



The South Africa I know, the home I understand

Guideline for the Implementation of the Data Access and Publication Development Policy

Compiled by: Publication Services Division

Date: June 2015

1. Introduction

The Statistics Act (Section 7) specifies one of the responsibilities of the Statistician General as being to collect, process, compile, analyse and disseminate official statistics. Section 17 of the act requires Statistics South Africa (Stats SA) to ensure that confidentiality is retained when users are provided with statistical data in the dissemination phase.

The United Nations Economic and Social Commission, on the 24th July 2013, adopted the Fundamental Principles of Official Statistics on recommendation from the UN Statistics Division. Principle 6, in support of the Stats Act, requires that users be provided with statistical information that is confidential.

In addition, Principle 1 puts a requirement that users be provided with statistical information in an impartial manner. This, then, implies that all users should be treated equally as far as access to data is concerned.

The data access policy, as approved in February 2015, addresses the following;

- What data can be published;
- Access to unpublished data;
- Ownership of products developed from unpublished data and
- Classification of the publication

This document provides guidelines on the implementation of the areas mentioned above.

2. What data can be published?

Statistics South Africa publishes data into the public domain from Administrative records and surveys (sample and censuses) in the form of summary data sets and unit records. The summary data are table and graphs indicating time series that are included in releases or spreadsheets. The unit records are confidentialised information of the unit of observation (person, household or unit of economic activity).

The data owner or Executive Manager responsible for data set (survey or administrative record) must compile a list of variables in the data set that are available to the public without a user having to make a special request to the Statistician General for access.

Examples of the format of the lists of variables are shown in Appendix A. This list will be published in support of the requirement by the Promotion of Access to information (PAIA) and will form part of marketing activities.

3. Access to unpublished data

A situation may arise that a user requires data that is not in the public domain. This applies to the summary data or unit record. This may be as a result of the need for:

- lower geographical levels

-
- lower levels of coding in variables such as industry and occupation
 - a variable that has not been published and is on the list of variables mentioned in section 2.
 - combination of variables that are not in the summary data set that is in the public domain.
- Examples of such summary are the community profiles from the census data sets.

The user in this instance applies to both external and internal persons. The internal user includes those persons who do not have access to the data set in performing their day to day duties. Even when they do, the variable they work on should be on the list of publishable variables. E.g. a survey statistician has access to a data set. She then decides, in her private capacity, to do a paper or article based on the data. If the paper includes one or more variables that are not on the list of publishable variables, then she should ask for permission for access.

3.1. Summary data

The summary data bases, also known as cubes or marts, may yield cells that have frequency tabulation yields a small number. This is more likely to happen in instances where the census data is tabulated at low geographical units such as sub-place. In the case of sample surveys, this is not likely to occur as the geographical unit of dissemination is national or province. Should a situation arise that a frequency in a cell is 5 or less, then a perturbation algorithm has to be performed on the summary data base. The purpose of the perturbation is to enhance confidentiality in the summary data base.

Where actual values, such as revenue or income, are published then a perturbation is not required irrespective of the value in the cell.

3.2. Unit records

A unit record contains information about the unit of observation. This information is defined by variables or data items contained in the form (survey questionnaire or administrative record).

The practice at Stats SA is that the entire set of unit records in our household sample surveys and the administrative records such as causes of death are published. For Housing and Population Censuses, only 10% of the unit records of the households, and all members of those households, are published. The variables that may compromise confidentiality are removed from these unit records. In addition, the variables with hierarchical coding such as industry and occupation are published at higher levels.

The unit records from the business sample surveys (economic statistics) are not published. What gets published are indices, e.g. consumer price index, and aggregate values of variables such as revenue and value of sales.

In considering a request for access to unpublished data, The Statistician General will determine whether the risk of providing the unit records will result in compromising confidentiality.



The South Africa I know, the home I understand

If the SG determines that there is a high risk and there is compelling motivation for access to the requested data, then permission should be granted with a proviso that access will only be had on the premises of Stats SA. This will result in the user not being provided with the copy of the data set but rather the results from statistical processing of the unit records.

On the other hand if the risk is deemed to be low, then the user will be granted permission to have a copy of the requested data set and can access the data outside the premises of Stats SA. The unit records will, of course, be confidentialised.

4. Classification of Publications

The entry level of most posts in the core functional areas is a graduate qualification. In some instances our intake on interns is postgraduate level. On the other hand Stats SA has a wealth of data from household surveys, business surveys and administrative records. The qualification of staff and the availability of data provide an opportunity for our members of staff to engage in research as well.

Stats SA has entered into the era of seeking to promote the writing and publishing of research papers by its staff members. The intention is to create time for the internal staff to engage in research work. This is in addition to the production of the mandatory release and reports that are published in fulfilling our work programme as a national statistics office.

The copyright law in South Africa, Section 5(2) of the Copyright Act of 1968 (as amended), dictates that if work is done under the control and funding of a particular government institution, then copyright subsists in that government institution.

The following classification is intended to facilitate the balance between Stats SA's publication and those for which assistance is provided to members of staff.

4.1. Structural publications

These are publications that are produced in fulfillment of our work programme. This includes corporate publications, newsletters and promotional material. Individuals will not be acknowledged as authors in these publications.

4.2. Specialised publications

Research reports or papers that members of staff have written as an individual or a group. These will not be part of our work programme and are part of our structural publication and may be. While these may be published under the Stats SA brand, it is in these publications that authors will be acknowledged individually.

4.3. Strategic publications

Researched organisational papers or reports that are intended to either to guide the organisation or influence targeted stakeholders. Most, if not all, of the strategic publications will remain internal



The South Africa I know, the home I understand

documents. Authors may or may not be mentioned individually. The branding will be that of Stats SA and may incorporate that of the strategic partner/s.

5. Other Incidental Items

5.1. Citation or acknowledgement

Stats SA publishes about 255 releases and reports every year. There are in addition, 192 data sets that we publish on a yearly basis. Users who use the statistical information in their reports and papers have to acknowledge Stats SA as the source of such information. The format of the citation is contained in the Report Standard

5.2. Permission granted to publishers

Publishers of school and university textbooks have approached Stats SA to request permission to include published material such as tables and graphs. Such requests have to be forwarded to the Manager: Publishing Services, who will issue the appropriate permission.



The South Africa I know, the home I understand

Appendix A: List of Variables to Published

Name of Dataset		Quarterly Labour Force Survey	
Variable Name	Unit of Measure	Level of coding	No of Decimal Places
Age	Completed years	NOT Applicable	0
Occupation	Not Applicable	1	Not applicable
Person Weight	Not Applicable	Not Applicable	10

Name of Dataset	Annual Financial Statistics		
Variable Name	Unit of measure	Level of coding	No of Decimal Places
Advertising Spend	Rands (000 000.00)	Not Applicable	0
Nature of Industry	Not applicable	1	Not applicable



The South Africa I know, the home I understand

Appendix B: Request for access to unpublished data

Surname	Name	Identity or Passport		Signature
		Number	Country of Issue	

Organisation:

Data Set:

Reference Period:

List of Variables required:

Data requested for the purpose of:

Recommended/not recommended

Executive Manager: _____

Date:

Recommended/not recommended

Deputy Director General: _____

Date:

Approved/Not approved

If approved, copy of the data to be given to the user/access to data on Stats SA premises.

Statistician General: _____

Date: