GOVERNMENT NOTICE

DEPARTMENT OF BASIC EDUCATION

No. 486 **6 June 2011**

SC008: DATA QUALITY STANDARD FOR SURVEYS

I, Angelina Matsie "Angie" Motshekga, Minister of Basic Education, after consultation with the Council of Education Ministers and in terms of section 3(4)(a) of the National Education Policy Act, 1996 (Act No. 27 of 1996), as read with section 59 of the South African Schools Act, 1996 (Act No. 84 of 1996) amend the Education Information Policy as read with paragraph 62 of the Education Information Policy published in Government Gazette No. 26710 of 27 August 2004, and amended 6 August 2010, gazette 33426, hereby publish the standard SC008: Data Quality Standard for Surveys as set out in the schedule.

MRS ANGIE MOTSHEKGA, MP
MINISTER
DATE: 25 February 2011

SCHEDULE

SOUTH AFRICAN EDUCATION INFORMATION STANDARDS

SC008

Data Quality Standard for Surveys

August 2010

basic education

Department:
Basic Education
REPUBLIC OF SOUTH AFRICA

Education Management Information Systems (EMIS)
Department of Basic Education
Private Bag X895
0001 PRETORIA

Points of Contact
Mr Enos Sello
Telephone: 012 357 3670
Email: sello.en@dbe.gov.za

Or

Ms Khabo Skosana
Telephone: 012 357 3667
skosana.k@dbe.gov.za

Table of contents

List of Tables

List of Figures

List of Acronyms

SASQAF	South African Statistical Quality Framework
QR	Quality Requirements
DBE	Department of Basic Education
EMIS	Education Management Information System
LURITS	Learner Unit Record Information and Tracking System
WBS	Work Breakdown Structure
SASAMS	South African School Administration Management System
HEDCOM	HEADS OF PROVINCIAL EDUCATION DEPARTMENTS COMMITTEE
MLI	Master List of Intuitions
EMIS Officer	The official of the education department charged with certain responsibilities regarding education management information

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 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 Basic Education (DBE). While it may be argued that DBE would not necessarily want their statistics approved as official statistics, **it would certainly be an advantage to DBE if the Department could ensure that the quality of all the DBE 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 DBE. 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 DBE 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 Basic Education and all its provincial counterparts.

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 DBE. 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 DBE.

1.4 How to use the standard

It is recommended that the SASQAF data quality indicators be used as a standard to assess existing DBE 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 DBE 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 DBE, 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 DBE context for the SASQAF indicators, as well as to provide support throughout the survey process. (This process is referred to as the statistical cycle.) 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 cycle. 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 DBE. Although the SASQAF indicators were contextualised to some extent for application by the DBE, 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 DBE.

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 DBE**. 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. Although many of the principles remain the same when data is collected from an administrative system, it is anticipated that some aspects, mentioned in this document, would change when the national Learner Unit Record Information and Tracking System (LURITS) has been successfully implemented at all (or the majority of) institutions reporting to the DBE. The Data Quality Standard on Surveys document will be updated or supplemented accordingly at such time.

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), *1st ed.*** / Statistics South Africa-Pretoria: Statistics South Africa, 2007, ISBN 0- 0-621-37108-4.

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 cycle, 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 DBE 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 DBE: 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 (ordinary schools, SNE schools, 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 EMIS 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[1] 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 EMIS. 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 Master List of Institutions 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

[1] See documents SC006, SC009 and SC010 in this regard.

	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 school) 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: Department Basic Education. Before information is referred to the DG: DBE for approval, it must be approved by the relevant staff internal to the department responsible for the survey, as well as by the 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 cycle

Any survey has to go through a number of crucial steps. For the purpose of this document, this process is called the statistical cycle 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.

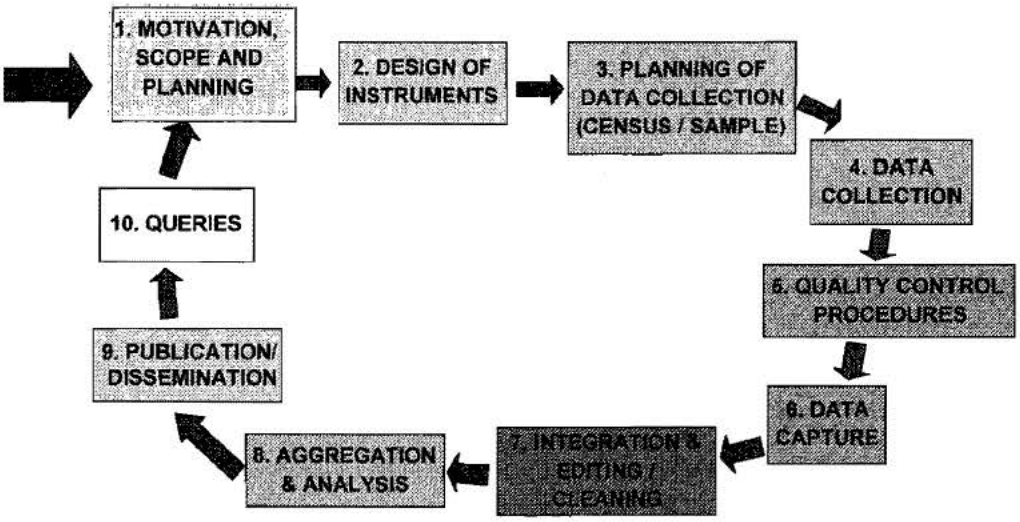

Figure 1: Steps in the statistical cycle

The guidelines are linked to the various steps in the statistical cycle, 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	**Motivation and scope**	
1.1.1	Determine whether the Department of Basic Education 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.*	0.1
1.1.2	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.*	
1.1.3	Compile a document providing: • a justification for the survey; • a motivation for the survey; • the objectives of the survey; and • the proposed scope of the survey.	
1.1.4	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).	7.2

1.2	**Planning timelines**	
1.2.1	Make a list of all the activities relevant to this specific study, which would be required at each step of the statistical cycle, 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).*	
1.2.2	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.	3.2, 3.5
1.2.3	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	Plan for interaction between the national department and the provincial departments. Allow sufficient time for different circumstances peculiar to the different provinces, which would impact on whether they are able meet certain timeframe deadlines. *For example, certain provinces have more schools in remote rural areas and so data collection takes longer than in provinces with more urban schools that are connected via the Internet.*	
1.2.5	Carefully consider timelines related to the development of software tools. This is of particular importance if the information will be collected via administrative systems *(such as SASAMS)*, 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.	
1.2.6	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?	3.1, 3.2
1.2.7	Ensure that an appropriate length of time is allowed between the end of the data collection and the date of publication of the results.	3.1, 3.2
1.2.8	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.	1.1, 1.3

1.3	**Interaction with users**	
1.3.1	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.*	1.2
1.3.2	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.	1.4, 1.5
1.3.3	Ensure that advance notice is given to users with regard to major changes in methodology, source data and/or statistical techniques.	8.3

1.4	**Planning resources**	
1.4.1	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.	0.5
1.4.2	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.	0.5
1.4.3	Put in place systems to monitor and manage the efficient use of resources in accordance with the project plan.	0.6
1.4.4	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.*	0.5
1.4.5	Check whether adequate resources have been planned for data/statistical services in the overall budget planning.	0.5
1.4.6	Determine whether adequate resources are available at provincial level.	0.5

1.5	**Ethical and governance issues**	

1.5.1	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.	8.6
1.5.2	Compile a document to describe the conditions under which policy-makers, government in particular, may have access to data before the actual release date.	8.2
1.5.3	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: Basic Education, etc.)* Adapt the project plan and timelines accordingly.	8.5

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	**Contents of data collection instrument(s)**	
2.1.1	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.	1.4
2.1.2	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.	2.6, 6.1
2.1.3	The collection instruments should take into account changes in both the education system and in policy priorities, particularly where repeated surveys are concerned.	2.6
2.1.4	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.	

