South Africa Microdata Scoping Study

Helen Barnes, Michael Noble, Chris Dibben, Charles Meth, Gemma Wright and Lucie Cluver

Working Paper No 6

Centre for the Analysis of South African Social Policy
University of Oxford

Working Paper No 6

South Africa Microdata Scoping Study

By Helen Barnes, Michael Noble, Chris Dibben, Charles Meth, Gemma Wright and Lucie Cluver

Centre for the Analysis of South African Social Policy
Department of Social Policy and Social Work
University of Oxford, England

January 2007

This working paper will be published in due course by the Economic and Social Research Council following the Foundation Conference of the International Data Forum, Beijing, 5-7 June 2007, where this study will be presented.

Contents

1 Introduction....1
2 Census and survey data....2
2.1 Statistics South Africa....2
2.2 Human Sciences Research Council....3
2.3 Data archives....5
2.3.1 South African Data Archive....5
2.3.2 Data Management and Information Delivery....6
2.4 Data catalogues....7
2.4.1 DataFirst....7
2.4.2 HRD Data Warehouse....8
2.5 Key datasets....8
2.5.1 Demography....9
2.5.2 Housing....9
2.5.3 Social welfare....9
2.5.4 Economy....10
2.5.5 Labour market....10
2.5.6 Education....10
2.5.7 Transport....11
2.5.8 Crime....11
2.5.9 Health....11
3 Administrative data....12
3.1 Overview of administrative data....12
3.2 Administrative data in South Africa....13
3.3 Key datasets....14
3.3.1 Demography....14
3.3.2 Housing....15
3.3.3 Social welfare....15
3.3.4 Economy....16
3.3.5 Labour market....17
3.3.6 Education....18
3.3.7 Transport....20
3.3.8 Crime....20
3.3.9 Health....21
3.3.10 Children....24
4 Conclusion....24
4.1 Summary....24
4.2 Recommendations....25
Appendix 1-4....27
Glossary....104
References....106

Acknowledgements

CASASP would like to thank all those who kindly contributed information for this report. A full list is provided in Appendix 2.

1 Introduction

This report, by the Centre for the Analysis of South African Social Policy (CASASP) at the University of Oxford, provides information on the availability of social and economic microdata resources in South Africa, which are not currently accessible to researchers. A broad definition of accessibility - or lack of access - is adopted. As highlighted in the *Review of International Data Resources and Needs*, access to data depends to a certain extent on knowledge of what is actually available and where to look. This report attempts to document the main data holders and data sets that are available in South Africa, both those that are easy to access and those where access may need to be negotiated. In both cases, procedures for access are given where possible. This report will guide the Economic and Social Research Council (ESRC) in undertaking, for South Africa, two key recommendations from the *Review*, namely to 'acquire, document and preserve national microdata resources from other countries' and to 'provide improved information on access procedures and increased support for researchers attempting to negotiate international access procedures' (Desai and Cowell, 2006: 54).

A range of officials in government departments and other organisations carrying out social and economic research were interviewed for this scoping study, and this added to existing knowledge of microdata in South Africa from projects currently being carried out by CASASP.

Microdata are data about individual objects (such as persons, companies, events, transactions). Objects have properties which are often expressed as values of variables of the objects. For example, a 'person' object may have values of variables such as 'name', 'address', 'age', 'income'. Microdata represent observed or derived values of certain variables for certain objects. National microdata is usually available from censuses, surveys and administrative and register data. These data are most commonly collected by the national government or statistical office and access provided by the statistical office or the national archive. The data are collected at an individual, household, or institution level as appropriate (Desai and Cowell, 2006).

Although the focus of this report is on microdata, the availability of **macrodata** is also briefly discussed. Macrodata are data aggregated to a country or regional level. Macrodata are estimated values of statistical characteristics concerning sets of objects (or 'populations'). A statistical characteristic is a measure that summarises the values of a certain variable of the objects in a population (Desai and Cowell, 2006). For this scoping study, macrodata are of interest as the context or background to the microdata sources.

In **Section 2**, Census and survey data are discussed in broad terms. A list of the main sources of data in the principal areas of interest to the ESRC is given in both sections, with a full account of the nature, utility, conditions and costs of the data provided in **Appendix 1**. Administrative data is reported on in **Section 3**. In both sections, two key questions are addressed: how easy is it for researchers to access data at present and what are the main channels of access? In **Section 4**, the findings of the report are summarised and recommendations for the ESRC to progress data access are discussed.

While every effort has been made to provide a comprehensive account of the availability of social and economic microdata in South Africa, it is inevitable that some datasets and information will have been unintentionally missed out. It is recommended that this document is regularly updated and therefore any additions or amendments could be incorporated. This report does not, on the whole, cover international datasets with a South Africa component, for example, surveys conducted by the World Health Organisation and other international institutions.

2 Census and survey data

2.1 Statistics South Africa

Statistics South Africa (Stats SA) is the national statistics agency in South Africa and is mandated to collect and process data and produce official statistics. Stats SA's mission is to provide a relevant, timely, reliable and accurate body of statistics to inform users on the dynamics in the economy and society through the application of internationally acclaimed practices. Stats SA seeks the broadest possible dissemination of the statistical data it collects, and the services it offers.

Stats SA produces a variety of statistics (macrodata) and microdata. In terms of macrodata, areas of interest include demography, health and vital statistics, national accounts, labour market, employment, industry and trade, prices, public sector spending, private sector finances and transport. The main microdata sources are the Population Census and household surveys such as the Income and Expenditure Survey and General Household Survey.

Publications containing the statistics produced by Stats SA are listed in a catalogue of publications which is updated bi-monthly and issued six times per year[1]. Many publications are available free of charge or can be downloaded from the Stats SA website[2]; for others there is a small charge. Certain statistics can be accessed from the homepage of the Stats SA website, and for some data sources (e.g. 2001 Census) it is possible, from the website, to compile tables according to your own specifications. The annual *Statistics in Brief* publication presents a wide range of the latest official statistical information in a pocket-size booklet.

Some organisations take Stats SA's data and make it available as macrodata in a user friendly format. One such example is the newly launched Data Corner, created by the Development Policy Research Unit at the University of Cape Town (UCT)[3], which provides labour market and social and poverty indicators for various time points between 1995 and 2004. The data on employment, unemployment, the labour force, labour force participation rates and household access to services largely comes from household surveys conducted by Stats SA.

Stats SA can be contacted directly for publicly available datasets, although as will be explained in **Section 2.3.1**, many of Stats SA's datasets can be accessed from the

[1] See http://www.statssa.gov.za/orderpublications/OrderPublications.pdf for the latest release.
[2] See http://www.statssa.gov.za or http://www.statssa.gov.za/publications/findpublication.asp to search for publications.
[3] http://www.commerce.uct.ac.za/Research_Units/DPRU/DataCorner.htm

national data archive. Under section 7 (3) (d) of the South African Statistics Act 1999, most datasets are provided free of charge from Stats SA; the only charges are those required to cover the costs of the medium of dissemination (i.e. CD or DVD). Stats SA states that 'As a general principle, Stats SA does not seek to recover any of the costs of data collected, products developed or standard services provided, as those costs are met from an allocation voted by Parliament' (Statistics South Africa, 2002: 1).

Once a copy of a dataset has been secured, it is possible to disseminate the data further, providing no charge is made and Stats SA is acknowledged as the supplier and owner of the data and copyright.

For Stats SA data that is not publicly available, access would need to be negotiated with the Statistician General. An example of an institution which has negotiated access to specific datasets for their own research is the African Census Analysis Project at the University of Pennsylvania[4]. They are not allowed to disseminate this data more widely however.

An explanation of Stats SA's organisational structure is given in **Appendix 3**. This is useful for finding out quickly which division of Stats SA is responsible for a particular type of data.

2.2 Human Sciences Research Council

The Human Sciences Research Council (HSRC)[5] is a statutory body which was established in 1968. It is the premier social science research organisation in South Africa, and has carried out various national and sub-national surveys. According to the HSRC website[6], the HSRC 'supports development nationally, in the Southern African Development Community and in Africa. It primarily conducts large-scale, policy-relevant, social-scientific projects for public-sector users, non-governmental organisations and international development agencies, in partnership with researchers globally, but specifically in Africa. Over the last couple of years the HSRC underwent major restructuring, aligning its research activities and structures to South Africa's national development priorities: notably poverty reduction through economic development, skills enhancement, job creation, the elimination of discrimination and inequalities, and effective service delivery. The HSRC also seeks to contribute to the research and development strategy of the HSRC's parent Department of Science and Technology, especially through its mission to focus on the contribution of science and technology in addressing poverty […] It recognises that if its research is to make a significant difference to aspects of national life, it must be responsive to the needs of users and relevant to topical social issues, while at the same time advancing a theoretically informed understanding of society in all its aspects.' An explanation of the HSRC's organisational structure is given in **Appendix 3**.

The Knowledge Systems (KS) unit within HSRC conducts primary and secondary research on socio-economic and governance issues generally, and the national system of innovation specifically, that enables evidence-based decision-making by its users.

[4] http://www.acap.upenn.edu/
[5] http://www.hsrc.ac.za
[6] http://www.hsrc.ac.za/about/mission/index.html

KS works both with external users and the research programmes of the HSRC. The key sections of KS are the Centre for Science, Technology and Innovation Indicators (CeSTII); the Socio-Economic Surveys (SES) unit; and the Geographic Information Systems (GIS) centre. CeSTII is not particularly relevant to this scoping study, but the SES unit is discussed further below and the GIS centre in **Appendix 1**. The principal activities of these sections involve the design, implementation and analysis of quantitative and qualitative data.

The SES unit 'endeavours to develop innovative survey techniques, statistical methodologies and data management processes that are relevant to social science research. In addition, it aims to lead and/or support HSRC research initiatives in the social sciences locally and on the African continent'. SES also provides services to external clients on a contract basis. All research contracts must fall within the HSRC's strategic framework. Selection criteria for projects are that the project must be large-scale, national or regional, and be multi-disciplinary in nature. SES always tries to collaborate with relevant research programmes within HSRC on subject-specific projects and only does small-scale projects if they are of strategic value to the unit or the HSRC. While SES is the focal point of national surveys in the HSRC, some research programmes do undertake their own (predominantly sub-national) surveys which are driven by both internal and external agents.

Although the HSRC conducts high quality research, the majority of the datasets are not officially in the public domain. HSRC public opinion surveys from the 1990s (the Omnibus and National Survey) are housed in SADA, but with the exception of the 2001-02 Migration Survey, no HSRC datasets have been deposited in SADA since 2000. There is a new drive within the HSRC for data to be released as soon as possible after data collection, cleaning and preliminary analysis. As such, the HSRC is currently in discussions about a way of releasing the data through Stats SA, with the possibility of Stats SA acting as the official archive of HSRC data. There is also a plan to establish a data repository for HSRC data, but this is still at an early stage. A pilot project to make some data available to the social sciences community is underway within the Information Technology Services directorate, and one of the priorities for 2007 is to make more data available. As there are data from the 1970s onwards, it will be a huge task, so it is likely that they will start with the most recent data and then concentrate on data which could be used in time series.

In addition to the HSRC, there are various other organisations which carry out research for clients. One such example is the Community Agency for Social Enquiry (CASE). CASE is an applied research non-governmental organisation (NGO) which was established in 1985 and operates nationally. CASE undertakes research in a wide range of areas, covering different aspects of the economy, society and politics in South Africa and the southern African region. Clients include national and provincial government departments, international agencies, parastatals, NGOs and community based organisations. Work includes national, provincial and local surveys, focus groups, project and organisational evaluations, programme impact assessments, policy and literature reviews. Most of the datasets belong to the clients for whom the work has been carried out, but there are some that could be made available to the public.

2.3 Data archives

2.3.1 South African Data Archive

Currently the majority of Stats SA's datasets are placed in the South African Data Archive (SADA) at the National Research Foundation[7]. SADA is a broker between various data providers and the research community. It safeguards datasets and related documentation and attempts to make them easily accessible.

SADA's objectives are[8]:

- To acquire and catalogue survey data and related information.
- To preserve such data against technological obsolescence and physical damage.
- To provide originators or depositors of data with necessary information in order to ensure high standards of data documentation.
- To re-disseminate such information for use by other researchers, for re-analysis of data, longitudinal and comparative studies, research training, teaching and policy-making decision purposes.
- To formulate policies for the scope and content of data and data preservation.
- To promote the optimal use of data.

SADA is a member of organisations such as the International Federation of Data Organisations, the Council of European Social Science Data Archives (of which it is an associate member), the Inter-University Consortium for Political and Social Research and the International Association for Social Science Information Service and Technology, and so can quite easily send data and information stored in its databases to interested researchers worldwide. The Networked Social Science Tools and Resources (NESSTAR) system has been implemented at SADA, in order to make the catalogues and data available through the internet[9].

SADA's data include Censuses and household surveys, demographic and health related studies, substance abuse, crime, income and poverty, labour and business, education and training and political perceptions and attitudes. A sample record from the warehouse is provided in **Appendix 4**. Datasets can be requested by researchers worldwide by filling in an online form. The data (in SPSS format) and accompanying documentation is then provided either by File Transfer Protocol (FTP) or sent on a CD. Within South African working hours, FTP will be within one hour. All datasets are currently provided free of charge (even those sent by CD).

It is not a comprehensive data archive for all social and economic related datasets in South Africa. However it contains a large majority of official government surveys. There is an agreement between Stats SA and SADA that Stats SA data will be deposited in the archive. There is a time lag in this process before passing it on to SADA to disseminate. There is however no standard time between Stats SA releasing the data and it being given to SADA. SADA does actively look for new datasets to add to the archive, but due to staff shortages, this is perhaps not carried out as fully as

[7] http://www.nrf.ac.za/sada
[8] http://www.nrf.ac.za/sada/introduction.html
[9] The NESSTAR system allows users of social science data resources to search for, analyse, bookmark and download data from several data archives and data providers over the Internet. NESSTAR can be described as a virtual data library offering global access to locally supported holdings.

possible. There also appears to be some reluctance to deposit data in the archive; the reasons for this are not entirely clear, but include concern over intellectual property, desire to keep track of data users, and restrictions imposed by project funders. It is certainly the case that recently (since 2002) the only datasets deposited in the archive are from Stats SA.

SADA has recently compiled a directory of organisations in South Africa that produce data[10]. These are organisations where data is available for use by researchers outside of the organisation concerned, rather than a list of all data producers. To facilitate access by individual researchers, SADA has provided details for the key contact within each organisation.

2.3.2 Data Management and Information Delivery

A new initiative within Stats SA to build a data and metadata repository is currently underway. This repository - End to End Statistical Data Management Facility (ESDMF) - will house all Stats SA data, gathered together from the various divisions who are the current data custodians, accompanied by metadata and documentation. The immediate aim is to standardise all datasets and metadata and provide a declaration of quality. Data quality will be assessed on eight dimensions: relevance, accuracy, timeliness, accessibility, interpretability, coherence, methodological soundness and integrity[11]. Initially access will be for researchers within Stats SA only, but eventually will be made available for researchers in South Africa and internationally.

The ultimate aim is to gather together data from government departments who are partners in the National Statistics System (NSS)[12] and house it centrally in a standardised way. Currently, NSS partners are Stats SA, the Departments of Health, Education and Home Affairs, and three provincial governments (Gauteng, KwaZulu-Natal and Limpopo).

The data available will in general be aggregated statistics only. It will be based on unit records but the public will not be able to access this microdata due to very strict confidentiality legislation. The government departments will still be the data custodians and so access to micro datasets (unit records) would need to be negotiated with the originating department, as is currently the case (see **Section 3.3**). The aggregate data will be accessed by researchers through an internet interface, but they will also be able to run some of their own analyses with front end tools. Researchers will also be able to submit special requests if they need more advanced analysis.

---

[10] See http://stardata.nrf.ac.za/html/fastlinkDatap.html for the directory.

[11] Further details can be found on the Stats SA website:
http://www.statssa.gov.za/news_archive/Statistics_SA_Data_Quality_Policy.pdf
http://www.statssa.gov.za/news_archive/Statistics_SA_Statistical_Quality_Assessment_Framework.pdf

[12] The NSS is a partnership of users, producers and suppliers of national statistics which will provide a framework for producing statistical outputs that meet user needs. The system will have more than one producer of legally recognised national statistics. Thus, in practice, each agency will be expected to produce its own statistics to meet its mandate. However, production will be centrally coordinated under the direction of the Statistician-General who will make sure that standards are met and that there is no duplication in the production process.

Some samples of Stats SA data (for example the 10% sample of the 2001 Census) will be made available.

The ESDMF is in its early stages, but the Data Management and Information Delivery (DMID) team responsible for building the ESDMF is keen to actively engage with potential users at this point in the process: in building the interface and developing templates for the metadata and quality declarations.

2.4 Data catalogues

2.4.1 DataFirst

DataFirst[13] at UCT is a survey research facility where clients can access data from surveys on South and Southern Africa. The unit provides information and training related to these social surveys. DataFirst has a collection of datasets from all major South African surveys for use by researchers at the university. These can be accessed in the unit's Resource Centre, which is also open to researchers from outside the university (at a cost of R100 – approximately £10 - per day). The exception to this is studies carried out by UCT, for which there are no copyright restrictions and datasets and accompanying documentation can be downloaded from the website free of charge.

DataFirst's goals are[14]:

- To act as a portal to digital resources and specialised published material available to users internationally.
- To facilitate the collection, exchange and use of datasets on a collaborative basis
- To provide basic and advanced training in data analysis.
- To develop and operate a web portal to promote the dissemination of data and research output via the internet.

For researchers outside South Africa, DataFirst will principally be a data catalogue facility. Documentation for a wide range of surveys, many of which are also housed in SADA, can however be downloaded from the website. There are also online bibliographies of publications which have made use of the survey data, which can be accessed via the DataFirst website.

Both DataFirst and SADA archive datasets from South African surveys, although DataFirst's main focus is on social survey data, while SADA has a broader collection. However, the main difference between the two archives is their purpose. SADA has a preservation and distribution role, whereas DataFirst is tasked with encouraging and facilitating secondary analysis of survey data, and promoting quality survey research in the country (by educating researchers in the accurate analysis of this data, and providing feedback to primary investigators of future surveys with respect to data quality issues).

The Mellon Foundation has just awarded a grant to DataFirst to clean some of Stats SA's datasets including the six October Household Surveys, the six-monthly Labour

[13] http://www.datafirst.uct.ac.za
[14] http://www.datafirst.uct.ac.za/about.html

Force Surveys, Census data from 1996 and 2001, Income and Expenditure Surveys, and special one-off surveys such as Stats SA's Time Use Survey. The team will look at how the changes in questionnaire and sampling design affect survey results and what kinds of corrections can be made to make the data comparable over time. The project will take three years to complete and the findings will be made available through the DataFirst website, with recommendations on how the datasets can be used and what their respective strengths and weaknesses are. DataFirst will also hold workshops on the shortfalls of the various datasets, teaching researchers techniques to analyse the information.

2.4.2 HRD Data Warehouse

The HSRC has a data catalogue – the Human Resources Development (HRD) Data Warehouse - which contains information about data in the broad areas of education, employment and skills. It was established in 2003 by the HSRC through a grant from the then Department of Arts, Culture, Science and Technology, now the Department of Science and Technology. It was developed from a study by the HSRC on HRD, which defined key education and training, labour market and macro-economic indicators in South Africa. Socio-economic and political contexts, as well as the demand, supply and scarcity of skills were examined to assess HRD trends. The data warehouse is a web-based information and data storage of the resources used in the study, which aims to provide an information resource to assist government in their HRD decision-making and planning. It is also useful for researchers as a source of information on available education, employment and skills data.

The warehouse provides information on the datasets (name, year, format and description, including variables, sampling design etc.), relevant reports and statistics, and details on how to gain access. A sample record from the warehouse is provided in **Appendix 4**. No actual datasets are held in the warehouse however. Dataset information includes some of the HSRC's own surveys, but also some from government departments and other organisations.

Although useful, the database only contains information on the datasets that were used in the HRD review and has not been added to since. Without regular updates, it will soon become an out-of-date resource.

2.5 Key datasets

The main Census and survey datasets that have been identified as easily accessible in each of the key areas of interest are listed below. The organisation listed as the principal investigator is not necessarily the data holder. Full details of the datasets are provided in **Appendix 1**. Attitude surveys are also included in the Appendix and there are a number of other datasets, for which full details were not available, which are listed in a table in **Section A1.4** of the Appendix.

2.5.1 Demography

Data source	Year(s)	Principal investigator
Africa Centre Demographic Information System	Began 2000	Africa Centre for Health and Population Studies
Agincourt Health and Demographic Surveillance System	Began 1992	Agincourt Health and Population Unit
Birth to Twenty	Began 1990	University of the Witwatersrand Medical School
Census of Population (10% sample)	2001, 1996	Stats SA
Demographic and Health Survey	1998	Department of Health
General Household Survey	2002-2005	Stats SA
October Household Survey	1994-1999	Stats SA
South African Migration and Health Survey	1999/2000	Centre for Population Studies, University of Pretoria
The 2001-2002 HSRC Migration Survey	2001/2002	HSRC

2.5.2 Housing

Data source	Year(s)	Principal investigator
Census of Population (10% sample)	2001, 1996	Stats SA
All Media Products Survey	1998-2005	South African Advertising Research Foundation
General Household Survey	2002-2005	Stats SA
KwaZulu-Natal Income Dynamics Study	1993, 1998, 2004	School of Development Studies, University of Kwazulu-Natal
October Household Survey	1994-1999	Stats SA
Rural Survey	1997	Stats SA

2.5.3 Social welfare

Data source	Year(s)	Principal investigator
General Household Survey	2002-2005	Stats SA
KwaZulu-Natal Income Dynamics Study	1993, 1998, 2004	School of Development Studies, University of Kwazulu-Natal

2.5.4 Economy

Data source	Year(s)	Principal investigator
Census of Population	2001, 1996	Stats SA
Financial Diaries Project	2003/2004	Southern Africa Labour and Development Research Unit, UCT
Greater Durban Metropolitan Area Large Manufacturing Firms Survey	2002/2003	School of Development Studies, University of Kwazulu-Natal and World Bank
Income and Expenditure Survey	1995, 2000	Stats SA
KwaZulu-Natal Income Dynamics Study	1993, 1998, 2004	School of Development Studies, University of Kwazulu-Natal

2.5.5 Labour market

Data source	Year(s)	Principal investigator
Cape Area Panel Study	Began 2002	Southern Africa Labour and Development Research Unit, UCT
Census of Population (10% sample	2001, 1996	Stats SA
General Household Survey	2002-2005	Stats SA
Khayelitsha/Mitchell's Plain Survey	2000	Southern Africa Labour and Development Research Unit, UCT
KwaZulu-Natal Income Dynamics Study	1993, 1998, 2004	School of Development Studies, University of Kwazulu-Natal
Labour Force Survey	2000-2006	Stats SA
October Household Survey	1994-1999	Stats SA
Rural Survey	1997	Stats SA
Survey of Activities of Young People	1999	Stats SA
The 2001-2002 HSRC Migration Survey	2001/2002	HSRC
Time Use Survey	2000	Stats SA

2.5.6 Education

Data source	Year(s)	Principal investigator
General Household Survey	2004	Stats SA
Cape Area Panel Study	Began 2002	Southern Africa Labour and Development Research Unit, UCT
Census of Population	2001, 1996	Stats SA

(10% sample)		
General Household Survey	2002-2005	Stats SA
KwaZulu-Natal Income Dynamics Study	1993, 1998, 2004	School of Development Studies, University of Kwazulu-Natal
October Household Survey	1994-1999	Stats SA
Student Choice Behaviour Project Phase 1	2001	HSRC
Survey of Activities of Young People	1999	Stats SA
Survey of Independent Schools	2002	HSRC
Survey of Private and Further Education in South Africa	2002	HSRC
Time Use Survey	2000	Stats SA

2.5.7 Transport

Data source	**Year(s)**	**Principal investigator**
All Media Products Survey	1998-2005	South African Advertising Research Foundation
Census of Population (10% sample)	2001, 1996	Stats SA
National Household Travel Survey	2003	Department of Transport
Time Use Survey	2000	Stats SA

2.5.8 Crime

Data source	**Year(s)**	**Principal investigator**
All Media Products Survey	1994-2005	South African Advertising Research Foundation
National Victims of Crime Survey	1998	Stats SA
National Victims of Crime Survey	2003	Institute for Security Studies
National Youth Victimisation Survey	2005	Centre for Justice and Crime Prevention

2.5.9 Health

Data source	**Year(s)**	**Principal investigator**
Agincourt Health and Demographic Surveillance System	Began 1992	Agincourt Health and Population Unit
Africa Centre Demographic Information System	Began 2000	Africa Centre for Health and Population Studies

Birth to Twenty	Began 1990	University of the Witwatersrand Medical School
Cape Area Panel Study	Began 2002	Southern Africa Labour and Development Research Unit, UCT
Census of Population (10% sample)	2001, 1996	Stats SA
Demographic and Health Survey	1998	Department of Health
General Household Survey	2002-2005	Stats SA
HIV and Sexual Behaviour Among Young South Africans	2003	Reproductive Health Research Unit, University of the Witwatersrand
KwaZulu-Natal Income Dynamics Study	1993, 1998, 2004	School of Development Studies, University of Kwazulu-Natal
National Food Consumption Survey	1999	Department of Health
National HIV and Syphilis Antenatal Sero Prevalence Survey	1991-2005	Department of Health
South African Migration and Health Survey	1999/2000	Centre for Population Studies, University of Pretoria
South African National HIV Prevalence, HIV Incidence, Behaviour and Communication Survey	2002, 2005	HSRC
Time Use Survey	2000	Stats SA

3 Administrative data

3.1 Overview of administrative data

Administrative data is defined here as data collected for administrative purposes and not primarily intended for research or statistical purposes. In the UK, such data has been collected for a very long time, and often formed the basis for traditional statistical information. The latter part of the 1980s saw significant steps in computerisation of administrative data. This resulted in a more systematic method of creating individual records, standardising definitions and collating data centrally. Initially, this facilitated service delivery by providing routine reports and audit trails as well as the necessary computer programming languages to interrogate the databases. Administrative data is now routinely extracted for research purposes, using these very same reporting tools. South Africa is behind the UK in its collection of administrative data, but there are some systems in operation that could and do provide data for research purposes.

The strengths of administrative data include:

- It is already being collected and therefore there are no additional costs of collection (although there will be costs of extraction and cleaning).
- It is regularly (sometimes continuously) updated.
- It can (in theory) relate to a very recent time period.
- It can also (if historical extracts are retained) provide information about past periods of time that pre-date the formation of the research question.
- If it forms part of a national system it is likely to be collected in a consistent way.
- It is subject to rigorous quality checks.
- The collection process is not intrusive to the target population (although there are data protection issues).
- Administrative systems will contain 100% of the records in question, and if adequately geocoded, allow small area analyses to be undertaken.

The very reason for the existence of administrative data is also in effect its major weakness.

- The information collected is restricted to data required for administrative purposes. In the case of social security data, although demographic and financial information relevant to the claim is held, there is no information on extraneous factors such as educational qualifications. People who do not take up the social security benefit do not form part of the administrative record.
- If there is a change in the administrative procedures (e.g. a change in entitlement) then the definition of the information collected could change. This in turn could make comparisons over time problematic.
- Administrative data may be subject to (some) fraudulent claims.

However these problems have to be offset against the cost of data collected by other means, particularly surveys with relatively poor response rates.[15]

3.2 Administrative data in South Africa

It was apparent from interviews with various government departments that it is difficult to gain access to administrative data. Government departments appear to have difficulty obtaining datasets from other government departments as there seems to be a fear that the data will be used to monitor performance or that the data collection and processing will be criticised. It is not clear whether such barriers would exist for non governmental researchers trying to access the data. In any case, researchers would, in most instances, have to negotiate access on a project by project basis. The general impression was that researchers would need to demonstrate that the research will be beneficial to the work of the organisation, rather than some purely academic exercise. However, this may not be true of all departments/organisations, and some, for example the Department of Education, do have datasets that are made available to bona fide researchers.

Statistical disclosure control (SDC) concerns safeguarding the confidentiality of the information that is held about people and organisations. While SDC techniques are usually applied to Census and survey data before release, such concerns have rarely

[15] For a more in depth discussion of administrative data see Smith and Noble (2000) and Smith et al. (2004).

been addressed with regard to administrative data in South Africa. This is largely because the potential of administrative data as a resource for researchers has not been realised and so the data have not been made available. In some of the discussions with government departments it was clear that the release of individual level data was considered impossible as confidentiality had to be maintained, and the possibility of applying measures to reduce the risk of identifying individuals was not entertained.

Macrodata based on administrative systems is quite widely available in reports or statistical releases published by government departments and Stats SA. Generally these can either be downloaded from the website or details of how to obtain the publication are given on the website. Some departments also publish statistics on their website (not in a report). A list of government departments and websites is given in **Appendix 2**.

3.3 Key datasets

The key administrative datasets that have been identified are detailed below, with some information on the specific situation with regards to data access in the government department or institution.

3.3.1 Demography

The main source of demographic administrative data in South Africa is that collected by the Department of Home Affairs. For example, the Population Register is a central computerised database which maintains a life profile of South African citizens. Records are updated on an on-going basis. Information is collected on births, civic marriages, customary marriages, deaths, adoptions, and naturalisation and resumption of citizenship[16]. However, such data is only released to Stats SA for analysis and public release as national and provincial statistics.

Another project, the Home Affairs National Identification System (HANIS) seeks to complement this demographic information with citizen identification data. The Department has always used a manual fingerprint record system to classify individual records mapped to the corresponding computerised demographic data. The fingerprint data on the manual cards is maintained by experts who have physical access to the cards whenever necessary. As the population grows however, so the number of cards grows. The Department therefore implemented an automated identification system based on the same fingerprint data. HANIS consists of an Automated Fingerprint Identification System and an ID card (with a single unique identity number which is used as a key to identify individuals on numerous systems that are used within the public and private sectors). However, it seems that such data are only released to Stats SA for analysis and public release as national and provincial statistics. The same is true of administrative data on divorces from the Department of Justice and Constitutional Affairs.

[16] See http://home-affairs.pwv.gov.za/forms.asp for the forms used by the Department of Home Affairs to collect information.

Stats SA does make available death notification data (currently for 1997-2003). It is understood that this is individual level data, but repeated attempts to find out the exact nature of the data have been unsuccessful.

It is worth noting that the current South African population register lacks various useful pieces of information. It does not record all non-citizens in the country and does not include some important demographic characteristics, such as population group, place of birth or place of residence. The records of addresses are also not regularly updated. In some cases, only postal addresses are recorded (which can be quite different to physical addresses), which limits attachment of information obtained to geographical area. It would therefore not be the most useful dataset for demographic analysis below national or provincial level.

3.3.2 Housing

The main datasets used by the Department of Housing are the 2001 Census, the Housing Subsidy System (HSS) database and GIS data. The housing subsidy is the primary assistance measure of the National Housing Programme. Households with an income of R 0 to R 3,500 (approximately £0 - £350) per month, who have not owned property previously, and who satisfy a range of other criteria, can apply for the subsidy and use it to obtain housing, either to own or to rent. It is intended to help households to access housing with secure tenure, at a cost that they can afford, and of a standard that satisfies the minimum health and safety requirements. A beneficiary may only receive the subsidy once, except where the scheme allows for deviations from this provision. There are six subsidy mechanisms that together comprise the housing subsidy scheme: project-linked, individual, consolidation, institutional, relocation assistance and the rural subsidy. Housing subsidies are paid out of the nine Provincial Housing Development Funds after approval by Provincial Housing Departments or accredited municipalities. The HSS manages information on both beneficiaries of the housing subsidy (either receiving or approved) and housing projects carried out (includes information on project budget, budget expenditure, number of applicants, and number of planned units). It is populated and maintained by the nine provinces. Unit records are not made available, but specific analysis can be requested and will be carried out by the Department. Requests should be directed to the Department of Housing's Director of Information Management. There are no costs for the information provided.

3.3.3 Social welfare

The South African Social Security Agency (SASSA) is responsible for service delivery regarding social assistance. It is a new agency which came into operation in April 2006, taking over responsibility for providing social security services from the Department for Social Development. The main administrative dataset is the SOCPEN system which has individual records of all beneficiaries in South Africa. As part of CASASP's programme of work with the Department of Social Development, extracts of individual anonymised records for various timepoints were obtained. This was part of a contract between the Department and CASASP. Policies around general access to data from SASSA have been drafted but there are still numerous issues to be resolved, largely because the agency has only recently been established.

SASSA developed a Needy Learner Database (NLD) in 2004 as there were concerns that many poor children were not captured on the SOCPEN system. Data (learner name, ID, address, information on grant receipt etc) relating to the poorest children was collected by school principals in public schools in the country. However, only 27% of schools and only 10% of learners were covered, and the poorest provinces had the worst coverage. There are plans to further develop the NLD, but such a project is not in the remit of SASSA and so would probably be undertaken by the Department of Social Development. The aim is to use data from the Department of Home Affairs as the spine of the database and link together data from various government departments, for example, Education, Health and Labour.

3.3.4 Economy

The major collectors of economic data are Stats SA, the South African Reserve Bank (SARB) and the South African Revenue Services (SARS). Stats SA produces a variety of economic and financial statistics, generally derived from sample surveys of enterprises, for example the Economic Activity Survey and the Quarterly Financial Statistics Survey. Between them, Stats SA and the SARB share responsibility for the compilation of the national accounts. Doing so requires collection of data from a huge variety of sources (the Tourism module in the General Household Survey, for example, is an input into the national accounts). The National Treasury holds various administrative datasets. This includes all the financial databases on expenditure and budgets, revenue, and personnel for national departments and provincial departments, and to a lesser extent local authorities and public entities. Little, if any of the data collected by these institutions is available to researchers as microdata (household, individual or firm level data). It is possible that exceptions may be made for projects of outstanding merit, but there is little likelihood of a general relaxation of policy in this regard.

The economic data collected by these institutions is released at various levels of aggregation. This may be illustrated using the Consumer Price Index. It is available at the national level as macrodata. Published information on its components, however, is also available, for example, monthly estimates of prices of vegetables in various urban areas. Meso level data of this sort may be used in a variety of analytical exercises (the comparison, for instance, of democratic and plutocratic price indices in urban and rural areas). Almost all of the economic and econometric modelling, and all of the time-series analysis carried out in South Africa, relies on economic data that originates in the institutions listed above.

Researchers may gain access to the mesodata in a variety of ways. The Stats SA and SARB websites allow direct access, free of charge, to most series. There are various other gateways for accessing economic data, mainly that generated by Stats SA and SARB, sometimes subjected to 'cleaning' to render it consistent. Three examples are the Trade and Industry Policy Strategies website[17], the Global Insight website[18] and the Quantec website[19]. The latter, a commercial operation (with commensurate charges), is one the prime sources of data used by researchers in the financial sector.

[17] http://www.tips.org.za
[18] http://www.globalinsight.co.za
[19] http:///www.quantec.co.za

3.3.5 Labour market

The main source of administrative data that could be made available to researchers is from the Unemployment Insurance Fund (UIF). The UIF is a social security fund within the Department of Labour to which employers and employees are obliged, in terms of legislation, to contribute to, and from which, under certain circumstances, employees may claim benefits. Contributions are calculated at a rate of 1% of earnings up to a ceiling amount and payable by both employer and employee (1% each). There are five benefits types: normal unemployment benefits (when an employee's services are terminated), maternity, illness, adoption and death benefits (payable to a deceased employees' dependents).

