Government Gazette
Staatskoerant

REPUBLIC OF SOUTH AFRICA
REPUBLIEK VAN SUID-AFRIKA

Vol. 566 Pretoria, 2 August Augustus 2012 No. 35560

CONTENTS • INHOUD

GENERAL NOTICE

NOTICE 610 OF 2012

GENERAL NOTICE

DEPARTMENT OF HIGHER EDUCATION AND TRAINING

I, Bonginkosi Emmanuel Nzimande, MP, Minister of Higher Education and Training, in terms of section 31 of the National Qualifications Framework Act, 2008 (Act 67 of 2008); read with section 45 of the Further Education and Training Colleges Act, 2006 (Act 16 of 2006); section 35 of the Adult Education and Training Act, 2000 (Act 52 of 2000); and sections 6 and 4 of the Higher Education Act, 1997 (Act 101 of 1997); hereby publish the standard DHET002: Data Quality Standard for Surveys as set out in the Schedule.

Dr BE Nzimande, MP
Minister of Higher Education and Training

Date: 23/05/12

SCHEDULE

SOUTH AFRICAN EDUCATION INFORMATION STANDARDS

DHET 002

Data Quality Standard for Surveys

higher education & training

Department:
Higher Education and Training
REPUBLIC OF SOUTH AFRICA

Education Management Information Support (MIS)
Department of Higher Education and Training
Private Bag X174
0001 PRETORIA

Point of contact:
Ms Hilda Herbst
Telephone: 012 312 5414
E-mail: herbst.h@dhet.gov.za

Table of contents

List of Tables

List of Figures

1. Purpose of the Data Quality Standard for Surveys

1.1 Outline of contents

Although it may seem easy to make a general statement about the perceived quality of a dataset, to provide a comprehensive assessment of the quality of a dataset can be quite challenging. This is because such a comprehensive assessment requires one to take into account more than just the accuracy of measurements contained in the data. Data quality is contextual: to provide a true assessment of the quality of a specific dataset, this dataset must be assessed in terms of the data collection protocol to determine whether it does in fact measure that which it was intended to measure. Therefore, although the title of this document refers to data quality in datasets, it addresses the wider context of data collection accuracy and the end result of such data collection, namely the statistical product produced as a result of the data collection and from the collected dataset.

Statistics South Africa (Stats SA) developed an instrument, which may be used to assess the quality of statistical products obtained via data collection, namely the South African Statistical Quality Assessment Framework (SASQAF). Not only was SASQAF developed to ensure comparability between different data sets collected in South Africa in terms of quality, it was also guided by international standards, such as the Data Quality Assessment Framework (DQAF) of the International Monetary Fund (IMF) and the Fundamental Principles of Official Statistics, provided by the United Nations' Statistics Division.

Using SASQAF would present the following advantages to any department or agency:

- SASQAF provides a comprehensive framework for evaluating quality. It focuses on issues pertaining to how accurately the data represents that which it was intended to measure, but also considers aspects such as relevance and accessibility.
- SASQAF is already a fully developed standard, and it is more practical to use an existing set of standards than to develop an entirely new set.
- Since SASQAF is available to all agencies in South Africa via Stats SA, using SASQAF could facilitate coherence between various national datasets.
- The indicators provided by SASQAF for assessment of quality are also intended as specifications for the approval of statistics as official statistics. Official statistics are those statistics that were certified by the Statistician-General as being official in terms of section 14(7) (a) of the Statistics Act [No 6 of 1999]. If the intention is to obtain classification as official statistics for a statistic produced by a government department or agency outside Stats SA, a necessary step would be to align the statistic with the requirements set in SASQAF.

The assessment in SASQAF places a specific statistic in one of four categories, namely:

- Quality statistics;
- Acceptable statistics;
- Questionable statistics; or
- Poor statistics.

Although SASQAF may be used for various purposes, it could be specifically applied to assess the quality of any statistical product, and can therefore be used to assess the quality of data and statistics produced by the Department of Higher Education and Training (DHET). While it may be argued that DHET would not necessarily want their statistics approved as official statistics, **it would certainly be an advantage to DHET if the Department could ensure that the quality of all the DHET data is such that any of its statistics could be classified as official statistics.**

SASQAF is therefore recommended as the standard against which data quality should be assessed in the DHET. This document is based on the use of SASQAF for this purpose. However, the document goes further than just reiterating SASQAF, since it also provides requirements and supporting guidelines, based directly or indirectly on SASQAF requirements or supporting guidelines, within a DHET context.

1.2 Scope and Applicability

While a standard on data quality would be of benefit to all information acquisition ventures, this particular Standard deals strictly with data collecting, processing and publishing by the Department of Higher Education and Training and all its entities.

It is envisaged that adherence to the data quality standard proposed, would result in a declaration of official statistics as provided by the National Statistics System (NSS).

1.3 Conventions followed in the standard

Any reference to the contents of SASQAF was taken from the official SASQAF documentation and no attempt has been made to change terminology or numbering systems to be customised for the DHET. This is done so that references in this document to SASQAF could be updated directly if any changes are made to SASQAF in the future, without the need for re-customising such changes for the DHET.

1.4 How to use the standard

It is recommended that the SASQAF data quality indicators be used as a standard to assess DHET products that have been compiled via data collection.

However, it should be noted that the intention of SASQAF is for it to be a generic standard that could be applied within any data context. The indicators, therefore, use terminology that is not contextualised for DHET processes and data concepts. Furthermore, SASQAF was intended as an assessment tool for a completed statistical product. It was therefore not designed to provide guidelines for establishing or ensuring quality throughout the process of planning, collecting, analysing and publishing data. The SASQAF indicators are not organised in terms of the chronological process underlying the data collection, and it is therefore difficult to use them as guidelines throughout the entire process.

Therefore, in order to implement the SASQAF standards within the context of DHET, there is a need for **ensuring** quality, as well as a way to **assess the** quality of a completed product and process. This document provides ways to ensure that quality processes would result in quality products – in fact, to ensure that a SASQAF assessment would result in a positive measure for data quality.

This is done in two parts. The first part provides a list of **requirements** that represent the absolute minimum that has to be done to ensure quality (in subsection 2.2 of this document). **Every data collection survey must conform to these requirements in order to ensure the good quality**. The second part provides more detailed **guidelines** that may be used to assure quality throughout the complete data collection process (in subsection 2.3 of this document). Although it may be possible that some of these guidelines may not be relevant for a particular study, *it is proposed that the guidelines be followed as closely as possible and a motivation be provided to indicate why a specific guideline is considered to be irrelevant.*

The requirements and guidelines are presented so as to provide the DHET context for the SASQAF indicators, as well as to provide support throughout the survey process. (This process is referred to as the Statistical Value Chain.) It should be noted that the requirements and guidelines include references to the relevant SASQAF indicators they support. It should furthermore be noted that, although the requirements and guidelines address the processes underlying the data and therefore cover the same aspects, there is no direct link between the structure of the requirements and that of the guidelines. The requirements represent high-level issues that should be address throughout the entire data collection process, while the guidelines provide the details to be considered during each individual step of the data collection process.

1.5 How to use this document

Details of the standard of data quality required for surveys are provided in section 2 of this document. Section 2 starts by providing a brief overview of what is actually measured by SASQAF. (It should be noted that the complete set of SASQAF indicators has not been included as part of the body of this document, but it is provided in Appendix A for reference purposes.) The ten **requirements** that provide aspects to be enforced in order to ascertain quality data, quality are included in subsection 2.2. Finally, the document provides **guidelines**, linked to SASQAF indicator prescriptions, for ensuring that quality is maintained throughout the Statistical Value Chain. These detailed guidelines are provided in subsection 2.3.

The requirements and guidelines were compiled in order to support the wide range of different data collection activities currently undertaken by the DHET. Although the SASQAF indicators were contextualised to some extent for application by the DHET, by removing indicators that were not deemed to be relevant, and by focussing on those indicators that were considered important, it would not be possible to provide detailed prescriptions in this document that would be applicable to each individual survey carried out in the DHET.

It is recommended that **the requirements** specifically**, but also the guidelines be used as support during the planning and execution of surveys undertaken within the DHET**. The team responsible for carrying out a survey could to refer to the guidelines for assistance, but must be **forced to comply** with the principles set out in the requirements. Once a survey has been completed and compulsive a quality declaration is made on data quality, then **the guidelines, in conjunction with the SASQAF indicators, may be used to support such quality declaration pertaining to the completed statistical product** emanating from the survey.

It should be pointed out that this document was designed to fit the current process of collecting information directly from institutions via a dedicated collection instrument.

