Statistics**
Key components	**Indicator**	**Level 4**	**Level 3**	**Level 2**	**Level 1**
• International norms and standards on methods • Data compilation methods employ acceptable procedures • Other statistical procedures employ sound statistical techniques • Revision policy, transparent, and those studies of revisions are done and made public	7.4 Revisions schedule followed (explain the extent to which it is regular and transparent).	Revisions schedule is always followed.	Revisions schedule is sometimes followed.	Revisions schedule is seldom followed.	No revisions schedule.
	7.5 Preliminary and revised data are identified in the metadata.	Preliminary and revised data are always identified and explained in metadata.	Preliminary and revised data are sometimes identified and explained in metadata.	Preliminary and revised data are seldom identified and explained in metadata.	Preliminary and revised data are not identified and explained in metadata.
	7.6 Studies of revisions and their findings are made public.	Studies of revisions and findings are always made public.	Studies of revisions and findings are sometimes made public.	Studies of revisions and findings are seldom made public.	Studies of revisions and findings are never made public.

Table 9: Integrity

Quality dimension		Description			
8. Integrity		Integrity refers to values and related practices that maintain users' confidence in the agency producing statistics and ultimately in the statistical product.			
		Assessment levels			
		Quality Statistics	**Acceptable Statistics**	**Questionable Statistics**	**Poor Statistics**
Key components	**Indicator**	**Level 4**	**Level 3**	**Level 2**	**Level 1**
• Professionalism and ethical standards in guiding policies and practices, which should be reinforced by their transparency standards • Assurances that statistics are produced on an impartial basis • Ethical standards are guided by policies and procedures	8.1 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.	The terms and conditions, including confidentiality, under which statistics are collected, processed and disseminated, are available to the public and completely follow the UN principles of official statistics.	The terms and conditions, including confidentiality, under which statistics are collected, processed and disseminated, are available to the public and by and large follow the UN principles of official statistics.	The terms and conditions, including confidentiality, under which statistics are collected, processed and disseminated are available to the public and to some extent follow the UN principles of official statistics.	The terms and conditions, including confidentiality, under which statistics are collected, processed and disseminated are not available to the public and/or do not follow the UN principles of official statistics.
	8.2 Describe the conditions under which policy-makers, specifically government, may have access to data before release. Are the conditions published?	Policy-makers always get the statistics at the same time as everyone else and this is publicly stated.	Policy-makers in exceptional cases get the statistics before everyone else and this is publicly stated.	Policy-makers often get the statistics before everyone else and this is not publicly stated.	Policy-makers routinely get the statistics before everyone else and this is not publicly stated.
	8.3 Advance notice is given of major changes in methodology, source data and statistical techniques.	Advance notice of major changes in methodology, source data and statistical techniques is always given.	Advance notice of major changes in methodology, source data and statistical techniques is sometimes given.	Advance notice of major changes in methodology, source data and statistical techniques is seldom given.	Advance notice of major changes in methodology, source data and statistical techniques is never given.
	8.4 Ministerial commentary, when data are released, should be identified as such, and not be seen as part of the official statistics.	Ministerial commentary, when data are released, is always identified as such, and is not seen as part of the official statistics.	Ministerial commentary, when data are released, is sometimes confused to some extent with the official statistics.	Ministerial commentary, when data are released, is often confused with the official statistics.	There is no clear distinction between Ministerial commentary, when data are released, and official statistics.

Table 9: Integrity (concluded)

Quality dimension		Description			
8. Integrity		Integrity refers to values and related practices that maintain users' confidence in the agency producing statistics and ultimately in the statistical product.			
		Assessment levels			
		Quality Statistics **Level 4**	**Acceptable Statistics** **Level 3**	**Questionable Statistics** **Level 2**	**Poor Statistics** **Level 1**
Key components	**Indicator**				
• Professionalism and ethical standards in guiding policies and practices, which should be reinforced by their transparency standards • Assurances that statistics are produced on an impartial basis • Ethical standards are guided by policies and procedures	8.5 Choice of source data, techniques and dissemination decisions are informed solely by statistical considerations (without political interference). 8.6 Ethical guidelines for staff behaviour are in place and are well known to the staff (professional code of conduct).	Source data, techniques and dissemination decisions are informed solely by statistical considerations without any political interference. Ethical guidelines for staff behaviour are in place, are well known to the staff and are adhered to.	Source data, techniques and dissemination decisions are informed solely by statistical considerations as well as limited political interference. Ethical guidelines for staff behaviour are in place, are known to the staff and are adhered to.	Source data, techniques and dissemination decisions are informed by statistical considerations with political interference. Ethical guidelines for staff behaviour are in place, are not well known to the staff and to some extent are adhered to.	Source data, techniques and dissemination decisions are informed solely by political interference. Staff do not know ethical guidelines for staff behaviour.

6. 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

Statistics South Africa, Internal document, (2002), Framework for Certification of National Statistics in South Africa, Statistics South Africa, 2002

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: Protocol specifying the procedure for the Statistician-General to designate statistics as official

Introduction

Purpose of this protocol

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

Mandate

Section 14(7)(a) of the Statistics Act (No. 6 of 1999) empowers to the Statistician-General to 'designate as official statistics any statistics or class of statistics produced from statistical collections by Statistics South Africa; or other organs of state, after consultation with the head of the organ of state concerned'.

In certain circumstances, statistics which are not produced by an organ of state can be evaluated and certified in the same way as statistics produced by an organ of state.

Definitions

Organ of state:

'*(a)*   any department of state or administration in the national, provincial or local sphere of government; or

(b) any other functionary or institution:

   (i)   exercising a power or performing a duty in terms of the Constitution or a provincial constitution; or

   (ii) exercising a public power or performing a public duty in terms of any legislation, but does not include a court or a judicial officer'.

Official statistics:

Statistics produced by an organ of state which the Statistician-General has designated as official in terms of Section 14(7)(a) of the Statistics Act (No. 6 of 1999).

National statistics:

Statistics produced by an organ of state that are within the public domain, and have not been designated as official statistics.

Member of NSS:

"An organ of state or other organisation that produces, supplies or uses statistics, and has signed a memorandum of understanding with the Statistician-General committing to adhere to common statistical quality criteria, standards, and procedures as set down by the Statistician-General in terms of Sections 7(2) (e) and 14 (6) of the Statistics Act."

Scope of the protocol

The protocol covers the role of the Statistician-General, the function of the National Statistics System (NSS), the general principles that guide evaluation of statistics, and the procedure for evaluating, and designating statistics as official by the Statistician-General.

Statistical principles

The Statistics Act empowers the Statistician-General to coordinate statistics and develop standards for both Statistics South Africa and other organs of state that produce statistics. Section 3(2) of the Act specifies that official statistics must be

'(a) relevant, accurate, reliable and timeous;

(b) objective and comprehensive;

(c) compiled, reported and documented in a scientific and transparent manner;

(d) disseminated impartially;

(e) accessible;

(f) in accordance with appropriate national and international standards and classifications; and

(g) sensitive to distribution by gender, disability, region and similar socio economic features.'

For statistics to be certified as official, they need to be aligned with the above principles. The principles that underpin the certification process, and provide the framework for the conduct of the parties involved in the certification process, are presented below.[2]

Principle 1: The aim of the National Statistics System (NSS) is to provide a framework for effective and comprehensive coordination of statistical output, quality, and standards.

Principle 2: It is a long-term goal of the NSS that only official statistics will be used to inform government policies, programmes and projects, including the Government-wide Monitoring and Evaluation (GWM&E) system.

Principle 3: It is a long-term goal of the NSS that all national statistical collections that meet the relevant criteria should be accorded official status.

Principle 4: National statistics qualify as official statistics when they are relevant beyond the organ or agency that collected them; when their production is sustainable; when they meet quality criteria and standards specified by the Statistician-General; and when they are accessible as a public good.

Principle 5: Assessment of statistical quality is guided by the South African Statistical Quality Assessment Framework (SASQAF).

Principle 6: Statistics declared as official will be reviewed at regular intervals, to be determined jointly by the Statistician-General and the head of the relevant organ or agency, in order to ensure that they remain relevant and of specified quality.

Principle 7: Official and national statistical series and other statistical products in the public domain may be evaluated for effectiveness, efficiency and comparative benchmarking at periodic intervals, determined jointly by the Statistician-General and the head of the producing organ or agency. Evaluations may incorporate peer reviews.