Employers are required by law to register with the UIF at commencement of activities, declare the details of their employees as and when changes take place in their employment (e.g. salary fluctuations, terminations, special leave such as maternity and illness, and the death of an employee). Employers also pay contributions on or before the 7th of each month, based on those declarations.

Unusually for a social insurance scheme in a middle-income country, the UIF in South Africa covers domestic and farm workers. Extension of the scheme to cover these workers coincided (roughly) with the introduction of minimum wage legislation in these sectors. Although compliance has been quite encouraging, casualisation is proceeding apace, a common occurrence the world over. By definition, the UIF covers workers in the 'formal' economy only (government employees are excluded), which limits the usefulness of the data for understanding labour market dynamics.

At 30 October 2006, the Fund had 529,237 active domestic employers[20] and 627,958 distinct, active domestic employees on its database. There were 509,280 active commercial employers[21] and 6,571,785 distinct, active commercial employees on the database. The UIF has various details of all the domestic and commercial employers ranging from contact details to the nature of the business, the wages/salaries paid and linkages with the SA Receiver of Revenue where applicable. The details of individual employees ranges from residential address at the time they claim, wage/salary, occupation, age, gender and employment history with various employers, if applicable.

The data, in an Oracle database, is updated in real time through the Siyaya system (an in-house application developed by the UIF) and the AXSOne Financial Management system as and when employers register, declare, pay contributions and when employees claim benefits and these benefits are paid. Both systems are integrated. The UIF uses various validation processes within the applications to ensure the validity of the data. For example, each distinct ID number (whether of an employer, employee or dependent) is validated against a table in the database, which is updated daily from the Department of Home Affairs Population Register.

The data is collected by the UIF purely for its own operational purposes and is often used by other organisations. The UIF regularly makes subsets of the data available to

[20] A domestic employer is an employer who employs a domestic worker in a private household.
[21] A commercial employer is an employer in any type of commercial activity.

other South African government departments in Excel spreadsheets, MS Access or in flat file formats depending on the size and according to the requirements of the recipient. The file(s) can be transferred to an address, sent as an attachment to an email if small enough, or copied to a CD/DVD and couriered to the recipient. A formal request for data should be made to the Unemployment Insurance Commissioner and Director General of the Department of Labour. In the request, the applicant must specify what the data will be used for and provide an undertaking regarding confidentiality. The technical requirements relating to the dataset(s) are usually finalised in discussions after approval has been granted by the Director General and Commissioner. Costs are born by the applicant and relate to the programmer's time in discussions and writing the queries (usually R 400 per hour) and the cost of the medium of transfer.

3.3.6 Education

Education Management Information System
The Education Management Information System (EMIS) Directorate of the Department of Education is charged with the development and elaboration of a national Education Management Information System which covers ordinary schools (Grades 1 –12) , Further Education and Training (FET) colleges, Higher Education institutions, Adult Basic Education and Training (ABET), Early Childhood Development (ECD) and Special Needs Education (SNE). EMIS is responsible for ensuring that education data and information (on learners, staff and institutions) relevant to education planning is collected, analysed, and reported using the EMIS.

EMIS conducts - or will conduct from 2007 - two surveys for five education sectors (ordinary schools, ABET, FET, SNE and ECD) on a yearly basis. The first is a Headcount Survey, which so far is only carried out for schools - both public and private - and is conducted in January on the tenth school day of the academic year. The second is a full Census, conducted in March (usually referred to as the Annual Survey). Strictly speaking both the Headcount and Annual Surveys are censuses in that the whole sector institutional population is surveyed.

The Headcount Survey contains data on learners (pupils) by grade and gender and educators by gender. It is not as comprehensive as the Annual Survey and only serves to give a quick snapshot of the education system in each province at the beginning of the year. The full Census has been conducted since 1996 for schools, for the past three years for ABET institutions and for the past two years for the other three types of institution. The Annual School Survey database, for example, contains information on each school in the country, its learner profile in terms of grade, age, home language and race. It also contains numbers of secondary learners enrolled by subject in each school as well as detailed information on educators.

The data collection and processing of data is carried out by the provincial education departments. Although each provincial education department has quality control procedures, the quality does vary by province. There are some concerns within the Department of Education about the quality of the data, particularly for 1997 to 2002 and so these are currently being resolved. EMIS has started to carry out audits on a sample of schools in order to verify whether accurate information has been provided in the surveys by individual schools. In July and September 2006, a sample survey of

two percent of schools was conducted, and this will be compared with the data from the Headcount and Annual School Surveys provided by the schools. From this it will be possible to establish the accuracy of the two datasets.

To access the data from the Headcount and Annual Surveys, requests have to be made to the Director of EMIS who will assess the request. Bona fide researchers should not have any difficulty accessing the data. Once a confidentiality form has been signed, the data, in Access format, is sent by CD. With regard to Annual Surveys, there is usually a 12 month lag from the data being collected to becoming available. Headcount Surveys have a much shorter turnaround time. The 2006 Headcount Survey for ordinary schools, for example had a 30 days turnaround time, although only about 90 percent of surveyed schools returned the forms. The situation is improving.

Other data that has been collected by EMIS in the past includes the School Register of Needs Survey (1996 and 2000), which is a survey of the physical facilities, services, equipment and resources of each institution and learner, educator and non-educator information. As of 2006, the Physical Infrastructure section of the Department of Education is now responsible for this data. Data from a further round is currently being compiled, but is only due for completion within the next two years. A similar survey was also carried out for special schools in 2003.

South Africa School Administration Management System
A Schools Administration Management System (SAMS) has been in operation in some schools since 2003. The SAMS application is provided free to any school that would like to implement the system. Other management systems are sold to schools by independent vendors. SAMS adheres to national standards of information gathering and was recently upgraded to be a national level South Africa School Administration Management System (SA-SAMS). Currently approximately 35% of schools in the country use the SA-SAMS, and it is hoped that in time it will be implemented in all schools that request its use. SA-SAMS is a self-standing management system which collects information on learners and educators.

The Department of Education/EMIS recognises the importance of collecting data on individual learners and has begun to re-engineer the whole process of education information acquisition. A Learner Unit Record Information and Tracking System (LURITS) is being developed in parallel to SA-SAMS. SA-SAMS will be a platform for LURITS, from which information on all individual learners in the country can be extracted. The data will be used to track progression of learners at individual schools, more detailed analysis of the drop out rate in high schools, as well as movement between schools and movement in and out of a province.

LURITS is currently being implemented in the Western Cape province and will be a good case study for the project. Full roll-out nationally is unlikely to be for a number of years. A database containing information on approximately 900,000 learners from Grades R (reception – aged 6/7 yrs) to 12 (final school year – Matriculation examinations) and teachers at more than 1,500 public schools in the Western Cape is now up and running. The data includes fields containing information on each learner, including name, ID number, date of birth, gender, address, telephone number, parents (alive or deceased), province of origin, grade, whether receiving state funding for R

grade, class, school attendance last year, whether receiving a social grant, whether boards at hostel or at home, whether benefits from school feeding scheme, special school or mainstream school, disabilities, language (home, teaching preferred), languages taken as subjects, matriculation subjects (for grades 10-12), and extra-mural activities. Each learner has a unique identifier, and in future this could be linked to examination results data.

Examination results
The Senior Certificate Examination Results database contains data on each candidate registering for the examinations (name, date of birth, address, school, race) as well as their results in each subject (individual marks and whether a pass or fail). Data is available for higher grade or standard grade and for whether the candidate passed with or without exemption. The data capture and output of results is managed by each province, while the database itself is run on a central mainframe at the Department of Education. Specific requests from the database can potentially be made available for research purposes with permission from the Director General of the Department. A written request is necessary which clearly indicates the reasons the data is required and exact details of the data needed.

3.3.7 Transport

The main administrative database at the Department of Transport is the National Traffic Information System, which houses all driver and vehicle licenses of registered drivers and vehicles in South Africa as well as details of traffic officers, testing stations, warrants of arrest and other information. Such data is unlikely to be of use to researchers and in any case is considered highly confidential and is not released to researchers.

3.3.8 Crime

Recorded crime
The South African Police Service (SAPS) publishes recorded crime statistics on its website[22]. Data extracted from the Crime Administration System are published for 27 crime categories at national, provincial, police area and police station or precinct levels. There are 43 police areas in the country and between 20 and 36 precincts in each area. The police areas and stations are unique geographies that do not match any other boundaries (e.g. magisterial district) in the country. The only geography they nest within is province boundaries. It is likely that police areas will be phased out in the near future, and there are moves to ensure that police geographies fit with other key geographies in the country.

A dataset of crimes recorded at all police stations in the country is currently made available to the HSRC GIS unit for mapping. It is possible that a similar dataset could be made available to other researchers. A written request would need to be made to the SAPS Head of Information Management.

[22] http://www.saps.gov.za

The South African Banking Risk Information Centre[23] collects data on bank and cash-in-transit crimes to provide clients and partners with crime risk information and liaison services. It is not known whether data is or could be made available to researchers.

National Injury Mortality Surveillance System
The National Injury Mortality Surveillance System (NIMSS)[24], managed by the University of South Africa's Institute for Social and Health Studies and the Medical Research Council, produces information on non-natural deaths from 37 mortuaries in six provinces. In essence, this system involves the active collation and centralisation of routinely-kept data. The data is currently biased to urban areas with rural mortuaries only included for the province of Mpumalanga. However, the NIMSS is being expanded annually, and will eventually process information from all mortuaries performing medico-legal post-mortems. Data collected includes race; age; sex; date and time of injury; date and time of death; province, town, suburb and district of injury; scene of injury; external cause of injury; manner of death; samples taken; and also details following court investigation (for homicides and suicides). It is understood that data are available by specific request[25].

3.3.9 Health

There are four main sources of administrative health data:

1. Primary healthcare sector
2. Secondary healthcare sector
3. Notifiable disease registers
4. Special monitoring systems

None of these sources provide very comprehensive microdata. They are typically limited in the information they hold on the individual and rarely contain very precise location information. They are also sensitive datasets, containing highly confidential information on individuals. Researchers from outside South Africa may therefore have both difficulties in finding suitable data but also in accessing this data. Partnership with South African institutions may help facilitate access but will not necessarily overcome the problem of suitability.

Primary healthcare sector
The main source of primary healthcare information comes from the District Health Information System (DHIS), a routinely collected anonymised patient dataset. The DHIS is based on a bottom-up, open-source data collection system, developed as part of the Health Information System Programme. Data are collected on a monthly basis from primary healthcare clinics, and more recently the system has also been expanded to take quarterly tuberculosis (TB) data, data on emergency medical services, environmental data and the hospital minimum dataset (a key set of secondary care indicators, at the hospital level). Users are encouraged to develop new functionality or new uses for the software and share this with others and this has resulted in the development of new functionalities being captured on the software.

[23] http://www.sabric.co.za/
[24] See http://www.sahealthinfo.org/violence/nimss.htm and http://www.unisa.ac.za/default.asp?Cmd=ViewContent&ContentID=18965 for more information.
[25] See http://www.sahealthinfo.org/violence/nimss.htm for contact details.

This data is typically presented at the district level, although it is actually held with clinic and hospital codes. It would therefore be possible to attribute it to areas, if the catchment area of the clinic is known. It has some simple population characteristics (e.g. age and gender) and therefore limited micro-analysis would be possible.

Secondary healthcare sector
There is no comprehensive system of secondary healthcare data collection in South Africa. Although a majority of the hospitals hold patient records electronically, there are a variety of information management systems used and many hospitals have difficulty extracting data for purposes other than the daily operation of their own institutions.

There is an Electronic Patient Record Working Group in the national Department of Health which is tasked with conceiving and designing a comprehensive national health information system for South Africa, and implementing it in stages, primarily for the management of the national health care services and facilities at the district, provincial and national levels. However this group has not as yet solved the problem of national compatibility being compromised by the implementation of the different information management systems across the country.

Researchers may well be able to negotiate access to data with individual hospitals but will of course run into the type of problems outlined above as well as the broader issue of bias and validity because of the selection process inherent in a in a healthcare system with both private and public provision where many of the population may not be able to afford care in particular institutions.

Notifiable disease registers
Local health authorities, under government law (Health Act No. 63, 1977), have to report specific diseases to the local, provincial and national Health Departments. This data provides case and death data as well as geographic and age trends. However there are issues of under reporting; while some diseases such as haemorrhagic fevers are very likely to be reported because they are alarming and have a highly visible impact, others, such as, food poisoning, are less likely to be reported, as people are less likely to seek medical attention. When they do seek attention, health workers are less likely to report these less severe conditions.

Special monitoring systems
Three issues of health concern have generated particularly focused datasets. These are: the Prevention of Mother To Child Transmission [of HIV] (PMTCT), the roll out of Anti-Retroviral Therapy (ART) and the treatment of TB.

The Health Systems Trust, of South Africa, has been collecting data as part of a process of monitoring and evaluation of the PMTCT programme. This has involved two separate research projects. First, routine data has been collected from the 18 pilot sites using DHIS software. Second, a National PMTCT Cohort Study has been set up with the aim of investigating infant feeding patterns and behaviours of mothers, both those HIV positive and HIV negative, and to describe and measure the impact of the PMTCT programme on the health of infants born to HIV positive mothers. This study

recruited women in September 2002 and is being conducted in Paarl (Western Cape), Umlazi (KwaZulu-Natal) and Rietvlei (Eastern Cape).

A number of provinces are monitoring the rollout of ART. In the Western Cape, for example, a system of clinical record-keeping and routine monitoring of ART has been created. It involves the archiving and linking of clinical records and the routine reports to a form a series of cohorts for monitoring. The cohort monitoring tool involves:

- Dedicated stationary for clinical record keeping, including a patient-retained record, a patient summary, and a visit summary designed to assist clinicians in providing optimal care to patients by integrating longitudinal data in a coherent manner.
- Facility-based registers as the minimum requirement for enumerating and documenting outcomes for all patients starting ART, allowing for up to 36 months of follow-up, and facilitating the extraction of routine reports.
- Standardised routine monthly and quarterly cohort reports, as well as adverse drug reporting forms.
- A set of stock management and reporting tools for pharmacists.

A similar form of monitoring takes place as part of the National Tuberculosis Control Programme. Here the various notification and follow-up forms required from those caring for patients with TB are collated and linked. The collation, however takes place at a facility level (i.e. the clinic or hospital), and is then passed on to the district which aggregates the data to be exported into the DHIS. The district held data is therefore an aggregated dataset, though maintaining some longitudinal elements. Data is then aggregated to a provincial level and passed on to the national department. The nationally held data is therefore not a microdata source.

Food insecurity

A research team, led by the HSRC, under the auspices of the Department of Agriculture, has developed a tool (the Food Insecurity and Vulnerability Information Management System) to help better understand food security, hunger and vulnerability. It is a decision-making information management support tool that provides up to date information on poverty and hunger, and from which the underlying causes can be analysed. This system will better inform planners and decision makers regarding relevant food security interventions within needy communities in the country. The information can be used to guide decision-makers in allocating resources to the most food insecure and vulnerable people and areas in the country.

An inventory of existing information sources from government departments such as Education, Health, Social Development, Public Works, Water Affairs and Forestry, and Provincial and Local Government will be compiled. This inventory will, for example, contain information on nutrition programmes and social grants. A key indicator database can then be derived from the inventory, and will be followed by a concise livelihood survey and the development of a food security model.

3.3.10 Children

Of relevance to this scoping study is a study that researchers within the Child, Youth, Family and Social Development research programme of the HSRC are carrying out for the Department of Social Development to identify and evaluate the usability and quality of government administrative data systems on orphans and vulnerable children (OVC). The main purpose of this study is to identify, describe and assess administrative data systems relating to OVCs and children in South Africa. Research of this type is important because a vital requirement for an optimal response to the OVC situation is accurate data systems that include information on who the children are, where they live, services that they are able to access and still require, as well as their unfulfilled rights and needs. There are several administrative data systems relevant to children in South Africa and other systems which, while not primarily addressing OVCs could nonetheless provide very useful information. The HSRC will provide an audit and brief qualitative description of the available government administrative data systems relating to children and OVCs in South Africa. Based on this audit, a small sample of active data systems will be selected for a more detailed and comprehensive qualitative examination (that is follow-up in-depth interviews). This study will therefore explore if and how administrative data systems have been or can be used to develop a comprehensive system for the care and support of OVCs at national, provincial or district level in South Africa.

4 Conclusion

4.1 Summary

This report has attempted to document the availability of social and economic microdata (and to a lesser extent macrodata) in South Africa. Key recommendations from the *Review of International Data Resources and Needs* were taken into account in the construction of the report, particularly those focusing on the acquisition of international microdata resources. This report provides a sound base for two important action points highlighted in the *Review*:

- A project within ESDS Access and Preservation, with support from ESDS International and the Data Centres to acquire, document and preserve national microdata resources from other countries.
- Where it is not possible to hold data within the UK, ESDS and Data Centres to provide improved information on access procedures and increased support for researchers attempting to negotiate international access procedures.

The general impression is that a variety of microdata resources are fairly easy to access - usually for little or no cost and without overly-demanding conditions of use - either from the national statistics agency or from the national data archive. Specific research groups/centres also make their data available, sometimes through collaborative work, but often not. Census and survey data are widely available, but accessing administrative data is generally not a straightforward process, and in many instances is simply not possible at the present time. Details of how to access the different datasets that are available have been provided.

A few difficulties were encountered in conducting this scoping study. Some organisations were not accustomed to the term 'microdata' and were unused to people expressing an interest in non-aggregated data. Careful efforts also had to be made to clarify the fact that explorations were being made about availability for any researcher, and that these were general requests for information on specific datasets rather than specific requests for the actual datasets. Without a specific proposal for a project using the data, it was difficult to obtain a complete picture of what is actually available. This was particularly true of government administrative data. In some instances the data sources did not have a publicly available main contact, and it was therefore a time-consuming process to identify the contact person. There is no centrally convened microdata user group in South Africa and so issues relating to quality of some of the datasets were dependent on descriptions in the data documentation or personal accounts from users.

4.2 Recommendations

A relatively easy first step for the ESRC would be to obtain South African macrodata. This report has provided brief details on various sources of macrodata.

The *Review* states that, 'A concerted effort should be made to acquire, process and disseminate microdata resources from less-developed countries… The ideal situation would be to acquire microdata resources countries [sic] to disseminate via the Archive. The Data Archive's high processing, preservation and documentation standards could be applied to the data, providing a valuable resource for researchers.' (Desai and Cowell, 2006: 27) The ESRC has already been in contact with SADA and the HSRC, but it is important to build on these links, and also to make contact with Stats SA (particularly Data Management and Information Delivery) and DataFirst. These four organisations cover the majority of microdata that is easily accessible in South Africa, and all are trying to make data resources more widely available. ESRC expertise in data management should be shared with the four organisations

SADA could potentially accommodate more datasets and it would be an extremely useful resource if most data could be held in one repository in this way, as is the case with the UK Data Archive. SADA is a small unit, however, and may struggle if there was huge demand from individual researchers. It might therefore be useful to provide technical support to SADA and assist in developing an automated system for providing datasets rather than the current manual process. Alternatively, it may be possible to acquire the datasets and documentation from SADA and elsewhere and deposit in an archive in the UK (i.e. ESDS International) which has the capacity to disseminate the data to researchers. This should only be undertaken if it is not possible to expand SADA or other archives in South Africa, in order for researchers to have easy access to good data and documentation. In both scenarios, the archiving of documentation would need to be rigorous and applied in a standardised way. This is highlighted in the *Review*, which states that 'the next best strategy would be to acquire access to as much documentation as possible. Where possible existing documentation could be enhanced, DDI standards applied and metadata incorporated into catalogues such as C-CAT' (Desai and Cowell, 2006: 27) Again, the four organisations mentioned above are good sources of documentation, although none have comprehensive coverage.

There seems to be some duplication of effort and it is not clear what the channels of communication are between the four organisations. There certainly does not seem to be any formal link between the various data warehouses and resource centres. It may be helpful therefore, to share the information in this report with the four organisations.

Indeed, this report will be a useful resource for anyone wishing to make use of microdata in research on South Africa. As such it is important that the report is disseminated widely. A useful forum for this would be the South African Regional Poverty Network (SARPN) website[26]. SARPN is a non-profit organisation that promotes debate and knowledge sharing on poverty reduction processes and experiences in Southern Africa.

This report is only a snapshot in time of the available microdata resources. It is hoped that this document could be regularly updated and so remain useful to researchers in years to come. Related to this is the need to establish a microdata user group, where local knowledge and experiences of using the different datasets can be shared. It is only through such a group that the true utility of the datasets to researchers can be assessed.

Unlike with Census and survey data, progressing access to administrative data is more complicated. There are specific areas where data are already made available to certain researchers and it may be that these would be the most sensible targets in the first instance. It would perhaps be useful also to share with government departments the UK experience of using administrative data for research purposes (Smith and Noble, 2000). It is also important to keep abreast of developments in the use of administrative data for research purposes in South Africa. The outcomes of the project to evaluate the usability and quality of government administrative data systems on orphans and vulnerable children, described in **Section 3.3.10**, will be useful in this regard, and also the proposed work by CASASP to harness administrative data to update the Provincial Indices of Multiple Deprivation for South Africa, which currently make use of 2001 Census data (Noble et al., 2006).

The table below lists the recommendations in order of achievability.

	Recommendation
1	Disseminate report widely.
2	Acquire macrodata.
3	Contact (or re-contact) SADA, HSRC, Stats SA and DataFirst and share report and ESRC technical expertise. Assist in building a comprehensive data archive in South Africa.
4	Acquire microdata and documentation that are already available and create archive of South African microdata in UK (only if not possible in South Africa). Ensure data and documentation are standardised.
5	Update report on a regular basis.
6	Establish microdata user group.
7	Acquire microdata that are less accessible.

[26] http://www.sarpn.org.za/index.php

Appendix 1 - 4 Contents

Appendix 1 Key datasets 28
A1.1 Introduction 28
A1.2 Census data 29
Census 2001 10% Sample 29
Census 2001 Community Profiles 31
Census 2001 Small Area Statistics 34
Census 1996 36
A1.3 Survey data 39
Afrobarometer South Africa 39
Africa Centre Demographic Information System 40
Agincourt Health and Demographic Surveillance System 41
All Media and Products Survey 42
Birth to Twenty 44
Cape Area Panel Study 44
Community Survey 47
Demographic and Health Survey 48
Financial Diaries Project 49
General Household Survey 51
Greater Durban Metropolitan Area Large Manufacturing Firm Survey 53
HIV and Sexual Behaviour Among Young South Africans: A National Survey of 15-24 Year Olds 54
HSRC National Survey 55
Income and Expenditure Survey 56
Khayelitsha/Mitchell's Plain Survey 58
KwaZulu-Natal Income Dynamics Study 60
Labour Force Survey 62
National Food Consumption Survey 64
National HIV and Syphilis Antenatal Sero Prevalence Survey 66
National Household Travel Survey 67
National Income Dynamics Study 68
National Victims of Crime Survey 69
National Victims of Crime Survey 71
National Youth Victimisation Survey 73
October Household Survey 74
Omnibus 76
Rural Survey 77
South African Migration and Health Survey 79
South African National HIV Prevalence, HIV Incidence, Behaviour and Communication Survey 81
South African Social Attitudes Survey 82
Student Choice Behaviour Project Phase 1 83
Survey of Activities of Young People 84
Survey of Independent Schools 87
Survey of Private and Further Education and Training in South Africa 88
The 2001-2002 HSRC Migration Survey 88
Time Use Survey 89
A1.4 Other Census and survey data 91
A1.5 GIS data 93
Appendix 2 Contacts 96
A2.1 List of persons interviewed (face to face, by telephone or through email correspondence) 96
A2.2 Government department websites 98
Appendix 3 Organisational structure: Stats SA and HSRC 99
A3.1 Stats SA 99
A3.2 HSRC 100
Appendix 4 Sample records 101
A4.1 South African Data Archive 101
A4.2 HRD Data Warehouse 103

Appendix 1 Key datasets

A1.1 Introduction

Detailed information on the key South African datasets is given in two main sections: Census data and Survey data. There is also a section containing information on other datasets for which it was not possible to find out detailed information, and a section on GIS data. The descriptions of survey data are organised in alphabetical order. Where possible the following information is provided:

1. Name and principal investigator;
2. Year(s);
3. Area(s) of interest (general i.e. demography, housing conditions, social welfare, economy, labour market, education, transport, crime, and public health, as specified in the tender document[1]; and more specific);
4. Brief description of the data source (summary; purpose; methodology e.g. variables/ questions, sample size and design, weighting, geography;
5. Assessment of the utility of the data to researchers;
6. Format in which the data are held;
7. Availability of associated documentation and data descriptions;
8. Nature of any conditions that are stipulated by data owners (disclosure control, restrictions on access, whether licensing agreements required);
9. Costs of preparing data for acquisition (e.g. data extraction and preparation costs);
10. Tabled outputs (macrodata); and
11. Contact details for information and data.

This information reflects the current state of South African microdata and documentation. All documentation relating to datasets is in English. Only data from 1994 onwards have been included (for both single/one-off surveys and series data) as this is the point at which South Africa became a democracy. Though every effort has been made to provide a comprehensive list of datasets, there may be some additional surveys that have not been listed here. It is hoped that this document will be regularly updated and that any additions or amendments could therefore be incorporated.

It is difficult to make an assessment of the documentation without making use of the data and accompanying data descriptions. The report does, however note when datasets are not fully documented. It is also difficult to assess how useful the data will be to researchers without first making use of the data. However an attempt has been made to highlight the most obvious strengths and weaknesses. This is not an endorsement or otherwise of the dataset in question and it is strongly recommended that individual researchers make their own assessment.

The contents of the tables are largely drawn from documentation that is in the public domain relating to the listed datasets. Except for the 'how useful to researchers?' section of the tables, most of the other sections in each of the tables are direct quotes from the metadata or related documentation. For ease of use, quotations and page numbers are not given, though references to the sources of information are supplied.

Notes:

1. Although DataFirst is noted as a contact for data in many instances, the majority can only be accessed in the Resource Centre at UCT (see **Section 2.4.1**). However, where documentation is available from the DataFirst website, this is noted. Where data is only available in the DataFirst Resource Centre, specific conditions may apply, but these are not noted in the table. Four datasets and documentation can all be accessed from the DataFirst website.
2. Where SADA is noted as a data holder, both datasets and documentation can be accessed via the website. All datasets obtained from SADA have the following conditions:

[1] A further category, attitudes, has been included as there are a number of attitude surveys conducted in South Africa.

- No attempt will be made to use the data to derive information on specifically identified individuals or households in the data.
- The data and documentation received from SADA will not be duplicated without prior approval of the Director of SADA.
- Any publication or other presentation based in whole or part on the data and documentation supplied by SADA must use the recommended citation (both SADA and the Depositor will be acknowledged in all published works based on the data and documentation).
- At least one copy of any published work or report based in whole or part on the dataset must be deposited with SADA.
- These are not re-produced in each table, and therefore where it states that there are no conditions attached to use of the data, it means none in addition to the above.

3. The data for two large studies, the Community Survey and the National Income Dynamics Study have not been collected. They are included here as they will be important datasets in the next few years.

A1.2 Census data

Name	**Census 2001 10% Sample**
Principal investigator	Stats SA
Year(s)	2001
Area(s) of interest	**General:** demography; housing; labour market; economy; education; transport; health **Specific:** age; gender; marital status; population group[2]; language; religion; citizenship; migration; disability; school attendance; level of education; work in previous 7 days; employment status; occupation; births; travel to school or work; income; deaths; household size; dwelling type; number of rooms; access to piped water; toilet facility; energy source for cooking, lighting, heating; ownership of household goods; access to telephone; refuse or rubbish disposal; annual household income
Brief description	**Summary** A10% unit level sample drawn from Census 2001. People present in the country who were living in households or communal living quarters or who were homeless on the night of 9-10 October were counted. **Dataset** The 10% sample comprises six files: households, persons, mortality, geography, household imputation flags and person imputation flags. All variables as per the questionnaire are included in the 10% sample, as well as derived variables and imputation flags (see below). Enumeration Area numbers are excluded to preserve confidentiality. Geographic type is excluded from the final sample. Instead two additional geographical variables are supplied: urban and rural Census '96 classification and size and density of locality. **Sampling** The sample was drawn from the full Census as follows: ▪ *Households* (948,592 records) - a 10% sample of households in housing units, and a 10% sample of collective living quarters (both institutional and non-institutional) and the homeless. ▪ *Persons* (3,725,655 records) - a sample consisting of all persons in the households and collective living quarters, and the homeless, drawn from the samples. ▪ *Mortality* (36,267 records) - a sample consisting of all mortality information for the households in housing units drawn in the 10% sample of households. **Weighting** Both the 10% household and person sample files contain a weight variable. This weight variable is the adjustment factor for undercount (for households or persons as appropriate) multiplied by 10 to inflate the 10% samples to the relevant population. In the person records, aggregated totals of sparsely populated codes, such as very old ages, might differ substantially from real totals due to sampling fluctuations – no scaling of the weights was done. In the household records aggregated totals will be approximately equal to real totals. Mortality was not adjusted for undercount and therefore there is no weight variable. An Excel file is available with four worksheets showing the

[2] The South African term for ethnic group. The four categories are black African, coloured, white and Indian/Asian.

	adjustment factors for persons and households at municipality and provincial level, which can be used to calculate the 100%. If required, standard errors for each variable can be calculated by Stats SA. **Geography** The South African geographical structure for the 10% sample consists of the following geographical entities, which fit into different geographical hierarchical levels: South Africa, Province, District Council (Category C) or Metropolitan Area (Category A), Magisterial Districts, Local Municipality (Category B), or District Management Area. While the structure is intended to be hierarchical, South Africa's geography has cross-boundary entities, which complicate the picture. For example, there are eight municipalities which lie across provincial boundary lines. Users are advised to bear this in mind when choosing the appropriate hierarchy. For example, for the City of Tshwane, which lies in two provinces, one would not use the provincial hierarchy. **Imputation** Imputation was used to allocate values for unavailable, unknown, incorrect or inconsistent responses. A combination of both logical imputation and hot deck imputation (dynamic imputation) was used. Undetermined values were used for only a few variables in a few cases, such as industry and occupation. **Confidentiality** In order to preserve confidentiality the lowest geographical level that unit records can be linked to is municipality. As further assurance of the confidentiality of the data, municipalities with 200 or fewer households are logically grouped with adjacent municipalities. Extract from the *Report of the Census Sub-Committee to the South African Statistics Council on Census 2001* "Preliminary investigations indicate that the 2001 census probably resulted in: • an underestimate of the number of children below age five* • an over-estimate of the number of teenagers aged between 10 and 20 • an underestimate of the number of men relative to the number of women* • an underestimate of the number in the white population • higher than expected numbers aged 80 and older, in the African population • an underestimate of the number of foreign-born, since some identified themselves incorrectly as being South African-born • age misstatement in the range 60-74 • an overestimate of the extent of unemployment • an underestimate of those who were employed for only a few hours per week • an underestimate of household income • an overestimate the number of paternal orphans and the number of fathers missing from the household. * This is a common feature of censuses, particularly in developing countries. In addition: • Scanning problems caused some births to be recorded in the wrong province. The number of cases is relatively small and should not lead to too much distortion for most purposes for which these data are used; however, it does produce obviously erroneous results when one tries to estimate the extent of inter-provincial migration of those born since the previous census. • The fertility data (numbers of children ever born, children surviving) are problematic. For further details of these investigations see the full report of the *Census Sub-Committee.*"
How useful to researchers?	A very useful dataset that comes with good documentation. Robust results can be produced down to municipality level for a range of areas of interest. Work has been undertaken to impute values for missing and implausible zero incomes, in addition to Stats SA's imputation work (Ardington et al., 2005; Barnes et al., 2006).
Data format	The six files are available in ASCII format from Stats SA and SPSS format from SADA.
Availability of data descriptions	The following detailed documentation is available from the Census 2001 website (http://www.statssa.gov.za/census01/html/default.asp): ▪ Metadata - for households, persons, mortality, geography and imputation files ▪ Code lists - country of birth and citizenship, religion, occupation and industry ▪ Questionnaires ▪ Record layouts ▪ Concepts and definitions

	▪ How the count was done
Conditions	The information products and services of Stats SA are protected in terms of the Copyright Act, 1978 (Act 98 of 1978). As the State President is the holder of State copyright, all organs of State enjoy unhindered use of the Department's information products and services, without a need for further permission to copy in terms of that copyright. Where a copy of the information is made available to any third party outside the State, the third party must be made aware of the existence of State copyright and ownership of the information by the State. The State (through Stats SA) retains the full ownership of its information, products and services at all times. Access to information does not give ownership of the information to the client. The use of any data is subject to acknowledgement of Stats SA as the supplier and owner of copyright. Users may apply or process the data, provided Stats SA is acknowledged as the original source of the data; that it is specified that the analysis is the result of the user's independent processing of the data; and that neither the basic data nor any reprocessed version or application thereof may be sold or offered for sale in any form whatsoever without prior permission from Stats SA. SADA conditions also apply to data obtained from SADA.
Costs	R 300 for a CD from Stats SA or free from SADA.
Tabled outputs	The following are available to download from the Census 2001 website (http://www.statssa.gov.za/census01/html/C2001publications.asp): ▪ *Census 2001 Census in Brief* - over 80 tables and graphs at national and provincial level using an extensive range of individual and household variables. ▪ *Census 2001 Key Results* - information on how the Census was done and highlights of Census results (national and provincial population, language, education and household goods). ▪ *Primary Tables: 1996 and 2001 compared* - one national and nine provincial reports with detailed tables from both Census 1996 and Census 2001 giving number and percentage information for each of 23 person (language, country of birth, citizenship, age, religion, marital status, disability, education, economic activity, full-time or part-time work, overall labour market data, occupation) and household (dwellings, electricity and other energy sources, access to clean water, toilet facilities, refuse removal, access to telephones, age distribution of household heads) variables broken down by sex and population group. For each topic there is a comparative graph and a short narrative. ▪ *Key Municipal Data* - key Census findings for each metropolitan, district and local municipality. It consists of one page per municipality, with tables showing information for selected person (population group by sex, age groups, present school attendance, highest level of education, labour market data, mode of transport to travel to school or work) and household (population group and sex of head of household, monthly imputed household income, type of housing unit, tenure status, access to water, access to electricity, toilet facilities, refuse removal) variables.
Contact	**For information:** ***Census 2001*** website: http://www.statssa.gov.za/census01/html/default.asp ***Stats SA User Information Services*** Email: info@statssa.gov.za Telephone: +27 (0)12 310 8600 ***Mr Kevin Parry***, Population Census Division, Stats SA Email: kevinp@statssa.gov.za Telephone: +27 (0)12 310 2111 **For data:** ***Stats SA - Mr Kevin Parry*** Email: kevinp@statssa.gov.za Telephone: +27 (0)12 310 2111 ***SADA*** website: http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0139 ***DataFirst*** (data available in resource centre)

Name	**Census 2001 Community Profiles**
Principal investigator	Stats SA
Year(s)	2001
Area(s) of interest	**General:** demography; housing; labour market; education; transport; health **Specific:** demography; disability; dwellings; education; family; head of household; household services; labour force; language; migration; mortality; persons and services; transport; welfare
Brief description	**Summary** 14 databases on various topics, which enable users to do cross-tabulations and mapping on different geographical hierarchies for each of the 21,219 sub-places (suburb or village level). **Databases** 1. *Descriptive* database covers all persons, and contains basic person variables such as highest

	educational level, present school attendance and language, which can be broken down by population group, gender, age and enumeration area type (that is, the type of area in which people live, such as tribal settlement, commercial farm, smallholding, urban settlement, informal settlement, etc). 2. *Disability* database covers all persons, and contains information about the disabled, which can be provided by highest educational level, employment status, occupation, income and enumeration area type. 3. Dwellings database provides information on dwellings inhabited by households, as well as annual household income. The universe of this database is all households, including those in workers hostels, residential hotels, student residences and homes for the independent aged. The homeless and persons in institutions are excluded. 4. *Education* database covers all persons, and enables the user to cross-tabulate highest educational level with employment status, occupation, and monthly income. (Information about present school attendance can be found in the Descriptive database). 5. *Family* database covers all persons and has information on marital status and relationship to the head of household. (Information about whether people's mothers and fathers are still alive can be found in the Mortality database). 6. *Head of Household* database contains some of the same variables as Education and Labour Force, but focusing on household heads. The variable 'work status' is unique to this database (work status describes whether an employed person is a paid employee, a paid family worker, self-employed, an employer or an unpaid family worker). The universe of this database is all households, including those in workers hostels, residential hotels, student residences and homes for the independent aged. The homeless and persons in institutions are excluded. 7. *Household Services* database contains information on services to households: water, energy, toilet facilities, refuse disposal and telephone facilities. The universe of this database is all households, including those in workers hostels, residential hotels, student residences and homes for the independent aged. The homeless and persons in institutions are excluded. 8. *Labour Force* database only covers people of working age (15–65 years of age), whether economically active or not economically active. The latter group is broken into subcategories. This database includes industry. 9. *Language* database covers all persons. It is the only database including information on religion. This can be cross-tabulated with language and highest educational level. 10. *Migration* database covers all persons. It contains information for migration studies, such as numbers and citizenship of non-South African residents, and inter-provincial migration over the last five years. 11. *Mortality* database covers all persons. It contains information about marital status and relationship to the head of household and about whether peoples' mothers and fathers are still alive. 12. *Persons and Services* database provides information about services to individuals rather than households. Percentages of certain groups of persons receiving services may therefore differ from percentages of households receiving services. The universe for this database is all persons (unlike the Household Services database). 13. *Transport* database covers all persons. It contains information about mode of travel to work or school. This information can be cross-tabulated with enumeration area type, present school attendance, highest educational level, employment status, and individual monthly income. 14. *Welfare* database covers all persons. It contains the variable enumeration area type together with highest educational level, employment status and individual monthly income (the Education and Labour Force databases do not have information on enumeration area type). **Geography** Only the field "Geography" has mappable hierarchical levels. There are five hierarchies: 1. South Africa by municipality District municipality Municipality Main place Sub-place 2. South Africa by province and municipality Province District municipality Municipality Main place Sub-place 3. South Africa by magisterial district Magisterial district 4. South Africa by province and magisterial district Province Magisterial district

	5. South Africa by electoral ward District municipality Municipality Ward **Confidentiality** All values less than 3 appear as either 3 or 0 (or –), in order to preserve confidentiality. As a result, totals generated in different databases may differ. **Under/over count** Extract from the *Report of the Census Sub-Committee to the South African Statistics Council on Census 2001* "Preliminary investigations indicate that the 2001 census probably resulted in: • an underestimate of the number of children below age five* • an over-estimate of the number of teenagers aged between 10 and 20 • an underestimate of the number of men relative to the number of women* • an underestimate of the number in the white population • higher than expected numbers aged 80 and older, in the African population • an underestimate of the number of foreign-born, since some identified themselves incorrectly as being South African-born • age misstatement in the range 60-74 • an overestimate of the extent of unemployment • an underestimate of those who were employed for only a few hours per week • an underestimate of household income • an overestimate the number of paternal orphans and the number of fathers missing from the household. * This is a common feature of censuses, particularly in developing countries. In addition: • Scanning problems caused some births to be recorded in the wrong province. The number of cases is relatively small and should not lead to too much distortion for most purposes for which these data are used; however, it does produce obviously erroneous results when one tries to estimate the extent of inter-provincial migration of those born since the previous census. • The fertility data (numbers of children ever born, children surviving) are problematic. For further details of these investigations see the full report of the *Census Sub-Committee*."
How useful to researchers?	The databases allow for pre-defined tabulations at a lower geographical level (wards) than with the 10% Census. However, the databases are not at an individual level and so are not microdata as such. They are included here as information on the different products available to manipulate Census 2001.
Data format	The databases use SuperTABLE software (data can be saved as Excel or ASCII files).
Availability of data descriptions	The following detailed documentation is available from the Census 2001 website (http://www.statssa.gov.za/census01/html/default.asp): ▪ Metadata - for households, persons, mortality, geography and imputation files ▪ Code lists - country of birth and citizenship, religion, occupation and industry ▪ Questionnaires ▪ Record layouts ▪ Concepts and definitions ▪ How the count was done The metadata for a specific database (topic) can be accessed by right-clicking in the general heading area of the open database, or via the Help menu. The metadata for a specific variable may be accessed by right-clicking when the cursor is on the name of that variable. In the case of derived variables, information about the underlying variables can be found in the "Metadata" referred to above. The available value set (list of options for any given variable) can be accessed by double-clicking on the particular variable in the 'Fields' dialogue box.
Conditions	The information products and services of Stats SA are protected in terms of the Copyright Act, 1978 (Act 98 of 1978). As the State President is the holder of State copyright, all organs of State enjoy unhindered use of the Department's information products and services, without a need for further permission to copy in terms of that copyright. Where a copy of the information is made available to any third party outside the State, the third party must be made aware of the existence of State copyright and ownership of the information by the State. The State (through Stats SA) retains the full ownership of its information, products and services at all times. Access to

	information does not give ownership of the information to the client. The use of any data is subject to acknowledgement of Stats SA as the supplier and owner of copyright. Users may apply or process the data, provided Stats SA is acknowledged as the original source of the data; that it is specified that the analysis is the result of the user's independent processing of the data; and that neither the basic data nor any reprocessed version or application thereof may be sold or offered for sale in any form whatsoever without prior permission from Stats SA.
Costs	R 300 for 12 CDs.
Tabled outputs	The following are available to download from the Census 2001 website (http://www.statssa.gov.za/census01/html/C2001publications.asp): ▪ *Census 2001 Census in Brief* - over 80 tables and graphs at national and provincial level using an extensive range of individual and household variables. ▪ *Census 2001 Key Results* - information on how the Census was done and highlights of Census results (national and provincial population, language, education and household goods). ▪ *Primary Tables: 1996 and 2001 compared* - one national and nine provincial reports with detailed tables from both Census 1996 and Census 2001 giving number and percentage information for each of 23 person (language, country of birth, citizenship, age, religion, marital status, disability, education, economic activity, full-time or part-time work, overall labour market data, occupation) and household (dwellings, electricity and other energy sources, access to clean water, toilet facilities, refuse removal, access to telephones, age distribution of household heads) variables broken down by sex and population group. For each topic there is a comparative graph and a short narrative. ▪ *Key Municipal Data* - key Census findings for each metropolitan, district and local municipality. It consists of one page per municipality, with tables showing information for selected person (population group by sex, age groups, present school attendance, highest level of education, labour market data, mode of transport to travel to school or work) and household (population group and sex of head of household, monthly imputed household income, type of housing unit, tenure status, access to water, access to electricity, toilet facilities, refuse removal) variables.
Contact	**For information:** ***Census 2001*** website: http://www.statssa.gov.za/census01/html/default.asp ***Stats SA User Information Services*** Email: info@statssa.gov.za Telephone: +27 (0)12 310 8600 ***Mr Kevin Parry***, Marketing Division, Stats SA Email: kevinp@statssa.gov.za Telephone: +27 (0)12 310 2111 **For data:** ***Stats SA - Mr Kevin Parry*** Email: kevinp@statssa.gov.za Telephone: +27 (0)12 310 2111

Name	**Census 2001 Small Area Statistics**
Principal investigator	Stats SA
Year(s)	2001
Area(s) of interest	**General:** demography; education; housing conditions **Specific:** population by age and gender; highest level of education; housing type; access to services (amenities)
Brief description	**Summary** The only Census product to be provided to users who request data at a level lower than sub-place name. The product is based on a small area layer (SAL) that was created by combining all Enumeration Areas (EAs) with a population of less than 500 with adjacent EAs within the same sub-place. The final SAL consists of 56,255 polygons. **Datasets** There are four tables: ▪ Population in five-year age groups by gender (with randomised small numbers. ▪ Percentage of population breakdown by highest level of education attained (no schooling, some primary, complete primary, some secondary, Grade 12/ Std 10, and higher) for age categories 0-4, 5-9, 10-14, 15-19, >20 (with randomised small numbers). ▪ Percentage of households living in informal and traditional housing (grouped). ▪ Percentage households without access to services (piped water, hygienic toilets, electricity for lighting, refuse removal, fixed line phone, cell-phone). Tables 1 and 2 contain information on the entire population (including collective living quarters), whilst Tables 3 and 4 contain information on households living in housing units only (excluding

	collective living quarters) **Spatial data** Spatial data, in shape file format, is provided to enable users to link the data spatially to the different geographical hierarchies. Apart from the spatial file, each table also contains codes and names to enable users to aggregate data to the different geographical hierarchies. **Geography** The South African geographical structure consists of the following geographical entities, which fit into different geographical hierarchical levels: South Africa, Province, District Council (Category C) or Metropolitan Area (Category A), Local Municipality (Category B), or District Management Area, Magisterial District (MD), Main place, Sub place, Small Area Layer (SAL), Enumerator Area (EA). While the structure is intended to be hierarchical, South Africa's geography has cross-boundary entities which complicate the picture. For example, there are eight municipalities which lie across provincial boundary lines. Users are advised to bear this in mind when using a specific hierarchy. For example, for the City of Tshwane, which lies in two provinces, one would not use the provincial hierarchy. The SAL has been generated to fill the gap between EAs and sub places. In some cases the SAL cuts through MDs. For this reason MDs are excluded from the tables. **Confidentiality** All values less than 3 appear as either 3 or 0 in order to preserve confidentiality. As a result totals may differ from the sum of the separate numbers, may vary from table to table, and may also differ slightly from published figures. This effect will be more noticeable where low values predominate, e.g. in categories such as males aged 85+ in sparsely populated areas. **Over/under count** All figures have been adjusted for undercount, in accordance with the findings of the post-enumeration survey. Resulting numbers are rounded to the nearest whole number. This also contributes to the totals from specific breakdowns not always matching published figures. Extract from the *Report of the Census Sub-Committee to the South African Statistics Council on Census 2001* "Preliminary investigations indicate that the 2001 census probably resulted in: • an underestimate of the number of children below age five* • an over-estimate of the number of teenagers aged between 10 and 20 • an underestimate of the number of men relative to the number of women* • an underestimate of the number in the white population • higher than expected numbers aged 80 and older, in the African population • an underestimate of the number of foreign-born, since some identified themselves incorrectly as being South African-born • age misstatement in the range 60-74 • an overestimate of the extent of unemployment • an underestimate of those who were employed for only a few hours per week • an underestimate of household income • an overestimate the number of paternal orphans and the number of fathers missing from the household. * This is a common feature of censuses, particularly in developing countries. In addition: • Scanning problems caused some births to be recorded in the wrong province. The number of cases is relatively small and should not lead to too much distortion for most purposes for which these data are used; however, it does produce obviously erroneous results when one tries to estimate the extent of inter-provincial migration of those born since the previous census. • The fertility data (numbers of children ever born, children surviving) are problematic. For further details of these investigations see the full report of the *Census Sub-Committee.*"
How useful to researchers?	Again, this data is not at individual level, and has only a limited number of indicators. However, it represents the smallest geographical area level dataset at which the 2001 Census is publicly available.
Data format	Data are available in Excel format.
Availability of data	The following detailed documentation is available from the Census 2001 website

descriptions	(http://www.statssa.gov.za/census01/html/default.asp): ▪ Metadata - for households, persons, mortality, geography and imputation files ▪ Code lists - country of birth and citizenship, religion, occupation and industry ▪ Questionnaires ▪ Record layouts ▪ Concepts and definitions ▪ How the count was done
Conditions	The information products and services of Stats SA are protected in terms of the Copyright Act, 1978 (Act 98 of 1978). As the State President is the holder of State copyright, all organs of State enjoy unhindered use of the Department's information products and services, without a need for further permission to copy in terms of that copyright. Where a copy of the information is made available to any third party outside the State, the third party must be made aware of the existence of State copyright and ownership of the information by the State. The State (through Stats SA) retains the full ownership of its information, products and services at all times. Access to information does not give ownership of the information to the client. The use of any data is subject to acknowledgement of Stats SA as the supplier and owner of copyright. Users may apply or process the data, provided Stats SA is acknowledged as the original source of the data; that it is specified that the analysis is the result of the user's independent processing of the data; and that neither the basic data nor any reprocessed version or application thereof may be sold or offered for sale in any form whatsoever without prior permission from Stats SA.
Costs	R 300
Tabled outputs	None
Contact	**For information:** ***Census 2001*** website: http://www.statssa.gov.za/census01/html/default.asp ***Stats SA User Information Services*** Email: info@statssa.gov.za Telephone: +27 (0)12 310 8600 ***Mr Kevin Parry***, Marketing Division, Stats SA Email: kevinp@statssa.gov.za Telephone: +27 (0)12 310 2111 **For data:** ***Stats SA - Mr Kevin Parry*** Email: kevinp@statssa.gov.za Telephone: +27 (0)12 310 2111

Name	**Census 1996**
Principal investigator	Stats SA
Year(s)	1996
Area(s) of interest	**General:** demography; economy; education; labour market; housing; health **Specific:** gender; age; marital status; population group; language; religion; citizenship; migrant worker; migration; disability status; births; highest education level; employment status; occupation; individual income (monthly); household/hostel/institution size; number of rooms; type of dwelling; dwelling ownership; fuel used for cooking, heating, lighting; main water supply; toilet facilities; refuse disposal; telephone facilities; additional money guaranteed; remittances received; highest income
Brief description	**Summary** A 10% unit level sample drawn from Census '96, the first count of all citizens of South Africa. All people present in the country on the night of 9-10 October were counted. Questionnaires were made available in all 11 official languages. **Dataset** A 10% unit level sample of all households (excluding special institutions and hostels) and all persons as enumerated in Census '96 in South Africa. This sample of actual Census records reflects some 50 categories at individual level and 25 at household level, including weights. **Sampling** The sample was drawn as a 10% systematic sample of households from the Census household file. The 10% person level sample was obtained by including all persons in these households plus the persons drawn in independent 10% systematic samples of all persons in special institutions and hostels. The Census household records were explicitly stratified according to province and District Council. Within each District Council the records were further implicitly stratified by local authority and EA type. Within each implicit stratum the household records were ordered according to the unique seven-digit census Enumerator Area number, of which the first three digits are the (old) Magisterial District number.

	Different terms are used for the local authority boundaries in different parts of the country. There are Transitional Local Councils (TLCs); Transitional Rural Councils (TRCs); Local Authority Councils; Metropolitan Sub-Structures; Metropolitan Local Councils; Rural Local Councils; District Councils (DCs); Transitional District Councils and Regional Councils. To ensure confidentiality within the 10% sample, a local authority had to have a minimum of 2,000 households. As many local authorities had fewer than this number, they had to be grouped together to ensure that the minimum number of households was met. For this purpose, hostel dwellers were treated as single person households. Local authorities with less than 2000 households were pooled with other local authorities based on the following principles: *All provinces except KwaZulu/Natal and North West*: A TLC with less than 2,000 households was grouped with the TRC within which the TLC was located. In cases where the TRC was big enough to stand on its own but the TLC's within its boundaries were too small, the sample was drawn in such a way that the TRC can be analysed either on its own or together with other TLCs within its boundaries. Where a TRC plus all the TLCs within its boundaries were less than the minimum of 2,000 households the TRC (including the TLCs within its boundaries) was pooled with the adjacent TRC. In a few cases the required minimum of 2,000 households could not be achieved when all the local authorities within a DC were pooled together. In such a case no further implicit stratification within the DC was done. *KwaZulu/Natal*: The equivalent to a DC in KwaZulu/Natal is known as a Regional Council. There are no rural councils in KwaZulu/Natal. Smaller local authorities could therefore not be pooled with the rural council within which boundaries it falls. Where such TLCs were adjacent to another TLC they were pooled to form one stratum. In two cases three TLCs were pooled to form one stratum. In all cases the TLCs that were pooled are adjacent to each other. ***North West***: The TRCs in the North West do not encompass TLCs as is the case in other provinces. The area between TLCs/TRCs in the North West is part of the relevant DC. Small TLCs/TRCs in the North West were either pooled with adjacent local authorities or they were pooled with the relevant DC. **Weighting** Both the 10% household sample file and the 10% person sample file contain a weight variable. This weight variable is the adjustment factor for undercount (for households or persons as appropriate) multiplied by 10 to inflate the 10% sample to the population.
How useful to researchers?	A useful comparator to the more recent Census. However, geographical boundaries have changed and research should be undertaken with caution. See Statistics South Africa (2005) for a comparison between 1996 and 2001 and also Bhorat et al. (2004) and Cronje and Budlender (2004).
Data format	Data are available in ASCII format from Stats SA and SPSS format from SADA.
Availability of data descriptions	The following detailed documentation is available on the Census 1996 website (http://www.statssa.gov.za/census01/Census96/HTML/default.htm): ▪ Metadata - for households, persons and migrant files ▪ Questionnaires ▪ Definitions ▪ How the count was done ▪ Calculating undercount in Census '96 - a description of how Stats SA measured undercount in the Census by means of a post-enumeration survey and how it used this information to weight the final data. ▪ Fertility and morbidity trends in South Africa - paper prepared by Dr Eric Udjo, Director of Thematic Demographic Analysis at Stats SA, as part of an exercise evaluating the quality of data gathered in Census '96. ▪ 10% sample documentation - description of the sampling methodology used to draw the 10% sample, record layout for households and persons, code lists Some documentation is also available on the DataFirst website (http://www.datafirst.uct.ac.za/surveys.html).
Conditions	The information products and services of Stats SA are protected in terms of the Copyright Act, 1978 (Act 98 of 1978). As the State President is the holder of State copyright, all organs of State enjoy unhindered use of the Department's information products and services, without a need for further permission to copy in terms of that copyright. Where a copy of the information is made available to any third party outside the State, the third party must be made aware of the existence of State copyright and ownership of the information by the State. The State (through Stats SA) retains the full ownership of its information, products and services at all times. Access to information does not give ownership of the information to the client. The use of any data is subject to acknowledgement of Stats SA as the supplier and owner of copyright. Users may apply or process the data, provided Stats SA is acknowledged as the original source of the data; that it is specified that the analysis is the result of the user's independent processing of the data; and that neither the basic data nor any reprocessed version or application thereof may be sold or offered for sale in any form whatsoever without prior permission from Stats SA. SADA conditions also apply to data obtained from SADA.

Costs	R 300 for a CD from Stats SA or free from SADA.
Tabled outputs	See http://www.statssa.gov.za/census01/Census96/HTML/default.htm for the following tabled outputs: ▪ *Census 1996 Census in Brief* ▪ *Census 1996 by ward* - marital status, literacy, formal dwellings, age, employment, language, services. ▪ *Provincial level data: persons* e.g. Table: Population Census 1996 by province, gender, age group and population group; Table: Population Census 1996 by province, gender, highest education level and population group. ▪ *Provincial level data: households* e.g. Table: Population Census 1996 by province, gender of head of household, fuel used for cooking and population group of head of household; Table: Population Census 1996 by province, gender of head of household, type of dwelling and population group of head of household. Contact Stats SA User Information Services section (contact details below) to obtain the following tabled outputs: ▪ Census information is available on an ad hoc basis from the complete Census '96 database. Users can request tabulations of any Census variables at specific geographical levels ▪ *Age tables* - National tables are presented in both single years and five-year groups, broken down into urban and non-urban areas, gender and population group, and displayed at both national and provincial levels. Actual numbers and percentages are shown for each table. ▪ *Primary tables* - An extensive range of census variables broken down by urban and non-urban residents, gender and population group is available in the national primary tables, which display the information for the country as a whole and for each province. The same set of variables is available for each Local Authority in the country on a provincial CD. It is also possible to obtain hard copies of selected tables for any set of geographic areas.
Contact	**For information:** ***Census 1996*** website: http://www.statssa.gov.za/census01/Census96/HTML/default.htm ***Stats SA User Information Services*** Email: info@statssa.gov.za Telephone: +27 (0)12 310 8600 ***Mr Kevin Parry***, Marketing Division, Stats SA Email: kevinp@statssa.gov.za Telephone: +27 (0)12 310 2111 **For data:** ***Stats SA - Mr Kevin Parry*** Email: kevinp@statssa.gov.za Telephone: +27 (0)12 310 2111 ***SADA*** website: http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0129 ***DataFirst*** (available in resource centre)

Note 1: Community Profiles are also available for Census 1996.

Note 2: Population Census data are available for 1991, 1985, 1980 and 1970 from Stats SA, SADA and DataFirst. The documentation is not in soft copy format and nor of such good quality as the later Censuses, and other than in 1970, the Censuses did not enumerate the whole population.

For information:
Stats SA User Information Services
Email: info@statssa.gov.za
Telephone: +27 (0)12 310 8600
Mr Piet Alberts, Data Integration, Statistical Information Services Division, Stats SA
Email: pieta@statssa.gov.za

For data and documentation:
Stats SA - Ms Joan Lindeque
Email: joanl@statssa.gov.za
SADA website: http://www.nrf.ac.za/sada
DataFirst (available in resource centre)

A1.3 Survey data

Name	**Afrobarometer South Africa**
Principal investigator	Institute for Democracy in South Africa (IDASA)
Year(s)	2000, 2002, 2004 (data available December 2006) and 2006 (data available mid-2007)
Area(s) of interest	**General:** attitudes **Specific:** democracy; governance; livelihoods; macro-economics and markets; social capital; conflict and crime; participation; national identity
Brief description	**Summary** A comparative series of public attitudes surveys measuring the social, political and economic atmosphere in African countries, including South Africa. **Purpose** The objectives of Afrobarometer surveys in general are: ▪ To produce scientifically reliable data on public opinion in sub-saharan Africa. ▪ To strengthen institutional capacity for survey research in Africa. ▪ To broadly disseminate and apply survey results. **Methodology** There have been 4 rounds of Afrobarometer for South Africa: Round 1 - 2000 Round 2 - 2002 Round 2.5 - 2004 Round 3 - 2006 Afrobarometer surveys are face-to-face interviews by trained interviewers in the language of the respondent's choice. National probability samples that represent an accurate cross section of the voting age population are used. Random selection is used at every stage of sampling and the sample is stratified to ensure that all major demographic segments of the population are covered. For South Africa, the sample size is 2,400 people (2,200 in Round 1). ***Sample design*** There is a standard protocol for drawing a national probability sample for an Afrobarometer survey and a new sample has to be drawn for each round of Afrobarometer surveys. The sample is designed as a representative cross-section of all citizens of voting age in South Africa. The goal is to give every adult citizen an equal and known chance of selection for interview. This is achieved by (a) strictly applying random selection methods at every stage of sampling and by (b) applying sampling with probability proportionate to population size wherever possible. A randomly selected sample of 2,400 cases allows inferences to national adult populations with a margin of sampling error of no more than plus or minus 2 percent with a confidence level of 95 percent. Excluded are areas determined to be either inaccessible or not relevant to the study, such as those experiencing armed conflict or natural disasters, as well as national parks and game reserves. People living in institutionalised settings, such as students in dormitories and persons in prisons or nursing homes are excluded. The sample design is a clustered, stratified, multi-stage, area probability sample. In a series of stages, geographically defined sampling units of decreasing size are selected. To ensure that the sample is representative, the probability of selection at various stages is adjusted as follows: The sample is stratified by key social characteristics in the population such as sub-national area (e.g. region/province) and residential locality (urban or rural). The area stratification reduces the likelihood that distinctive ethnic or language groups are left out of the sample. And the urban/rural stratification is a means to make sure that these localities are represented in their correct proportions. Wherever possible, and always in the first stage of sampling, random sampling is conducted with probability proportionate to population size. The purpose is to guarantee that larger (i.e., more populated) geographical units have a proportionally greater probability of being chosen into the sample. The sampling design has four stages: 1. Stratify and randomly select primary sampling units. 2. Randomly select sampling start-points. 3. Randomly choose households. 4. Randomly select individual respondents. ***Weighting*** The data are weighted to correct for either deliberate (e.g. to provide an adequate sample of

	specific sub-groups for analytical purposes) or inadvertent over- or under-sampling of particular sample strata. In these cases, a weighting variable is included as the last variable in the data set, with details described in the codebook. These weighting factors should be used when calculating all national-level statistics.
How useful to researchers?	This dataset serves as a useful comparator between different countries in sub-Saharan Africa.
Data format	Data from Rounds 1 and 2 are available in Excel or SPSS format.
Availability of data descriptions	The following documentation is available from the Afrobarometer website (http://www.afrobarometer.org/index.html): ▪ Round 1 and Round 2 codebooks - for each question/field in the dataset, the following information is provided: question number, question, variable label, values, value labels, source, notes ▪ Round 1 questionnaire ▪ Round 1 documentation is also available on the DataFirst website (http://www.datafirst.uct.ac.za/surveys).
Conditions	Because the Afrobarometer is funded from public resources, its data are a public good. All data are released via the Afrobarometer website and other outlets, along with relevant codebooks. But, to allow initial in-house analysis and publication, data will not be published in the two-year period following the first release of any survey's results. Afrobarometer data are protected by copyright. Authors of any published work based on Afrobarometer data or papers are required to acknowledge the source including, where applicable, citations to data sets posted on the Afrobarometer website.
Costs	None
Tabled outputs	See http://www.afrobarometer.org/results.html See http://www.afrobarometer.org/southafrica.htm
Contact	**For information:** ***Afrobarometer*** website: http://www.afrobarometer.org/index.html ***Prof. Robert Mattes***, Deputy Director of Afrobarometer Email: Robert.Mattes@uct.ac.za **For data:** ***Afrobarometer*** website (Rounds 1 and 2 only): http://www.afrobarometer.org/index.html ***DataFirst*** (Round 1 data available in resource centre)

Name	**Africa Centre Demographic Information System** (ACDIS)
Principal investigator	Africa Centre for Health and Population Studies
Year(s)	Began 2000, ongoing
Area(s) of interest	**General:** demography; health **Specific:** births; deaths; migration; water and sanitation; education; employment; asset ownership; expenditure; self-assessed financial status; hunger; religion; HIV/AIDS
Brief description	**Summary** A demographic surveillance system which maps over 10,000 inhabited households in a 435 square kilometre area, following on average 90,000 persons. Births, deaths, family composition, economic circumstances and health are monitored, and key social and health changes examined. **Purpose** ▪ To describe the demographic, social, and health impacts of a rapidly spreading HIV epidemic in this population. ▪ To measure the burden of disease (including HIV/AIDS) and to measure the effect of HIV intervention strategies. ▪ To better understand the two-way links between health status and economic status. ▪ To estimate prevalence and incidence of HIV at population level within the ACDIS. ▪ To identify cohorts of HIV-infected and uninfected individuals within ACDIS for further prevention and intervention studies. **Methodology** The ACDIS collects demographic, socio-economic and health data from 11,000 households in southern portion of the Umkhanyakude District in KwaZulu-Natal. Household members who are not resident in the area are also tracked and studied. The total population surveyed is 90,000. Data collection began in February 2000, with homesteads visited every six months to register new individuals and households, update demographic variables, and record all births, deaths and migrations. Three survey rounds (two household and one individual surveillance) are conducted each year. Since 2001, socio-economic data has also been collected (once a year), and since 2003 the Centre has offered every adult in the demographic surveillance area the opportunity to be tested for important health problems including HIV. This individual surveillance is repeated annually. The data system processes large, continual flows of data organised in 'episodes' that

	reflect each household and person over time. The type of data collected includes: ▪ Quantitative data on demography, socio-economic status and health. ▪ Qualitative ethnographic surveillance, in which researchers live with families and describe their lives in detail. ▪ Sample surveys using specially designed questionnaires and randomly selected households to yield statistically valid in-depth data. ▪ Verbal autopsy to attempt to estimate cause of death of all members of the demographic surveillance system. ▪ Longitudinal HIV testing: population-based annual serological survey using blood samples. Persons eligible for HIV testing are all resident women aged 15 to 49 and resident men aged 15 to 54. 28% of persons aged 15 to 54 are non-residents members of a household within the ACDIS and 12.5% of this non-resident population was sampled additionally. Data on health status and sexual risk behaviour are collected using structured questionnaire.
How useful to researchers?	Demographic surveillance data is a useful resource for researchers. The detailed and multifaceted types of data collected allow for exploration of multiple risk factors often excluded from survey data. The longitudinal survey design also allows patterns of, for example, disease to be followed within and between households over time. Longitudinal anonymous linked HIV surveillance allows for accurate HIV incidence determination.
Data format	Data is kept in a SQL Server database and provided for analysis in various formats including Stata.
Availability of data descriptions	Metadata and other documentation are available with the data but is not always sufficiently explicit when describing the data collection methods, and some of the subtleties may need to be explained.
Conditions	Due to a new director at the Africa Centre, the data access procedures are undergoing change. Full details of the process of access to ACDIS data will be provided on the Africa Centre website (http://www.africacentre.ac.za) in due course.
Costs	If no new data needs to be collected there are no costs associated with obtaining data, unless special data extraction needs to be done (and cannot be done by the external team) in which case the time of the Africa Centre staff will be charged for.
Tabled outputs	See http://www.africacentre.ac.za/home2.asp?layout=90
Contact	**For information:** ***Africa Centre*** website: http://www.africacentre.ac.za Email: info@africacentre.ac.za **For data:** ***Dr Kobus Herbst***, Deputy Director, Africa Centre for Health and Population Studies Email: kherbst@africacentre.ac.za ***Ms Caterina Hill***, Secretary, Africa Centre for Health and Population Studies Email: chill@africacentre.ac.za Telephone: +27 (0)35 550 7500

Name	**Agincourt Health and Demographic Surveillance System** (AHDSS)
Principal investigator	Agincourt Health and Population Unit, Health and Population Division, School of Public Health, University of the Witwatersrand
Year(s)	Began 1992, ongoing
Area(s) of interest	**General:** demography; health **Specific:** non-communicable disease; persisting malnutrition; child well-being; social support; violence and Injury; population, the environment and health; short and longer-term impacts of illness and death; migration; HIV and households
Brief description	**Summary** Longitudinal health and demographic surveillance, established in 1992 in Agincourt, a rural sub-district of Bushbuckridge District, Mpumalanga Province, South Africa, close to the Mozambique border. The total population then numbered some 58,000 persons, rising to about 70,000 today in 11,700 households across 21 villages covering about 400 sq km. It was part of the Gazankulu 'homeland' during the pre-1994 apartheid era; post-1994 the geographic borders remained politically contested, with the area originally assigned to Limpopo Province and recently moved to Mpumalanga. **Purpose** Due to an absence of vital registration in rural South Africa, a health and demographic surveillance system was introduced in 1992 into the Agincourt sub-district, as part of a programme of decentralised health systems research and development that aimed to inform the rural efforts of a new South African Department of Health. This evolved in 1997 to a site for advanced community-based research.

	Methodology A wide range of health and demographic surveillance data is collected, as well as qualitative data (through in-depth interviews and focus groups). The AHDSS includes annual census and special events updates (systematic recording of all births, deaths and in- and out-migrations). Over the 1992-1996 period the research focused on decentralised health systems R&D; thereafter the work transitioned to a primary research focus on health, population, and social transitions. Examples of data collected include: ▪ Verbal autopsy to determine probable cause of death ▪ Health care utilisation ▪ Asset survey ▪ Prevalence of stroke and related disability through screening questions added to the census update, followed by comprehensive clinical and disability assessment ▪ Cardiovascular risk factor prevalence in a sample over 35 years ▪ Physical and cognitive function in individuals over 55 years ▪ Decision-making and coping strategies through household interviews ▪ Risk factors for severe malnutrition in children ▪ Meanings of life and death in a high HIV setting ▪ Changing patterns of natural resource use in death-affected households.
How useful to researchers?	Demographic surveillance data is a useful resource for researchers. The detailed and multifaceted types of data collected allow for exploration of multiple individual, household and community factors often excluded from survey data. The longitudinal survey design allows temporal analyses.
Data format	Once requests for data are approved (see below), data are available in flat file or relational model format, as a cross-sectional or panel dataset, and usually in Stata.
Availability of data descriptions	Metadata can currently be accessed through research collaborations or a generic '1 in 10' dataset and its accompanying data dictionary.
Conditions	Policies and operating procedures regarding data availability are in development. Currently data can be accessed through research collaborations or the '1 in 10' dataset. Agincourt has a strong track record of collaboration and welcomes collaborations with external parties. Generally proposals are developed in partnership with an external collaborating group and the Agincourt scientific leadership team (contact details below).
Costs	Costs are incurred for data access and time associated with preparation of tailored datasets. These depend on the extent and depth of data requested.
Tabled outputs	See http://hermes.wits.ac.za/www/Health/PublicHealth/Agincourt/pages/publications_workingpapers.htm
Contact	**For information:** ***Agincourt Health and Population Unit*** website: http://hermes.wits.ac.za/www/Health/PublicHealth/Agincourt/index.htm **For data:** Interest in collaboration or the '1 in 10' dataset should be communicated in writing to: ***Prof. Stephen Tollman***, Director, MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt Health and Population Unit) Email: tollmansm@sph.wits.ac.za ***Mr Mark Collinson***, Field Research Manager Email: mark@agincourt.co.za ***Dr Kathleen Kahn***, Senior Researcher Email: Kathleen.Kahn@wits.ac.za

Name	**All Media and Products Survey** (AMPS®)
Principal investigator	South African Advertising Research Foundation (SAARF)
Year(s)	1994 to 2003 (twice per year), 2004 onwards (once per year)
Area(s) of interest	**General:** housing; transport; crime **Specific:** use of the mass media (radio and television, newspapers, consumer magazines, cinema and outdoor); advertising; internet usage; ownership of motor vehicles, ownership of durable items; travelling patterns; use and purchasing of personal and household products; use of financial services; holidays, sport and shopping patterns; personal and household details; activities and lifestyles; life stages; living standards measure
Brief description	**Summary** A single source survey collecting information on media usage, product/brand consumption and demographics, using an in-home face to face computer assisted personal interview plus a leave behind questionnaire. **Purpose** To provide comprehensive, unbiased, valid, reliable and credible media audience and

	product/brand consumption measures. **Methodology** The sample is currently about 12,500 adults (16 years and older) per annum and the interviewing is now conducted during the first half of the year. Previously research was carried out in two waves of interviewing each year during the first and second half of the year. The results were published on a 12-month rolling basis by combing the most recent two interviewing waves. The SAARF Teen AMPS® survey which is carried out periodically is similar to the adult survey and also measures mass media consumption, product usage and activities among South African teenagers aged 12 to 15 years. A simplified study of children of 7 to 11 years old is carried out from time to time. From 2001 the measurement of product data has been via self-completion questionnaires which are left behind so that respondents can fill them in, in their own time. This new methodology is an improvement on the old. In addition, the product categories are now measured on the basis of units used or purchased compared to the self-classification system that was used in the past. ***Sample design*** The sample is designed by using multi-stage area stratified systematic sampling, pre-stratified by population sector (coloured, Indians and whites form one stratum and black another), province, community size, gender and age. The residential addresses of almost all urban communities are listed on Nielsen Media Research's geo-frame. The addresses are arranged within each geographic stratum (area) alphabetically by suburb name, then by street name within suburb and numerically by street number within street. Multiple dwelling units such as flats, cluster houses with the same street number are listed individually. The sample is selected in urban areas by using a random starting point, and then selecting systematically with a fixed interval every Nth address number. Two addresses are selected at each primary sampling point to save travelling costs. In rural areas (villages and farms) sampling points are selected using maps and global positioning systems. Two addresses are selected. One respondent is selected at every address, using gender and age to ensure a proportionate sample by these two variables. Inhabitants at mines and hostels and domestic workers are sampled differently in accordance with their gender composition. *SAARF Crime Monitor*: Crime has an impact on various aspects of marketing and advertising. For example, it influences advertising of personal, home and motor vehicle security. Following this influence on marketing and advertising, SAARF has included the following two questions on crime since SAARF AMPS® 2003: ▪ Have you been a victim of violent crime in the past 12 month (with examples)? ▪ Have you been a victim of non-violent crime in the past 12 month (with examples)?
How useful to researchers?	Some useful data on consumption, ownership of goods, housing conditions, amenities, and the crime module may also be of interest.
Data format	Data are available in ASCII format on CD from SAARF and in SPSS format from SADA.
Availability of data descriptions	Technical reports for 2004, 2005 and 2005 available from the SAARF website (http://www.saarf.co.za). They include information on definition of terms; universe and sample; sampling method; the interview; fieldwork and results; treatment of the data; confidence limits; the questionnaire and other materials used in the interview and the interviewers' instructions
Conditions	Unknown
Costs	The 2006 data cost R13,695.
Tabled outputs	Results from the SAARF AMPS® database are made available in PDF format on CD (at a cost of R864). The use of the mass media is cross tabulated with other variables such as demographics as well as the use of services and products, while top line media results are published on the SAARF website (http://www.saarf.co.za). The SAARF AMPS® Trend Pocket Edition contains a comparative summary of adult data for the most recent five years.
Contact	**For information:** ***SAARF*** website: http://www.saarf.co.za Telephone: +27 (0)11 463 5340/1/2 ***Mr Paul Haupt***, Chief Executive Officer, SAARF Email: paul@saarf.co.za ***Ms Fiona Lister***, Administrator, SAARF Email: fiona@saarf.co.za **For data:** ***SAARF - Ms Fiona Lister*** Email: fiona@saarf.co.za ***SADA*** website (1995 data only): http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0045

Name	**Birth to Twenty** (BT20), also known as 'Mandela's Children'
Principal investigator	University of the Witwatersrand Medical School
Year(s)	Began 1990, ongoing
Area(s) of interest	**General:** health; demography **Specific:** psychosocial development during childhood and adolescence; growth and bone health; infant, child and adolescent well-being; nutrition; urban migration patterns of children and their families; genetic interaction; emergence of sexual and lifestyle risk factors during adolescence; risk behaviour, exposure to sexually transmitted infection, unwanted pregnancy; precursors to the metabolic syndrome; socio-economic status
Brief description	**Summary** The largest and longest running study of child and adolescent health and development in Africa, and one of the few large-scale longitudinal studies in the world. The study documents and explores the socio-economic, socio-political, demographic and nutrition transition that is underway within South Africa and its impact on children and their families. **Purpose** To understand the holistic determination of child and adolescent health and development within Johannesburg-Soweto. **Methodology** For seven weeks between March and June 1990, 3,273 children were born in the metropolitan area of Johannesburg-Soweto and enrolled into a long-term birth cohort study that will follow them and their families for the next 20 years. The following aspects are measured: ▪ Socio-economic status (home structure type; consumer durables; services i.e. electricity, sanitation) ▪ Psychosocial development (questionnaires adjusted for age groups i.e. Bailey's, Child Behaviour Checklist Youth Self Report (Achenbach)) ▪ Growth and bone health (height, weight, circumference; infant, child and adolescent well-being; fatherhood, parenting etc) ▪ Nutrition (dietary intake; food frequency; food security) ▪ Urban migration patterns ▪ Genetic interaction (DNA samples stored in DNA bank for all children and biological caregivers blood samples from age 9) ▪ Risk behaviour in adolescence (questionnaires on STIs and HIV; smoking; sexual activity; exposure to violence; insulin resistance, obesity, cholesterol) Data is collected on 80-90% of the cohort at each stage and there is still contact with 72% of the original birth cohort.
How useful to researchers?	A powerful research tool for researchers wishing to study the socio-economic, socio-political, demographic and nutrition transition that is underway within South Africa and its impact on children and their families. It only relates to children living in a deprived area of Gauteng province.
Data format	Data are available in an Access database.
Availability of data descriptions	Currently, metadata are only accessible through research collaborations. However, the BT20 battery of questionnaires and instruments should be made available in November 2006 on the BT20 website: http://www.wits.ac.za/birthto20/index.php?menuvar=106
Conditions	Data and metadata are only available to researchers engaging in collaborations with the BT20 team which culminate in a publication. The BT20 team welcomes collaborations with external researchers. All proposals are considered by a committee (contact details below).
Costs	There are costs attached to accessing the data depending on the number of variables, data staff time, and general administrative costs involved.
Tabled outputs	See http://www.wits.ac.za/birthto20/index.php?menuvar=102
Contact	**For information:** ***Birth to Twenty*** Website: http://www.wits.ac.za/birthto20/ Telephone: +27 (0)11 488 3602/4 **For data:** To access the data researchers need to send a written proposal to ***Dr Shane Norris***, Project Manager, Birth To Twenty Email: san@global.co.za

Name	**Cape Area Panel Study** (CAPS)
Principal investigator	Southern African Labour and Development Research Unit (SALDRU), University of Cape Town (UCT)
Year(s)	Began 2002, ongoing
Area(s) of interest	**General:** labour market; education; health

	Specific: education; employment; incomes; labour market; population; family dynamics; health; unemployment; youth; HIV/AIDS
Brief description	**Summary** A longitudinal study of the lives of youths and young adults in metropolitan Cape Town. CAPS began in 2002 as a collaborative project of the Population Studies Center in the Institute for Social Research at the University of Michigan and the Centre for Social Science Research at UCT. Other units involved in subsequent waves include SALDRU and the Research Program in Development Studies at Princeton University. Primary funding is provided by the National Institute of Child Health and Human Development of the U.S. National Institutes of Health. **Purpose** To transcend the limits of existing data on adolescence in South Africa in a number of ways: ▪ Data needed to be collected on a range of topics – including schooling, health and relationships – that had been neglected as far as young people are concerned. ▪ Young people need to be understood within a wider range of relationships than simply the co-residential household. Data was required on relationships with other kin, including especially those close kin who were not co-resident with the adolescent. ▪ Qualitative and quantitative research needed to be conducted in synchrony, each informing the design of the other and interpretation of the resulting data collected. ▪ Data needed to be collected on a longitudinal basis, at least through more thorough retrospective questions and ideally through a panel study. **Methodology** Wave 1 collected interviews from about 4,800 randomly selected young people age 14-22 in August-December, 2002. Wave 1 also collected information on all members of these young people's households, as well as a random sample of households that did not have members age 14-22. The youth sample has been interviewed again in 2003-04 (Wave 2), 2005 (Wave 3) and 2006 (Wave 4). Wave 4 also includes a sample of individuals age 50 and over. The study covers a wide range of outcomes, including schooling, employment, health, family formation, and intergenerational support systems. The household questionnaire collected data on all members of the household, covering basic social and demographic variables, education, migration to Cape Town, work and income. The household questionnaire also collected data on the house and on household finances, including income, assets and debt. There was also a short module on shocks - including death, illness or injury, retrenchment and bankruptcy - and on how the household coped financially with these shocks. The young adult questionnaire collected a mix of current and retrospective data on the lives of the respondents. Much of the retrospective data were collected in the form of a life-history calendar, on which was recorded details of schooling, with whom the respondent had lived, and pregnancies and births, year-by-year from birth to the present. Detailed data on schooling, including results in the matriculation (school-leaving) examination, aspirations, homework, transport to schools and other issues was collected. The employment module asked detailed questions about current work, the most recent work that a respondent had done if he or she was not working currently, and the first work that he or she ever did. The module also asked about job search and aspirations. Questions on health and fertility were asked, focusing primarily on past and present sexual behaviour, as well as pregnancy and birth histories and HIV/AIDS. The questionnaire also included modules on kin and the young adults' relationships with them. Modules asked about biological parents who did not live in the same household (and who were therefore not included in the household roster in the household questionnaire); grandparents; young adults' relationships with parents, step-parents and others; home and family environment in which young people had grown up; and limited data on what young people spent their time doing. The literacy and numeracy evaluation (or skills test) comprised a set of 45 questions to test the literacy and numeracy skills of the young adult respondents. The questions ranged from simple to complex. Almost all of the young adults who completed a young adult interview also completed the literacy and numeracy evaluation. ***Sample design*** The CAPS sample is two overlapping samples: a sample of young adults aged between 14 and 22 years-old, and a household sample including both households with and households without young adult members. The CAPS household sample was drawn through a two-stage process. The first stage entailed selecting areas of Cape Town. First, the Enumeration Areas (EAs) used for the 1996 Population Census were divided into three strata according to whether the population of each was predominantly African, coloured or white. EAs with fewer than 25 households were combined with nearby EAs to produce primary sampling units (PSUs) with at least 25 households. A sample of PSUs was selected within each stratum with probability proportional to size. The probability of

	selection was roughly twice as high in African and white areas as in coloured areas. This was based on a target of producing roughly equal numbers of African and coloured young adults, and about half as many white young adult respondents. The second stage of the sampling design entailed selecting households within the selected PSUs. Within each PSU a sample of 25 screener households was drawn using aerial photographs combined with on-site inspection and updating. Secondary households such as backyard shacks on the same property as screened households were added to the screened sample and treated in the same way as all other screened households. All screened households with members aged 14-22 were selected into the final sample of interviewed households. Households without any members aged 14-22 were selected into the final sample with probability around 0.5 in African and coloured areas and with probability around 0.3 in white areas. The sample of young adults was drawn from the sample of households. Up to three young adults were selected for the young adult sample from each household. In cases where there were more than three young adults, the three with the most recent birthdays were selected. Household response rates were high in predominantly African and coloured areas, but were disappointingly low in white areas (a problem common to surveys in South Africa). Once households were interviewed, response rates among young adults were high for all three population groups. ***Weighting*** The public release data sets include sample weights that should be used to adjust for the sample design. Three sample weights are included in the data: ▪ *weightsd* adjusts for three critical elements of the sample design: 1) the intentional oversampling of African and white households; 2) the intentional differential sampling of households with and without young adult household members; and 3) the addition of secondary households (backyard shacks) into the sample of screener households in the field. This weight is incorporated into the other two sample weights. It can be considered as either a household or individual weight if there is no concern with adjustment for non-response. ▪ *weighthr* begins from the first weight and adds additional adjustments for unit non-response at the level of PSUs. In order to avoid unusually large weight being given to households in PSUs with low response rates, a number of PSUs in the same population group cluster which were in close geographic proximity were combined for purpose of adjusting for unit non-response. This should be considered the appropriate household weight if it is considered desirable to adjust for non-response. The implicit assumption in the adjustment for non-response is that the households that responded to the interview do not differ systematically from the households in the same PSU that did not respond. While this assumption is unlikely to be strictly true, it is true that most EAs are relatively homogenous neighbourhoods. Since there is no information on households that did not respond to the interview, there is no way to explicitly examine the extent to which they differ from responding households. ▪ *weightyr* is an individual young adult weight that adds additional adjustment for individual non-response. This adjustment is made by calculating response rates for each combination of single years of age, gender, and population group using the information provided on the household questionnaire. The small number of individuals classified as Indian and other were merged with the coloured group. This approach is taken as an alternative to using young adult response rates at the PSU level, based on the assumption that there is more homogeneity among all white 18 year-old males in Cape Town than there is among the 14-22 year-olds in a given PSU. As discussed above, the response rates were lower for older white and coloured males, so the non-response adjustment is greatest for those groups. This weight makes the same implicit assumption about household level non-response as the previous weight, and adds the additional assumption that within a given age/population group/gender cell there are no systematic differences between respondents and non-respondents. Using the third weight, *weightyr*, the weighted distribution of 14-22 year-olds by population group is within one percentage point of the population group distribution in Cape Town in the 1996 Census. This weight should therefore provide results that are reasonably representative of the young adult population of Cape Town. One of these sample weights should always be used with the CAPS data in order to adjust for the key features of the sample design. Most importantly, the weights all adjust for the systematic oversampling of African and white households, and for the differential sampling probabilities for households with and without young adults. The second and third weights make particular assumptions about non-response, which users may or may not want to assume. When properly weighted, the wave 1 data is representative of all households or all individuals aged 14 to 22 in Metropolitan Cape Town in 2002.
How useful to researchers?	A powerful research tool for researchers wishing to study a wide range of issues affecting young adults aged 14-22. It only relates to children living in metropolitan Cape Town.
Data format	Wave 1 data are available in SPSS, SPSS Portable, NSDstat, Statistica, Stata (version 7 or 8),

	DIF, DBase or SAS format.
Availability of data descriptions	The following documentation is available from the CAPS website (http://www.caps.uct.ac.za/index.html) or when downloading the data from the DataFirst website (http://www.datafirst.uct.ac.za/data_caps.html): ▪ User Guide - gives background and objectives of the CAPS project as well as major themes in the data. ▪ Technical Document for wave 1 - detailed description of methodology. ▪ Metadata for wave 1. ▪ Questionnaires for all 4 waves. ▪ Tips and Do-Files - some tips and do files for use in STATA for working with the CAPS data (only available on the CAPS website). Additional file descriptions and metadata are available when you register to access the data.
Conditions	SALDRU believe that ready access to data is fundamental both for ensuring the ongoing vitality of democracy within countries but also to facilitate impartial scientific collaboration around the world. The data set is in the public domain, but the following conditions apply: ▪ The dataset and its funders should be cited in any publication as follows: The CAPS (Cape Area Panel Study) is produced and distributed by the universities of Michigan and Cape Town, with funding from the National Institutes of Health and the Andrew W. Mellon Foundation. In references, it should be cited as: Cape Area Panel Study, Wave 1 (2002), public use dataset. Produced and distributed by the universities of Michigan and Cape Town with funding from the National Institutes of Health and the Andrew W. Mellon Foundation. ▪ Copies of any seminar/conference papers or published work should be sent to Lynn Woolfrey at Data First: Email: lynn.woolfrey@uct.ac.za ▪ Postal address: Centre for Social Science Research, University of Cape Town, Private Bag, Rondebosch, Cape Town 7701, South Africa. ▪ Users should not attempt to identify specific individuals in the CAPS data. ▪ Users should not redistribute the data to other users – all users should register on the CAPS web site. ▪ Users should notify CAPS staff regarding errors in the data or any features of the data that could compromise respondent confidentiality.
Costs	None
Tabled outputs	See http://www.datafirst.uct.ac.za/resource/b_caps_02-03.pdf for list of publications using CAPS data.
Contact	**For information:** ***CAPS*** website: http://www.caps.uct.ac.za/index.html ***DataFirst*** website: http://www.datafirst.uct.ac.za/data_caps.html **For data:** ***DataFirst*** website: http://www.datafirst.uct.ac.za/data_caps.html

Name	**Community Survey**
Principal investigator	Stats SA
Year(s)	2007
Area(s) of interest	**General:** demography; housing conditions; labour market; social welfare; education; health **Specific:** households; disability; fertility; mortality; education; the expanded public works programme; economic activity; access to service delivery
Brief description	**Summary** A large scale Census replacement/inter-censal survey, representative at municipality level containing similar questions to 2001 Census, which in addition to key demographic indicators such as fertility, mortality and migration, will generate some of the key indicators that the government prioritised in its 10-year review document. It will also generate some of the variables required for budget allocation to municipalities, as well as some of the UN's MDG indicators. Although not yet available, it is worth highlighting as a dataset for the future. **Purpose** Some statistical agencies are able to use administrative records to gather information. However, in South Africa these sorts of records cannot be used to provide information at the municipal level. In particular, geographic referencing on these records is not consistent and compatible with Stats SA's geographic boundaries. Thus it is difficult to use to attach information to specific municipalities. In some countries, a combination of the population register and household sample surveys is used to obtain municipal-level data between Censuses. However, the current South African population register does not record all non-citizens in the country; does not allow for geographical analysis; and does not include some important demographic characteristics, such as population group, place of birth or place of residence. The records of addresses are also not regularly updated. In some cases, only postal addresses are recorded, which further limits

	attachment of information obtained to geographical area. In the absence of viable alterative sources of data for use between Censuses, Stats SA has introduced a community survey to provide information (at municipal level) in the inter-censal period. **Methodology** Although the community survey is based on a sample, the nature and size of the sample (approximately 280 000 households), as well as the sophisticated methodology employed to create it, will enable key information to be attached to each of South Africa's municipalities. Demographic information will include date, month and year of birth; age, sex, marital status and population group; relationship to head of household; type of disability, if any, and whether receiving a disability grant; children born, children alive and last child born; and whether anybody in the household has died in the 12 months prior to the survey and, if so, sex and age of the deceased and cause of death. Questions will also be asked about school attendance, level and field of education, and participation in the expanded public works programme. In terms of economic activity, the survey will probe whether the respondent has had any work in the prior week; reason for not working; active steps to seek work; availability to start work; status in employment; whether in the formal or informal sector; occupation; industry; participation in public works; and income. Measuring access to social services is an important output of the survey and questions will be asked about type of housing and number of rooms; access to water, energy, sanitation and refuse removal; and access to different modes of communication. ***Sample design*** A systematic simple random sampling procedure was used to select the enumerator areas (EAs). The EAs within each municipality were ordered by type (farm, hostel, industrial area, informal settlement, smallholding, tribal settlement, urban settlement, or vacant land). They were then further differentiated by geographic type (rural formal, tribal area, urban formal and urban informal). To select enumerator areas for the survey: ▪ All EAs in municipalities with less than 30 areas were automatically selected; ▪ For those municipalities with more than 30 EAs, the sample selection used a fixed proportion of 19% of the total number of areas. However, if less than 30 areas in a municipality were selected, the sample in the municipality was increased to 30 enumerator areas; ▪ The selection of the subsample will be based on a fixed proportion of 10%. A constraint is imposed on small enumerator areas of less than 10 dwellings, in which case the selection is increased up to 10 dwellings units; and ▪ All households within the selected dwelling units will be covered.
How useful to researchers?	N/A
Data format	N/A
Availability of data descriptions	N/A
Conditions	N/A
Costs	N/A
Tabled outputs	N/A
Contact	**For information:** ***Stats SA User Information Services*** Email: info@statssa.gov.za Telephone: +27 (0)12 310 8600 **For data:** ***Stats SA*** (when available)

Name	**Demographic and Health Survey**
Principal investigator	Department of Health (DoH)
Year(s)	1998
Area(s) of interest	**General:** demography; health **Specific:** age; gender; level of education; marital status; use of health services; economic activity; unemployment, employment and self-employment; fertility; childhood mortality; maternal and child health; health behaviours; knowledge of contraceptives; oral health; occupational hazard; chronic diseases; blood pressure; lung function
Brief description	**Summary** A national survey designed to measure various aspects of health and healthcare. **Purpose** ▪ To collect data as part of the National Health Information System of South Africa. The survey results were intended to assist policymakers and programme managers in evaluating and

	designing programmes and strategies for improving health services in the country. ▪ To continue the information base for health and population development programme management through accurate and timely data on a range of demographic and health indicators. ▪ To provide baseline data for monitoring programmes and future planning. ▪ To build research and research management capacity in large-scale national demographic and health surveys. **Methodology** Data on a sample of women in households and a sub-sample of an additional adult in half of the households was collected by survey questionnaire. ***Sample design*** The sample was designed to be a nationally representative probability sample of approximately 12,000 interviews with women between the ages of 15 and 49. The country was stratified into the nine provinces and each province was further stratified into urban and non-urban areas. In addition the Eastern Cape was stratified into five health regions, with each health region stratified into urban and non-urban areas. The sampling frame was the list of approximately 86,000 enumeration areas (EAs) created for the 1996 Census. Within each stratum a two-stage sample was selected. The Primary Sampling Units (PSUs) corresponded to the EAs and were selected with probability proportional to size, the size being the number of Census visiting points in the EA. This led to a total of 972 PSUs being selected for the survey (690 in urban areas and 282 in non-urban areas). In urban EAs, ten households were selected, while in non-urban EAs 20 households were selected. This resulted in a total of 12,860 households being selected throughout the country. Every second household was selected for the adult health survey. In this second household, in addition to interviewing all women aged 15-49, interviewers also interviewed all adults aged 15 and over. ***Geography*** Data can be analysed at national, provincial level and urban and the non-urban areas.
How useful to researchers?	A useful dataset on demography and health, with an innovative adult health module added to the more typical module on maternal health.
Data format	Data are available in SPSS format.
Availability of data descriptions	Little documentation is available on the SADA website. The full report describing the methodology is on the DoH website: http://www.doh.gov.za/facts/1998/index.html
Conditions	SADA conditions of use apply (see notes). DoH conditions may also apply.
Costs	None
Tabled outputs	See http://www.doh.gov.za/facts/1998/index.html for the full report.
Contact	**For information:** ***DoH - Dr Lusanda Mahlasela*** Email: MahlaL@health.gov.za **For data:** ***SADA*** website: http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0033 ***DataFirst*** (available in resource centre)

Note: A further round of the Demographic and Health Survey was conducted in 2003 but data have not yet been released.

Name	**Financial Diaries Project**
Principal investigator	Southern African Labour and Development Research Unit (SALDRU), University of Cape Town
Year(s)	2003/2004
Area(s) of interest	**General:** economy **Specific:** income; expenditure; financial management; financial decision making; financial markets in townships
Brief description	**Summary** A household survey which gives a picture of the financial inflows and outflows of poor households through data on income, consumption, savings, lending and investment. The level of detail is much higher than is in larger surveys and complements the information available from larger data bases. The project is funded by FinMark Trust, the Ford Foundation and the Micro Finance Regulatory Council. **Purpose** To help government and the financial industry provide financial services to the poor by providing an understanding of the financial needs of the poor.

	Methodology The study is a year-long household survey based on fortnightly interviews in Diepsloot (Gauteng), Langa (Western Cape) and Lugangeni (Eastern Cape). In total, 166 black households were involved, drawn from three different areas, from a range of dwelling types and wealth categories. The study uses a combination of closed and open-ended questionnaires. The first three initial questionnaires are structured and gather information on household demographics, physical assets, typical income and expenditure patterns, historical and current employment, and lastly current and previous use of financial instruments. There are roughly 28 pre-defined financial instruments that each has their own questionnaires that define different aspects of the instrument. Each existing financial instrument receives its own financial device code against which cash flows are captured in the future. These three interviews not only allow the household to become more comfortable with the fieldworkers, but are also used to create an initial balance sheet position, as well as a typical monthly cash flow statement. From that point, the households are interviewed every other week for a year, capturing every cash flow come into and out of the household, including income, expenditure, changes in physical assets, servicing financial instruments, initiating financial instruments, etc. To facilitate the collection of data, the data from the first initial questionnaires will be used to produce a diaries questionnaire that is specific to each household. This is used to both prompt memory, aid data collection and to save the respondents time and patience. The diary process is enabled by a specially conceived and built Access database and the consistent weekly capture of data. Each week the fieldworker also fills in a journal whereby he notes various observations, events or comments made by the respondent that are not captured elsewhere in the diary questionnaire. ***Sample design*** A participatory wealth ranking method was used. The first step was to consult with community leaders and ask how they would divide their community. Within each type of areas, representative neighbourhoods of about 100 households each were randomly chosen. Townships in South Africa are organised by street – with each street or zone having its own street committee. The street committees are meant to know everyone on their street and to serve as stewards of all activity within the street. Each street committee in each area was invited to a central meeting and asked to map their area and give a roster of household names. Following the mapping, each area was visited and the maps and rosters were checked by going door to door with the street committee. Two reference groups were then selected from the street committee and senior members of the community with between four and eight people in each reference group. Each reference group was first asked to explain how they define a poor household versus those that are well off. Following this discussion, each reference group then ranked each household in the neighbourhood according to their perceived wealth. Households could be put into as many different wealth piles as felt appropriate. Only households known by both reference groups were kept in the sample. A score was assigned to each household in a particular pile. The scores were created by dividing 100 by the number of piles multiplied by the level of the pile. This means that if the poorest pile was number 1, then every household in the pile was assigned a score of 100, representing 100% poverty. If the wealthiest pile was pile number 6, then every household in that pile received a score of 16.7 and every household in pile 5 received a score of 33.3. An average score for both reference groups was taken for the distribution. Analysis of how consistent the rankings were between the two reference groups was carried out. The sample was divided into three different wealth categories depending on the household's overall score. Making a distinction between three different categories of wealth prevents over-stratifying the sample. A sample of 60 households was then drawn randomly from each wealth category.
How useful to researchers?	A useful resource for those wishing to explore the financial inflows and outflows of poor households using the diary method.
Data format	Data are available in SPSS, SPSS Portable, NSDstat, Statistica, Stata (version 7 or 8), DIF, DBase or SAS format.
Availability of data descriptions	The following documentation is available on the Financial Diaries website (http://www.financialdiaries.com) or on the DataFirst website (http://www.datafirst.uct.ac.za/data_fdiaries.html): ▪ Diary methodology - literature review and description of survey methodology. ▪ Questionnaires ▪ Also available only on the DataFirst website are: ▪ User guide ▪ Metadata Additional file descriptions and metadata are available when you register to access the data.
Conditions	SALDRU believe that ready access to data is fundamental both for ensuring the ongoing vitality of democracy within countries but also to facilitate impartial scientific collaboration around the world. The data set is in the public domain, but the following conditions apply: ▪ The dataset and its funders should be cited in any publication as follows: Financial Diaries

	Survey 2004, public use dataset. Produced and distributed by the University of Cape Town with funding from the Ford Foundation, FinMark Trust and the Micro Finance Regulatory Council of South Africa. ▪ Copies of any seminar/conference papers or published work should be sent to Lynn Woolfrey at Data First: Email: lynn.woolfrey@uct.ac.za ▪ Users should not attempt to identify specific individuals in the Financial Diaries data. ▪ Users should not redistribute the data to other users – all users should register on the DataFirst website. ▪ Users should notify Financial Diaries staff regarding errors in the data or any features of the data that could compromise respondent confidentiality.
Costs	None
Tabled outputs	See http://www.financialdiaries.com/key_findings.htm or http://www.datafirst.uct.ac.za/data_fdiaries.html for key findings.
Contact	**For information:** ***Financial Diaries*** website: http://www.financialdiaries.com ***DataFirst*** website: http://www.datafirst.uct.ac.za/data_fdiaries.html ***Ms Daryl Collins***, Research Associate, SALDRU or ***Ms Brenda Adams***, SALDRU Email: dcollins@commerce.uct.ac.za or badams@commerce.uct.ac.za **For data:** ***DataFirst*** website: http://www.datafirst.uct.ac.za/data_fdiaries.html

Name	**General Household Survey** (GHS)
Principal investigator	Stats SA
Year(s)	2002, 2003, 2004, 2005
Area(s) of interest	**General:** demography; housing; labour market; education; health; social welfare **Specific:** demographic information (name, sex, age, population group); biographical information (education, health, disability, welfare); activities related to work and unemployment; trips undertaken in the 12 months prior to the survey interview; household information (type of dwelling, ownership of the dwelling and other assets, electricity, water and sanitation, environmental issues, services, transport, expenditure); comprehensive coverage of living conditions
Brief description	**Summary** A national annual survey designed to measure various aspects of the living circumstances of South African households. The five broad areas covered by the GHS are: education, health, activities related to work and unemployment, housing and household access to services and facilities. **Purpose** To measure the level of development and the performance of government programmes and projects. The indicators measured in the 13 nodal areas identified for the Integrated Rural Development Strategy (IRSD) formed the subject matter for the survey. **Methodology** Information was collected on various aspects of the living circumstances of members from over 30,000 households across the country. The sampled dwelling units in each of the nine provinces were visited by field staff employed and trained by Stats SA, and a questionnaire was completed through face-to-face interviews for each household visited. ***Sample design*** Enumeration Areas (EAs) that had a household count of less than 25 were omitted from the Census frame that was used to draw the sample of Primary Sampling Units (PSUs) for the Master Sample. Other omissions from the Master Sample frame included all institution EAs except workers' hostels, convents and monasteries. EAs in the Census database that were found to have less than sixty dwelling units during listing were pooled together to form PSUs. The Master Sample is a multi-stage stratified sample. The overall sample size of PSUs was 3,000. The explicit strata were the 53 district councils. The 3,000 PSUs were allocated to these using the power allocation method. The PSUs were then sampled using probability proportional to size principles. The measure of size used was the number of households in a PSU as calculated in the Census. The sampled PSUs were listed with the dwelling unit as the listing unit. From these listings systematic samples of dwelling units were drawn. These samples of dwelling units formed clusters. The size of the clusters differed depending on the specific survey requirements. The GHS used one of the clusters that contained ten dwelling units. ***Weighting***

	A two-stage weighting procedure was applied to the GHS. The PSU inclusion probability is given by $P_{PSU} = n_{PSU} / N_{PSU} * n_S$ where n_{PSU} is the number of households constituting the selected PSU during census fieldwork, n_S is the number of PSUs per stratum, in this case the District Council (DC) and N_{PSU} is the number of households constituting the selected stratum during census fieldwork. The household inclusion probability per PSU is given by $P_{HH} = n_{HH} / H_{HH}$ Where n_{HH} is the number of selected dwelling units per PSU, H_{HH} is the number of dwelling units in the PSU in question at a particular time different from the census time. The non-response adjustment factor is given by $1 / r_{HH}$, where r_{HH} is the response rate and is given by $r_{HH} / n_{RESP} / n_T$ where n_{RESP} is the number of responding households and n_T is the total number of visited households (in the sampled dwelling units) per PSU. The design weights adjusted for nonresponse are now given by $W_{HH} = 1 / (P_{PSU} * P_{HH} * r_{HH})$ A SAS macro called CALMAR was used to benchmark W_{HH} to the population estimates. The midyear population estimates were adjusted to give population estimates for September 2004 (when survey fieldwork took place). ***Geography*** In early surveys, data are available for analysis at national and provincial level, with a rural/urban split. From 2005 the GHS also presents data at district municipality (42 areas), cross-border district municipality (5) and major metropolitan area (6) levels. ***Data revision*** Stats SA has revised the population model to produce mid-year population estimates in the light of recent mortality data released during 2005 per five-year age group. The benchmarks for all previous GHSs have been adjusted accordingly. Data for GHS 2002, 2003, 2004 and 2005 presented in the 2005 GHS release are therefore comparable. Care must be exercised when using the data for years prior to 2005.
How useful to researchers?	A useful resource if caution is exercised. Consult the literature on weaknesses in the surveys (Meth, 2006a; Meth, 2006b) before drawing hard and fast conclusions. The data do not contain sections on migrant workers. Non-response distribution by income/class is not specified. There are reliability problems with the sub-provincial data.
Data format	Data are available in flat, ASCII, fixed field format, with one line of given length per record from Stats SA and in SPSS format from SADA.
Availability of data descriptions	A comprehensive metadata document is available on the data CD from Stats SA. Little documentation is available on the SADA website, but more may be available when you request data.
Conditions	The information products and services of Stats SA are protected in terms of the Copyright Act, 1978 (Act 98 of 1978). As the State President is the holder of State copyright, all organs of State enjoy unhindered use of the Department's information products and services, without a need for further permission to copy in terms of that copyright. Where a copy of the information is made available to any third party outside the State, the third party must be made aware of the existence of State copyright and ownership of the information by the State. The State (through Stats SA) retains the full ownership of its information, products and services at all times. Access to information does not give ownership of the information to the client. The use of any data is subject to acknowledgement of Stats SA as the supplier and owner of copyright. Users may apply or process the data, provided Stats SA is acknowledged as the original source of the data; that it is specified that the analysis is the result of the user's independent processing of the data; and that neither the basic data nor any reprocessed version or application thereof may be sold or offered for sale in any form whatsoever without prior permission from Stats SA. SADA conditions also apply to data obtained from SADA.
Costs	R 300 for a CD containing the data and documentation from Stats SA or free from SADA.
Tabled outputs	See http://www.statssa.gov.za/publications/P0318/P0318July2005.pdf for 77 pages of tables on population; education; attendance at educational institution; health; social welfare; population of working age; dwellings and services; assets; income and expenditure; transport.