2. Data Quality Standard for Surveys

The information in this document uses the latest version of the SASQAF document available at the time. The complete reference for this latest version is as follows:

South African Statistical Quality Assessment Framework (SASQAF), *2nd* ed. / Statistics South Africa - Pretoria: Statistics South Africa, 2010, ISBN 978-0-621-39105-3.

The SASQAF indicators, as obtained from this reference, are provided for reference purposes in Appendix A of this document.

2.1 SASQAF data quality dimensions

According to the SASQAF document, data quality may be assessed in terms of specific prerequisites for quality, as well as eight dimensions of quality. **Prerequisites for quality** are defined by SASQAF as referring to the institutional and organisational conditions that have an impact on data quality. These include the institutional and legal environment, and the availability of human, financial and technological resources.

The ***eight quality dimensions*** cover the following issues:

- **Relevance:**

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

- **Accuracy:**

  The degree to which the output correctly describes the phenomena it was designed to measure. It relates to the similarities 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.

- **Timeliness:**

  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.

- **Accessibility:**

  The ease with which it may 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 via which the information may be accessed. The cost of the information may also constitute an aspect of accessibility for some users.

- **Interpretability:**

  The ease with which users can understand statistical information via the 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.

- **Coherence:**

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

- **Methodological soundness:**

  The application of international, national or peer-agreed standards, guidelines, and practices, aimed at producing statistical outputs. Application of such standards fosters national and international comparability.

- **Integrity:**

  Values and related practices that maintain users' confidence in the agency producing statistics and, ultimately, in the statistical product.

These dimensions of quality may overlap and are interrelated. Furthermore, not all elements of the dimensions may be equally relevant to all surveys. There is a need to pay attention to the various aspects within the dimensions at various times throughout the Statistical Value Chain, and to address those aspects within the context of the specific survey. The requirements and guidelines provided in this document describe practical ways of applying the SASQAF principles throughout the survey.

2.2 Data acquisition, processing and reporting requirements

In order to ensure that a data collection process (census or sample survey), which is carried out within the DHET will deliver data of an appropriate quality; the ten requirements listed below **must** be met.

Table 1: Requirements

QR1	Sufficient time must be allowed for planning the collection process, as well as for approval to be obtained for the plans. The following should be adhered to in planning and documenting the survey process: 1.1. The way in which this data collection initiative fits into the overall set of educational information within DHET: the way it relates to other data already collected and the specific needs addressed by this collection, which are not addressed by another data source. 1.2. The intended scope and target (institutions, etc.) of the data collection. It should be clearly specified whether data will be collected by the national department or whether it will be collected at provincial level and integrated at national level. 1.3. The intended users of the data must be identified, as well as the way in which this study will address their information needs. This includes documentation of the formal process used to consult with the intended key users, comments on how the needs will be met, and the feedback provided to the key users.

	1.4. Methods that will be used to meet the objectives and needs, including data collection processes, analyses to be done and information to be disseminated. Documentation must be provided in terms of how overarching issues, such as the confidentiality of data, the ethics of data collection and the verification of the results will be addressed. *(See SASQAF indicators 7.2, 1.1, 1.2, 1.3, 1.4, 7.3, 8.1, 8.5, 8.6 in support of these issues.)*

QR2	A project plan must be drawn up to allow for the planning of resources required for the census/survey. These include: 2.1. timelines; 2.2. type and number of staff members required; 2.3. the development and dissemination of computer software required; and 2.4. the estimated budget that is required. The plan should include specifications pertaining to the responsibility for tasks and, in particular, the role of provincial staff members and budgets, as well as quality control of data by circuit/district officials, and it should take the different circumstances in provinces into account. *(See SASQAF indicators 0.2, 0.5, 3.1, 3.2, 3.5 in support of these issues.)*

QR3	When data collection instruments (questionnaires) are designed, the following must be considered during the design process: 3.1. The prescriptions by the national EMIS Officer pertaining to the identification of institutions, as well as the cover page must be followed. Prescriptions in terms of subject codes, PERSAL pay points and other administrative data, incorporated into the data collection, must also be followed. 3.2. Standards prescribed for concepts, definitions, codes and classifications must be followed. 3.3. Development of software tools to support the collection, capture and analysis of data and its impact on the design of the instrument must take place. 3.4. The content of the instruments must conform to acceptable questionnaire design principles. *(See SASQAF indicators 6.1, 6.3, 6.5, 7.1 in support of these issues.)*

QR4	If computer software tools are required to support the data collection, or if data is to be extracted from existing systems to support the survey, then existing software and platforms should be used wherever possible, or systems must be designed that are compatible with existing systems. The inefficient use of state resources due to the duplication of tools and systems must be avoided. *(See SASQAF indicator 2.6 in support of these issues.)*

QR5	Approval of the planning documentation, the project plan and the proposed budget must be obtained from the relevant branch conducting the data collection before the collection project could be launched. Before the instrument could be used for data collection, it must be field-tested (piloted). A report must be compiled on the execution of the pilot project. Once the feedback from the pilot project has been included in the instruments, the final instruments must be

	approved by the national EMIS Officer. If provincial staff members will be involved in the data collection, HEDCOM should approve the planning documentation, the project plan and the proposed instrument, and it should be informed on the budget and resource requirements before the start of the project. *(See SASQAF indicator 7.3 in support of these issues.)*

QR6	The sample frame to be used as the basis for the collection must be carefully considered. This includes both the full master list and any other information related to the master list. A census would be based on the entire sample frame, while a sample would be drawn from the sample frame. Since such sample frame information is dynamic, attention should be paid to quarterly update schedules when requesting the information. The sample frame information must be date-stamped. If a sample is selected, the sampling methodology must be adequately documented. The sample design must be specified in advance and the size of the sample must be planned by taking cognisance of acceptable levels for sampling errors in key variables. Decisions on the extent of quality checks, such as audits or "back-checks" must be implemented before the data collection gets underway. *(See SASQAF indicator 2.5 in support of these issues.)*

QR7	Processes must be put in place to monitor data quality at various points throughout the survey project. The quality should be verified at the point of collection (e.g. at the institutions) and formal procedures for the verification of the quality of electronically captured data should be carried out before the data is processed or analysed. All aspects of the project execution should be assessed against the plans. This includes the following specific aspects: 7.1. the intended sample size and/or return rates; 7.2. the results from quality checks at the different points; and 7.3. the achievement (or failure to achieve) of time and budgetary targets. *(See SASQAF indicators 0.7, 3.3, 5.2 in support of these issues.)*

QR8	When the results of the data collection project are published and disseminated, the following must receive the necessary attention: 8.1. No information may be supplied in tables or graphs without an accompanying description and explanation. 8.2. All reports must contain definitions of concepts and explanations of acronyms used, so that a user of the information will understand clearly what the results represent. 8.3. Whenever statistical results are published, they must be accompanied by measures of accuracy, and by any information relating to data quality that will enable the user to understand the limitations inherent in the data collection and analysis processes. 8.4. Response rates must always be reported, especially where a census survey is concerned. 8.5. No information must be released before approval by the Director-General: Higher Education and Training. Before information is referred to the DG: HET for approval, it must be approved by the relevant staff internal to the department responsible for the survey, as well as by the national EMIS Officer. 8.6. The content must be checked for accuracy and metadata should be checked for completeness.

	8.7. If the results originate from a repeat survey, the report should refer to information from a previous period, where possible, and be discussed. Any changes in either the methodology or instruments between the current and previous studies should be clearly pointed out in order to improve comparability between the current and previous periods. *(See SASQAF indicators 5.3, 5.1, 2.1, 2.2, 2.3, 4.13, 6.3 in support of these issues.)*

QR9	Data must be prepared for release by finalising data formats, compiling specifications for the release of data (including protecting confidential information from being released) and adding metadata and descriptions to facilitate interpretation and proper utilisation of the data. Data that is released from a specific survey must contain references to the specific survey. The date on which the survey was finalised, as well as the date on which the data was supplied must be indicated on the dataset. The survey is finalised when the data is disseminated to potential data users. *(See SASQAF indicators 4.1, 4.4., 4.13, 5.1 in support of these issues.)*

QR10	Formal feedback on published results must be obtained from key users. As a form of feedback results should also be communicated to the individuals or institutions who supplied the data. *(See SASQAF indicator 1.5 in support of these issues.)*

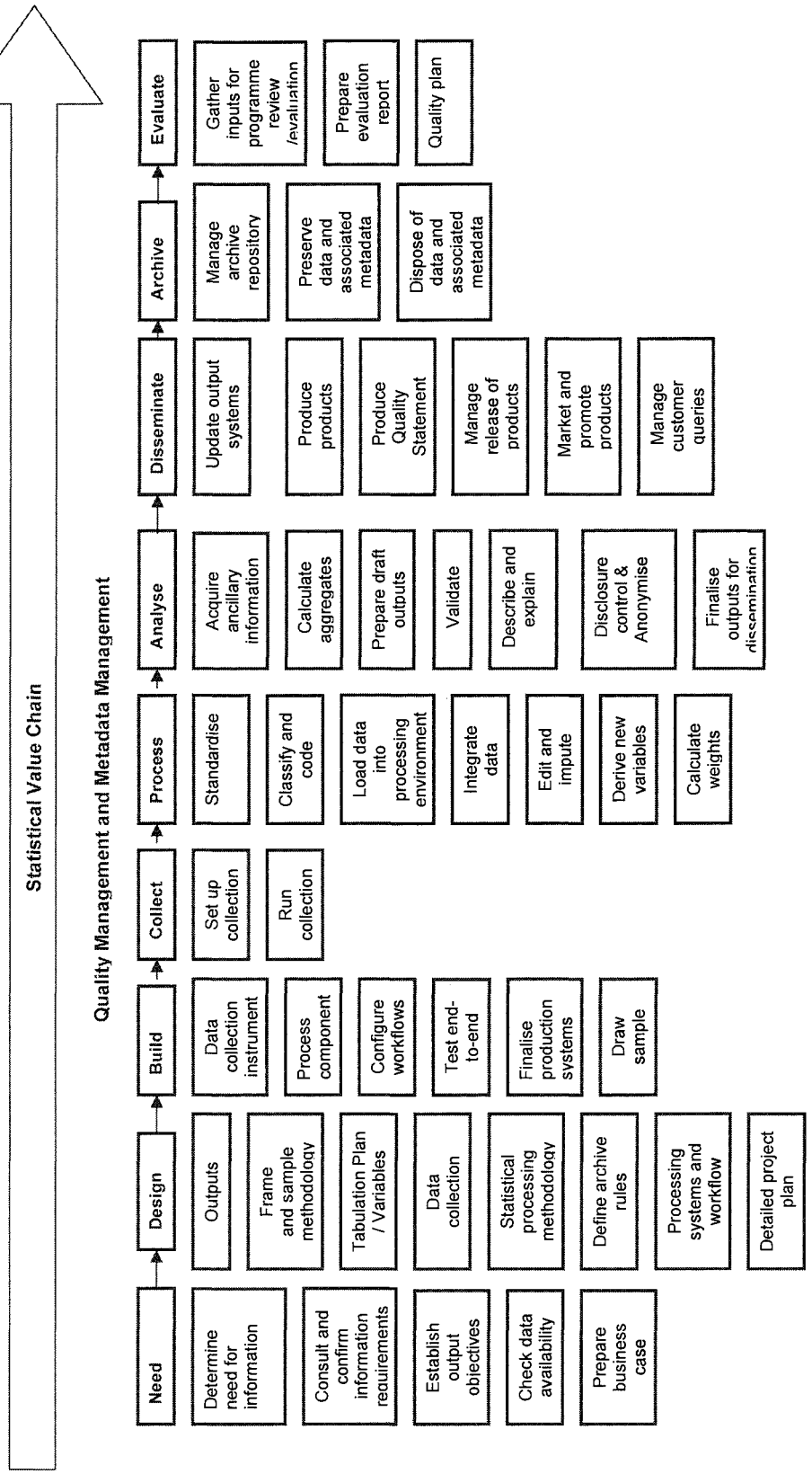
2.3 Detailed guidelines

The **requirements listed in subsection 2.2, above, must be met** in order to ensure that data of acceptable quality will be available. However, the requirements do not provide enough detailed guidance to support the whole life cycle of a survey project. The following section provides such details. Although the details are provided as guidelines and cannot all be enforced in the same way as the requirements above, **the general principle to follow is to comply with all the guidelines if at all possible and, if not possible, to provide a good motivation why it is not possible to comply**.

The intention is to provide guidelines throughout the whole survey process, and therefore the guidelines are linked to the various steps in the process. These steps are briefly summarised in the subsection below.

2.3.1 Statistical Value Chain

Any survey has to go through a number of crucial steps. For the purpose of this document, this process is called the Statistical Value Chain and it is illustrated in Figure 1 below. Although Figure 1 seems to indicate that the steps follow each other sequentially, in practice it is a more complex, interlinked process.



The guidelines are linked to the various steps in the Statistical Value Chain, and these are discussed in the subsection to follow.

2.3.2 Data quality guidelines and principles to follow

In order to provide practical guidelines, so as to ensure that the principles of data quality, embodied in SASQAF, are applied during a survey, guidelines are provided for each step that has to be carried out in a survey. The guidelines are provided in the form of action items, with comments and suggestions indicated in *italic* script, and these guidelines are based on the principles contained in specific relevant SASQAF indicators. (Refer to Appendix A for the description pertaining to the relevant SASQAF indicator.)
During a survey, the team conducting the survey may consider whether each specific action item needs to be carried out. If a specific action item is not carried out, the reason for the omission must be properly motivated.

2.3.2.1 STEP 1: Motivation, scope and project management/planning

Objectives to be reached during this step:

The initial planning and scoping stage of any survey is of crucial importance to the success of the survey. This planning underlies all survey activities to follow and should therefore be given sufficient time and consideration.

Specific attention should be paid to the aspects that the survey should cover, the methods that would be used in the various steps of the process, and what will be done with the results of the survey. It is important to establish who needs this survey and what knowledge they need to gain from it. The question should also be asked whether this could actually be provided. It is furthermore of crucial importance to plan for the resources required by the survey, in order to ensure the availability of the resources necessary for the successful completion of the survey.

QG1: Guidelines pertaining to actions carried out during this step:

		Action points: project motivation, scope and project management/planning	Related SASQAF indicator
1.1	**SVC Phase**	**Motivation and scope**	
1.1.1	Need	Determine whether the Department of Higher Education and Training is mandated to collect the information envisaged in this study. Determine whether this mandate extends to the appropriate provincial education departments. *Ideally, a legal framework, such as those provided by clause 13 of the Educational Information Policy, should specify this.*	1.1
1.1.2	Need	Conduct a review of related studies and surveys to ensure that the required information is not already available from an existing source, or could not be obtained by adding questions to an existing survey. *An example of such a review would be the information contained in a concept note for the study.*	7.3
1.1.3	Need	Compile a document providing: • a justification for the survey; • a motivation for the survey;	2.1 2.2 2.4 8.2

		Action points: project motivation, scope and project management/planning	Related SASQAF indicator
		• the objectives of the survey; and • the proposed scope of the survey.	
1.1.4	Design	Determine whether there are standards, guidelines and good practices that exist internationally or nationally for similar studies, and how these would influence the design of the study, as well as the instrument(s).	8.1

1.2	**SVC Phase**	Planning timelines	
1.2.1	Design	Make a list of all the activities relevant to this specific study, which would be required at each step of the Statistical Value Chain, i.e. as related to the information collection process. Remember to include actions relating to interaction with users, as well as the publication of the final results. *Such a list could be compiled in the format of tasks for a project plan or as a Work Breakdown Structure (WBS).*	1.6
1.2.2	Design	Attach timelines to each of the listed activities. These timelines should allow enough time for each activity, but also ensure that the final results are obtained at a suitable end-date.	4.3
1.2.3	Design	Consider the risks involved in each of the activities in terms of both the ability to keep to timelines and the impact on data quality.	
1.2.4	Design	Plan for interaction between the national department and the provincial departments/institutions. Allow sufficient time for different circumstances peculiar to the different provinces/institutions, which would impact on whether they are able to meet certain timeframe deadlines. *For example, certain provinces/institutions have more institutions in remote rural areas and so data collection takes longer than in provinces/institutions with more urban institutions that are connected via the Internet.*	4.3
1.2.5	Design	Carefully consider timelines related to the development of software tools. This is of particular importance if the information will be collected via administrative systems, since the time required for changes to such software, as well as for the decentralised distribution of the software, has to be taken into account.	1.6
1.2.6	Design	Determine the results required at the end of the project. Adjust timelines accordingly. Ask questions such as: Will there be a preliminary, as well as a final release of data? Would these releases come from the same source? Will a 100% return rate be required before the survey is considered complete?	4.1, 4.2
1.2.7	Design	Ensure that an appropriate length of time is allowed between the closing of the data collection and the date of publication of the results.	4.3
1.2.8	Need	Identify the key users of the information, both internal and external to the Department *(for example international agencies)* whose inputs are required for the survey. Decide who would need this information and how they would utilise it.	2.1, 2.2