	Action points: Design of instruments	Related SASQAF indicator
2.1.5	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.	
2.1.6	All relevant policies, pertaining to the objectives and contents of the survey, must be taken into account.	

2.2	**Formulation of questions and instructions**	
2.2.1	Methodologies used must follow accepted standards, guidelines or good practices (national, international, peer-agreed), with regard to questionnaire design.	7.3
2.2.2	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 to describe the objectives of the survey, as well as to ensure that the correct person completes the form.	7.3
2.2.3	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.	8.5
2.2.4	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.*	1.6
2.2.5	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.)*	
2.2.6	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	**Coordination of information collected**	
2.3.1	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 School Survey they may be added into other surveys, or questions may be added into the Annual School Survey instead of conducting a separate survey.*	
2.3.2	The collection instruments must be checked for comparability of data from one year to the next and from one province to another province.	6.1

2.3.3	Checks must be done to ensure consistency between information contained in the Master List of Institutions[2] (MLI) 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.	6.3

2.4	**Common definitions, concepts and classifications**	
2.4.1	The available common concepts, definitions and classifications[3], must be used.	6.3, 7.1
2.4.2	Ensure synchronisation between operational data systems (surveys) and administrative systems in terms of information, as well as concepts, definitions and.classification.	6.1
2.4.3	The national EMIS number must be used as an identifier and the front page, as prescribed by EMIS, must be used as far as possible for all surveys.	6.5
2.4.4	There must be a good reason for any departure from standards. Such reasons must be documented, so as to be available for metadata purposes.	6.3

2.5	**Piloting the instruments**	
2.5.1	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.	7.3
2.5.2	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.*	
2.5.3	The group on which the instruments are piloted should have similar characteristics to those of the intended sample population, but it does 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	**General**	
2.6.1	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	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.	
2.6.3	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.	
2.6.4	The survey instrument(s) must be approved by EMIS, since the Educational Information Policy designates EMIS as the custodian of educational information.	
2.6.5	The final instrument(s) must to be approved by the relevant stakeholders *(For example: approval by Head of Education Departments HEDCOM).*	

[2] See documents SC006, SC009 and SC010 in this regard
[3] See documents SC006, SC009 and SC010 in this regard

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 schools), the sample frame will to be used to identify all the schools, 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	**Sample frame**	
3.1.1	Since the Master List of Institutions (MLI) is used as a sample frame for the majority of surveys, it is important that the contents of the MLI and the procedures for updates to the MLI are confirmed for applicability to the survey. The same applies to other information related to the MLI, which provides information with regard to the selection of the sample frame.	2.5
3.1.2	The MLI is dynamic and updates at provincial level are provided quarterly to the national MLI. Information extracted from the MLI to be used as a sample frame must be date-stamped, so as to indicate when the "snap shot" of the MLI for the sample was taken. Information used for a census of all schools, which is based on the MLI, must be similarly date-stamped. *Assessments, such as response rate, must be compared to the MLI at the time of planning and not measured against changes occurring at a later stage.*	2.5
3.1.3	When using information from the MLI and other sources as a sample frame, attention must be paid to the common concepts and definitions contained in the MLI, and these must correspond to those used in the survey. *(For*	6.1

	Action points: Planning of data collection (census / sample)	Related SASQAF indicator
	example, the definition of a school as defined in the MLI and in surveys within and outside of the DBE).	
3.1.4	Non-sampling errors relating to the sample frame should be documented. *An example would be any errors found in the sample frame.*	2.2