Contact	**For information:** ***Stats SA User Information Services*** Email: info@statssa.gov.za Telephone: +27 (0)12 310 8600 **For data:** ***Stats SA User Information Services*** Email: info@statssa.gov.za Telephone: +27 (0)12 310 8600 ***SADA*** website: http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0135 (2002) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0136 (2003) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0142 (2004) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0145 (2005) ***DataFirst*** (available in resource centre)

Name	**Greater Durban Metropolitan Area Large Manufacturing Firm Survey**
Principal investigator	School of Development Studies (SDS), University of KwaZulu-Natal and the World Bank
Year(s)	2002/2003
Area(s) of interest	**General:** economy **Specific:** firm information; production; finance; purchase; sales and marketing; human resources; administration and legal issues; exports
Brief description	**Summary** A representative sample of 225 manufacturing firms in the Greater Durban Metropolitan Area (GDMA). This data, with similar data collected in the Johannesburg area, represents the only firm level data of manufacturing in South Africa. The large firm survey is the result of an agreement entered by the Durban Unicity Council in 2000 (with USAID) to fund a World Bank technically-supported survey of firms in the greater Durban region. **Purpose** ▪ To provide an up-to-date assessment and refined identification of the current issues facing manufacturing firms, the extent to which these matter and the ways in which firms adapt to these. ▪ To shed light on how firms select their factors of production and allocate their resources over an immediate-, short- and medium-term horizon. ▪ To capture how the firms respond to specific developments at national and local levels. **Methodology** The methodology applied for the large manufacturing firms survey undertaken in the GDMA between May 2002 and April 2003 by the Bureau of Market Research (BMR) at the University of South Africa closely follows that of the Johannesburg survey. However, additional modules (e.g. a port questionnaire) and sectors (e.g. the informal sector and tourism) were covered in the GDMA survey exercise. The sample is highly representative, with about one in three existing large manufacturing firms interviewed in the GDMA. ***Sample design*** BMR used a series of sources to draw sampling frames (business registers, the South African Chamber of Commerce database, the Unemployment Insurance Fund database, the KwaZulu-Natal Tourism database, a port database compiled by Prof. Jones, University of Natal, and the Durban Yellow Pages). Because no sample frame is comprehensive enough to include all firms operating in the Durban Metropolitan Area, various sample frames had to be utilised. These sample frames were used to contact firms randomly by telephone in order to set up appointments with the managing directors, managers or owners of the firms. The study was constructed to stratify industry by type, employment size group (small: 50-99 employees, medium: 100-199, and large: 200 and above) and geographic area. The sample frames had some limitations in this regard: they lacked employment size group classifications, they had a limited number of firms for certain sectors and they showed geographic location problems. In some cases information on firms was outdated (i.e. non-existing or a change in contact details). Within these multi-strata, simple random sampling was performed. This sample design introduces design effects that must be removed using appropriate statistical weights. ***Weighting*** Weights have to be applied to ensure that the observations from the 225 GDMA firms interviewed across sectors are representative of the GDMA population of firms. Further adjustments had to be made after fieldwork to take into account relocations and downsizing of firms. These arise from discrepancies between databases and *expost* observations. This type of difficulty appeared for two sectors, Leather and Footwear and Iron and Steel. For the former, downward adjustments to a first

	set of weights were made for mid- to large-size firms to take into account an overall real downsizing of 80%. For the latter sector, an adjustment was made to shift a mid-size firm (within the large firm sectoral subset) into the largest firms group. This is because no firm was surveyed in the Iron and Steel sector that had a number of employees equal to or in excess of 200. Here only the size class, not the number of employees (173), was changed. An analysis by size without a case for this sector would have been unreliable otherwise. The final set of weights was arrived at further by following a process of consultation with industry experts. The total frame universe has been tied to 600 final firms and weights derived accordingly.
How useful to researchers?	One of only two datasets about manufacturing in South Africa. It is only for the GDMA.
Data format	Data are available in SPSS version 11.5 for Windows format.
Availability of data descriptions	The following documentation is available from the SDS website (http://sds.ukzn.ac.za/default.php?11,0,0,0,0): ▪ Background to available data ▪ Coding instructions ▪ Questionnaires
Conditions	None specified
Costs	None
Tabled outputs	Devey et al. (2005) See http://sds.ukzn.ac.za/files/RR64.pdf
Contact	**For information:** ***SDS - Mr Imraan Valodia*** Email: valodia@ukzn.ac.za ***SDS - Dr Myriam Velia*** Email: veliam@ukzn.ac.uk **For data:** ***SDS*** website: http://sds.ukzn.ac.za/default.php?11,0,0,0,0

Name	**HIV and Sexual Behaviour Among Young South Africans: A National Survey of 15-24 Year Olds**
Principal investigator	Reproductive Health and HIV Research Unit (RHRU), University of the Witwatersrand
Year(s)	2003
Area(s) of interest	**General:** health **Specific:** sexual risk behaviour; risk behaviour; HIV prevalence; knowledge of and attitudes towards HIV/AIDS
Brief description	**Summary** A nationally representative household survey on HIV and sexual behaviour among young people. **Purpose** ▪ To establish the prevalence of HIV and related risk behaviours among young people aged 15-24 years. ▪ To examine the extent of young people's exposure to LoveLife (a national programme providing information, services and media-based campaigns about HIV). ▪ To monitor changes in HIV prevalence and sexual behaviour among South African youth. **Methodology** Of eligible and enumerated youth, 77.2% completed an interview. This sample of 11,904 is 68.2% of all eligible youth. The survey examines: ▪ Socio-demographic information ▪ Comprehensive sexual behaviour information and sexual history information ▪ Other behavioural information ▪ Knowledge, attitudes and norms towards HIV and AIDS ▪ Information on young people's exposure and responses to LoveLife ▪ HIV testing by collecting an oral fluid sample using the Orasure Oral Specimen Collection Device. This sample was then tested for HIV-1/2 antibodies using the Vironostika ELISA. HIV testing was anonymous although results were linked through a unique id number. ***Sample design*** A three-stage, disproportionate, stratified sample of all young people aged 15-24 years in the nine provinces of South Africa was used. The 2001 Census Enumeration Areas (EAs) were used as the primary sampling unit. One eligible youth in each household was randomly selected and interviewed.
How useful to	One of several important datasets relating to HIV/AIDS in South Africa.

researchers?	
Data format	Data (with unique id number) are available in Stata format.
Availability of data descriptions	The following documentation is available when accessing the data: ▪ Full questionnaire with coding ▪ Detailed report of sample design, fluid sampling etc
Conditions	All analysis of data must take place as part of a collaboration with RHRU researchers. Standard operating procedures on terms and conditions of data use must be signed. RHRU welcomes collaborations with researchers, and if requested will send a list of topics which would benefit from secondary analysis of the survey data.
Costs	None
Tabled outputs	See http://www.rhru.co.za/images/Docs/Fact%20Sheet.pdf
Contact	**For information:** ***RHRU*** website: http://www.rhru.co.za ***Dr Catherine MacPhail***, Senior Researcher, RHRU Email: c.macphail@rhrujhb.co.za Telephone: + 27 (0)11 989 9200 or +27 (0)83 441 5415 **For data:** External researchers should send a protocol, brief hypothesis and variables requested to ***Dr Catherine MacPhail***. If the research changes or different variables are required, an update letter should be sent.

Name	**HSRC National Survey**
Principal investigator	HSRC
Year(s)	1998 (November), 1999 (March)
Area(s) of interest	**General:** attitudes **Specific:** gender, age, highest educational qualification, employment status, marital status, population group, household income, trends in political affiliation, democratic development, national priorities, perceptions of economic issues, government performance and service delivery, gender issues and socio-political issues
Brief description	**Summary** A public opinion survey in the same style as the Omnibus surveys (see below). **Purpose** ▪ To investigate public attitudes about national priorities, social issues, political parties and the government's service delivery programme, and to reflect the extent of and attitudes towards the transition from the previous apartheid system to a constitutional democracy. ▪ To provide regular and reliable data and analysis of various national social priority issues. ▪ To help government, parliament, civil society and the various political parties in formulating policy priorities for the future. **Methodology** 2,200 people aged 18 years and older were interviewed using survey questionnaires. ***Sample design*** The population was stratified according to nine socio-economic area types. The allocation was roughly proportional to the adjusted 1991 Census figures. Multistage cluster sampling was used to draw the respondents, using the adjusted 1991 population census figures as a sampling frame. Census enumerator areas and similar areas were used as the clusters in the penultimate sampling stage, from which an equal number of households were drawn. All the clusters were drawn from the final clusters with equal probability. The respondents were drawn at random from qualifying household members.
How useful to researchers?	This provides historical data about attitudes in South Africa and complements the more recent Social Attitudes Surveys.
Data format	Data are available in SPSS format.
Availability of data descriptions	Documentation is only available in hard copy.
Conditions	SADA conditions of use apply (see notes).
Costs	None
Tabled outputs	Unknown
Contact	**For information:** Unknown **For data:** ***SADA*** website: http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0105 (1999) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0106 (1998) ***DataFirst*** (available in resource centre)

Name	**Income and Expenditure Survey** (IES)
Principal investigator	Stats SA
Year(s)	1995, 2000
Area(s) of interest	**General:** economy **Specific:** demographic information (name, sex, age, population group, employment); area of purchase of goods and services; information regarding dwellings; cost of housing; cost of domestic workers; food; alcoholic and non-alcoholic beverages; cigarettes; personal care; other household goods; household services; household fuel; clothing and footwear; furniture and appliances; health services and medical requisites; transport; computing and telecom equipment; cost of; communication; education; reading matter and stationery; recreation, entertainment, sport; household production and consumption of home produce; debts; income from all sources
Brief description	**Summary** A national household survey tracking all the details of receipts of cash, goods and services and those related to the purchase of goods and services for the household's own consumption. The survey forms the basis for the determination of the basket of consumer goods and services used for the calculation of the Consumer Price Index (CPI). Although primarily intended to provide weights for the CPI, the IES gathers so much information on incomes and expenditure, that, inevitably, much use of it has been made in poverty and inequality studies. **Purpose** ▪ To determine the average expenditure patterns of households in different areas of the country. ▪ To provide an independent source of information on private final consumption, which can be used for the generation of National Accounts such as the Gross Domestic Product. **Methodology** (descriptions refer to the 2000 IES, but 1995 is similar) The survey was conducted in October 1995 and October 2000. The 2000 IES was based on the sample for the rotating panel of the twice yearly Labour Force Survey. The survey was done by means of an interview with the household head or a responsible adult and the questionnaire was completed by the enumerator during this interview. In cases where the household requested to complete the questionnaire themselves, it was dropped off by the enumerator, and the completed questionnaire was collected at a second visit. Depending on the nature of the transaction, respondents were required to recall over periods ranging from 1 month (non-durable consumption, for example) to 12 months (durables and other major expenditure), ***Sample design*** The 2000 IES used a Master Sample based on the 1996 Census of enumeration areas (EA) and the estimated number of dwelling units from the 1996 Census. All 3,000 PSUs included in the Master Sample were used in the IES. A PSU is either one EA or several EAs when the number of dwelling units in the base or originally selected EA was found to have less than 100 dwelling units. Each EA had to have approximately 150 dwelling units but it was discovered that many contained less. Thus, in some cases, it has been found necessary to add EAs to the original EA to ensure that the minimum requirement of 100 dwellings, in the first stage of forming the PSUs, was met. The size of the PSUs in the Master Sample varied from 100 to 2,445 dwelling units. Special dwellings such as prisons, hospitals, boarding houses, hotels, guest houses (whether catering or self-catering), schools and churches were excluded from the sample. Explicit stratification of the PSUs was done by province and area type (urban/rural). Within each explicit stratum, the PSUs were implicitly stratified by District Council, Magisterial District and, within the magisterial district, by average household income (for formal urban areas and hostels) or EA. The allocated number of EAs was systematically selected with probability proportional to size in each stratum**.** Once the PSUs included in the sample were known, their boundaries had to be identified on the ground. After boundary identification, the next stage was to list accurately all the dwelling units in the PSUs. The second stage of the sample selection was to draw from the dwelling units listing whereby a systematic sample of 10 dwelling units was drawn from each PSU. As a result, approximately 30,000 households (units) were interviewed. However, if there was growth of more than 20% in a PSU, then the sample size was increased systematically according to the proportion of growth in the PSU. ***Weighting*** For the 2000 IES, the initial weights (household weights), based on the sample design, were equal to the inverse of the probability of selection. That is: Household weight = $1 / (P_1 * P_2)$

	Where P_1 = ((Census no. of households in PSU) * (No. of PSUs in stratum)) / (Census total no. of households per stratum) And P_2 = Sample size [ie, 12 dwelling units per PSU] / No. of dwelling units in the selected PSU The initial weight for each member of the household was the same as the weight for the household itself. Further adjustment factors were then calculated within PSUs to account for non-response. To adjust for under-enumeration and to align survey estimates with independent population estimates, the weights were calibrated against person benchmarks. A software package called CALMAR was used to perform this calibration. Using an iterative procedure, CALMAR adjusted the weights so that person estimates conformed as closely as possible to external person benchmarks. Gender, race and age group parameters were used for the person cross-classification of the population.
How useful to researchers?	Although useful, handle with caution as non-response at the top end of the distribution is quite serious and correction for non-response is probably inadequate. Under-reporting of income and expenditure elsewhere is also a problem. See Meth (2006a; 2006b) and Simkins (2003) for further details.
Data format	The IES data are available as flat, ASCII, fixed field format files, with one line of given length per record from Stats SA and in SPSS format from SADA.
Availability of data descriptions	A comprehensive metadata document is available on the data CD from Stats SA. Little documentation is available on the SADA website, but more may be available when you request data. Questionnaires and code lists are available on the DataFirst website (http://www.datafirst.uct.ac.za/surveys.html).
Conditions	The information products and services of Stats SA are protected in terms of the Copyright Act, 1978 (Act 98 of 1978). As the State President is the holder of State copyright, all organs of State enjoy unhindered use of the Department's information products and services, without a need for further permission to copy in terms of that copyright. Where a copy of the information is made available to any third party outside the State, the third party must be made aware of the existence of State copyright and ownership of the information by the State. The State (through Stats SA) retains the full ownership of its information, products and services at all times. Access to information does not give ownership of the information to the client. The use of any data is subject to acknowledgement of Stats SA as the supplier and owner of copyright. Users may apply or process the data, provided Stats SA is acknowledged as the original source of the data; that it is specified that the analysis is the result of the user's independent processing of the data; and that neither the basic data nor any reprocessed version or application thereof may be sold or offered for sale in any form whatsoever without prior permission from Stats SA. SADA conditions also apply to data obtained from SADA.
Costs	R 300 for a CD from Stats SA or free from SADA.
Tabled outputs	▪ See http://www.statssa.gov.za/publications/P0111/P01112000.pdf for 80 pages of tables on the 2000 IES, mainly cross-tabs of expenditure by various household or personal characteristics. ▪ See http://www.statssa.gov.za/publications/statsdownload.asp?PPN=EarningSpending&SCH=2293 for the 1995 IES.
Contact	**For information:** ***Stats SA User Information Services*** Email: info@statssa.gov.za Telephone: +27 (0)12 310 8600 **For data:** ***Stats SA User Information Services*** Email: info@statssa.gov.za Telephone: +27 (0)12 310 8600 ***SADA*** website: http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0073 (1995) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0140 (2000) ***DataFirst*** (available in resource centre)

Note: A further round of the IES is currently underway, but uses 'diary' rather than 'recall' methodology. Approximately 24,000 households have been visited countrywide and all money spent each day is recorded in a diary. The underlying assumption in the five-year frequency is

that consumption patterns of households remain constant over such a period. However, in the light of technologically rapid changes, resulting in the introduction of new goods and services on the market, that assumption can no longer be justified. In line with international best practice, Stats SA has decided to move to a three-year cycle for the updating of the CPI basket. Therefore, an IES will be conducted every three years. The 2005/06 IES will be followed by a second round in 2008.

Name	**Khayelitsha/Mitchell's Plain Survey** (KMP)
Principal investigator	Southern Africa Labour and Development Research Unit (SALDRU), University of Cape Town
Year(s)	2000
Area(s) of interest	**General:** labour market **Specific:** age; racial classification; educational attainment; language; religion and health; migration (place of origin, relocation and destination); intergenerational mobility; employment history; wage employment; unemployment; self-employment; non-labour force participants; casual work; reservation wages; savings, borrowing and grants and investment income attempted to capture income derived from sources other than work; perceptions of distributive justice
Brief description	**Summary** A small survey with a special focus on labour market issues. This was a precursor to the Cape Area Panel Study planned for 2002 which focused on youth in the area. It was decided to target the magisterial district of Mitchell's Plain within the Cape Metropole for the survey. This decision was informed by data gleaned from the 1996 Census which revealed that Mitchell's Plain contained almost 30% of the population in the Cape Metropolitan Council area. It straddled the two cities of Cape Town and Tygerberg and housed nearly 74% of the African population and over 20% of the coloured metropolitan population. It included the townships of Khayelitsha, Langa, Gugulethu and Nyanga as well as informal settlements such as Crossroads and Browns Farm. **Purpose** To capture the multiple activities in which household members engage as part of their livelihood strategies, given the high poverty and unemployment rates in South Africa. National sample surveys tend not to have the flexibility to do this. **Methodology** The sample was designed to represent all adults (18 years of age and older) in the Mitchell's Plain Magisterial District. Two questionnaires were administered. The *household questionnaire* was aimed at establishing the household roster with the usual questions on age, gender and relationships. It was divided into two sections covering those aged 18 and older and those younger than 18. For the latter a separate set of questions covering education, health and work status was included. The *adult questionnaire* aimed to fit the international standard approach on the labour force by allocating the labour market status of 'employee' to all those 'at work' (for profit or family gain, in cash or in kind). Respondents were asked about all income-earning activities and were not allocated into particular labour market categories during the process of the interview. The *emergency questionnaire* was a severely truncated version of the adult questionnaire covering age, educational achievement, employment status and income**.** It was devised to capture respondents reluctant to spend time answering the full adult questionnaire. This increased the numbers but did not record information vital for useful labour market information. ***Sample design*** The sample is based on the 1996 Census which yielded a total population of 2,496,672 people in the nine magisterial districts which constitute the Cape Metropolitan area. According to the Census, Mitchell's Plain, the site of the survey, had a total population of 728,916, of whom 474,859 were African (65%) and 240,299 were coloured (33%). A two-stage cluster sample was used. The first stage of this sample entails selecting clusters of households and the second stage entails the selection of the households themselves. Enumerator Areas (EAs) as defined by Stats SA for the 1996 Census were used. These EAs are neighbourhoods of roughly 50 to 200 households. Before selecting the EAs, all non-residential and institutional EAs (except for hostels) were excluded from the sample frame. EAs were selected systematically in such a way as to ensure that their probability of selection was proportionate to their population size. The Mitchell's Plain Magisterial District, as defined in the 1996 Census, consists of 1,486 populated EAs. Using the 1996 Census results, the average number of adults per household was calculated to be 2.66. To administer 2,875 questionnaires, 1,081 households needed to be selected. To interview at least 10 households from each selected EA, the number of households to be interviewed per EA was 108. All the EAs were listed in geographical order and by housing type. By doing this, the sample was implicitly stratified by location and housing type. To select the 108, a cumulative total was calculated. A sampling interval was then calculated by dividing 169,884 (the total number of households in the Mitchell's Plain Magisterial District) by 108 giving an interval for selection of 1,573. A number between 1 and 1,573 was randomly selected (this was 723) and the first EA with

	a cumulated total equal to or greater than 723 was selected. The process was repeated by adding the sampling interval of 1,573 to the random number and the EA with a cumulated total greater than or equal to this number was selected. To ensure that adults in at least 10 households in every EA were interviewed, 13 households were interviewed at the second stage of the sample to fit the expected response rate of 80%. The households were selected using a systematic sampling method with a random start. ***Weighting*** The probability that each household in the population has of being selected into the sample can be calculated in order to draw conclusions about the population based on the sample drawn. The overall probability of selecting a household into the sample is the product of the probabilities at each selection stage. The probability of selecting a household into the sample is constant. The overall probability of a household being selected into the sample should be equal to the number of households selected from a specific EA, in our case 13, multiplied by the reciprocal of the sampling interval, in our case, 1/1573. This gives an overall probability of a household being selected into the sample of 0.00826. Households can then be weighted by the reciprocal of there inclusion probabilities, resulting in a constant weight of 1/0.00826 = 121. This means that each household in the sample represents 121 households in the total population. This weight is called *pweight1* in the data. Non-response can lead to an increase in sample errors and bias in estimates. The weight *pweight2*, included in the data, is designed to correct for households who were selected into the sample, but were not interviewed. This could have been due to refusals, failure to locate the household or not being able to make contact with the residents. Pweight2 was constructed by adjusting the original weight, pweight1 by the inverse of the response rate in each EA. For example, if the response rate is 80 percent then a suitable weight would be 1/0.80 =1.25. This was done for each EA and applied to each responding household. Under such a method of dealing with non-response it is assumed that all households selected into the sample have the same probability of responding. Non-response, in the survey, also occurs at the individual level i.e. there are cases where not all adults within a household were interviewed. Non-response at the individual level is corrected for by post stratifying the data according to known age, gender and race proportions as reflected in the 1996 Census. The raking ratio method of post stratifying weight adjustment was used to calculate and make adjustments to the pweight2 variable in the data. This results in the variable, *adultrakingweight***,** in the data. When analysing data in the adult file of the data the post stratified weight, adultrakingweight, should be used to adjust for adult non-response in the survey. It is suggested that the post stratified weights be applied when doing most analysis.
How useful to researchers?	A precursor to CAPS, this provides information about people's livelihood strategies in an area of the Cape Town metropole.
Data format	Data are available in SPSS, SPSS Portable, NSDstat, Statistica, Stata (version 7 or 8), DIF, DBase or SAS format.
Availability of data descriptions	The following documentation is available on the DataFirst website (http://www.datafirst.uct.ac.za/data_kmp.html): ▪ KMP survey report and baseline information ▪ Questionnaires Additional file descriptions and metadata are available when you register to access the data.
Conditions	SALDRU believe that ready access to data is fundamental both for ensuring the ongoing vitality of democracy within countries but also to facilitate impartial scientific collaboration around the world. The data set is in the public domain, but the following conditions apply: ▪ The dataset and its funders should be cited in any publication as follows: In the text, it should be cited as: Khayelitsha/Mitchell's Plain Survey 2000 is produced and distributed by the universities of Michigan and Cape Town, with funding from the Andrew W. Mellon Foundation. In references, it should be cited as: Khayelitsha/Mitchell's Plain Survey 2000, public use dataset. Produced and distributed by the universities of Michigan and Cape Town with funding from the Andrew W. Mellon Foundation. ▪ Copies of any seminar/conference papers or published work should be sent to Lynn Woolfrey at Data First: Email: lynn.woolfrey@uct.ac.za ▪ Users should not attempt to identify specific individuals in the KMP data. ▪ Users should not redistribute the data to other users – all users should register on the DataFirst website. ▪ Users should notify KMP staff regarding errors in the data or any features of the data that could compromise respondent confidentiality.
Costs	None
Tabled outputs	▪ See http://www.datafirst.uct.ac.za/resource/kmps_2000.pdf for survey report and baseline information.

	▪ See http://www.datafirst.uct.ac.za/resource/b_kmp_2000.pdf for list of publications using KMP data.
Contact	**For information:** All queries regarding the data set should be addressed to ***Mr Matthew Welch***, Data First Email: mwelch@commerce.uct.ac.za Telephone: +27 (0)21 650 5710 All queries about the design of the questionnaires should be addressed to ***Prof. Nicoli Nattrass***, School of Economics, University of Cape Town Email: nnattrass@commerce.uct.ac.za Telephone: +27 (0)21 650 3567 **For data:** ***DataFirst*** website: http://www.datafirst.uct.ac.za/data_kmp.html

Name	**KwaZulu-Natal Income Dynamics Study** (KIDS)
Principal investigator	School of Development Studies (SDS), University of KwaZulu-Natal (UKZN)
Year(s)	1993, 1998 and 2004
Area(s) of interest	**General:** housing; economy; labour market; education; health; social welfare **Specific:** household demographics; household environment; education; food and non-food expenditures; remittances; employment and income; agricultural activities; health and anthropometry; school availability; health care facilities; prices for various commodities
Brief description	**Summary** A ten year, three wave longitudinal household survey based on the national Project for Statistics on Living Standards and Development (PSLSD). The third wave of KIDS is being undertaken in collaboration with the South African Department of Social Development with a consortium comprising the Universities of KwaZulu-Natal and Wisconsin-Madison, the International Food Policy Research Institute, the London School of Hygiene and Tropical Medicine and the Norwegian Institute for Urban and Regional Studies. Within UKZN, the Department of Public Health, School of Medicine and the Food Security Programme, School of Agricultural Sciences and Agribusiness, have been partners in this project. **Purpose** The purpose of the 1993-1998 survey was to collect hard statistical information about the conditions under which South Africans live in order to provide policy makers with the data required for planning strategies to implement their goals. An important adjunct of apartheid has been the absence of credible and comprehensive data on which policy, such as poverty reduction strategies, can be grounded. The previous regime had little interest in collecting information of this nature and, indeed, often suppressed data that depicted conditions in the former bantustan areas. This automatically excluded a large proportion of the poor from official statistics. It was not until the 1993 PSLSD that a comprehensive household database for development was created. Despite its usefulness, a cross-sectional study such as the PSLSD is unable to address a variety of questions, particularly those concerning dynamic processes, important to policy researchers and practitioners. KIDS 1993-1998 aimed to fill that gap. **Methodology** In 1998, a consortium of South African and international researchers re-surveyed 1,100 of the households that were first surveyed in 1993 in KwaZulu-Natal province as a part of the PSLSD. In 1993, the KwaZulu-Natal portion of the PSLSD sample was representative at the province level, conditional on the accuracy of the 1991 Census and other information used as the sampling frame, and contained households of all races (see details on PSLSD, below). It was decided however, not to re-survey white and coloured households in 1998. In the KwaZulu-Natal province, Africans represent 85% of the population and Indians represent 12%. Compared with their representation nationally, whites and coloureds are underrepresented in KwaZulu-Natal. The sample size of white and coloured households was small in 1993 (112 and 53 respectively) precluding comparative ethnic analyses. Moreover the households in these groups are too geographically concentrated in a few clusters (due to the general lack of spatial integration of the population), to permit meaningful inference. The KIDS study has thus been limited to the first two population groups. To ensure comparability, the 1998 household questionnaire largely followed the 1993 version (with questions on household composition, expenditure on food as well as on other durable and non-durable goods, education, health, agricultural production, employment, and additional sources of labour and non-labour income), though there were some important changes. One of these was a greater focus on individual (as opposed to household) ownership of assets and control over their use so that gender-differentiated analysis is possible. A second underlying change was an expanded emphasis on the set of individuals not living in the household but economically linked to

	it. Finally, four new sections were added including economic shocks (both positive and negative), social capital (including group membership, kin networks, civic engagement, and trust), assets brought to marriage, and household decision-making. To the extent possible, the new sections on economic shocks and social capital were replicated in 67 community level surveys. Given the various purposes for the study (e.g. income generation, child health, etc.), the identification of main decision-makers within households was very important to enable the collection of longitudinal data on them. A household member was designated ex-ante as a Core person (key decision-maker) if he/she satisfied *any* of the following criteria: ▪ A self-declared head of household (from 1993) ▪ Spouse/partner of self-declared head of household (from 1993) ▪ Lived in a three generation household *and all of the following were true:* Child, child-in-law, or niece/nephew of self-declared head At least 30 years old Have at least one child living in household ▪ Spouse/partner of person satisfying criterion. Therefore all heads of households and spouses of heads are automatically included and in some three generation households adult children are included. Prior to beginning fieldwork, a list of the Core persons in the household was identified, to be targeted for additional information and tracking purposes. This methodology meant fewer key decision makers in the household were missed than if the focus had been on the self-declared heads. In addition, guidelines were provided for designating new core persons to avoid missing other key decision makers in the household. Another important aspect of the 1998 re-survey is that when possible households who had moved were tracked, followed, and re-interviewed. While the tracking procedures were somewhat more involved, the main elements were that Core persons were to be followed if they had moved and were no longer household members. The combination of Core persons, of which there were often more than one in an original 1993 household, and tracking movers meant that it was possible for original households to split and for the split-offs to remain in the sample. In the field 1,132 (84%) original 1993 households were re-interviewed in 1998 and for 36 of these two (or more) interviews were completed in 1998. As a result, it is possible to analyse the sample as a panel of households (ignoring or possibly recombining the split-off households) or as one of Core persons. *Updates to data (Release Version 2, September 2001)* During related follow-up field research in May 2001 it was discovered that all 39 household interviews in clusters 217 and 218 had been fabricated in both 1993 and 1998; these households are dropped in the updated release of the data). Also, cluster 206 is now coded as urban as it was incorrectly coded as rural in the 1993 released data. Thus the 1998 target sample consisted of 1,354 African and Indian households. Triangulating different methodologies, households were re-visited in 2001 and 2004 using qualitative methodologies. The third wave of KIDS interviewed 867 households first surveyed in 1993 and new households established by adult children of the original respondents. In 2004, due to the ageing of the core members and the HIV/AIDS prevalence in South Africa, the study incorporated a complementary way to keep track of the dynasties. Known as the *children of the core* or the *next generation*, these are sons and daughters of core members older than 18, who started a new household since 1993. To ensure comparability, the 2004 questionnaire also largely followed the 1993 and 1998 versions. In addition to the socio-economic data previously gathered, the survey collected information concerning the impact of social security grants, particularly the Child Support Grant, recent deaths of household members, work history for those aged between 24 and 30 at the time of interview, and a module with learning tests for children between the ages of seven and nine.
How useful to researchers?	An important longitudinal resource about people's living conditions in one of South Africa's poorest provinces. While KwaZulu-Natal should not be considered a typical or representative province, many of the underlying social and economic conditions are similar to those found in other provinces in which a substantial proportion of the population reside in the former homeland areas.
Data format	Data are available in SPSS or Stata format from SDS and SPSS format from SADA.
Availability of data descriptions	The following documentation is available from the oDS website (http://sds.ukzn.ac.za/default.php?11,0,0,0,0): **For 2004:** ▪ Overview ▪ Codebook ▪ Questionnaire **For 1993-1998:**

	▪ Overview and description of data files ▪ Fieldworker's manual ▪ PSLSD coding manual ▪ 1998 coding manual Questionnaires and codebooks for 1993 and 1998 are also available from the DataFirst website (http://www.datafirst.uct.ac.za/surveys.html) and SADA website (http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0050).
Conditions	The team are committed to making the data from all three waves of the study available to researchers. It is necessary to fill in a short online form (with name, position, organisation and intended use) in order to access the data and metadata from SDS. The information supplied by researchers is for internal use only. SADA conditions also apply to data obtained from SADA.
Costs	None
Tabled outputs	See http://www.datafirst.uct.ac.za/resource/b_kids_93-98.pdf for list of publications using KIDS 1993-1998 data.
Contact	**For information:** ***SDS*** website: http://sds.ukzn.ac.za/default.php?7,12,9,4,0 ***Prof. Julian May***, Head SDS E-mail: mayj@ukzn.ac.za Telephone: +27 (0)31 260 2841 **For data:** ***SDS*** website (all three waves): http://sds.ukzn.ac.za/default.php?11,0,0,0,0 RECOMMENDED ***SADA*** website (1993-1998): http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0050 ***DataFirst*** (1993-1998 available in resource centre)

Note: The PSLSD is a nationally representative, multi-purpose household survey which contains information on a series of subjects relating to standard of living. The sample consists of approximately 9,000 and was undertaken in the nine months prior to the country's first democratic elections in April 1994. The main instrument used in the survey was a comprehensive household questionnaire. The sample design adopted for the study was a two-stage self-weighting design in which the first stage units were Census Enumerator Sub-districts (ESDs, or their equivalent) and the second stage were households. The sampling frame was drawn up on the basis of small, clearly demarcated area units, each with a population estimate. The nature of the self-weighting procedure adopted ensured that this population estimate was not important for determining the final sample, however.
In the sample design chosen, the area stage units (generally ESDs) were selected with probability proportional to size, based on the Census population. Systematic sampling was used throughout, that is, sampling at fixed interval in a list of ESDs, starting at a randomly selected starting point.

Name	**Labour Force Survey** (LFS)
Principal investigator	Stats SA
Year(s)	2000, 2001, 2002 (February and September), 2003, 2004, 2005, 2006 (March and September)
Area(s) of interest	**General:** labour market **Specific:** demographic information (name, age, sex, age, population group); biographical information (marital status, language, migration, education, training, literacy); activities related to work in the past seven days; unemployment and non-economic activities; main work activities in the past seven days; agricultural activities; uncompensated activities in the past seven days; migrant workers; type of dwelling; ownership of the dwelling and other assets; services; transport; social grants; expenditure
Brief description	**Summary** A national household survey measuring the dynamics of employment and unemployment in the country. It also provides insight into a variety of issues related to the labour market, including the level and pattern of unemployment and the industrial and occupational structure of the economy. The design of the questionnaire and definitions used conform to the requirements set by international bodies such as the International Labour Organisation. **Purpose** To meet the demands of the International Monetary Fund's Special Data Dissemination System for regular labour market information. Starting life as a slimmed-down version of the October Household Surveys, with the labour market institutional information sections extended, the survey was supposed to allow for piggybacking of supplementary surveys. That has not been done very often, if at all.

Methodology (descriptions refer to the September 2004 LFS, but the others are similar)
Detailed information was collected about the labour market situation of approximately 68,000 adults of working age (15–65 years) living in over 30,000 households across the country. The sampled dwelling units in each of the nine provinces were visited by field-staff employed and trained by Stats SA, and an LFS questionnaire was completed through face-to-face interviews for each household visited. A twice yearly rotating panel methodology was used. This obtains a better picture of movements into and out of the labour market over time. The rotating panel methodology involves visiting the same dwelling units on a number of occasions (in this instance, five at most). After the panel is established, a proportion of the dwelling units is replaced each round (in this instance, 20%). New dwelling units are added to the sample to replace those that are taken out. The advantage of this type of design is that it provides the basis for monitoring changes in the work situation of members of the same households over time, while retaining the larger picture of the overall employment situation in the country. It also allows for both longitudinal and cross-sectional analysis.

Sample design
Enumeration Areas (EAs) that had a household count of less than twenty-five were omitted from the census frame that was used to draw the sample of PSUs for the Master Sample. Other omissions from the Master Sample frame included all institution EAs except workers' hostels, convents and monasteries. EAs in the Census database were pooled in two stages, before and after sampling. Before sampling, the criterion that was used to pool EAs was that they should contain a minimum of 100 households. However, during listing it was discovered that there were discrepancies between the information on the database and what was on the ground. Therefore, in the second stage of pooling, EAs that were found to have less than sixty dwelling units during listing were pooled. The Master Sample is a multi-stage stratified sample. The overall sample size of PSUs was 3,000. The explicit strata were the 53 district councils. The 3,000 PSUs were allocated to these using the power allocation method. The PSUs were then sampled using probability proportional to size principles. The measure of size used was the number of households in a PSU as calculated in the Census. The sampled PSUs were listed with the dwelling unit as the listing unit. From these listings systematic samples of dwelling units per PSU were drawn. These samples of dwelling units form clusters. The size of the clusters differs depending on the specific survey requirements. The LFS uses one of the clusters that contains ten dwelling units.

Weighting
A two-stage weighting procedure was applied to the LFS. The PSU inclusion probability is given by

$P_{PSU} = n_{PSU} / N_{PSU} * n_S$

where n_{PSU} is the number of households constituting the selected PSU during census fieldwork, n_S is the number of PSUs per stratum, in this case the District Council (DC) and N_{PSU} is the number of households constituting the selected stratum during census fieldwork.

The household inclusion probability per PSU is given by

$P_{HH} = n_{HH} / H_{HH}$

Where n_{HH} is the number of selected dwelling units per PSU, H_{HH} is the number of dwelling units in the PSU in question at a particular time different from the census time.

The non-response adjustment factor is given by $1 / r_{HH}$, where r_{HH} is the response rate and is given by $r_{HH} / n_{RESP} / n_T$ where n_{RESP} is the number of responding households and n_T is the total number of visited households (in the sampled dwelling units) per PSU. The design weights adjusted for nonresponse are now given by

$W_{HH} = 1 / (P_{PSU} * P_{HH} * r_{HH})$

A SAS macro called CALMAR was used to benchmark W_{HH} to the population estimates. The midyear population estimates were adjusted to give population estimates for September 2004 (when survey fieldwork took place).

Geography
In early surveys, the geographical levels available for analysis are national and provincial, with a rural/urban split. Surveys from September 2004 also present data at district municipality (42 areas), cross-border district municipality (5) and major metropolitan area (6) levels.

Data revisions

	Stats SA has revised the population model to produce mid-year population estimates in the light of recent mortality data released in 2005. The benchmarks for the LFS have been adjusted accordingly. Data for March 2004 and September 2004 are therefore comparable. The re-benchmarking of earlier surveys using the new mortality information is currently underway and Stats SA will in due course release the historical datasets based on the revised information. Caution must be exercised when interpreting the results of the LFS at low levels of disaggregation since the revisions to the LFS data based on the new population estimates involved benchmarking at the national level in terms of age, sex and population group while at the provincial level benchmarking was by population group only.
How useful to researchers?	It is not possible to disaggregate to sub-provincial level because the sample size is so small. Anyone intending to look at the whole series should apply to Stats SA for revised weights for surveys prior to September 2003. Informal economy worker and subsistence agricultural worker estimates are weak. A Survey of Employers and Self-Employed conducted in 2001 caused a large spike in informal employment estimates. See Meth (2006a) and Meth (2006b) for further details.
Data format	Data are available in flat file, ASCII, fixed field format, with one line of given length per record from Stats SA and SPSS format from SADA.
Availability of data descriptions	A comprehensive metadata document is available on the data CD from Stats SA. Metadata and questionnaires are available on the SADA website (http://www.nrf.ac.za/sada) for some rounds of the LFS. Metadata, code lists and questionnaires for 2000, 2001 and 2002 are available on the DataFirst website (http://www.datafirst.uct.ac.za/surveys.html).
Conditions	The information products and services of Stats SA are protected in terms of the Copyright Act, 1978 (Act 98 of 1978). As the State President is the holder of State copyright, all organs of State enjoy unhindered use of the Department's information products and services, without a need for further permission to copy in terms of that copyright. Where a copy of the information is made available to any third party outside the State, the third party must be made aware of the existence of State copyright and ownership of the information by the State. The State (through Stats SA) retains the full ownership of its information, products and services at all times. Access to information does not give ownership of the information to the client. The use of any data is subject to acknowledgement of Stats SA as the supplier and owner of copyright. Users may apply or process the data, provided Stats SA is acknowledged as the original source of the data; that it is specified that the analysis is the result of the user's independent processing of the data; and that neither the basic data nor any reprocessed version or application thereof may be sold or offered for sale in any form whatsoever without prior permission from Stats SA. SADA conditions also apply to data obtained from SADA.
Costs	R 300 for a CD from Stats SA or free from SADA.
Tabled outputs	See http://www.statssa.gov.za/publications/P0210/P0210March2006.pdf for March 2006 tables.
Contact	**For information:** ***Stats SA User Information Services*** Email: info@statssa.gov.za Telephone: +27 (0)12 310 8600 **For data:** ***Stats SA User Information Services*** Email: info@statssa.gov.za Telephone: +27 (0)12 310 8600 ***SADA*** website: http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0125 (February 2000) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0127 (September 2000) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0126 (February 2001) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0128 (September 2001) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0131 (February 2002) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0132 (September 2002) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0133 (March 2003) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0134 (September 2003) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0137 (March 2004) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0141 (September 2004) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0143 (March 2005) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0144 (September 2005) ***DataFirst*** (available in resource centre)

Name	**National Food Consumption Survey**
Principal investigator	Department of Health
Year(s)	1999
Area(s) of interest	**General:** health **Specific:** child malnutrition; food insecurity; dietary intake; anthropometric status
Brief description	**Summary**

	A nationally representative cross-sectional survey investigating the food consumption of children under 10 years. **Purpose** ▪ To determine usual food consumption of children aged 1 - 9 years in South Africa. ▪ To assess the usual nutrient intake of children aged 1 - 9 years in South Africa. ▪ To identify factors impacting on food consumption. ▪ To determine anthropometric status. ▪ To propose/recommend appropriate food(s) for fortification and appropriate nutrition education material. **Methodology** A nationally representative sample with provincial representation was selected using the Census 1996 information. This implied that each child in the country had a known probability of being selected (selfweighting for the strata selected). This initial sample was adapted by means of 50% oversampling to accommodate for children that would not be at home at the time of the survey (approximately 25%), and for allowing an overrepresentation of the children living in high-risk areas (approximately 25%). The following questionnaires were used in the survey: *Socio-demographic questionnaire* provided information on factors relevant to the household regarding the environment in which the child lived. *24-hour recall questionnaire* provided information on the current diet and eating pattern of the child. *Quantitative food frequency questionnaire* provided information on the eating patterns and intake over the previous six months for children older than two years and over one month for children aged between 12 and 23 months. Consequently seasonality effects could be observed. *Food procurement and household food inventory questionnaire* provided information on purchasing patterns and storage of food, which is believed to be essential for policy formulation on food fortification. The food procurement section of the questionnaire was completed in all households (HHs), whereas the food inventory section was completed in all HHs in high-risk areas as well as one randomly selected HH in all other enumeration areas (EAs). *Hunger scale questionnaire* provided information on the caregivers' perception of whether hunger was experienced in the HH and by the child. ***Sample design*** The stipulated number of children to be studied was originally 2,200. However, in order to have a minimum of 50 observations per province and per urban/rural strata for the 24-H-R questionnaire, the number of children to be studied was increased to 2,440 children. The sample was further increased to 3,120 children to over represent children from high-risk areas and to allow for children that would not be at home at the time of the survey. EAs drawn up for the 1996 Census were used. For financial and practical reasons, in formal/informal urban and tribal areas only EAs with at least 16 qualifying households were considered for inclusion in the sample, whereas in commercial farms only EAs with at least 6 qualifying households were considered for inclusion in the sample. A qualifying household was defined as any household with at least one child aged between 1 to 9 years in it. All other qualifying EAs for the survey were randomly selected. A total of 156 EAs were included in the survey, 82 of which were urban and 74 non urban. The distribution of EAs per province was determined proportionately to the distribution of the total population and the urban/non-urban distribution in each province. After the maps of the relevant EA's were obtained, they were passed on to the respective fieldwork teams. An estimate was made of the total number of households (HHs) in each EA required in order to determine the approximate number of qualifying HHs with children within the prescribed age interval in the EA. An adapted version of snowball sampling was used to set up a partial sampling frame. The snowball sampling method entailed the random selection of a number of HHs in each EA in which it was asked whether there were other HHs in the vicinity with children in the prescribed age range of the survey. These HHs were then recorded on the EA map. Ultimately, a list of qualifying HHs in the EA with children in the prescribed age range was drawn. From this list, the required number of households for the survey was randomly selected. *For urban and peri-urban EAs*: A random starting point within each EA was selected. If this randomly selected HH had no children in the desired age range, then the fieldworker moved on to the next nearest household that had children in the prescribed age range. All the occupants of the HH were listed. If there was more than one child in the age range of 1-9 years, a random number table was used to select ONE child in a given HH to be included in the survey. All the required questionnaires were filled in and all the required anthropometric measurements were taken. This HH was then considered as completed. *For commercial farms and other rural EAs*: A small number of farms with equal probability were drawn using a random number table. From these farms, all qualifying HHs were listed, and from

	these HHs a simple random sample of a total minimum of 30 HHs was drawn so as to ensure that 20 HHs were included in the survey. One child only in each randomly selected HH was included in the survey. If there was more than one child present in the prescribed age interval in a HH, then all children in the HH in age order were numbered, so that a single child could be selected at random. This list was then used to randomly select one child for inclusion in the survey using a specially designed random number table.
How useful to researchers?	An important resource for food insecurity studies.
Data format	Unknown
Availability of data descriptions	The technical report (http://www.sahealthinfo.org/nutrition/foodconsumption.htm) includes: ▪ General methodology - includes study objectives, sampling strategy, questionnaire descriptions, dietary aides and equipment. ▪ Questionnaires ▪ Consent form ▪ Protocol ▪ Training manuals ▪ Video script It is not known what metadata/data descriptions will be provided with the data.
Conditions	The data is in the public domain, and thus available to any researcher, with permission from a committee of survey directors (contact details below).
Costs	None
Tabled outputs	See http://www.sahealthinfo.org/nutrition/foodconsumption.htm for technical report.
Contact	**For information:** ***SA Health Info*** website: http://www.sahealthinfo.org/nutrition/foodconsumption.htm ***Prof. Demetre Labadarios***, Head, Human Nutrition, University of Stellenbosch E-mail: demetre@sun.ac.za Telephone: +27 (0)21 938 9259 **For data:** ***Prof. Demetre Labadarios***, Head, Human Nutrition, University of Stellenbosch E-mail: demetre@sun.ac.za Telephone: +27 (0)21 938 9259

Name	**National HIV and Syphilis Antenatal Sero Prevalence Survey**
Principal investigator	Department of Health (DoH)
Year(s)	1991-2005
Area(s) of interest	**General:** health **Specific:** HIV prevalence; syphilis prevalence
Brief description	**Summary** One of the major South African surveys used to estimate HIV prevalence. **Purpose** ▪ To estimate HIV prevalence in South Africa and present trends in HIV progression. ▪ To describe HIV trends in sub population groups (province) and describe trends in different age groups. ▪ To estimate HIV infection rates in the general population (men, women and children) using the UNAIDS spectrum model. ▪ To estimate syphilis prevalence among women and describe trends in syphilis. **Methodology** All women attending a sample of public ante-natal clinics for their first check-up during pregnancy are surveyed. Demographic details are collected using standard data collection forms. These exclude personal identifiers. Blood samples are collected from all women and lab tested. In 2005 and 2004, approximately 16,000 women were surveyed. ***Sample design*** Clinics were selected by probability proportional to size sampling. In 2005, 399 clinics were selected.
How useful to researchers?	A useful resource for HIV related studies.
Data format	Data are available in SPSS and Stata format.
Availability of data descriptions	Unknown
Conditions	The data from this set of studies is available to researchers, 6 months after the Department of Health report is published. Requests for data should be addressed to Dr Lindiwe Makubalo (contact details below). A standard data users agreement has to be signed and researchers must

	agree to share research findings with the DoH.
Costs	None
Tabled outputs	See http://www.doh.gov.za/docs/reports/2005/hiv.pdf for the 2005 survey report.
Contact	**For information:** ***DoH - Dr Lusanda Mahlasela*** Email: MahlaL@health.gov.za **For data:** In order to access data, a written request must be sent to ***Dr Lindiwe Makubalo***, Chief Director for Health Information, Evaluation and Research, DoH Email: MakubL@health.gov.za Telephone:+27 (0)12 312 0774/ 0767/ 0776

Name	**National Household Travel Survey** (NHTS)
Principal investigator	Department of Transport (DoT)
Year(s)	2003
Area(s) of interest	**General:** transport **Specific:** worker and learner travel; modes of travel; periods of travel; travel times; travel costs
Brief description	**Summary** The first national survey of the travel habits of individuals and households. Stats SA executed the survey on behalf of the DoT. **Purpose** ▪ To assist with the effective targeting of subsidies for public transport. ▪ To assist in identifying disadvantaged regions for investment in transport infrastructure. ▪ To measure the Key Performance Indicators for land passenger transport as required by the National Land Transport Transition Act (Act No. 22 of 2000) and the National Land Transport Strategic Framework. ▪ To understand the transport needs and habits and/or behaviour of all household members at all times of day and for all purposes. ▪ To ascertain the cost of transport for individuals and households and to assess the extent to which they can afford to pay for the mobility which is essential for their survival. ▪ To assess customer attitudes towards transport services, service providers and the quality of transport facilities which they are required to use. ▪ To measure existing car ownership and uses. ▪ To understand the travel choices of different market segments. ▪ To determine the extent of accessibility to opportunities such as work, health facilities, education and markets for social interaction and all other social needs. **Methodology** The NHTS sampled about 0.5 per cent of all households in South Africa in May/June 2003. Metropolitan and district municipal boundaries, as determined by the Demarcation Board in 2000, were used as the basis for the determination of the analysis zones. The sample was made proportional to the population of each of the municipalities. In order to measure movements between different parts of a metropolitan or district municipality, each municipality was divided into a number of travel analysis zones. The minimum number of households per analysis zone was 100 households. The NHTS relied on selected household members (worker and learner) recalling all trips taken on the day prior to the survey. ***Sample design*** The explicit strata were the 342 Travel Analysis Zones (TAZ). A sample of 5,000 Enumerator Areas (EAs) was allocated using the power allocation method. The first step was to take out vacant, industrial, institution, and recreational EAs. EAs were selected with probability proportional to size, using the total number of households as enumerated during Census 2001 as a measure of size. EAs which had less than 80 dwelling units were pooled together with another EA with similar characteristics to form primary sampling units (PSU). An EA with 80 or more dwellings automatically qualified to become a PSU. Census listings of the selected PSUs were updated where necessary and then a systematic sample of 10 dwellings units was selected in each PSU. Because there is sometimes more than one household at each dwelling unit, the sample of 50,000 dwelling units produced a sample of 52,376 households. Section 7 of the questionnaire required the selection of one person aged 15 years and above to answer the attitude questions. This person was randomly selected using a grid.

	Weighting A two stage weighting procedure was used: adjusting for non-response and benchmarking where population totals were adjusted at municipality level and gender, five-year age group and race were taken into consideration at national level. The only variables which caused some concern were travel cost to work and personal income. Doubtful values were set to "missing" in the dataset. Missing values were largely confined to cost of travel to work, personal income, cost of travel to education and household income. Two main causes were identified for the missing and doubtful cost values. These were apparently incorrect recording of costs either in Rands rather than in cents as instructed, or in the wrong units e.g. per trip instead of per month. The first was probably a fieldworker error, while the second was more likely to have been a respondent error. In any event, the incidence of missing data can be regarded as very low, due to the stringent checking procedures followed, and the quality of the data is of a high standard.
How useful to researchers?	Very useful if wishing to take into account travel time, cost and choices.
Data format	Data are available in ASCII format.
Availability of data descriptions	See http://www.dot.gov.za/projects/nts/framesPage.htm for the technical report.
Conditions	The data is available to any researcher. Details of the proposed research project would be of interest to the DoT but are not essential. The DoT must be acknowledged as the data source. The data can be further disseminated providing no charge is made and the DoT is acknowledged.
Costs	None
Tabled outputs	See http://www.dot.gov.za/projects/nts/keyresults.pdf for key results. See http://www.dot.gov.za/projects/nts/framesPage.htm for the technical report which contains detailed results as well as methodology descriptions.
Contact	**For information:** ***DoT*** website: http://www.dot.gov.za/projects/nts/framesPage.htm ***DoT - Ms Laverne Dimitrov*** Email: DimitroL@dot.gov.za Telephone: +27 (0)12 309 3533 **For data:** ***DoT - Ms Laverne Dimitrov*** Email: DimitroL@dot.gov.za Telephone: +27 (0)12 309 3533

Name	**National Income Dynamics Study**
Principal investigator	Southern Africa Labour and Development Research Unit (SALDRU), University of Cape Town are undertaking this study for the Policy Coordination and Advisory Services of The Presidency
Year(s)	2007/2008 onwards
Area(s) of interest	**General:** labour market; health; education; housing; economy **Specific:** income and expenditure dynamics; determinants of changes in poverty and well-being; household composition and structure; reliance on migration and migrant strategies; labour market participation and economic activity; human capital formation, health and education; vulnerability and social capital
Brief description	**Summary** A household panel survey providing, in the first wave, a nationally representative survey of South African well-being, and generating information on some of the dynamics of well-being through subsequent waves. Although not yet available, it is worth highlighting as a dataset for the future. **Purpose** ▪ To track across time how well South Africans are faring in terms of a number of clearly defined poverty, wealth and well-being indicators. ▪ To get a sense across time of the extent to which South Africans are able to engage in successful strategies for 'creating wealth', accumulating assets, participating effectively in the economy, and reducing their poverty and vulnerability. **Methodology** The sample size will be approximately 8,000 households. Four survey instruments will be designed for Wave 1. These are: *Household form* - will contain confidential household information such as the GPS location of the household, the address, a description of the dwelling and a sketch map of its location, a detailed log of all attempts to contact the household and its members, basic information (name, gender, date of birth) about all household members, regardless of whether they are ultimately interviewed or not. It is not expected that any of the data in the household form will be included in the public

	release version of the data. *Household questionnaire* - one questionnaire will be completed per household, typically by the oldest woman in the household or another person knowledgeable about household affairs and particularly household spending. This questionnaire will cover topics such as consumption expenditure, access to services, housing quality, involvement in agricultural home production, non-business assets, positive and negative shocks experienced by the household, migrants, existence and characteristics of non-co-resident household members (parents, siblings, and children), and whether money, goods, or services were transferred between these household members during the year before the interview. *Person questionnaire* - will collect detailed employment, income, family and background information from household members aged 15 years or over. Questions on personal asset ownership and non-labour income (especially social grants), household decision-making and (for women) fertility and child-bearing will be included. It is possible that the person questionnaire will collect additional data such as residential histories, personal expenditure data and information on shocks (as distinct from the information on shocks collected at the household level). *Child questionnaire* - will collect information about children younger than 15. It is possible that children above a certain age (say 12) might answer the questions on their own behalf, but generally the child's mother, female guardian or primary caregiver will answer the questions. The focus will be on the child's educational history and access to grants, but it is possible that other modules (e.g. child work) might be included. ***Sample design*** It is proposed to use a two stage sampling procedure where Enumerator Areas (EAs) are selected, followed by 20 visiting points within each EA. If more than one household is residing at a visiting point, every household at that visiting point will be included in the sample. Provided that at least one member of the household agrees to participate in the study, each household member at that point in time then becomes a member of the indefinite life panel. This applies regardless of whether that particular individual agrees to participate in Wave 1. For the first stage sampling, an existing master sample will probably be used.
How useful to researchers?	N/A
Data format	The data from the first wave of NIDS will be made public at the end of the third year at the agreement of the NIDS Steering Committee. All of the surveys that SALDRU has conducted have been made available for public release through the DataFirst website in a variety of commonly used statistical package formats.
Availability of data descriptions	N/A
Conditions	N/A
Costs	N/A
Tabled outputs	N/A
Contact	**For information:** ***Prof. Murray Leibbrandt***, Director, SALDRU or ***Dr Ingrid Woolard***, SALDRU Email: mleibbra@commerce.uct.ac.za or iwoolard@commerce.uct.ac.za **For data:** ***DataFirst*** website (when available)

Name	**National Victims of Crime Survey**
Principal investigator	Stats SA
Year(s)	1998
Area(s) of interest	**General:** crime **Specific:** experiences of crime; patterns and extent of crime; types of crime; reporting of crime; perceptions of police and police services
Brief description	**Summary** The first national survey examining victims of crime. In 1998, Stats SA was commissioned by the Department of Safety and Security and the United Nations Interregional Crime and Justice Research Institute (UNICRI) to carry out the survey. **Purpose** Prior to this study diverse research had been conducted on victims of crime, but these studies did not cover the entire country. In particular, they did not include rural areas. The only information available on criminal activities in the country as a whole was through police and court records. While these records were highly valuable resources of statistics about criminal activity, they relied largely on the extent to which the reporting of crimes took place. Those incidents, which were not reported by victims, remained unknown and unrecorded. Police and court records allowed for the monitoring of policing and of the criminal justice system. An important supplementary way of

	gaining understanding of crimes, and of improving planning for crime prevention, was by speaking to citizens to find out about their experiences of crime and the criminal justice system. The main objectives of the survey were therefore to: ▪ Conduct the first national benchmark household victimisation survey in South Africa; ▪ Collect information to feed into the National Crime Prevention Strategy on the nature, extent and patterns of crime from the perceptive of victims; ▪ Examine the extent of crime reporting for different types of crimes; ▪ Obtain a picture of the perceptions that members of the public hold of police and policing; ▪ Look at the type of support structures that exist for victims of crime; ▪ Determine the crime prevention measures that South Africans use; and ▪ Compare, where possible, South African crime patterns and reporting behaviour with those in other countries. **Methodology** The survey was conducted between 16-27 March 1998 following an extensive consultation and design process. The instrument and methodology was built on that of the UNICRI international surveys, with some adaptations to the South African context. A survey questionnaire was used for data collection. The questionnaire was based on the standard international questionnaire with certain modifications. The international questionnaire covered eleven main crimes (theft of a car or other motor vehicle, theft from a car or other vehicle, car vandalism, theft of a motor cycle or scooter, theft of a bicycle, burglary or housebreaking, attempted burglary, robbery with force, personal theft, sexual incidents and assault, consumer fraud and corruption). The following crimes were added to meet specific South African needs: theft of livestock, poultry and other animals, hijacking or attempted hijacking of vehicles, deliberate damage, burning or destruction of dwellings and deliberate killing or murder. ***Sample design*** The sample consisted of 4,000 people aged 16 years or more and was drawn in three stages. ▪ A probability sample of 800 enumerator area (EAs) was drawn, as demarcated for the 1996 population census. This sample was stratified explicitly by province, and implicitly by the 42 police districts of the country. ▪ Within each of the 800 EAs, five households were selected for interviewing, using systematic sampling. ▪ One respondent aged 16 years or more was selected to be interviewed in each of the five households in each sampled EA. This person was chosen using a table of random numbers. Once a respondent had been selected, fieldworkers were instructed to make sure that they interviewed only that specific person and nobody else. In case of non contacts with that person, repeated call backs (at least three) had to be made. There were no substitutions for refusal or non-contact. ***Weighting*** The 1996 Census formed the basis for weighting the data to population of the country. Two different sets of weights were used for this study, household and individual. The questions posed on crimes committed against the households were weighted to the population of households in the country, while those concerning crimes committed against individuals were weighted to the population of individuals aged 16 years or more. Factors taken into account in weighting households were province, police area and EA type (urban formal, urban informal, non-urban traditional, commercial farms and other types of non-urban areas). Weights were calculated by using the reciprocal of the inclusion probabilities.
How useful to researchers?	National victim surveys provide useful information on victimisation rates and vulnerable groups, because they focus on the victims of crime (rather than the perpetrators as is the case with police and court data). The surveys also provide an understanding of public perceptions of crime and safety, and the fear of crime, as well as victims' actual experiences of specific types of crime. They also offer insight into the underreporting of crime.
Data format	Data are available in SPSS format.
Availability of data descriptions	The following documentation is available from the SADA website (http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0101): ▪ General notes - information on the sample design, how the weights were calculated and the classification of crime ▪ Questionnaire ▪ Metadata ▪ Codes for open ended questions
Conditions	The information products and services of Stats SA are protected in terms of the Copyright Act, 1978 (Act 98 of 1978). As the State President is the holder of State copyright, all organs of State enjoy unhindered use of the Department's information products and services, without a need for

	further permission to copy in terms of that copyright. Where a copy of the information is made available to any third party outside the State, the third party must be made aware of the existence of State copyright and ownership of the information by the State. The State (through Stats SA) retains the full ownership of its information, products and services at all times. Access to information does not give ownership of the information to the client. The use of any data is subject to acknowledgement of Stats SA as the supplier and owner of copyright. Users may apply or process the data, provided Stats SA is acknowledged as the original source of the data; that it is specified that the analysis is the result of the user's independent processing of the data; and that neither the basic data nor any reprocessed version or application thereof may be sold or offered for sale in any form whatsoever without prior permission from Stats SA. SADA conditions also apply to data obtained from SADA.
Costs	None
Tabled outputs	Hirschowitz et al. (1998) See http://www.statssa.gov.za/publications/VictimsOfCrime/VictimsOfCrime1997.pdf
Contact	**For information:** ***Stats SA User Information Services*** Email: info@statssa.gov.za Telephone: +27 (0)12 310 8600 **For data:** ***SADA*** website: http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0101 ***DataFirst*** (available in resource centre)

Name	**National Victims of Crime Survey** (NVCS)
Principal investigator	Institute for Security Studies (ISS)
Year(s)	2003
Area(s) of interest	**General:** crime **Specific:** victimisation, fear of crime, police and courts, community and official response to crime, community cohesion, corruption, victim support and other interventions
Brief description	**Summary** In 2003, ISS undertook the second victims of crime survey in South Africa. The study was prompted by the need for an accurate picture of crime levels in the country to complement that provided by the official crime statistics published annually by the South African Police Service (SAPS). While a similar need exists in countries throughout the world, the moratorium on the release of police crime statistics in South Africa intensified the requirement for an independent and reliable national study on crime levels. The onset of the moratorium coincided with claims by the SAPS and the Department of Safety and Security that crime levels were stabilising. However in the absence of any statistical information on the extent of crime nationally, South Africans became increasingly sceptical of the motives and pronouncements of the police. A national victim survey provided a reliable means for obtaining an alternative picture of crime. A similar study was conducted in 1998 which allows for a comparison of crime levels over time, and thus provides a means of assessing police claims that crime has stabilised. The survey was designed to ensure comparability with the 1998 NVCS conducted by Stats SA for the Department of Safety and Security. The ISS was assisted throughout the study by a group of stakeholders from government and civil society, many of whom participated in the 1998 survey. **Purpose** ▪ To present reliable data on the levels of crime throughout South Africa. ▪ To compare current crime levels, as recorded by the victim survey, with those of the 1998 NVCS. ▪ To establish the key risk factors that predispose certain people to becoming crime victims ▪ To assess public perceptions of crime, safety and the criminal justice system. **Methodology** The survey was conducted between September and October 2003. Households were randomly selected across the country based on the census data. An independent research company, Development Research Africa, with prior experience conducting victim and crime related surveys, was appointed to undertake the data collection, quality control, coding and capture process. In order to ensure comparability with the 1998 NVCS, as well as ISS surveys in other African countries that were being conducted at the same time as the national survey, specific aspects of the survey design were predetermined. These included certain components of the questionnaire as well as the sample design. The following issues were discussed in relation to crimes against the household and individual: when, where, violence/weapons used, injuries, reported to the police, police response, arrest and

	conviction, knowledge of perpetrators, number of perpetrators, motive, drugs and alcohol, priorities after victimisation, feel safer as a result, reporting to others. *Household crimes*: deliberate damage, burning or destruction of dwellings; deliberate killing or murder; hijacking of motor vehicles; housebreaking and burglary; theft of livestock, poultry and other animals; theft of motor vehicles; theft of goods from vehicles; motor vehicle vandalism; theft of motor cycles; theft of bicycles. *Individual crimes*: sexual offences (including rape and domestic sexual abuse); robbery involving force; assaults and threats of assault (including domestic violence); theft of personal property; consumer fraud. *Additional questions asked about crimes*: theft of car – type of car, what security measures did the car have; burglary – were items stolen, was anyone at home, additional security added; theft out of car – what was stolen; robbery – did victim resist, what was taken; hijacking – type of car, inside or outside vehicle, alone or in company at the time; assault – prior attacks, pressurised not to report; sexual assault – classification of incident, awareness of help available after incident; murder – how many died, victim a source of household income, relationship to victim, part of ongoing problem; theft – what was stolen, when realised; theft of livestock – type of livestock stolen; theft of crops – type of crops taken; damage to buildings – what buildings were damaged; damage to motor vehicles – what part of vehicle was damaged. Other questions included those on crime and safety; police and courts; community and official response to crime; community cohesion; corruption; victim support and other interventions; personal and household information ***Sample design*** Multi-stage cluster sampling was utilised, with Enumerator Areas (EAs) from the 2001 Census selected at the first stage of the sampling, households within the EAs at the second stage, and individuals within the household at the third stage. Based on the total number of households in South Africa (identified by the 2001 Census as 11,205,705), a total of 80,787 EAs were allocated. The total sample size was determined to be 4,050 households. The sample was calculated at a 95% confidence interval, and with a design effect of two. Ten interviews were collected from each EA. The distribution of the sample through the provinces is reflected in the table below. In total, a sample of 4,860 was realised. Households were selected randomly in the following manner. Maps of the EAs were obtained prior to entry of the EA, and random starting points selected. The household nearest to the starting point was selected, and a household interval randomly selected by the supervisor using the day, week and month of the interview. Every nth household was then completed until the target number of interviews was obtained. Respondents over the age of 16 years were then randomly selected for interview using a KISH grid. **Weighting** In order to correct slight discrepancies between Census data (based on the 2001 Census) and the demographics of the sample achieved, the data was weighted by province, race, gender, age and employment status. This is consistent with the process followed in the 1998 NVCS and ensured that the findings are truly representative of the South African population.
How useful to researchers?	National victim surveys provide useful information on victimisation rates and vulnerable groups, because they focus on the victims of crime (rather than the perpetrators as is the case with police and court data). The surveys also provide an understanding of public perceptions of crime and safety, and the fear of crime, as well as victims' actual experiences of specific types of crime. Change over time can be monitored if used with the 1998 NVCS.
Data format	Data are available in SPSS or Stata format.
Availability of data descriptions	The questionnaire will be made available with the data.
Conditions	In principle the ISS is happy to make the data available for non-commercial purposes. The only condition is to provide written details of what the data will be used for, how it will be used and how the ISS will be acknowledged (the ISS would need to be fully and clearly acknowledged as the originator of the data).
Costs	None
Tabled outputs	Burton et al. (2004) See http://www.issafrica.org/index.php?link_id=3&slink_id=371&link_type=12&slink_type=12&tmpl_id=3
Contact	**For information:** ***Ms Antoinette Louw***, Senior Research Fellow, Crime and Justice Programme, ISS Email: alouw@issafrica.org **For data:** ***ISS - Ms Antoinette Louw***

	Email: alouw@issafrica.org ***DataFirst*** (available in resource centre)

Name	**National Youth Victimisation Survey**
Principal investigator	Centre for Justice and Crime Prevention (CJCP)
Year(s)	2005
Area(s) of interest	**General:** crime **Specific:** demography; home environment; feelings on, and nature of, neighbourhood; general thinking and beliefs on crime; experiences at school; exposure to violence; access to information; social activities; perceptions of authorities; morality and values; access and exposure to drugs; view of the future; experience of, and the nature of, crime
Brief description	**Summary** The first national survey on youth victimisation. **Purpose** ▪ To obtain a more comprehensive understanding of the crimes committed against youth in South Africa. Crime and violence is pervasive in South African society; however, research studies investigating criminal victimisation have generally been confined to the perspectives and experiences of adults. The discourses pertaining to young people and crime have largely centred on the offending of the youth. Only in recent years have researchers become concerned with the escalating exposure of young people to both violent and non-violent forms of victimisation. ▪ To assess the extent of victimisation experiences of young people. ▪ To assess the nature and correlates of these experiences. **Methodology** Participants were asked a series of questions pertaining to their exposure to, and experiences of, violence in their homes, at school and within their broader communities. The types of crimes explored in the study included general violent crimes (robbery, assault, sexual assault, car hijacking and exposure to murder) as well as property crimes (housebreaking and theft of personal property). Information was also obtained about their perceptions of their communities, home environments, involvement in social activities, as well as their knowledge of support structures. Respondents were specifically asked about their experiences in the past year, in order to minimise recall limitations. ***Sampling*** The sample was designed to be proportionately representative in order to make it reflective of the South African population. The sample frame was provided by Stats SA's 2001 Census, and the sample was stratified by province and race. The total population between the ages of 12 and 22 years was identified. Based on this, a sample of 333 enumerator areas (EAs) were randomly selected, with 13 households identified to be interviewed in each. Each EA was mapped, each household within the EA assigned a number, and a list of all houses within the EA was compiled. Households were then randomly selected from this numbered list and visited by enumerators. Where a youth between the ages of 12 and 22 lived in the household and was available and willing to participate in the study, an interview was conducted. Where no respondent falling within the required age cohorts lived in the house, the next house on the list was visited. The final sample comprised 4,409 young people recruited from all provinces in South Africa. ***Weighting*** The final data was weighted by province, race and gender using the marginal totals drawn from the 2001 Census. This was done to ensure the most accurate representation of the experiences of young people throughout South Africa.
How useful to researchers?	This survey provides useful information on victimisation rates for young people. It also provides an understanding of young people's perceptions of crime and safety, and the fear of crime, as well as their actual experiences of specific types of crime.
Data format	Data are usually available in SPSS format (but could be converted to Excel or Stata on request).
Availability of data descriptions	At present, there is limited documentation available. The best sources of information are the reports detailed below.
Conditions	The Centre for Justice and Crime Prevention must be cited as the source.
Costs	The data is available free of charge in its entirety, via email. Specific data runs can be carried out for researchers without access to or knowledge of SPSS. This is normally free of charge, provided it is limited to no more than a half days work.
Tabled outputs	The following are available from CJCP (contact details below): ▪ Leoschut and Burton (2006) ▪ Burton, P. (2006)

Contact	**For information:** ***Mr Patrick Burton***, Research Director, CJCP Email: patrick@cjcp.org.za **For data:** ***Ms Wendy Chetty***, Administrator, CJCP Email: wendy@cjcp.org.za Telephone: +27 (0)21 687 9177. ***Mr Patrick Burton***, Research Director, CJCP Email: patrick@cjcp.org.za

Name	**October Household Survey** (OHS)
Principal investigator	Stats SA
Year(s)	1994 to 1999
Area(s) of interest	**General:** demography; housing; labour market; education **Specific:** age; gender; level of education; marital status; migration; use of health services; economic activity; employment; unemployment; informal sector; internal migration; services available by type of dwelling; access to health and social services; safety and well-being of household; households by average household size and type of dwelling; level of education; quality of life; health statistics; vital statistics
Brief description	**Summary** An annual survey based on a probability sample of a large number of households, providing detailed information about the living conditions and life circumstances of all South Africans and covering a range of development and poverty indicators. **Purpose** ▪ To fill the gap created by the cessation, in 1986, of the Current Population Surveys (CPS). First introduced in 1978 as the debate over unemployment in South Africa intensified, the CPSs became increasingly unreliable and unrepresentative, as the abortive project of dividing the country into so-called 'independent' states fragmented the national statistical system. ▪ To describe, through a series of household surveys, the situation in a country at a given point in time. ▪ To measure change in people's life circumstances as and when new government policies are implemented. **Methodology** Altogether, seven OHS have been conducted. ▪ The first OHS was undertaken in October 1993. This survey is not directly comparable with the later surveys, since it excluded the former Transkei, Bophuthatswana, Venda and Ciskei (TBVC states). ▪ The 1994 OHS was the first to cover the entire country, including the former TBVC states. Interviews were conducted with respondents in 30,000 households in 1,000 enumeration areas (EAs). Thirty households were visited in each EA. ▪ In 1995, the OHS was also conducted among 30,000 households. However, the sample was more widely dispersed throughout the country. Three thousand, rather than 1,000 EAs were sampled, and interviews were conducted in 10 households in each EA. ▪ In 1996, the survey was conducted in November, since enumeration for the 1996 Census took place in October. Due to time and financial constraints, 16,000 households were visited in 1,600 EAs. The EAs were less dispersed than in previous years, in that the survey was conducted in 800 pairs of adjacent EAs. ▪ In 1997, the sample size was once again increased to 30,000 households, selected from 3,000 sampled EAs. ▪ In 1998, due to budget constraints, the sample size was reduced to 20,000 households in 2,000 EAs. ▪ In 1999, the sample size was again increased to 30,000 households. This was the first time that a master sample was used to select the sample of households to be interviewed. The survey was funded by the Department for International Development, UK. ***Sample design*** The OHSs of 1994, 1995, 1996, 1997, 1998 and 1999 were independent cross-sectional surveys, and different samples were designed for each of them. The OHS of 1999 was drawn from a master sample, in which households sampled from the same primary sampling units (PSUs) can be visited for a variety of surveys, including the twice-yearly Labour Force Survey. The database of EAs, as established during the demarcation phase of Census 1996 and finalised after the enumeration phase, constituted the sampling frame for selecting EAs for the 1997 and 1998 OHSs. It also formed the sampling frame for OHS 1999. The surveys prior to 1996 were based on selecting

	areas within magisterial districts. In 1999, as part of the master sample, small EAs consisting of fewer than 100 households were combined with adjacent EAs to form PSUs of at least 100 households, to allow for repeated sampling of households within each PSU. The sampling procedure for the master sample involved explicit stratification by province and within each province, by urban and non-urban areas. Independent samples of PSUs were drawn for each explicit stratum. A disproportionately larger number of PSUs were allocated to the smaller provinces than the bigger provinces. Altogether, 3,000 EAs were drawn in 1999, by means of probability proportional to size principles in each stratum. ***Weighting*** The 1999 OHS, in common with 1997 and 1998, was weighted to estimates of the population size. The estimates are based on the 1996 Census, as adjusted by a post-enumeration survey (PES), using post-stratification by province, sex and five-year interval age groups. In 1998 and 1999, relative scaling was also done, to cater for population group and urban/non-urban splits. The 1996 OHS was also weighted to the PES-adjusted count of Census 1996. However, because of the smaller sample size and the more clustered sample of households that was drawn, different weighting procedures were used. The previous OHS releases, both provincial and national, were based on weights derived from the 1991 Census. The data in the 1999, 1998, 1997 and 1996 OHS and those in the 1998, 1997 and 1996 are therefore not directly comparable with the previously published OHS figures for 1994 and 1995. Stats SA are supposed to have re-weighted the earlier surveys to reflect estimates of the population size based on the 1996 Census but these appear not to have been published. When the process is complete, comparisons between 1994, 1995, 1996, 1997, 1998 and 1999 should be possible. Early sample frames are regarded with some suspicion, and consequently, results generated using the 1995 OHS are to be handled cautiously.
How useful to researchers?	The data can only be reliably used at the provincial level and above as the lower levels of data are not representative. The reliability of the sample frame in earlier surveys has been questioned. There is also a structural break between the last OHS figures and the first LFS results which is caused by weighting the LFS figures to the 2001 Census and the OHS figures to the 1996 Census.
Data format	Data are available in flat, ASCII, fixed field format, with one line of given length per record from Stats SA or SPSS format from SADA.
Availability of data descriptions	A comprehensive metadata document is available on the CD from Stats SA. The following documentation is also available from the SADA website (http://www.nrf.ac.za/sada), although the later surveys have better documentation than the earlier ones: ▪ Metadata ▪ Questionnaire ▪ Code lists for occupation, districts, countries of birth, provinces and industries ▪ Record layout Questionnaires and code lists for 1995 to 1999 are also available from the DataFirst website (http://www.datafirst.uct.ac.za/surveys.html).
Conditions	The information products and services of Stats SA are protected in terms of the Copyright Act, 1978 (Act 98 of 1978). As the State President is the holder of State copyright, all organs of State enjoy unhindered use of the Department's information products and services, without a need for further permission to copy in terms of that copyright. Where a copy of the information is made available to any third party outside the State, the third party must be made aware of the existence of State copyright and ownership of the information by the State. The State (through Stats SA) retains the full ownership of its information, products and services at all times. Access to information does not give ownership of the information to the client. The use of any data is subject to acknowledgement of Stats SA as the supplier and owner of copyright. Users may apply or process the data, provided Stats SA is acknowledged as the original source of the data; that it is specified that the analysis is the result of the user's independent processing of the data; and that neither the basic data nor any reprocessed version or application thereof may be sold or offered for sale in any form whatsoever without prior permission from Stats SA. SADA conditions also apply to data obtained from SADA.
Costs	R 300 for a CD from Stats SA or free from SADA.
Tabled outputs	Various reports containing aggregated statistics on the OHS are available on the Stats SA website (http://www.statssa.gov.za/publications/findpublication.asp). These include: ▪ Hirschowitz et al. (2001) See http://www.statssa.gov.za/publications/SAInTransition/SAInTransition1999.pdf ▪ Hirschowitz (1996) See http://www.statssa.gov.za/publications/LivingInSA/LivingInSA.pdf
Contact	**For information:** ***Stats SA User Information Services*** Email: info@statssa.gov.za Telephone: +27 (0)12 310 8600

	For data: ***Stats SA User Information Services*** Email: info@statssa.gov.za Telephone: +27 (0)12 310 8600 ***SADA*** website: http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0059 (1995) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0075 (1996) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0076 (1997) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0113 (1998) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0114 (1999) ***DataFirst*** (available in resource centre)

Name	**Omnibus**
Principal investigator	HSRC
Year(s)	1994 (October), 1995 (February and September), 1996 (February)
Area(s) of interest	**General:** attitudes **Specific:** politics; economy; police; non South African citizens; elections; image of the government; local government; quality of life; crime and victimisation; services; health; national symbols; education
Brief description	**Summary** A public opinion survey. **Purpose** ▪ To investigate public attitudes about national priorities, social issues, political parties and the government's service delivery programme, and to reflect the extent of and attitudes towards the transition from the previous apartheid system to a constitutional democracy. ▪ To provide regular and reliable data and analysis of various national social priority issues. ▪ To help government, parliament, civil society and the various political parties in formulating policy priorities for the future. **Methodology** The target population was taken as the whole of South Africa. The data were collected by trained interviewers by means of personal face to face interviews with a structured questionnaire, with scientifically drawn respondents. The respondents were visited at home and the interviews were usually done after hours to ensure that the qualifying person was present. If the qualifying respondent was not available but would be available during the period of the fieldwork, an appointment was made and the particular household was revisited. If the qualifying respondent was not available for the duration of the survey, the household was replaced with another one. The questions were in Afrikaans and English in the questionnaire, were translated into the particular black languages of the survey area and the interviews were conducted in the language preferred by the respondents. ***Sample design*** Respondents were drawn by means of a multistage stratified cluster (probability) sample design. The population of persons 18 years and older was stratified according to province and socio-economic strata. The sample allocation to these strata was proportional to the 1991 population census figures with some exceptions (disproportion was introduced to give a minimum number of 120 respondents per province; the minimum number of Asians in the overall sample was fixed at 120; an additional sub-sample was introduced for live-in domestic workers, based on the incidence of households employing domestic workers as found in previous Omnibus surveys). Census enumerator areas and similar areas were used as the clusters in the penultimate sampling stage. The number of respondents per drawn enumerator area was either 4 or 8. All clusters were drawn with probability proportional to size, whilst households were drawn form the final clusters with equal probability (serial sampling procedure was applied). Respondents were drawn randomly (by applying a grid) from qualifying household members. The sampling interval was determined by dividing the number of visiting points in an EA by the sum of the two samples for that EA, e.g. 320 / 2x8 = 20. In order to ensure proper representation, the two questionnaires were then administered at alternate visiting points. ***Weighting*** Factor weighting: The factor weights to be applied to the captured data set were derived at the sampling stage. The aim of the factor weights was to correct the disproportions which were incorporated in the sample design (a minimum number of 120 respondents per province; a slightly over-sampled population of Asians). The only weighting targets were the stratification variables: 'province' and 'socioeconomic category.' RIM weighting: RIM weighting is generally used in two cases: ▪ When the purpose is to weight data according to various characteristics, but the relationship of

	the intersection of those characteristics is unknown, or ▪ When there are not enough respondents to fill all the possible cells (e.g. males who had passed standard 8, 35-44 years old, employed part-time.) As the RIM weighting process runs, it attempts to distort each variable as little as possible while still trying to attain all the desired proportions among the characteristics. The 'Root Mean Square' figure was used to determine how much distortion had been introduced. The following variables were submitted to the RIM weighting procedure at various stages: ▪ Age ▪ Sex ▪ Education ▪ Employment status ▪ Occupation ▪ Marital status ▪ Language ▪ Population group Close examination of outputs suggested that a satisfactory solution can not be obtained despite the increased number of iterations (too high between the weights ratio) and only 'Age', 'Sex', 'Education' and 'Population group' variables were retained. RIM weighting targets for the former TBVC states were estimated according to population characteristics drawn from the available Census 1991 data for the corresponding stratification variables. The 'rest of the RSA' component was weighted strictly according to the Census data.
How useful to researchers?	This provides historical data about attitudes in South Africa and complements the more recent Social Attitudes Surveys.
Data format	Data are available in SPSS format.
Availability of data descriptions	The following documentation is available from the SADA website (http://www.nrf.ac.za/sada): **For 1994:** ▪ Fieldwork report ▪ Questionnaire ▪ Code lists **For 1996:** ▪ Questionnaire Only hard copy documentation is available for 2005.
Conditions	SADA conditions of use apply (see notes).
Costs	None
Tabled outputs	Unknown
Contact	**For information:** ***Dr Mbithi wa Kivilu***, Director, Socio-Economic Surveys Section, Knowledge Systems Group, HSRC Email: jmkivilu@hsrc.ac.za **For data:** ***SADA*** website: http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0117 (October 1994) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0083 (February 1995) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0046 (September 1995) http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0039 (February 1996) ***DataFirst*** (available in resource centre)

Note: Other rounds of the Omnibus survey were conducted but these do not appear to be available from SADA or elsewhere. In addition to the four listed above, since 1994, the following rounds have been carried out: February and May 1994, May 1995, June and October 1996, February, June and September 1997. The two HSRC National Surveys, detailed above, are possibly part of the Omnibus series. A further three rounds were conducted (January 1998, June 1998 and February 1999). It is not known if these are available.

Name	**Rural Survey**
Principal investigator	Stats SA
Year(s)	1997
Area(s) of interest	**General:** housing; labour market **Specific:** dwellings; services; income; population structure; employment and unemployment; farming; land; activities and experience; farm processes and products; agricultural inputs and implements used; farming assistance, loans and operating costs
Brief description	**Summary** A sample survey of households in rural locations in the former homeland areas, i.e. Transkei, Bophuthatswana, Venda, and Ciskei (TBVC states); KaNgwane, KwaNdebele, KwaZulu,

	Gazankulu, Qwaqwa and Lebowa (self governing territories). **Purpose** ▪ To determine to what extent rural households in the former homelands had access to land and to income-generating activities. ▪ To obtain data on actual farming activities. **Methodology** ***Sample design*** The database that was established during the 1996 Census constituted the sampling frame for selecting EAs for the survey, A total of 600 enumeration areas (EAs) were drawn and 10 households were selected from each EA, yielding a sample of about 6, 000 households. Sample selection was carried out independently in each stratum applying a two stage sampling procedure; a systematic sample of EAs followed by a systematic sample of households. A power allocation scheme with 0.25 as the power was used to allocate the sample to each stratum. Former homelands in six provinces were regarded as natural occurring strata and as a result Western Cape, Northern Cape and Gauteng were not part of the strata of interest. Since the sample design was not necessarily a self-weighting design, weights had to be used. ***Weighting*** Two types of weights were calculated, person and household. Household weights were calculated by using the reciprocal of the product of an EA inclusion probability and a household inclusion probability. Relative scaling was done on the household weight due to the changes in the EA sizes from the initial database which was used for drawing of the sample to the final Census database which was used for weighting. Relative scaling was also used to take into account the growth rate of the population, from 1996 to 1997. Person weights were calculated by first post-stratifying the data by province (former homeland in a province) gender and 5 year age groups. The 1996 Census population figures were used as benchmarks for the post stratum. Since the target areas were former homelands, which are largely African populated, population group was implicitly taken into consideration in the benchmarks.
How useful to researchers?	Now ten years out of date, this dataset is unlikely to be of great use, except for historical purposes.
Data format	Unknown
Availability of data descriptions	Unknown
Conditions	The information products and services of Stats SA are protected in terms of the Copyright Act, 1978 (Act 98 of 1978). As the State President is the holder of State copyright, all organs of State enjoy unhindered use of the Department's information products and services, without a need for further permission to copy in terms of that copyright. Where a copy of the information is made available to any third party outside the State, the third party must be made aware of the existence of State copyright and ownership of the information by the State. The State (through Stats SA) retains the full ownership of its information, products and services at all times. Access to information does not give ownership of the information to the client. The use of any data is subject to acknowledgement of Stats SA as the supplier and owner of copyright. Users may apply or process the data, provided Stats SA is acknowledged as the original source of the data; that it is specified that the analysis is the result of the user's independent processing of the data; and that neither the basic data nor any reprocessed version or application thereof may be sold or offered for sale in any form whatsoever without prior permission from Stats SA. SADA conditions also apply to data obtained from SADA.
Costs	R 300 for a CD from Stats SA.
Tabled outputs	See http://www.statssa.gov.za/publications/P0360/P0360June1999.pdf
Contact	**For information:** ***Stats SA User Information Services*** Email: info@statssa.gov.za Telephone: +27 (0)12 310 8600 **For data:** ***Stats SA User Information Services*** Email: info@statssa.gov.za Telephone: +27 (0)12 310 8600 ***DataFirst*** (available in resource centre)

Name	**South African Migration and Health Survey**
Principal investigator	Centre for Population Studies (CENPOPS), University of Pretoria
Year(s)	1999/2000
Area(s) of interest	**General:** demography; health **Specific:** characteristics of community (geographic and population size, topographical features, and primary economic activities); origins of migrants; reasons for in-migration; reasons for out-migration; local infrastructure; availability of facilities and services; changes in agriculture and labour force needs; socio-demographic characteristics (age, sex, relation to head, education, labour force status, occupation, place of birth, date moved to current residence if not born there, children ever born, temporary absences from the household); housing characteristics (structure and facilities); income; migration experience; networks used in connection with migration; remittances to home place and visits; future mobility; fertility history; health; utilisation of health facilities; own health assessment
Brief description	**Summary** A national sample survey on the relationship between migration, economic change, women's status, reproduction and health for the black population of South Africa. The survey was a collaborative effort between CENPOPS, the HSRC and the Population Studies and Training Centre at Brown University, and was part of a four country study. Funding was provided by the Mellon Foundation and the HSRC. **Purpose** ▪ To enhance expertise on migration of personnel in national and local/regional research and planning agencies. This is consistent with the goals of South Africa's government to strengthen the domestic capacity to undertake research on population and its relation to the development process. ▪ To undertake field surveys that will provide both direct research experience for individual trainees and information on migration patterns and their causes and consequences. Such information will allow better insights on migration as a process. In so, doing, it should provide a stronger basis for policy development and evaluation on the interrelations among population redistribution development, women's status, health, and reproduction. ▪ To create the opportunity for comparative assessment of migration patterns and policies in the four countries in the UNFPA/Mellon Foundation project, in order to enhance understanding of migration in these countries in particular and in developing countries generally. **Methodology** The study was designed to gather information through a national sample survey of the African South African population 18 years or older at several levels: the community, the household, and the individual. To do so, three different questionnaires were developed to be used in the field surveys. *Community questionnaire*: A questionnaire was administered to selected knowledgeable respondents in each rural area where the survey was done. Information was sought on the basic characteristics of the community: geographic and population size, topographical features, and primary economic activities. The local situation with respect to migration was ascertained. If a receiving area, information was gathered on the origins of the migrants and their reasons for in-migration; if a sending area, destination of out-migrants was asked and their reasons for leaving. A section of the questionnaire was devoted to ascertaining features of the local infrastructure and changes in the availability of facilities and services during the five years preceding the survey. Detailed questions were also asked about any changes in agriculture and labour force needs. Parallel information was not asked for urban areas, since such data are available form the census and other statistical sources. Together, the data for rural and urban places can provide the background information on the contextual factors that are related to migration patterns and the decisions of individuals to migrate or not. *Household questionnaire*: For each household in the sample, either the head of the household or the spouse was asked a series of some 150 questions about persons in the household and the characteristics of the housing unit. A complete listing of all the usual residents of the household obtained information on their socio-demographic characteristics, including age, sex, relation to head, education, labour force status, occupation, place of birth, date moved to current residence if not born there, children ever born, and temporary absences from the household. Similar information was ascertained for all visitors to the household who had slept there the night preceding the interview. A separate section of the questionnaire focused on the housing characteristics, including kind of structure and facilities available. A listing of major possessions was also included, as were questions on sources of income. The final section of the questionnaire was devoted to persons who had moved out of the household since 1980 or since it was established, if after 1980. For all such persons, characteristics at time of move were ascertained, as were reasons for leaving,

	destination and current residence, and contacts with the surveyed household since the time of the move. *Individual questionnaire*: Finally, a detailed questionnaire was administered to a randomly selected adult within the household. The major focus of this questionnaire was the migration experience of the respondent. Included, therefore, was a complete migration history that ascertained origin and destination of each move, reasons for move, the work experience of the respondent just before and after the move and occupational change. Migrants were asked about networks used in connection with migration, and available resources, both the job search and living at destination. Remittances to home place and visits were also ascertained. All respondents were asked details about any moves of a week or more made during the year preceding the survey. The survey also obtained information on possible future mobility within the following three years. A full fertility history was also sought for each female respondent or the current spouse if a male respondent, including information on breastfeeding and antenatal care. Finally, a section was devoted to matters of health. Respondents were asked about utilisation of health facilities both for themselves and for their children, and they were asked to assess their own health in comparison to others their own age and, for migrants, in comparison to their health before migrating. The survey thus collected a wealth of data to assess the relation of migration to a broad array of other processes, and to put the migrants' experiences within the context of their communities of destination and within the time frame of their moves. This data set also allows great flexibility in defining migration per se. ***Sample design*** Because it was assumed that the underlying migration process was more complex than what would be reflected in a simple distinction between permanent and temporary migrants, a strictly random process was used in selecting respondents from within the sampled households. South Africa was stratified into three primary strata: (a) metropolitan areas, (b) other urban areas and (c) rural areas. Samples were then drawn independently from each of the three types of localities. Initially, in each of the three locality types, 800 respondents were to be drawn, resulting in a total sample size of 2,400. The 800 respondents in each stratum were to be drawn from 20 randomly selected Primary Sampling Units (PSUs), either a transitional local council (TLC) or a transitional rural council (TRC) by randomly selecting four Enumerator Areas (EAs) in each PSU. From each selected EA 10 households would be randomly selected, and finally, one adult respondent would be selected randomly in each household. It was later decided to draw 11 households in each EA, instead of 10, to ensure that there would be sufficient room to deal with refusals and non-responses. The first stage of the sampling process was the drawing of the 20 PSUs in each of the three primary strata. Four secondary strata were formed within this stratum. They were Gauteng, Durban Metro, Cape Metro and PE-Uitenhage-Kirkwood. The 20 PSUs that had to be drawn in these four strata were allocated using the proportion migrants as well as the total population in each of the strata, as observed during Census '96. This was done to ensure that not too many PSUs were allocated to a very small stratum purely because of a high proportion of migrants in that stratum. Using both the proportion migrants and the total population meant that in such a case the small population reduced the allocation to that stratum. In each of these four secondary strata, the following procedure was followed to draw the required number of PSUs. The PSUs were grouped into three categories according to the proportion migrants, resulting in a category that would contain PSUs with a high proportion of migrants, a category that would contain PSUs with a medium proportion migrants and a category that would contain PSUs with a low proportion migrants. The number of PSUs allocated to the stratum was then allocated to these three categories such that more PSUs were drawn in the 'high proportion migrants' category, fewer in the 'medium proportion migrants' category and the smallest number in the 'low proportion migrants' category. Any potential bias in this allocation would be rectified by weighting Within each of the three categories, the allocated number of PSUs was subsequently randomly drawn with probability proportional to their total population, using a systematic sampling procedure. This means that the PSUs were ordered in a specific way – geographically in this case – and every nth PSU was then drawn, with the starting point being selected randomly. The PSUs in the other urban areas were similarly assigned to one of the three categories. The 20 PSUs that had to be drawn were then allocated to these three categories, 12 to the 'high proportion' category, six to the 'medium proportion' category and two to the 'low proportion' category. The PSUs in each of these three categories were then drawn after an implicit stratification by province had been undertaken. This means that the provinces were first ordered meaningfully (and also the PSUs within each province) and then a systematic sample of PSUs (ignoring province) was drawn randomly with probability proportional to the total population. For rural areas, the drawing of the 20 PSUs was done in exactly the same way as for urban areas.

	After the PSUs had been selected, four EAs were drawn in each PSU. Instead of drawing the four EAs randomly in a particular PSU, it was decided to first group the EAs in each PSU into categories containing EAs with a high proportion migrants, with a medium proportion migrants, and with a low proportion migrants. Subsequently, two EAs were allocated to the 'high proportion migrants' category, one EA to the 'medium proportion' category and one EA to the 'low proportion' category. The EAs were subsequently drawn from the category to which they were allocated, using systematic random sampling with selection probability proportional to the size (i.e. the 1996 black South African population) of the EAs. The next step was to select 11 households in each selected EA. This was done using systematic random sampling after the fieldwork team had completed a re-listing of visiting points in the EA to update the 1996 information. In each of the selected households, one qualifying respondent (i.e. a person 18 years or older) was randomly selected and interviewed. ***Weighting*** Since the allocation of sample elements was not done proportionally, it was necessary to introduce unequal weights to the sample elements in order for the sample to be a true representation of the population.
How useful to researchers?	An important resource for those studying migration and health.
Data format	Data are available in SPSS format.
Availability of data descriptions	The following documentation is available from the SADA website (http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0130): ▪ Description of methodology ▪ Questionnaires
Conditions	SADA conditions of use apply (see notes).
Costs	None
Tabled outputs	Kok et al. (2006) See http://www.hsrcpress.ac.za/full_title_info.asp?id=2094
Contact	**For information:** ***Prof. Louis van Tonder***, Director, Centre for Population Studies, University of Pretoria Email: louis@postino.up.ac.za Telephone: +27 (0)12 420 3302 **For data:** ***SADA*** website: http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0130

Name	**South African National HIV Prevalence, HIV Incidence, Behaviour and Communication Survey**
Principal investigator	HSRC
Year(s)	2002, 2005
Area(s) of interest	**General:** health **Specific:** HIV incidence; HIV prevalence; risky behaviours; social and cultural determinants of HIV; HIV and mental health
Brief description	**Summary** The Nelson Mandela Foundation (NMF) commissioned the first national, household sero-prevalence survey of HIV/AIDS in 2002. Following the roll out of a comprehensive programme for the care and treatment of HIV-infected individuals and investment in mass media campaigns aimed at preventing new infections, the NMF realised that it was important to assess the extent to which these policies and practices had changed the shape of the pandemic in South Africa by following up on the first survey. The second national HIV survey was carried out in 2005 as part of a series of national household surveys conducted by the HSRC. The next survey is planned for 2008. **Purpose** ▪ To determine HIV incidence and prevalence by age, sex, province, race, and locality type. ▪ To examine the social, behavioural and cultural determinants of HIV infection. ▪ To assess relationship between HIV and mental health and identify a baseline. ▪ To asses public opinion on HIV, anti-retrovirals and vaccines. **Methodology** The methods used to collect data included questionnaires and testing of dry blood spot samples for HIV antibodies as well as HIV incidence. The 2005 questionnaire had the same questions as the 2002 version, as well as some extra modules. Of the 24,236 individuals eligible for the 2005 survey, 23,275 (96.0%) completed the interview and 15,851 (65.4%) agreed to HIV testing.

	Sample design The survey used a multi-stage, disproportionate, stratified sampling approach. The sample was based on a master sample of 1000 Enumerator Areas (EAs) used by Stats SA for the 2001 Census. The sample was stratified by province and locality type of the EAs. Locality types were urban formal, urban informal, rural formal and rural informal. In urban formal areas, race was used as a third stratification. Therefore urban formal areas were over-sampled in order to ensure sufficient sample of white and Indian racial groups. Within each household, 3 individuals could be selected, with only one from each of age groups: 2-14, 15-24, 25+. People living in educational institutions, old age homes, hospitals and uniformed service barracks were excluded, but those living in hostels were included.
How useful to researchers?	The survey allows for analysis of HIV against a wider selection of demographic factors (i.e. race, residence type, marital status) than the National HIV and Syphilis Antenatal Sero Prevalence Survey (see above). It also allows analysis of HIV against information collected on knowledge, attitudes, sexual behaviours etc. The 2005 survey is a repeat of 2002, which allows for comparisons over time.
Data format	Data are available in SPSS and Stata format.
Availability of data descriptions	Full sets of metadata (manuals, descriptions of the HIV testing process) are available with the data.
Conditions	HSRC is currently in negotiation with Stats SA to make the data from the 2002 and 2005 surveys freely available and in the public domain. Currently, the research group receives many requests for use of the data. They are willing to collaborate with external researchers, especially if the research proposal fits with the research areas of the Social Aspects of HIV/AIDS and Health programme.
Costs	There may be costs associated with data extraction, but this is dependent on the individual project.
Tabled outputs	Shisana et al. (2005) See http://www.hsrcpress.ac.za/full_title_info.asp?id=2134
Contact	**For information:** ***Prof. Thomas Rehle***, Director, Social Aspects of HIV/AIDS and Health, HSRC Email: trehle@hsrc.ac.za Telephone: +27 (0)21 466 7844 **For data:** A request for data should be sent in the form of a written research proposal to ***Prof. Thomas Rehle*** Email: trehle@hsrc.ac.za Telephone: +27 (0)21 466 7844

Name	**South African Social Attitudes Survey** (SASAS)
Principal investigator	HSRC
Year(s)	2003, 2004, 2005, 2006
Area(s) of interest	**General:** attitudes **Specific:** race; class; politics; poverty; inequality; service delivery; partner violence; moral attitudes
Brief description	**Summary** A nationally representative sample survey of approximately 5,000 adults that gathers information on the public's attitudes, beliefs, behaviour patterns and values. SASAS has been designed as a time series, with the British Social Attitudes Survey series in mind. **Purpose** To provide a long-term account of the social fabric of modern South Africa, and of how its changing political and institutional structures interact over time with changing social attitudes and values. **Methodology** (descriptions refer to SASAS 2003, but the others are similar) ***Sample design*** SASAS has been designed to yield a representative sample of adults aged 16 and older. The sampling frame for the survey is the HSRC's master sample, which was designed in 2002 and consists of 1000 primary sampling units (PSUs). The 2001 population census enumerator areas (EAs) were used as PSUs. These PSUs were drawn, with probability proportional to size, from a pre-census 2001 list of EAs provided by Stats SA. The master sample excludes special institutions (such as hospitals, military camps, old age homes, schools and university hostels), recreational areas, industrial areas and vacant EAs. It therefore focuses on dwelling units or visiting points as secondary sampling units (SSUs), which have been defined as 'separate, (non-vacant) residential stands, addresses, structures, flats, homesteads, etc'. As the basis of the 2003 SASAS round of interviewing, a sub-sample of 500 PSUs was drawn from the HSRC's master sample. Three explicit stratification variables were used, namely province, geography type and majority population group. Within each stratum, the above allocated number of PSUs was drawn using proportional to size probability sampling. In each of these drawn PSUs, two clusters of eleven dwelling units each

	were drawn. These 22 dwelling units in each drawn PSU were systematically grouped into three sub-samples of sizes 7, 7 and 8 to give the two SASAS samples and a client survey that was run in parallel. Interviewers called at each visiting point selected from the master sample and listed all those eligible for inclusion in the sample – that is, all persons currently aged 16 or over and resident at the selected visiting point. The interviewer then selected one respondent using a random selection procedure based on a Kish grid. ***Weighting*** The data were weighted to take account of the fact that not all the units covered in the survey had the same probability of selection. The weighting reflected the relative selection probabilities of the individual at the three main stages of selection: visiting point (address), household and individual. Visiting points in the Northern Cape and KwaZulu-Natal were oversampled, due to the small population size in the former and a desire to ensure a decent Indian sample in the latter. Because of this, weights had to be applied to compensate for the greater probability of being selected. The weights were then scaled down to make the number of weighted productive cases exactly equal to the number of unweighted productive cases.
How useful to researchers?	A very important resource for exploring South African attitudes on a range of issues.
Data format	Data, if made available, will be in SPSS, Stata and tab-delimited formats.
Availability of data descriptions	Unknown
Conditions	There is no official mandate to distribute data. At present, data is provided to those writing chapters for the SASAS volumes, as well as interested internal parties. Collaborative analysis between HSRC and non-HSRC researchers could possibly also be permitted outside the scope of the books, but this situation has not yet arisen.
Costs	Unknown
Tabled outputs	Pillay et al. (2006) See http://www.hsrcpress.ac.za/full_title_info.asp?id=2139 (the first volume in the series presenting the public's responses during interviews conducted in 2003)
Contact	**For information:** ***Mr Ben Roberts***, Research Specialist, Urban, Rural and Economic Development, HSRC Email: BRoberts@hsrc.ac.za

Name	**Student Choice Behaviour Project Phase 1**
Principal investigator	HSRC
Year(s)	2001
Area(s) of interest	**General:** education **Specific:** present education; choice of higher education course and institution; future work and life prospects
Brief description	**Summary** A survey of Grade 12 learners, investigating future choices with regard to education, commissioned by the Department of Education. The project is the first in a series of studies which investigate the transition from school to work. **Purpose** ▪ To understand the factors affecting the choices Grade 12 learners make with regard to entering higher education (HE), institution type (university or technikon) and specific institution, and programme of study. ▪ To facilitate planning at secondary education level in the areas of subject focus, career guidance, learning improvement and teacher upgrading – in particular, enabling Department of Education planners and school management to devise strategies to increase the number of Grade 12 learners entering HE in subjects that will lead to HE qualifications that address the national demand for specific skills; plan more effective career guidance programmes at the Grade 11 and 12 levels; and target specific subjects and subject areas for teacher improvement programmes. **Methodology** The survey was conducted amongst Grade 12 learners during a two-week period in August 2001. A total of 12,204 learners in 288 schools completed a questionnaire that generated the data for the study. Up to 48 learners per school participated in the survey. The questionnaire, which contains 65 questions (583 variables) has nine sections: 1. Present school situation. 2. One year from now (preferred and anticipated life situations one year hence). 3. Intention to study at a university or technikon.

	4. Application to study at a university or technikon. 5. Choice of university or technikon. 6. Choice of study programme. 7. Looking to the future (anticipated life – particularly work – situation ten years hence). 8. Personal information (biographical data). 9. Freedom of choice (unfettered choice with regard to higher education and work). Within these sections, which focus on three variables (entry into higher education, choice of institution, and choice of field of study), there are data on learners' subject choices and marks; career guidance; factors affecting learners' choices; application to study at institutions; future work and life prospects in South Africa; and various socio-economic indicators. ***Sample design*** In a two-stage sampling procedure, schools were stratified by province and by pass rate in the 2000 matriculation examinations, and, in cases where there were more than 48 learners per Grade 12 group in a school, learners were selected by a variety of sampling methods.
How useful to researchers?	A specialised dataset relating to people's choices about entering higher education.
Data format	Data are available in SPSS files (one for each section). Each file contains the weighting variable for descriptive analyses and the scaled weighting variable for inferential analyses.
Availability of data descriptions	See http://hrdwarehouse.hsrc.ac.za/datasource/hsrc/studentchoicebehavior/phase1/studentchoice1questionnaire.pdf for questionnaire.
Conditions	Requests for information must be addressed to the HRD Data Warehouse (contact details below). Requests will be processed depending on the type of the request and the availability of a researcher to process the request. Not all requests will be processed. The HSRC owns the database and any data received may be used provided that the HSRC is acknowledged as the source of the data and the data are used for non-profit purposes only.
Costs	None
Tabled outputs	Cosser and Du Toit (2002) See http://www.hsrcpublishers.ac.za/full_title_info.asp?id=1975
Contact	**For information:** ***Mr Michael Cosser****,* Chief Research Specialist, Research Programme on Education, Science and Skills Development, HSRC Email: mcosser@hsrc.ac.za **For data:** ***Mr Michael Cosser****,* Chief Research Specialist, Research Programme on Education, Science and Skills Development, HSRC Email: mcosser@hsrc.ac.za

Note: In Phase 2 of the project, a cross-section respondents who participated in the 2001 survey were traced into their various destinations (see http://www.hsrcpublishers.ac.za/full_title_info.asp?id=1948 for report). Data are also available from ***Mr Michael Cosser****,* Chief Research Specialist, Research Programme on Education, Science and Skills Development
Email: mcosser@hsrc.ac.za

The Phase 1 survey was repeated in 2005, obtaining a response of 20,659, and a tracer study will be conducted soon. Data are not yet available.