Principle 8: Collection, processing and analysis of data should be governed exclusively by scientific principles in accordance with international or peer-agreed best practice, within the parameters of available resources.

Principle 9: Statistical processes, procedures and methodology should be fully documented to enable users to assess fitness for purpose.

Principle 10: Custody of data designated as official statistics will normally be with Statistics South Africa, unless the Statistician-General and the head of the producing organ of state or agency agree otherwise. Ownership of data remains with the originating organ of state or agency.

---

[2] Some of the principles also appear in the South African Statistical Quality Assessment Framework (SASQAF). However, SASQAF provides an operational framework and more detailed criteria for assessment, while the principles are wider in scope and provide an 'environmental' framework.

Procedure for designating statistics from other organs of state as official

1. The Statistician-General will publish and regularly update SASQAF as a framework within producing agencies who may apply to have data designated as official statistics. The Statistician-General, in consultation with the head of the producing organ of state or agency, determines the elements or outputs of the producing organ of state or agency to be designated as official statistics. These could include a survey, a register, a dataset, indicators, a data table, etc.

2. An organ of state or agency will apply, through the division responsible for the NSS at Statistics South Africa, to the Statistician-General to have their statistics designated as *official statistics*.

3. Applications will be referred to a Data Quality Assessment Team (DQAT) constituted by the Statistician-General. DQAT will be constituted by members from Data Management and Information Delivery (DMID) project, NSS Division, Methodology and Standards, subject specialists and the applicant.

4. For assessment to begin, the submitting organ of state and the statistics under review need to comply with three initial criteria:

   • The producing agency must be a member of the NSS.

   • The statistics are used to meet user needs beyond those specific and internal to the producing agency.

   • The statistics produced should be part of a sustainable series, not a once-off collection.

5. DQAT will assess the quality of the product(s) in terms of SASQAF requirements, assigning a SASQAF quality level to the product. The assessment process is as follows:

   a. the applicant will identify all the SASQAF indicators that are relevant for the product under evaluation, and motivate why the remaining indicators are not relevant;

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

   b. once DQAT and the applicant reach agreement on which indicators are relevant, and the standard related to each indicator; they will sign an agreement to this effect;

   c. the applicant will then be asked to produce a quality declaration for their product, for all the agreed indicators; and

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

6. DQAT will recommend the overall SASQAF level of the product.

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

8. If the product satisfies the requirements of *quality statistics* set out in SASQAF, the SG will designate the product as *official statistics*.

9. Once the product has been designated as *official statistics*, it will be published with the Statistician-General's official seal of approval (the Official Statistics Mark), and stored in the NSS archive for public access.

10. The product then becomes subject to periodic reviews, determined by the SG in consultation with the head of the producing agency or department.

Annexure B: Fundamental principles of official statistics relative to Statistics South Africa

In its endeavour to fulfill the purpose of providing users with quality information, Statistics South Africa (Stats SA) has adopted the following principles developed by the Economics and Social Council Statistical Commission of the United Nations:

Principle 1: Relevance, impartiality and equal access
Official statistics provide an indispensable element in the information system of a democratic society, serving the Government, the economy and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis to honour citizens' entitlement to public information.

Principle 2: Professional standards and ethics
To retain trust in official statistics, Stats SA will decide, according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and presentation of statistical data.

Principle 3: Accountability and transparency
To facilitate the correct interpretation of data, Stats SA will present information according to scientific standards on the sources, methods and procedures of statistics.

Principle 4: Prevention of misuse
Stats SA is entitled to comment on erroneous interpretation and misuse of statistics.

Principle 5: Cost-effectiveness
Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Stats SA will choose the source with regard to quality, timeliness, costs and the burden on respondents.

Principle 6: Confidentiality
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.

Principle 7: Legislation
The laws, regulations and measures under which the statistical systems operate will be made public.

Principle 8: National coordination
Stats SA will promote coordination among statistical producers within South Africa in order to advance consistency and efficiency in the statistical system.

Principle 9: International standards
Stats SA will use international concepts, classifications and methods, where possible, to promote the consistency and efficiency of statistical systems between countries.

Principle 10: International cooperation
Bilateral and multilateral cooperation in statistics contributes to the improvement of systems of official statistics in all countries.

Annexure C: Certification process for the production of official statistics