1.3	**SVC Phase**	**Interaction with users**	
1.3.1	Need	Set up a systematic way to communicate with users for a specific survey. Provide users with opportunities to indicate what information is required and how it should be supplied to them. *Examples would be to send them the survey form and analysis plan, or to hold a workshop for them.*	2.2

1.3.2	Need	Formalise processes to analyse the relevance and applicability of the indicated user needs, to build in comments from key users into the survey planning, to test the satisfaction of users, and to provide users with explicit feedback about which needs could or could not met.	2.3, 2.4, 2.5
1.3.3	Dissem ination	Ensure that advance notice is given to users with regard to major changes in methodology, source data and/or statistical techniques.	9.3

1.4	**SVC Phase**	**Planning resources**	
1.4.1	Design	Identify the type and number of resources required. This includes number and types of people required, computer equipment required, total available time and total budget required.	1.6
1.4.2	Design	Determine whether sufficient resources are available. Also ascertain whether the available resources will be available in accordance with timelines taking into account that different surveys could be running concurrently.	1.6
1.4.3	Design	Put in place systems to monitor and manage the efficient use of resources in accordance with the project plan.	1.7
1.4.4	Design	Compile a budget linked to the activities to be carried out, as well as to the associated timelines. *This budget must have sufficient detail to allow for checks on the potential impact of budget cuts or limitations pertaining to the quality of the survey.*	1.6
1.4.5	Design	Check whether adequate resources have been planned for data/statistical services in the overall budget planning.	1.6
1.4.6	Design	Determine whether adequate resources are available at provincial level.	1.6, 1.7

1.5	**SVC Phase**	**Ethical and governance issues**	
1.5.1	All phases	Put in place ethical guidelines for staff behaviour *(for example a professional code of conduct)* and communicate this to all staff members participating in the survey.	9.6
1.5.2	Dissem ination	Compile a document to describe the conditions under which policy-makers, government in particular, may have access to data before the actual release date.	9.2
1.5.3	Dissem ination	Governance issues must be addressed in the planning stage, namely who legitimises and approves the results *(for example, the Minister: Basic Education, the Director-General: Higher Education and Training, etc.)* Adapt the project plan and timelines accordingly.	9.4

2.3.2.2 STEP 2: Design of instruments

Objectives to be reached during this step:

In this step, the important issue of **what** is to be collected must be addressed – both for once-off collections and for repeated surveys. It is important that the data collection instrument(s) should adhere to common concepts, definitions, classifications and standards. Furthermore, user needs must be considered and coordination between surveys must be done, in order to minimise unnecessary overlaps in data collection via different surveys.

It should be remembered that data collected from an official departmental survey will become official statistics within the department and therefore the planning of the collection of information must be carried out with care.

QG2: Guidelines pertaining to actions carried out during this step:

		Action points: Design of instruments	**Related SASQAF indicator**
2.1	**SVC Phase**	**Contents of data collection instrument(s)**	
2.1.1	Build	The design of the instrument(s) must take into account requirements by users. Communication with users, ensured by the planning in Step 1, as described above, is therefore necessary.	8.3
2.1.2	Build	In repeated surveys, revisions to the survey must take into account any new developments *(e.g. changes in definitions, classifications, etc.)*. The survey instrument must be kept up to date in accordance with current changes and requirements.	7.1
2.1.3	Build	The collection instruments should take into account changes in both the education system and in policy priorities, particularly where repeated surveys are concerned.	3.6
2.1.4	Build	Any changes made to the survey instrument(s) must be documented and the impact of these changes must be considered and documented; for instance in terms of comparability with historical data, data aggregation, and other pertinent aspects.	6.1, 7.1
2.1.5	Build	Changes to the survey instrument(s) must be implemented in the database(s) and other systems underlying the collection process. The impact of such changes on software development timelines must be incorporated into the planning.	3.6
2.1.6		All relevant policies, pertaining to the objectives and contents of the survey, must be taken into account.	1.2

2.2	**SVC Phase**	**Formulation of questions and instructions**	
2.2.1	Build	Methodologies used must follow accepted standards, guidelines or good practices (national, international, peer-agreed), with regard to questionnaire design.	8.3
2.2.2	Build	Good questionnaire design principles should be followed in compiling the instrument(s). These include: • instructions should be clear; • unnecessary questions should be eliminated; and • an introduction should be provided in an introductory letter to describe the objectives of the survey, as well as to ensure that the correct person completes the form.	8.3
2.2.3	Build	All questions must be checked to ensure that they are clear and concise and not ambiguous. The formulation of the questions must be motivated by statistical considerations.	8.3
2.2.4	Design	The instrument(s) must be checked to ensure that the data collected matches the needs in terms of the final results and statistical products required. In order to eliminate unnecessary questions, an analysis plan could be drawn up to indicate how the answer to each question would be used in the analysis. *Unnecessary questions, i.e. those that do not contribute to the answers or analysis required, must be eliminated, and questions must be asked in such a way that the intended results could be obtained from the answers.*	8.3

2.2.5	Design	In order to improve the clarity of questions and to ensure the appropriate format of answers, a tabulation plan could be drawn up. Such a tabulation plan would indicate how calculations from answers should be done. *If necessary, guidelines or a format specification should be provided for answers. (E.g., if the answer must be provided as a date, the format should be indicated as day-month-year or just month-year.)*	8.1
2.2.6	Build	The impact of the development of software on the design of the instrument(s) must be taken into account. *The use of software tools could impact on both the formulation of questions, as well as on the timelines required for instrument development and testing.*	

2.3	**SVC Phase**	**Coordination of information collected**	
2.3.1	Need	The information collected by related surveys must be considered. Duplication of collected information between different surveys must be avoided, and changes to other surveys must be considered. *For example, if questions are removed from the Annual Survey they may be added into other surveys, or questions may be added into the Annual Survey instead of conducting a separate survey.*	
2.3.2	Build	The collection instruments must be checked for comparability of data from one year to the next and from one province to another province.	7.1
2.3.3	Proces sing	Checks must be done to ensure consistency between information contained in the master list and that in the planned survey. If similar information is collected, find methods aimed at ensuring updates to the master list from information collected in the survey.	7.3

2.4	**SVC Phase**	**Common definitions, concepts and classifications**	
2.4.1	Design, Build	The common concepts, definitions and classifications, used must follow accepted standards, guidelines or best practices	8.1
2.4.2	Proces s	Ensure synchronisation between operational data systems (surveys) and administrative systems in terms of information, as well as concepts, definitions and classification.	7.1
2.4.3	Build	The national EMIS number must be used as an identifier and the front page, as prescribed by the national EMIS Officer, must be used as far as possible for all surveys.	7.5
2.4.4	Design, Build	There must be a good reason for any departure from standards. Such reasons must be documented, so as to be available for metadata purposes.	7.3

2.5	**SVC Phase**	**Piloting the instruments**	
2.5.1	Build	Once the content has been finalised, the instrument(s) must be field-tested (piloted). Such a pilot project involves asking a group similar to the intended user pool to complete the instrument in order to check for remaining ambiguities or errors in the questions, or to pinpoint practical problems in answering certain questions.	8.3
2.5.2	Build	A feedback report should be compiled, based on findings of the pilot project, and the instrument(s) should be revised based on the feedback contained in this report. *It should be noted that planning should take into account the possibility that an instrument could be substantially revised or even rejected, based upon the findings of the report.*	8.3
2.5.3	Build	The group on which the instruments are piloted should have similar characteristics to those of the intended sample population, but it does	8.3

		not need to be a very big group. *It should be noted that instruments completed as part of the pilot project must not be analysed as part of the survey results.*	

2.6	**SVC Phase**	**General**	
2.6.1	Design	Appropriate timeframes must be set for the design and finalisation of instruments, and checks should be done for adherence to timeframes, specifically in respect of timeframes concerning changes to instruments. These timeframes must also allow for logistical requirements in the different provinces, where appropriate.	3.3
2.6.2	Design	Attention should be paid to the development of software that will be used to implement the instruments, and sufficient time should be allowed for such development.	1.6
2.6.3	Design	The instruments, timeframes and any information motivating the design of the instruments or changes to instruments must be circulated amongst all the relevant role-players.	1.6
2.6.4	Build	The survey instrument(s) must be approved by the national EMIS Officer.	1.6
2.6.5	Build	The final instrument(s) must to be approved by the relevant stakeholders *(For example: approval by Head of Education Departments HEDCOM).*	1.6

2.3.2.3 STEP 3: Planning of data collection via census or sample

Objectives to be reached during this step:

This step includes all activities pertaining to the maintenance of the sample frame, as well as to the designing and compilation of the sample.

If a census has to be conducted (i.e. information collected from all institutions), the sample frame will to be used to identify all the institutions, and therefore sample frame choice and maintenance are applicable to a census.