3.2	**Sampling issues**	
3.2.1	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.	7.3
3.2.2	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.	7.3
3.2.3	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 during the scoping phase.	7.2
3.2.4	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.*	0.7
3.2.5	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.*	5.2, 4.13
3.2.6	The size of the sample must be planned by taking into account acceptable levels for sampling errors in key variables. The actual 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.*	2.1, 4.13

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	**Planning and monitoring of data collection**	
4.1.1	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.*	8.1
4.1.2	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, educator or school, which should not be released to the general public, such as personal contact numbers or fax numbers of a school. 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 MLI and data proposed to be incorporated into the learner tracking system in future.	0.4
4.1.3	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 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.*	7.3
4.1.4	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.	7.3
4.1.5	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.*	

4.2	**Timelines**	
4.2.1	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.	2.4, 3.3
4.2.2	The timeline should specify clear deadlines for the collection of data from	2.4, 3.3

	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 schools and the collection process should ensure that schools 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.*	
4.2.3	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.	2.4, 3.3
4.2.4	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 School Survey for a particular year end? How much time is allowed for such completion?*	

4.3	**Monitoring data collection**	
4.3.1	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.	0.7, 3.3
4.3.2	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.*	
4.3.3	The procedure for conducting back checks must be established and documented before the start of the data collection. If back checks 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.*	0.7, 2.5
4.3.4	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 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.)*	0.7
4.3.5	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.	2.2, 4.13
4.3.6	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.*	0.7, 2.5
4.3.7	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	2.6

	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.	
4.3.8	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	**Verifying data quality**	
5.1.1	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. 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.	0.7
5.1.2	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.	0.7
5.1.3	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.	0.2
5.1.4	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.*	0.7
5.1.5	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?*	2.3

	Action points: Quality control procedures	Related SASQAF indicator
	• *Does information match throughout the survey form?*	
5.1.6	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 schools and not just sign off that they have received the form.*	2.3
5.1.7	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.	3.3
5.1.8	Methods must be put in place to determine the existence of duplication in the data.	2.7
5.1.9	Methods must be put in place to check that institutions are identified correctly in accordance with the national EMIS number.	6.5
5.1.10	Correctness of data should be checked against other sources, where possible.	6.4

5.2	**Correcting errors in data**	
5.2.1	Processes should be put in place for providing feedback on errors in EMIS data or in the information on the MLI, as detected by other surveys.	6.4
5.2.2	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.	2.4
5.2.3	A procedure should be introduced to resolve discrepancies indicated by the back checks.	

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	**General**	
6.1.1	The data capture process should be completed within the required deadline.	3.3
6.1.2	Data capture should be planned to start as soon as the first data is	3.3

	Action points: Data capture	Related SASQAF indicator
	received, and not only once all data has been received, so as not to delay the data processing process.	
6.1.3	Data capture methodologies that are used should follow accepted standards, guidelines and good practices, and these must be consistent across the different provinces. Data capture software must also be consistent across provinces, so that consistent standards could be delivered.	7.3
6.1.4	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.	
6.1.5	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.	
6.1.6	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.*	
6.1.7	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.	7.3
6.1.8	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.	0.7
6.1.9	Data capturers should receive sufficient training.	
6.1.10	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.	
6.1.11	It should be noted that the 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 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	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.	0.4, 8.1

2.3.2.7 STEP 7: Integration and editing/cleaning of data[3]

Objectives to be reached during this step:

[4] For National Department

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 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 cycle, 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	**Integration**	
7.1.1	Procedures for integrating data from different provinces 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 DBE with information collected by the DBE.	0.3
7.1.2	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.*	3.3
7.1.3	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 DBE, 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.	6.5
7.1.4	The national EMIS number may also be used to check and confirm information obtained from a census against the current version of the MLI.	

7.2	**Editing/cleaning**	
7.2.1	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.*	0.7, 6.2
7.2.2	Data editing procedures should include a comparison of information from schools with their peers, and within the dataset as a whole. *For example,*	6.2