If a sample is required, this step deals with the sample frame maintenance, as well as with the way in which the sample is selected and the attributes of the sample. Designing a sample required for data collection firstly involves choosing an appropriate sample frame and method of distribution and, secondly, selecting a sample size based on the information that must be collected and the accuracy required for using such information. The sample design needs to be specified in detail before data collection could start.

QG3: Guidelines pertaining to actions carried out during this step:

		Action points: Planning of data collection (census / sample)	**Related SASQAF indicator**
3.1	**SVC Phase**	**Sample frame**	
3.1.1	Design	Since the master list is used as a sample frame for the majority of surveys, it is important that the contents of the master list and the procedures for updates to the master list are confirmed for applicability	3.5, 8.3

		Action points: Planning of data collection (census / sample)	Related SASQAF indicator
		to the survey. The same applies to other information related to the master list, which provides information with regard to the selection of the sample frame.	
3.1.2	Design	The master list is dynamic and updates at provincial level are provided quarterly to the national master list. Information extracted from the master list to be used as a sample frame must be date-stamped, so as to indicate when the "snap shot" of the master list for the sample was taken. Information used for a census of all institutions, which is based on the master list, must be similarly date-stamped. *Assessments, such as response rate, must be compared to the master list at the time of planning and not measured against changes occurring at a later stage.*	3.5, 8.3
3.1.3	Design	When using information from the master list and other sources as a sample frame, attention must be paid to the common concepts and definitions contained in the master list, and these must correspond to those used in the survey. *(For example, the definition of a institutions as defined in the master list and in surveys within and outside of the DHET).*	7.1, 8.1
3.1.4	Design	Non-sampling errors relating to the sample frame should be documented. *An example would be any errors found in the sample frame.*	3.2

3.2	**SVC Phase**	**Sampling issues**	
3.2.1	Build	The methodology used for sampling should be motivated in terms of sound statistical methodologies. As far as possible, the methodologies used should follow accepted standards, guidelines or good practices (national, international, peer-agreed), in respect of sampling methods, sample frame design, and sample frame maintenance.	8.3
3.2.2	Design	The available budget for the project must be checked to ensure that it will allow for the implementation of the intended sampling methodology. If the budget does not allow for a sufficient sample size, this must be pointed out and documented.	1.6
3.2.3	Design	The sampling methodologies must be checked to ensure that the scope of the study is consistent with accepted standards, guidelines or good practices, as well as with the scope specified.	
3.2.4	Design, Evaluate	Processes must be implemented to validate the quality of the methodologies used. *This may be survey specific – i.e. if the sampling is of crucial importance to the survey, it may be advisable to obtain approval from a person independent to the study, who could act as an outside expert to approve the design of the study.*	1.8
3.2.5	Design	The sampling methodology must be adequately documented (e.g. proportional, stratification, weighting, etc.). It is important that the sample design is specified in advance and meets the objectives, goals and scope of the study. Advice must also be provided on how the sample should be conducted, including specifications at provincial level. *It should be noted that this information must be available at the end of the study as part of the metadata. However, it is a practical approach to compile the documentation as the survey proceeds through the various phases.*	6.1, 5.12
3.2.6	Design	The size of the sample must be planned by taking into account acceptable levels for sampling errors in key variables. The actual	3.1

		sampling error obtained may be reported afterwards, both in terms of the errors that occurred and how these compare to the errors anticipated in the planning stage. *It should be noted that the metadata reported at the end of the study must include, not only the sampling error estimates based on the actual data collected, but also how the "actual" measurements compare to the "planned" measurements.*	

2.3.2.4 STEP 4: Data collection

Objectives to be reached during this step:

This step involves the actual collection of the survey information. The collection must be carried out in accordance with the objectives set for the survey and the planned collection strategy. The collection process must be monitored to detect and correct data collection errors, or to report on problems experienced during data collection, which could have an effect on data quality and which could not be rectified.

QG4: Guidelines pertaining to actions carried out during this step:

		Action points: Data collection	Related SASQAF indicator
4.1	**SVC Phase**	**Planning and monitoring of data collection**	
4.1.1	Collect	The terms and conditions, including confidentiality, under which statistics are collected, should be documented and made available to the people providing information. *This is especially necessary if the survey includes questions about perceptions and opinions.*	9.1
4.1.2	Collect	Measures must be in place to ensure that individual data is kept confidential and used for statistical purposes only. *(Individual data may refer here to data about an individual learner/student, educator/lecturer or institutions, which should not be released to the general public, such as personal contact numbers or fax numbers of a institutions. Usually there will be a first level of information that could be released, but there could also be an additional level of information that may not be released without further approval.)* Confidentiality needs to be ensured throughout the collection process. NOTE: Although the SASQAF indicator refers only to data collected via a survey, confidentiality must also be ensured in terms of administrative data, such as data on the master list and data proposed to be incorporated into the learner tracking system in future.	1.4
4.1.3	Collect	The process used for data collection must follow accepted standards, guidelines and good practices. • Detailed specifications, which should be followed to ensure that procedures do not differ between provinces, will be provided. Such specifications should include details about: o how the questionnaires are distributed; o who receives the questionnaires; o which source documents to use for information; o who should complete the questionnaire o who should provide information; and	8.3

		Action points: Data collection	Related SASQAF indicator
		o who should approve the information provided. *For example, for the collection of Snap data, it is important to specify what the source documents for the enrolment figures are; e.g. the admission register.*	
4.1.4	Design	The data collection process must be well documented and the data collection protocol must be written before the data collection process starts, so that it could be followed consistently throughout the entire data collection process.	6.1
4.1.5	Collect	Where necessary, data collectors should be properly trained, in order to ensure that they follow the correct procedures. *This is particularly important in those surveys where the data is collected by an interviewer and not by the completion of a form by the individual respondent concerned.*	1.6

4.2	**SVC Phase**	**Timelines**	
4.2.1	Design	Deadlines should be set for the various aspects related to data collection. These deadlines should allow sufficient time for data collection, and the data collection process should be monitored against these deadlines.	4.3
4.2.2	Design	The timeline should specify clear deadlines for the collection of data from the primary source. These deadlines must be met and should be monitored. *For example, with regard to the annual survey, there is one date for the collection of information from all institutions and the collection process should ensure that institutions keep to this date. The organisation of the data collection should also be done in such a way that it enables acquisition from sources to be done in accordance with deadlines, for example by distributing instruments on time.*	4.3
4.2.3	Design	If information is collected at provincial level, the timeline should specify clear deadlines for the transfer of information from the primary source to the provinces and from the provinces to the national department. These timelines should be agreed upon before starting the collection of data and they must be monitored.	4.3
4.2.4	Design	The planning of data collection should include specifications for when the data collection would be considered to be complete, and this should also be reflected in the timeline. *For instance: When does the data collection for Annual Survey for a particular year end? How much time is allowed for such completion?*	4.3

4.3	**SVC Phase**	**Monitoring data collection**	
4.3.1	Collect	All parts of the data collection processes should be monitored in terms of quality and timeliness, including the delivery of the survey instruments and their return. Monitoring should ensure that every unit is delivered, received and processed.	1.8, 4.3
4.3.2	Collect Process	In order to do quality checks on the accuracy of collected data, a planned process of auditing (sometimes called "quality checks" or **"back checks"**) may be done. This involves revisiting a respondent and verifying the correctness of all the answers provided. *Usually, a person who was not involved in the initial data collection would carry out such a back check.*	3.2
4.3.3	Collect	The procedure for conducting back checks must be established and documented before the start of the data collection. If back checks	1.8, 3.5

		are to be performed, the back check sample has to be drawn before the start of the data collection process, and must not be influenced by the actual collection of data. *For example, the back check must not merely be done on the first units to be collected.*	
4.3.4	Collect	Processes must be put in place to monitor data collection and to track the progress of such collection. If provinces are involved in the data collection process, the national department must prescribe the nature of this process and the process must be consistent for all provinces. *An example would be the use of control sheets to monitor data collection. Every province/institution could have a control sheet, indicating to which institutions the survey form had been sent, as well as to record when it was returned. (Or, if not returned, why.)*	1.8
4.3.5	Collect	The data collection process should include the monitoring of non-responses *(refusals, units omitted by accident, no feedback received)*. There should be plans in place to deal with such issues, in order to ensure that the required return rates are met. Data collection problems should be addressed where possible and details must be documented for metadata purposes.	3.2, 5.12
4.3.6	Collect	All roles and responsibilities during the data collection process should be clearly specified. Furthermore, roles and responsibilities should include checks for data accuracy by persons as close to the source as possible. *For instance, the person collecting the information from the institution must also verify the correctness of the information.*	1.8
4.3.7	Design	If electronic methods of data collection are used, the tools designed for such data collection should be open and flexible to allow for new developments and changes. If the national department develops a capture tool for data collection, then the use of this tool must be enforced. If alternative software is used or developed, this software product must conform to the departmental capture tool.	3.6
4.3.8	Evaluate	The process of data collection and the problems experienced should be documented to guide future survey process planning.	

2.3.2.5 STEP 5: Quality control procedures

Objectives to be reached during this step:

Although there must be procedures for quality control in place throughout the survey, specific actions should be taken to ensure checks for completeness, correctness and coherence of data. This step describes those actions and it may be carried out just after the completion of the data collection, or it may be included as the first step in data processing. Confirming data quality before processing or analysing the data must receive special attention, and must not be ignored due to time pressures.

QG5: Guidelines pertaining to actions carried out during this step:

		Action points: Quality control procedures	**Related SASQAF indicator**
5.1	**SVC Phase**	**Verifying data quality**	
5.1.1	Collect	Processes must be in place to focus on, monitor and verify data quality. There should be a data process flow diagram for data collection procedures at national and/or provincial level, and it should be indicated where the quality control processes are located.	1.8

		Action points: Quality control procedures	Related SASQAF indicator
		The quality control processes must also be checked to ensure that they are adequate for the survey, and that none of the procedures is redundant.	
5.1.2	All phases	The measures used to monitor quality at various points in the survey should be appropriate and should include measures of time and cost with regard to corrective actions. This will allow for measurement the effectiveness of the quality control procedures.	1.8
5.1.3	Pre-requisites of quality	Standards and policies are in place to promote consistency of methods and results. These are documented and used to ensure consistency between the different provinces/institutions.	1.2
5.1.4	Collect Process	Back checks should be carried out as planned (see discussion above for reference to the planning of "back checks"), and any problems detected during the back checks must be documented. *The aim would be to use the back checks to report both on the accuracy of the sample frame, as well as on the accuracy of the collected data.*	1.8
5.1.5	Collect Process	A process must be specified for checking and verifying the information that has been collected prior to the capture of the information. Checks must be carried out to determine coverage, timeliness and coherence. This process must be consistent between provinces. *For example:* • *If there is a blank space, why is it blank?* • *Was complete information provided?* • *Does information match throughout the survey form?*	3.2
5.1.6	Collect	Checks on data accuracy must be done at the institution or by the province concerned to make it easier to correct the data. Such checks must be nationally specified for consistent application, and responsibility for the checks must be clearly communicated. *For instance, circuit and district managers must check Snap forms when they receive them from the institutions and not just sign off that they have received the form.*	3.2
5.1.7	Collect	The deadlines for the completion of the data collection should allow sufficient time for back checks and for checks pertaining to the correctness and completeness of data.	4.3
5.1.8	Process	Methods must be put in place to determine the existence of duplication in the data.	3.2
5.1.9	Process	Methods must be put in place to check that institutions are identified correctly in accordance with the national EMIS number.	7.5
5.1.10	Analyse	Correctness of data should be checked against other sources, where possible.	7.4
5.2	**SVC Phase**	**Correcting errors in data**	
5.2.1	Analyse	Processes should be put in place for providing feedback on errors in the data or in the information on the master list, as detected by other surveys.	7.4
5.2.2	Design	Timelines for data collection and processing should be planned in such as way as to allow time for correcting any errors found in the data.	4.3
5.2.3	Design Build	A procedure should be introduced to resolve discrepancies indicated by the back checks.	8.3

2.3.2.6 STEP 6: Data capture

Objectives to be reached during this step:

The process of converting data that is in paper format into electronic format must be handled in such a way that it does not delay the overall survey time and does not compromise the quality of the data either.

QG6: Guidelines pertaining to actions carried out during this step:

		Action points: Data capture	Related SASQAF indicator
6.1	**SVC Phase**	**General**	
6.1.1	Process	The data capture process should be completed within the required deadline.	4.3
6.1.2	Design	Data capture should be planned to start as soon as the first data is received, and not only once all data has been received, so as not to delay the data processing process.	4.3
6.1.3	Process	Data capture methodologies that are used should follow accepted standards, guidelines and good practices, and these must be consistent across the different provinces/institutions. Data capture software must also be consistent across provinces/institutions, so that consistent standards could be delivered.	8.3
6.1.4	Build Process	Electronic data capture software must be standardised for a specific survey and data structures must be standardised within a survey. This will ensure comparability of data capture, data validation during capturing, and outputs of capture software.	1.6
6.1.5	Build	The data capture software should make use of appropriate and compatible technology. The software should be planned so as to utilise, or link up with existing systems.	3.6
6.1.6	Build	Software developed for data capture should include a facility for checking the correctness of data formats during the capturing process. This will assist in the quality assurance of captured data. *For instance, the software should check that a character is not entered in a field that should contain a number or a date, and that the number captured in a specific field falls within a specified range for that field.*	3.2
6.1.7	Process	Attention should be paid to the codes used to identify missing data. There should be a clear distinction between a value that is zero and a value that has not been filled in.	8.3
6.1.8	Process	Where possible, a double entry should be done for either the entire dataset or for a subset of the dataset. If the data capture software does not allow for this, manual checks may have to be carried out to confirm the correctness of the captured data.	1.8
6.1.9	Process	Data capturers should receive sufficient training.	1.6
6.1.10	Process	As a general principle, a data capturer should be concentrating on transcription and not on editing, in order to complete the data capture task quickly and accurately. It should therefore be possible to outsource the data capture task.	3.2?
6.1.11	Process	It should be noted that the quality checks done on forms prior to capture should not be considered part of the capturing task itself, but this must be managed as a separate activity. If a survey uses data capturers specialising in educational information to enable	

		Action points: Data capture	Related SASQAF indicator
		them to check and capture at the same time, this must be taken cognisance of regarding the deadlines planned for the data capture activity and the monitoring of progress with the data capture process.	
6.1.12	All phases	If the survey contains a large proportion of sensitive or confidential data, it may be good practice to put a requirement in place that data capturers should sign confidentiality agreements.	1.4; 9.1

2.3.2.7 STEP 7: Integration and editing/cleaning of data

Objectives to be reached during this step:

This step refers to the process of integrating all data that has been collected, in order to prepare such data for analysis. Typically, this may mean integrating data obtained from the various provinces/institutions into one national dataset. The aim would be to prepare the data for analysis and eventual publication.

Some basic editing (i.e. data checking and cleaning) should be carried out at this point in the Statistical Value Chain, in order to prepare the dataset for analysis. However, it should be mentioned that bad quality data cannot always be fixed at this point in the process. It is therefore preferable to put in place processes to prevent errors, or to ensure that such rectification takes places as early as possible in the cycle, and as close as possible to the source of the data.

QG7: Guidelines pertaining to actions carried out during this step:

		Action points: Integration and editing/cleaning of data	Related SASQAF indicator
7.1	**SVC Phase**	**Integration**	
7.1.1	Process	Procedures for integrating data from different provinces/institutions or from different directorates within an education department must be planned and communicated to the relevant role-players. Similarly, planning should be done before integrating data from sources outside the DHET with information collected by the DHET.	7.5?
7.1.2	Process	The processing of data should be monitored to ensure that it adheres to the set time deadlines. *Time should rather be spent on earlier quality control procedures than at this late stage in the cycle.*	4.3
7.1.3	Process	Integration of data must be done by using the national EMIS number as a unique identifier of data records. With regard to data collected by agencies on behalf of the DHET, the recommendation is that they must also use national EMIS numbers as the basis for their collection. Checks must be done to identify any possible duplication of records in the dataset and its elimination.	7.5
7.1.4	Analyse	The national EMIS number may also be used to check and confirm information obtained from a census against the current version of the master list.	1.8, 3.5, 7.3, 7.5

7.2	SVC Phase	Editing/cleaning	
7.2.1	Process	There should be formal data editing procedures in place to check on data consistency within a dataset, and these should be set up as far in advance as possible. These checks should identify data values that are **potentially in error**, namely data that that is missing, invalid, duplicated or inconsistent. Once identified, these values should be checked and either confirmed or corrected. *Ideally, some of these procedures must be automated into software for consistency in checking the electronic data, but for smaller surveys manual processes could be used.*	1.8; 7.2
7.2.2	Process	Data editing procedures should include a comparison of information from institutions with their peers, and within the dataset as a whole. *For example, checking for outliers compared to the average and standard deviations in the whole dataset. It should be noted that a value, which is identified as an outlier, does not necessarily indicate incorrect data.*	7.2
7.2.3	Process	Data editing procedures should include cross-checks between related values, and ideally, should also include checks on summary totals for such related fields. *For example, this would include checking subtotals in Snap tables.*	8.3
7.2.4	Process	For repeated surveys, editing procedures should compare values submitted previously for a specific institution. Ranges are set for acceptable changes from one year to the next, and changes in values for individual institutions are checked and verified against these ranges.	8.3
7.2.5	Process	Editing procedures should include checks on the validity of values, such as checks against list tables for codes, values falling outside acceptable ranges, as well as required fields or sections that have not been completed.	8.3
7.2.6	Build Process	Data cleaning procedures must be revised to take into account changes in the data collection instruments from one year to the next, which makes it impossible to successfully compare data from one year to the next. Such revisions must be documented for metadata purposes.	8.5
7.2.7	Build	Methodologies used for editing should follow accepted standards, guidelines or good practices (national, international and peer-agreed).	8.3
7.2.8	Analyse	Care should be taken if there are revisions to data *(for instance by a province that submitted an initial dataset and then submitted a revised dataset)*. If any data is submitted after publication of the results and such data is included in the national dataset, this additional data need to be tagged as such in the dataset.	8.5

2.3.2.8 STEP 8: Aggregation and analysis

Objectives to be reached during this step:

When the collected data is analysed, it is important to use sound methodologies for the analysis. The focus of the analysis should be on providing answers to key questions and not just to show all the results. Of further importance is to document all the procedures and methods used. Documentation of the procedures provides evidence of the acceptability and quality of the obtained results, and also ensures the repeatability of the results.

It should be noted that an analysis could involve providing descriptive statistics, such as aggregated totals and counts, but it could also comprise an in-depth (technical) analysis of the collected data. The guidelines cover general aspects that relate to both these kinds of analysis, as well as aspects that relate primarily to technical analysis.

QG8: Guidelines pertaining to actions carried out during this step:

		Action points: Aggregation and analysis	Related SASQAF indicator
8.1	**SVC Phase**	**General**	
8.1.1	Build Analyse	The methods planned for the analysis must be motivated by the objectives of the study and by statistical considerations, rather than by the data itself or by political interference. Therefore, the choice of the level of analysis, as well as the relevant analysis methods, should preferably be finalised before data collection takes place.	8.3 9.4?
8.1.2	Design	The analysis should, as far as possible, be aligned with the analysis plan that was drawn up during the planning and design steps, in order to ensure that the analysis provides answers that are in line with the objectives of the study.	8.3
8.1.3	Build	Electronic systems designed for data analysis must be designed so as to be compatible with the systems designed for data collection and capture. *If systems are developed by a third* party *(i.e. outside the Department of Higher Education and Training), then this software product must link up with the Department of Higher Education and Training capture and analysis tools and produce analysis outputs that are acceptable to the department.*	8.3
8.1.4	All phases	The terms and conditions, including confidentiality, under which statistics are collected, processed and disseminated, and which was planned for the survey, should be adhered to during the analysis.	9.1

8.2	**SVC Phase**	**Technical analysis**	
8.2.1	Analyse	All assumptions underlying the analysis should be stated clearly, and appropriate methods for the analysis should be selected accordingly.	8.3
8.2.2	Analyse	Assumptions, analysis methods and any limitations to the analysis should be documented for metadata purposes.	4.12
8.2.3	Analyse	Measures of sampling errors for key variables must be calculated from the actual data that was collected, and these measures must be compared to the targets set for these values during the survey planning. As far as possible, results should be reported in accordance with the relevant standard errors or confidence intervals. This information must also be provided in the metadata that reports on data quality.	3.1; 5.1
8.2.4	Analyse	Results should be reported in such a way that it supports appropriate conclusions from the analysis. Assumptions, limitations, data quality issues, response rates, potential biases and other important aspects that affect the analysis should be taken into account.	8.3

8.2.5	Analyse	If possible, measures of non-sampling errors must be calculated – such as errors picked up in the sample frame or during editing. It is also important to consider the effect of errors, such as deficiencies in sample frame or instruments, or other underlying biases to the analysis on the quality of the data and analysis. These errors and their impact should be documented for metadata purposes.	3.2
8.2.6	Analyse	With regard to census surveys, it is important to report on non-response. The extent of non-response, possible reasons for the non-response and measures (if any) taken to address the non-response should be documented and reported.	3.2
8.2.7	Analyse	Documentation must be provided to describe record-matching methods and techniques used (for instance matching of collected data to the master list), and other analysis and editing methods used. It is important to record what was done, what errors were found and how these errors were rectified.	3.7

2.3.2.9 STEP 9: Publication of results

Objectives to be reached during this step:

Publication is done in order to make results available to users in an accessible way. In order to ensure proper use of the results, metadata, explaining the contents of the data and the methods used to obtain and analyse it, should be provided with the results.

QG9: Guidelines pertaining to actions carried out during this step:

		Action points: Publication of results	**Related SASQAF indicator**
9.1	**SVC Phase**	**Before disseminating the data**	
9.1.1	Dissemination	Determine whether there is a data dissemination policy relevant to the survey. *For instance, this would be a publication policy that governs the contents, the form in which it is distributed, to whom it is distributed, etc.*	5.9
9.1.2	Dissemination	Determine whether there are clearance procedures ensuring that the document has been approved by the unit producing it and whether there are guidelines on the issues to be addressed before publication could be approved. Ensure that these prescriptions are followed before publication of the results.	5.9
9.1.3	Dissemination	The Director-General: Higher Education and Training will need to approve the contents before releasing it to the general public. Before information is referred to the DG: HET for approval, it must furthermore be approved by the relevant staff members internal to the department. These include the staff members responsible for conducting the survey, and also the national EMIS Officer. Determine the requirements and processes in obtaining approval for release and ensure that these are correctly followed.	????
9.1.4	Dissemination	For qualitative data *(e.g. financial planning)*, the publication should be in a format that would allow the data to be integrated and analysed.	5.4?
9.1.5	Dissemination	Adding or removing results due to political pressure should be avoided.	9.4

9.1.6	Dissemination	Ministerial commentary when data is released should be identified as such, and should not be seen as part of the official statistics.	9.4

9.2	**SVC Phase**	**Reporting on the quality of the results**	
9.2.1	Dissemination	Statistical measures of data quality should be reported with the results. It should also be indicated how the obtained accuracy differed from the planned levels, including the response rate that was actually achieved. It should be noted that it could be important to indicate if a particularly bad response was obtained with regard to certain aspects, even if the overall response rate was satisfactory. *This helps the user to understand what the results can and cannot be used for.*	3.2
9.2.2	Dissemination	The metadata collected throughout the survey should be collated. Issues recorded in the metadata that affect data quality or influence the use of the results should be published in conjunction with the results. *This could include objectives, methodology, data sources, accuracy, instruments, sampling plan, editing and imputation.*	6.1
9.2.3	Dissemination	There will be metadata that should be published with the data, metadata that may be released and metadata that will only be released if legally required to do so. Metadata for internal use carries a lot of detail, while metadata for publication carries less detail. Determine which portions of the recorded metadata should be released with the results. *A person who reads the report must understand how the study was conducted and what exceptions or unanticipated aspects were found in the data or the processes.*	6.1
9.2.4	Dissemination	Concepts, definitions and classifications underlying the data should be published in order to clarify the data. *This should form part of the reference section in all reports.*	6.1
9.2.5	Dissemination	Information on the scope, sampling, data sources and statistical techniques used should be published. Differences relating to accepted standards, guidelines or good practices should be pointed out and motivated.	6.1
9.2.6	Dissemination	Some information must also be published in terms of the processing of the data. For internal use, metadata should include the actual information one would need to process the data, such as the statistical code used to run the analysis, and this code must be available if requested.	6.1
9.2.7	Dissemination	All the statistical information published (tables, etc.) must be accompanied by an explanation of the contents, and clarification of the key findings. *There must be sufficient information to guide the user in the interpretation of the information.*	6.3
9.2.8	Dissemination	If there are revisions of data, or if data collection continues after publication of the results, information pointing this out must be included with the published results. *This must be done to warn users about differences between the published results and the survey data disseminated.*	8.4 – 8.6

9.3	**SVC Phase**	**General issues**	
9.3.1	All phases	All activities should be monitored and controlled in order to ensure that publication deadlines are met.	4.3
9.3.2	Dissemination	When the results are published, it should be indicated whether this is a repeated survey and how often the survey has been repeated.	5.1

9.3.3	Dissemination	Ensure that the results will reach key users by employing the appropriate types of media/channels. Make sure that the correct level of confidentiality is maintained in the published results. *Do not release raw data or data that allows individuals to be identified.*	5.4; 1.4
9.3.4	Dissemination	Make sure that the release of data before publication adheres to the planned timelines and conditions.	4.3, 9.2
9.3.5	Dissemination	Put systems in place to obtain feedback on the published results from key users. Furthermore, ensure that published results are provided to those who provided data for the survey.	2.5

2.3.2.10 STEP 10: Handling queries

Objectives to be reached during this step:

The survey is only completed once all data has been finalised and is ready for release. Furthermore, once a study has been completed and results are published, there may be requests from users for the data attached to the survey. In clause 24 of the Education Information Policy, it is stated that the needs of researchers to analyse data pertaining to the education sector should be anticipated and that data should be prepared accordingly.

It is important to disseminate data from the study according to the same principles as the release of the publication itself, and therefore many of the requirements for release of the data are similar to those for publication purposes. However, in releasing data is it also important to ensure that the correct context is maintained. The context provided in terms of the entire published report should also be clearly applied to sections of the report released in the form of data or as individual summary tables.

QG10: Guidelines pertaining to actions carried out during this step:

		Action points: Handling queries	Related SASQAF indicator
10.1	**SVC Phase**	**General**	
10.1.1	Dissemination	All the requirements in terms of clearance before release, accompanying metadata, confidentiality and comments on data quality listed in terms of publication of the results, should also be followed for the release of data. *For queries and for reports published on the Internet, make sure that the information retains the correct context by providing appropriate accompanying metadata, since extracts from the report may be misinterpreted if read out of context.*	3.2, 5.9, 5.12, 5.1, 6.2, 6.3, 9.5
10.1.2	Dissemination	The types of media/channels used for providing requested data should be adequate and should preserve confidentiality. *Also plan for a way to provide the relevant metadata in conjunction with the data.*	5.3
10.1.3	Dissemination	Catalogue systems to identify available information from the survey should be made available to users, and this catalogue should be regularly updated.	5.11
10.1.4	Dissemination	Data must be made accessible to users by ensuring that the data could be provided in a variety of formats that will satisfy the requirements of all users.	5.4

10.1.5	Dissemination	Make sure that data from more than one instance of a repeat survey is provided in a comparable format.	7.2; 7.3
10.1.6	Dissemination	Preliminary and revised data should also be identified in the data.	8.5
10.1.7	Dissemination	Data that is released from a specific survey should contain references to that survey. The date on which the survey was completed/finalised, as well as the date on which the data was supplied, must be indicated on the publication.	
10.1.8	Dissemination	Publication policies should address more than just the published hard copy and should include policies pertaining to electronic copies, Internet versions of reports, queries and all output provided to users.	5.9
10.1.9	Dissemination Process?	Make sure that part of the process for providing data includes a check on the methods or codes used to extract data in response to queries. *Someone must double-check queries used to extract data.*	1.8
10.1.10	Dissemination	Put in place systems to obtain feedback on the published results from key users.	2.5
10.1.11	Dissemination	Put in place systems to log all the queries in order to keep all queries responded to, on file.	2.1???

APPENDIX A: SASQAF Data Quality Indicators

Tables 2 to 10 that follow in this subsection provide the details on the SASQAF standard to be used to assess data quality, organises into the eight quality dimensions and the prerequisites for quality. Note that the numbering convention for the SASQAF indicators is to use the dimension as the first level of the number, so that the indicator numbered 3.2 refers to the second indicator under dimension 3. The table headings list the number of the dimension along with the name of the dimension.

Each table contains an indicator number, a description of the indicator, as well as the four possible assessment levels. When a certain statistic is measured in terms of an indicator, it is placed in one of the four levels, with level 4 indicating best and level 1 lowest quality.

Table 2 Prerequisites of quality (SASQAF dimension 0)

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
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.

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
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 3 Relevance (SASQAF dimension 1)

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
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.

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
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 4 Accuracy (SASQAF dimension 2)

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
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.

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
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 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 (questionnaire effects, data collection mode effects, interviewer effects, respondent effects), processing errors (data entry error rate, coding errors, editing failure rates, imputation rates), model assumption errors, non-response errors (overall response rate, item response rate, unit non-response, weighted and unweighted response rates)	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.

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
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.
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.

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
	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.
2.6	Are data collection systems sufficiently open and flexible to cater for new developments (e.g., changes in definitions, classifications, etc)?	Data collection programmes are sufficiently robust, with changes causing minimal impact on systems.	Data collection programmes are sufficiently robust, with changes causing significant impact on systems.	Although the data collection programmes are weak, the changes would result in significant system changes (not a major overhaul).	Data collection programmes are weak, with changes requiring an overhaul of the entire system.
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	Manual and electronic techniques used for matching records are thoroughly documented.	Although incomplete, a high degree of documentation exists on manual and electronic record-matching techniques used.	Some documentation exists on manual and electronic record-matching techniques used.	Manual and electronic techniques used for matching records are not documented at all.

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
	but relate to separate entities)				

Table 5 Timeliness (SASQAF dimension 3)

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
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 are 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 6 Accessibility (SASQAF dimension 4)

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
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.
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	Statistics not routinely	Statistics not routinely	Statistics not routinely	Statistics/ administrative

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
	not routinely disseminated are made available upon request.	disseminated are always available on request; or Administrative records not routinely shared are always available on request (where a legal framework is in place)	disseminated are usually available on request; or Administrative records not routinely shared are usually available on request (where a legal framework is in place).	disseminated are occasionally available on request; or Administrative records not routinely shared are occasionally available on request (where a legal framework is in place).	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.
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 7 Interpretability (SASQAF dimension 5)

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
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.	Some deviations from acceptable standards are annotated.	Few deviations from acceptable standards are annotated.	Deviations from acceptable standards are not annotated.
		All concepts used in administrative records are well defined and documented.	The vast majority of the concepts, definitions and classifications used in administrative records are well defined and documented.	Some of the concepts, definitions and classifications used in administrative records are well defined and documented.	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 8 Coherence (SASQAF dimension 6)

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
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.
6.4	Statistics are checked for consistency with those obtained through other data sources (identify comparable datasets and incomparable ones).	Statistics are always checked for consistency with those obtained through other data sources.	Statistics are sometimes checked for consistency with those obtained through other data sources.	Statistics are rarely checked for consistency with those obtained through other data sources.	Statistics are not checked for consistency with those obtained through other data sources.
6.5	A common set of identifiers (for the purpose of	A common set of identifiers (for the purpose of	Some identifiers exist, facilitating record matching,	Some identifiers exist, but is insufficient for	No common identifiers exist

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
	record matching) exist and have been agreed upon by the data producers.	record matching) exist and have been agreed upon by the data producers.	but have not been agreed upon.	accurate record matching	

Table 9 Methodological soundness (SASQAF dimension 7)

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
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,	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.

Indicator number	Indicator description	Assessment Levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
	standard analytical methods				
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 10 Integrity (SASQAF dimension 8)

Indicator number	Indicator description	Assessment levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
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.

Indicator number	Indicator description	Assessment levels			
		Quality Statistics	Acceptable Statistics	Questionable Statistics	Poor Statistics
		Level 4	Level 3	Level 2	Level 1
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.
8.5	Choice of source data, techniques and dissemination decisions are informed solely by statistical considerations (without political interference).	Source data, techniques and dissemination decisions are informed solely by statistical considerations without any political interference.	Source data, techniques and dissemination decisions are informed by statistical considerations as well as limited political interference.	Source data, techniques and dissemination decisions are informed by statistical considerations with political interference.	Source data, techniques and dissemination decisions are informed solely by political interference.
8.6	Ethical guidelines for staff behaviour are in place and are well known to the staff (professional code of conduct).	Ethical guidelines for staff behaviour are in place, are well known to the staff and are adhered to.	Ethical guidelines for staff behaviour are in place, are known to the staff and are adhered to.	Ethical guidelines for staff behaviour are in place, are not well known to the staff and to some extent are adhered to.	Staff does not know ethical guidelines for staff behaviour.

Printed by and obtainable from the Government Printer, Bosman Street, Private Bag X85, Pretoria, 0001
Publications: Tel: (012) 334-4508, 334-4509, 334-4510
Advertisements: Tel: (012) 334-4673, 334-4674, 334-4504
Subscriptions: Tel: (012) 334-4735, 334-4736, 334-4737
Cape Town Branch: Tel: (021) 465-7531

Gedruk deur en verkrygbaar by die Staatsdrukker, Bosmanstraat, Privaatsak X85, Pretoria, 0001
Publikasies: Tel: (012) 334-4508, 334-4509, 334-4510
Advertensies: Tel: (012) 334-4673, 334-4674, 334-4504
Subskripsies: Tel: (012) 334-4735, 334-4736, 334-4737
Kaapstad-tak: Tel: (021) 465-7531