Name	**Survey of Activities of Young People** (SAYP)
Principal investigator	Stats SA
Year(s)	1999
Area(s) of interest	**General:** labour market; education **Specific:** living conditions; demography; migration; household income; school attendance; economic and non-economic activities of children; adult employment status and income
Brief description	**Summary** A household-based survey gathering information on the extent, nature, patterns, determinants and consequences of the work-related activities of children aged 5-17 years other than formal learning and leisure. The survey was commissioned by the Department of Labour. **Purpose** ▪ To gather information necessary for formulating an effective programme of action to address the situation of working children in South Africa, including child labour. ▪ To produce comprehensive statistical data on the work activities of young persons at the

	national level. ▪ To provide a comprehensive analysis of the state of the nation's working children, identifying major parameters, priority groups and patterns, the extent and determinants of child work, conditions and effects of work, etc. ▪ To look closely at factors such as economic activity, excessive household chores and maintenance activities at school which may be affecting young people's abilities to attend school or engage in other childhood activities. ▪ To disseminate as widely as possible the results of the nationwide survey on the activities of young persons, in particular, the areas where such activities are most intensive. ▪ To formulate a module on the activities of young persons to be attached to one of the rounds of the half-yearly Labour Force Survey. **Methodology** Fieldwork for the survey took place in June and early July 1999 by means of face-to-face interviews. The survey gathered detailed information in two phases. In phase one, interviews were conducted in a probability sample of 26,081 households throughout the country, in order to determine the extent of work-related activities among children. Information was gathered on about 33,000 children between the ages of 5 and 17 years. In the second phase, follow-up interviews were conducted in a probability sub-sample of 4,494 of the households in which the first phase showed that at least one child was engaged in these activities. More extensive questions about the nature of work the children were doing were put to an adult in the family, and to the child or children involved in these activities. During this phase, information was gathered on about 10,000 children between the ages of 5 and 17 years. Two questionnaires were developed for the survey, one for each phase. The first phase questionnaire covered the following topics: ▪ Living conditions of the household, including the type of dwelling, fuels used for cooking, lighting and heating, water source for domestic use, land ownership, tenure and cultivation. ▪ Demographic information on members of the household (both adults and children) including age, gender and population group of each household member, marital status, relationships to each other, and levels of education ▪ Migration of the household in the two years prior to the survey. ▪ Household income. ▪ School attendance of children aged 5-17 years. ▪ Among the children aged 5-17 years, information on economic and non-economic activities in the 12 months prior to the survey, if any. The household was included in the selection process for phase two of the survey if there was any child in the household who was: (a) regularly engaged in non-economic work activities, fetching wood and/or water, or unpaid domestic work for one hour or more per day; or (b) engaged in any other economic work activities at all during the specified time period of one year. For the second phase questionnaire, one or more adults in the household answered the following questions: ▪ The employment status of all adults in the household aged 18 years or more. ▪ Details of the type of work in which the employed adults were engaged. ▪ Income earned by each adult in the past 12 months. ▪ The type of work-related activity each child aged 5-17 years in the household was engaged in, if any. ▪ Reasons for the child/children to engage in these activities. ▪ School attendance and problems at school, regarding children aged 5-9 years. ▪ Safetv and health, illness and injury related to work-related activities, regarding children aged 5-9 years. ▪ Each child in the household aged between 5 and 17 years was asked to answer the following ▪ questions: ▪ Whether engaged in work-related activities during the past year and during the past seven days and the type of activity, if any. ▪ Details of type of work, sector and occupation, among children engaged in economic activity for pay, profit or economic family gain. ▪ Times of the year and times during the day when those involved in these activities worked. ▪ Reasons for engaging in these activities. ▪ Conditions under which the work took place, including whether the child (if a paid employee) experienced sexual harassment or abuse at work. ▪ Income earned, and proportion of earnings paid to adults in the household. ▪ Safety and health, illness and injury, related to economic activity (asked only to children aged

	10 to 17 years). ▪ Whether looking for work. ▪ Main activity. ▪ School attendance and (if attending school) difficulties experienced at school. ▪ Reasons for missing school or not attending school. ***Sample design*** The sampling frame for the selection of primary sampling units (PSUs) was based on the results of the demarcation of the country into enumerator areas (EAs) for Census 1996. A PSU contains at least a hundred households, whereas an EA may be smaller than this. In cases where an EA is smaller than 100 households, it is combined with adjacent EAs to form larger units. A PSU can therefore consist of two or more combined small EAs, or a single larger EA. The sample for the survey was drawn in two phases. In the first phase, 900 PSUs were selected using probability-sampling techniques. Of these PSUs, 579 were situated in urban areas, and 321 were situated in non-urban areas. The main aim of the first phase of the survey was to examine the extent to which South Africa's children are engaged in activities which can be regarded as children's work. For sampling for phase 1, within each province, urban areas were further stratified into formal and informal settlements, while non-urban areas were further stratified into commercial farms and other non-urban areas (largely traditional rural areas). Within each PSU in urban areas, 25 households were interviewed, while within each PSU in non-urban areas, 50 households were interviewed. These households were selected by means of systematic sampling. The main aim of phase 2, on the other hand, was to explore work-related activity by children in more detail. A further sample was drawn, consisting of at most ten households in each PSU in rural areas and five households in each PSU in urban areas, among those households where there was evidence of at least one child per household being engaged in work-related activities (using the low cut-off points) in phase 1. Before systematic sampling was applied for the selection of households in Phase 1, all the dwelling units within a sampled PSU were listed. This was done in order to select the households to be visited for the first phase (all households in a dwelling unit were interviewed), and to enable a sub-sample of households to be drawn for the second phase to which the fieldworker could return. Dwelling units in South Africa, particularly in rural areas and informal settlements, do not necessarily have addresses. It is therefore important to do a complete listing of dwelling units in a particular PSU to ensure that the same unit can be identified on a map and on the ground, and re-visited. ***Weighting*** Two different set of weights were used for this study,(household and individual weights). The 1996 Census, as adjusted by a post-enumeration survey, was used as a basis for the weighting. Household weights were calculated by using the reciprocal of the inclusion probabilities. For the person weight, the inflated data were post-stratified by province, gender and age group (5-year age groups), and a population control adjustment based on the Stats SA population estimates (using the 1996 Census) was applied.
How useful to researchers?	This is an important resource that would complement data on young people such as CAPS and Birth to Twenty, and has a special focus on child labour.
Data format	Data are available in ASCII, SPSS or SAS format.
Availability of data descriptions	The following documentation is available on the DataFirst website (http://www.datafirst.uct.ac.za/surveys.html) and the International Labour Organisation (ILO) website (http://www.ilo.org/public/english/standards/ipec/simpoc/southafrica/index.htm): ▪ Metadata - includes a brief methodology ▪ Questionnaires ▪ Industry and occupation codes A technical report is provided in the full report on the survey on the Department of Labour website: http://www.labour.gov.za/useful_docs/doc_display.jsp?id=9506
Conditions	The ILO imposes certain conditions of use, but it is not known whether Stats SA or the Department of Labour have their own conditions. Users obtaining data (for personal, non commercial use) from the ILO are required: ▪ To act at all times so as to preserve the confidentiality of individuals and institutions recorded in the Materials. In particular the User undertakes not to use or attempt to use the materials to derive information relating specifically to an identified individual or institution nor to claim to have done so. ▪ To acknowledge the ILO in any publication, whether printed, electronic or broadcast, based wholly or in part on such materials. To declare in any such work that the ILO and those who carried out the original collection and analysis of the data bear no responsibility for their further

	analysis or interpretation. ▪ To deposit with the ILO an electronic copy of any published work or report based wholly or in part on such materials. The ILO at its own discretion may make it available to the public. ▪ At any time at the request of the ILO, to offer for deposit on a suitable medium any new datasets which have been derived from the materials supplied or which have been created by the combination of the data supplied with other data. The deposit of the derived datasets will include sufficient explanatory documentation to enable the new data files to be accessible to others. ▪ To notify the ILO of any errors discovered in the materials. ▪ To accept that the ILO or the country that conducted the actual survey bears no legal responsibility for their accuracy or comprehensiveness.
Costs	None
Tabled outputs	▪ Statistics South Africa (2001a) See http://www.labour.gov.za/useful_docs/doc_display.jsp?id=9506 ▪ Statistics South Africa (2001b) See http://www.labour.gov.za/useful_docs/doc_display.jsp?id=9505
Contact	**For information:** ***Stats SA User Information Services*** Email: info@statssa.gov.za Telephone: +27 (0)12 310 8600 **For data:** ***ILO*** website: http://www.ilo.org/public/english/standards/ipec/simpoc/southafrica/index.htm ***DataFirst*** (available in resource centre)

Name	**Survey of Independent Schools**
Principal investigator	HSRC
Year(s)	2002
Area(s) of interest	**General:** education **Specific:** independent schools; learners; staff; finance
Brief description	**Summary** A national quantitative survey profiling independent (private) schools. **Purpose** To fill a gap in data sources on independent schooling. **Methodology** A standardised instrument was mailed to a total of 1,287 registered independent schools in South Africa. A response rate of 31.9% was achieved yielding a sample of 410 schools. The sample of schools is fairly representative of the population. Each school in the sample completed a single instrument that captured mainly quantitative data on the school e.g. headcounts of learners and teachers. ***Weighting*** The 410 cases in the database are statistically weighted using locational attributes to be more representative of the population of 1,287 schools.
How useful to researchers?	Useful for researchers interested in taking private education into account.
Data format	Data are available in SPSS format. There are approximately 90 to 100 variables in the database ranging from administrative variables to indicators and calculations of indicators in the instrument.
Availability of data descriptions	See http://hrdwarehouse.hsrc.ac.za/datasource/hsrc/independentschools/independentschoolssurveyform.pdf for questionnaire.
Conditions	Requests for information must be addressed to the HRD Data Warehouse (contact details below). Requests will be processed depending on the type of the request and the availability of a researcher to process the request. Not all requests will be processed. The HSRC owns the database and any data received may be used provided that the HSRC is acknowledged as the source of the data and the data is used for non-profit purposes only.
Costs	None
Tabled outputs	Du Toit (2004) See http://www.hsrcpress.ac.za/full_title_info.asp?id=1987
Contact	**For information:** Unknown **For data:** ***HRD Data Warehouse*** Email: hrdwarehouse@hsrc.ac.za

Name	**Survey of Private and Further Education and Training in South Africa**
Principal investigator	HSRC
Year(s)	2002
Area(s) of interest	**General:** education **Specific:** further education and training
Brief description	**Summary** A survey of providers in the private post-school Further Education and Training (FET) sector in South Africa, providing details of the location, programmes offered, modes of delivery and participants in this sector. **Purpose** To develop a comprehensive picture of the private FET sector in order to achieve better management of provision. **Methodology** Since current information on this sector was either speculative or non-existent, data provided by the Department of Education (DoE) to pre-register private FET providers was used to define the sample. The DoE dataset provided information on provider size, location and typology, and this survey elaborated on the nature and form of provision, including ownership profile, learner and staff composition, and nature of programmes offered. From a database containing 864 providers that pre-registered with the DoE, 238 provided responses to the electronic survey (28% response rate).
How useful to researchers?	Useful for researchers interested in taking private provision of further education and training into account.
Data format	Data are available in an Access database.
Availability of data descriptions	See http://hrdwarehouse.hsrc.ac.za/datasource/hsrc/privatefet/privatefetsurveyform.pdf for questionnaire.
Conditions	Requests for information must be addressed to the HRD Data Warehouse (contact details below). Requests will be processed depending on the type of the request and the availability of a researcher to process the request. Not all requests will be processed. The HSRC owns the database and any data received may be used provided that the HSRC is acknowledged as the source of the data and the data is used for non-profit purposes only.
Costs	None
Tabled outputs	▪ See http://hrdwarehouse.hsrc.ac.za/datasource/hsrc/privatefet/privatefet.pdf ▪ Akoojee (2005) See http://www.hsrcpublishers.ac.za/full_title_info.asp?id=1995
Contact	**For information:** Unknown **For data:** ***HRD Data Warehouse*** Email: hrdwarehouse@hsrc.ac.za

Name	**The 2001-2002 HSRC Migration Survey**
Principal investigator	HSRC
Year(s)	2001/2002
Area(s) of interest	**General:** demography; labour market **Specific:** migration; population; migration intentions; economic activity; unemployment; employment; self-employment
Brief description	**Summary** A national sample survey investigating the causes of migration. **Purpose** ▪ To determine the importance of economic and non-economic factors in the explanation of migration trends in South and southern Africa. ▪ To pave the way for a likely future longitudinal study to determine the extent to which and the circumstances under which migration intentions lead to actual moves. **Methodology** A survey questionnaire was administered to one randomly selected individual aged 18-69 years in each selected household. The size of the envisaged sample (4,266 households) was intended to ensure an adequate number of cases of analytical purposes. ***Sample design*** A stratified, clustered random (complex) sampling design was used. The explicit strata used in the

	sampling design were the following: (a) Census-based population group (African, coloured, Indian or white); (b) Metropolitan/non-metropolitan location; (c) whether in a so-called spatial development area or not; and (d) Census-based locality type (rural/urban). In non-metropolitan areas an implicit stratification of local government level (city, other local municipality or district municipality) was also used. Clustering took the form of randomly selecting (using systematic sampling) six households in each of the randomly selected 1996 Census Enumerator Areas across the different strata. ***Weighting*** The over representation of women in the realised sample and the allocation of weights presented certain challenges. Initially the weighting was done purely on the basis of the sampling frame. This initial weighting resulted in excessive weights in a number of strata, mainly because of the large number of refusals and no contacts, and because not all fieldwork teams and interviewers recorded the required information (especially the number of visiting points per EA and the number of households per visiting point). When this problem was discovered, the weighting was redone and it was decided to undertake a benchmarking exercise as well, to reduce excessive over- and under-representation. This was done in terms of locality type, age, gender, educational attainment and population group. These final weights are essential for all further data analyses.
How useful to researchers?	An important resource for those studying issues relating to migration.
Data format	Data are available in SPSS format.
Availability of data descriptions	Little documentation is available. Appendix 1 and Chapter 7 of Kok et al. (2006) give a history of the data and basic methodology reports. See http://www.hsrcpress.ac.za/full_title_info.asp?id=2094
Conditions	SADA conditions of use apply (see notes).
Costs	None
Tabled outputs	Kok et al. (2006) See http://www.hsrcpress.ac.za/full_title_info.asp?id=2094
Contact	**For information:** ***Dr Pieter Kok***, Chief Research Specialist, Urban, Rural and Economic Development, HSRC Email: pckok@hsrc.ac.za **For data:** ***SADA*** website: http://www.nrf.ac.za/sada/ahdetails.asp?catalognumber=0138 ***DataFirst*** (available in resource centre)

Name	**Time Use Survey**
Principal investigator	Stats SA
Year(s)	2000
Area(s) of interest	**General:** labour market; health; education; transport **Specific:** household information; demographic information (gender particularly important); time use; activities; economic work; housework; care work; learning; leisure; personal care; travel
Brief description	**Summary** A national sample survey, using a diary methodology, to investigate how people aged 10 years and above spend their time. **Purpose** ▪ To provide information on the way in which different individuals in South Africa spend their time. Such information contributes to greater understanding of policymakers on the economic and social well-being of different societal groups. ▪ To provide new information on the division of both paid and unpaid labour between women and men, and greater insight into less well understood productive activities such as subsistence work, casual work and work in the informal sector. ▪ To improve concepts, methodology and measurement of all types of work and work-related activity. **Methodology** The fieldwork for the study was conducted in three rounds - February, June, and October 2000 - so as to catch possible seasonal variations in time use. The study used a 24-hour diary, divided into half-hour slots, as the core instrument to record activities. In each slot, a maximum of three activities could be recorded. The diary was administered face-to-face to the respondent by means of an interview. In addition to the diary, the questionnaire contained many of the standard questions of Stats SA household surveys. Thus one member per household provided basic information about the household as a whole, and, before administration of the diary, the respondent was asked for basic demographic information about themselves, such as age, sex, children and work situation. The questionnaire for the time use survey comprised five sections.

	Section one covered details of the household. Section two covered demographic details of the first person selected as a respondent in that household. Section three consisted of a diary in which to record the activities performed by the first person selected during the 24 hours between 4 am on the day preceding the interview and 4 am on the day of the interview. Sections four and five were for the second selected person in the household but were otherwise identical to sections two and three respectively. ***Sample design*** The sample for the time use survey was chosen so as to be representative of the country's population. Each round included households from all nine provinces and from four different strata, or types of settlement area. The strata were formal urban settlements, informal urban settlements, commercial farming areas, and other rural areas. The survey used a sampling frame that had been prepared for the Survey of Activities of Young People (see above). For this survey, the stratum variable was used in weighting both households and individual respondents so as to ensure that the reported figures more accurately reflected the population and its activities. The planned sample for the survey was 10,800 dwelling units, 3,600 per round. Each dwelling unit could contain more than one household, although this is unusual. The realised sample was smaller than planned, at 8,564 households and 14,553 respondents. The main reasons for non-realisation of the full sample were unoccupied dwelling units and dwelling units which were marked on the maps but not found on the ground. Some of the discrepancies, particularly in informal areas, were due to changes on the ground that occurred between mapping of the areas and enumeration. The number of refusals was very small. The survey aimed to collect information from two respondents aged 10 years and above from each selected household. Where the household contained only one person in this age group, only that person was interviewed. The interviewer first listed all household members from oldest to youngest. The respondents to be interviewed were chosen systematically through the use of a selection grid included in the interviewer's manual. If only those households with which contact was made are included, the survey captured diaries from 94% of targeted respondents. ***Weighting*** The raw data were weighted so as to adjust the responses collected to be representative of the underlying sample frame. Because the sample frame itself was designed so as to reflect the population of South Africa aged 10 years and above, the results reported should reflect the proportions in the total population of this age in terms of sex, population group, age group and settlement type. Because of the relatively small size of the sample, the results were weighted to reflect the 25,000 odd individuals aged 10 years and above whom one would have expected to find in 10,800 dwelling units rather than the number of people of this age in the full population. This approach was adopted to remind users and readers that the survey is smaller than the usual Stats SA surveys and thus slightly less reliable in terms of extrapolation. The proportions should, nevertheless, reflect the full population.
How useful to researchers?	A useful resource on time use relating to a number of themes, for which virtually no other information is available for South Africa or for most African countries.
Data format	Data are available in ASCII format.
Availability of data descriptions	The following documentation is available on the data CD: ▪ Metadata ▪ Fieldwork manual ▪ Questionnaire ▪ Code lists The questionnaire is available on the DataFirst website (http://www.datafirst.uct.ac.za/surveys.html).
Conditions	The information products and services of Stats SA are protected in terms of the Copyright Act, 1978 (Act 98 of 1978). As the State President is the holder of State copyright, all organs of State enjoy unhindered use of the Department's information products and services, without a need for further permission to copy in terms of that copyright. Where a copy of the information is made available to any third party outside the State, the third party must be made aware of the existence of State copyright and ownership of the information by the State. The State (through Stats SA) retains the full ownership of its information, products and services at all times. Access to information does not give ownership of the information to the client. The use of any data is subject to acknowledgement of Stats SA as the supplier and owner of copyright. Users may apply or process the data, provided Stats SA is acknowledged as the original source of the data; that it is specified that the analysis is the result of the user's independent processing of the data; and that neither the basic data nor any reprocessed version or application thereof may be sold or offered for sale in any form whatsoever without prior permission from Stats SA.
Costs	R 300 for a CD from Stats SA.
Tabled outputs	▪ Statistics South Africa (2001c) See http://www.statssa.gov.za/publications/TimeUse/TimeUse2000.pdf ▪ Chobokoane and Budlender (2002)

	See http://www.statssa.gov.za/publications/TechPaperActivitiesTime/TechPaperActivitiesTime.pdf ▪ See http://www.datafirst.uct.ac.za/resource/young/syp_tables.pdf for tables.
Contact	**For information:** ***Ms Debbie Budlender*** Email: debbieb@mail.ngo.za **For data:** ***Stats SA User Information Services*** Email: info@statssa.gov.za Telephone: +27 (0)12 310 8600 ***DataFirst*** (available in resource centre)

A1.4 Other Census and survey data

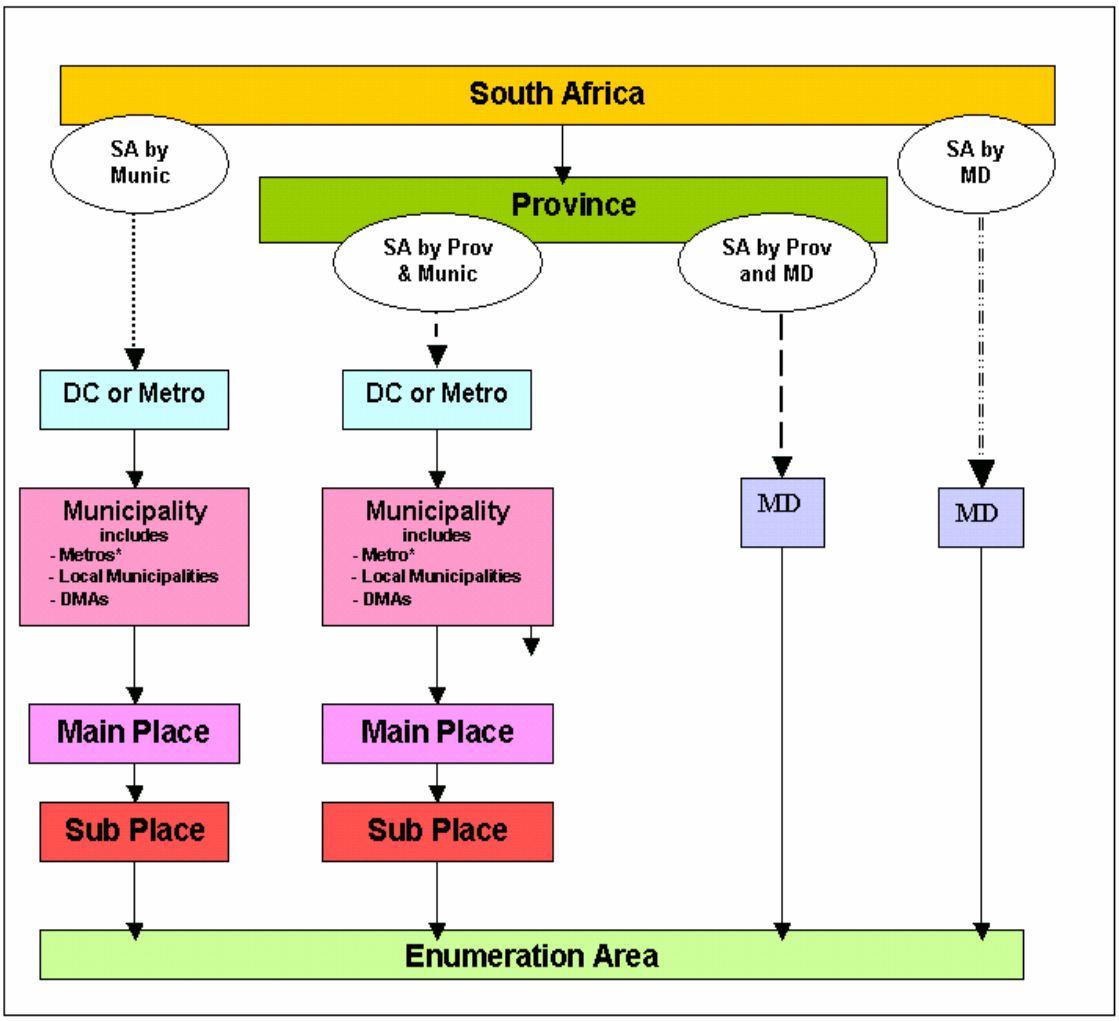
The table below lists Census and survey data for which it was not possible to obtain detailed information on content or availability. If, as is hoped, this report is regularly updated, further information about these datasets could be added, along with details of other datasets not included in this Appendix.

Data source	Year	Principal investigator	Notes
Anthropometric Survey in Primary Schools in the RSA	1994	Department of Health	Potentially available to researchers, but it was not possible to find out the required information about data availability. Contact: Prof. Demetre Labadarios, Head, Human Nutrition, University of Stellenbosch E-mail: demetre@sun.ac.za Telephone: +27 (0)21 938 9259
Census of Agriculture	2002	Stats SA	Stats SA has been contacted about the availability of this data, but it has not been possible to obtain further information.
Census of Local Government		Stats SA	
Census of Manufacturing	1996	Stats SA	
Ekhuruleni Manufacturing Survey		Corporate Strategy and Industrial Development, University of the Witwatersrand	
Evaluating Public Opinion Programme	1998-2001	HSRC	
Greater Johannesburg Metropolitan Area Manufacturing Firms Survey	1998	World Bank	Despite similarities in the method and framework between the GDMA survey and this survey, the surveys cannot be compared. One important difference is that, whereas two sets of weights have been applied in the former (so that the results are representative at the municipal and national levels), the data for the latter has been weighted so that the sample is representative at the municipal level only. In addition, four years separate the GDMA and the GJMA surveys and important policy changes occurred over the period. No specific details on how to access the data have been obtained. Contact: Matthew de Gale at Trade and Industrial Policy Strategies Email: matthew@tips.org.za
HIV/AIDS Educators Study	2005	HSRC	The aim of the study was to help the Education Labour Relations Council to understand the impact of HIV/AIDS on South African educators. The target population included all public education sector educators in grade R to 12, those in further education and training colleges, and student teachers. Specifically, this study examined how the supply of educators was affected by AIDS morbidity and mortality, general health status, TB, alcohol and drug use, and migration.
Household Livelihood	2002	Programme for Land and	The dataset is available but there is little documentation.

Survey		Agrarian Studies (PLAAS), School of Government, University of the Western Cape	Contact: Andries du Toit, Senior Researcher, PLAAS Email: presence@iafrica.com
Iodine Deficiency Disorder Survey of Primary School Learners in South Africa	1999	South African Institute for Medical Research	Potentially available to researchers, but it was not possible to find out the required information about data availability. Contact: Prof. Demetre Labadarios, Head, Human Nutrition, University of Stellenbosch E-mail: demetre@sun.ac.za Telephone: +27 (0)21 938 9259
Langeberg Survey	1999	Southern Africa Labour and Development Research Unit	Small survey conducted in the Langeberg health district of the Western Cape province of South Africa. Information was collected from a stratified sample consisting of 294 households (103 black; 126 coloured; and 65 white) by means of four questionnaires: one directed at 121 adults of 55 years and older; one at 573 younger adults between 18 and 54 years of age; and two focused on health (including anthropometric measurements) of 812 adults and 294 children. Contact: Brenda Adams, SALDRU Email: badams@commerce.uct.ac.za Send an e-mail request giving a brief description of your intended use of the data.
Manpower Survey (Occupational Survey)	1994, 1995, 1996	Stats SA	Stats SA has been contacted about the availability of this data, but it has not been possible to obtain further information.
Migration Surveys		Department of Sociology, University of Stellenbosch	The Department has a research programme on internal migration and has conducted at least two surveys. It has not been possible to find out whether these are widely available however, although it seems that one might be available from SADA and the other from the DataFirst Resource Centre.
National Election Survey, Local Government Election Study, Political Culture Study	1994, 1995, 1997	IDASA	The precedents to Afrobarometer in South Africa. Data are available from SADA (http://www.nrf.ac.za/sada). A fourth survey was also conducted by IDASA in 1998.
National Enterprise Survey	1999	HSRC	
National Skills Survey	2003	HSRC	
Sexual Violence and HIV/AIDS National Youth Survey	2001-2004	Community Information and Epidemiological Technologies (CIET) Africa	It is thought that the data are not available to external researchers, but this should be checked. Contact: Neil Anderson, Executive Director, CIET Canada Email: neil@ciet.org See http://www.ciet.org/en/documents/projects/200621015051.asp for the report.
South African Vitamin A Consultative Group (SAVACG) Survey	1998	SAVACG	Potentially available to researchers, but it was not possible to find out the required information about data availability. Contact: Prof. Demetre Labadarios, Head, Human Nutrition, University of Stellenbosch E-mail: demetre@sun.ac.za Telephone: +27 (0)21 938 9259
Survey of Manufacturing	2001	Stats SA	

A1.5 GIS data

The Geography Division of Stats SA provides digitised boundaries for use with Census and other data. The whole country was delimited into EAs according to municipality and province by Stats SA for Census 2001. The figure below shows four pathways or hierarchical structures of geographical entities built for Census 2001 spatial products dissemination. The first and second reflect the new municipal structure, while the third and fourth represent the old Magisterial District geography. The Magisterial District routes are alternatives which preserve the old geography structure of Census '96.



Source: Stats SA Census 2001 Metadata: Geography hierarchy and attributes. Report no. 03-02-25.

For information:
Stats SA Geography Division
Website: http://mapserver.statssa.gov.za/geographywebsite/
Email: geoinfo@stassa.gov.za
Telephone: +27(0)12 310 8470

In addition, the GIS Centre at the HSRC is engaged in the development of spatial databases. These databases are developed to meet the needs of HSRC researchers or external stakeholders. A list of the spatial databases available in the GIS Centre is given in the table below.

Category	Layers	Spatial units
Administration	District councils	Boundaries
	Magisterial districts	Boundaries
	Municipalities	Boundaries
	Provinces	Boundaries
	Suburbs	Boundaries
	Wards	Boundaries
Crime	Crime Statistics (1994-2000)	Magisterial district, police station
	Crime victimization survey	Police station, sample point location
Development funding	Financial budgets and expenditure (1994-1999)-FFC, RDP, Provincial, DBSA, District councils	Province, district council, magisterial district, project point location
Economic	Gross geographic product (1991,1996)	Magisterial district
Education	Adult basic education and training (ABET 1999)	ABET point location
	Education regions	Education region
	Pre-primary schools	Pre-primary school point location
	Schools (SRN 1996)	School point location
	Tertiary institutions (technical, private, technikons, universities, distace)	Point location
	Graduate statistics	Postal code
	Graduate business survey	Sample point location
Elections	Election results	Ward
Environmental	Dams	Polygon
	Geology	Polygon
	Land use	Polygon
	Rivers	Line
	Roads	Line
	Towns	Polygon, point location
	Vegetation	Polygon
Geographical placenames	HSRC placenames	Point location
	Stats SA placenames	Polygon, point location
	Surveyor General placenames	Point location
	US Gazetteer	Point location
Health	Clinics	Point location
	Epidemiology (i.e. measles, TB, HIV/Aids)	Provincial
	Health districts	Polygon
	Hospitals	Point location
Home affairs	Home affairs regions	Polygon
Income	Disposable income (1996)	Enumerator area
Justice	Magisterial district (1996)	Polygon
	Magistrate courts (1997)	Point location
Land	Land cover	Polygon
Lifestyle segmentation	Lifestyle segmentation (1991)	Small market area
	Lifestyle segmentation (1996)	Small market area
Living standard measures (LSM)	LSM's (1999)	Small market area, enumerator area
Map grids	1:500 000	Polygon grid
	1:250 000	Polygon grid
	1:50 000	Polygon grid
Migration	Migration numbers between districts (1996)	Magisterial district
Multi-purpose community centres (MPCC)	Geographic location	Point location
NGO's/CBO's	Prodder directory (1998)	Point location
Opinion surveys	SA EPOP (1999)	Sample point location
	SA EPOP (2000)	Sample point location, region, lifestyle, provincial
Police	Police station areas (1998)	Polygon

	Police district boundaries (1998)	Polygon
	Police station locations (1998)	Point location
	Crime statistics (1994-2000)	Magisterial district, police station
Political violence	Political violence (1990-1995)	Magisterial district
Postal	Post offices	Point location
	Postal codes (1998-2000)	Point location
Poverty	Poverty gap (1991)	Magisterial district
	Poverty gap (1996)	Small market area, magisterial district
	Gini coefficient (1996)	Municipal
Prisons	Prisons (1997)	Point location
Railway lines	1:500 000 scale railway lines	Line
Rivers	1:500 000 scale	Line
Roads	Surveyor General (1991)	Line
Rural/urban densification	Small market area	Polygon
Sea	Coastline	Line
Services and infrastructure	Data for 1991 - includes 5 social and 15 services variables	Magisterial district
	Data for 1996 - includes 5 social and 15 services variables	Magisterial district
Small market areas	1996 SMA's - 150 variables and 80,000 spatial features modified from the 1996 census	Small market area
Socio-economic	Census - 150 census variables(1991)	Enumerator area, suburb, local authority, census district, province
	Census - 14 community profile tables (1996)	Enumerator area, suburb, municipality, magisterial district, province
Streets	Computamaps streets (1998)	Streets
Substance abuse	Alcohol and drug abuse survey (1991)	Magisterial district, sample point location
	Alcohol and drug abuse surveys in prisons (1996)	Sample point location
	Arrestee drug abuse monitoring (SA-ADAM 2000)	Police station, sample point location
Telecommunications	Telephone dialing codes (2000)	Point location
Towns	Census local authorities (1991)	Point location
	Built-up areas	Point location
	Towns	Point location

For information:
Craig Schwabe, Director, GIS Centre
Email: caschwabe@hsrc.ac.za
Telephone: +27 (0)12 302 2507

Also please see the African Geo Information Researcher Network (http://www.agirn.org), a geo-information research portal that provides the research community with a mechanism to publish and access high quality work, to share in geo-information knowledge and to engage in discussion through a widely accessible web based medium.

Appendix 2 Contacts

A2.1 List of persons interviewed (face to face, by telephone or through email correspondence)

Name	Position	Organisation
Dr Kobus Herbst	Deputy Director	Africa Centre for Health and Population Studies
Prof. Robert Mattes	Deputy Director	Afrobarometer
Prof. Stephen Tollman	Director	Agincourt Health and Population Unit
Dr Kathleen Kahn	Senior Researcher	Agincourt Health and Population Unit
Mr Patrick Burton	Research Director	Centre for Justice and Crime Prevention
Mr Ian Macun	Director	Community Agency for Social Enquiry
Ms Debbie Budlender		Community Agency for Social Enquiry
Ms Lynn Woolfrey	Manager, DataFirst Resource Centre	DataFirst
Mr Rufus Poliah	[Examination Division]	Department of Education
Dr Lusanda Mahlasela		Department of Health
Mr Vusi Mkhise/Mr Thomas Sigama		Department of Home Affairs
Mr Kritzinger		Department of Home Affairs
Mr Victor Rajkumar	Deputy Director, GIS and Analysis, Directorate: Information Management	Department of Housing
Mr Francois de Villiers		Department of Labour
Mr Chris Mill	Manager, Declarations Database, Unemployment Insurance Fund	Department of Labour
Dr Shane Norris	Senior Researcher	Department of Paediatrics, University of the Witwatersrand
Mr Krzysztof Wojciechowicz		Department of Trade and Industry
Ms Laverne Dimitrov		Department of Transport
Ms Wendy Watson	Chief Director, Land Transport Regulation	Department of Transport
Mr Hilton Visagie		Education Management Information Systems
Mr Siza Shongwe	Director	Education Management Information Systems
Prof. Demetre Labadarios	Head	Human Nutrition, University of Stellenbosch
Dr Andrew Paterson	Research Director, Education Science and Skills Development	Human Sciences Research Council
Mr Ben Roberts	Research Specialist, Urban, Rural and Economic Development	Human Sciences Research Council
Ms Marie-Louise van Wyk	Manager, ITS System Development	Human Sciences Research Council
Dr Mbithi wa Kivilu	Director, Socio-Economic Surveys Section, Knowledge Systems Group	Human Sciences Research Council

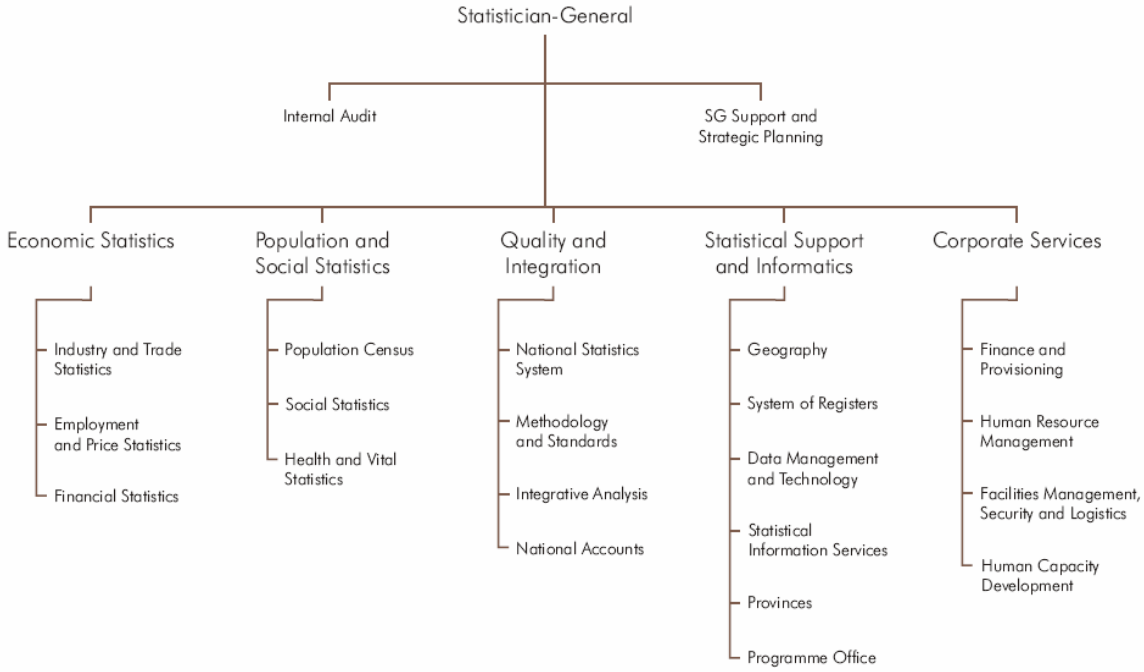
Mr Michael Cosser	Chief Research Specialist, Education, Science and Skills Development	Human Sciences Research Council
Dr Pieter Kok	Chief Research Specialist, Urban, Rural and Economic Development	Human Sciences Research Council
Ms Sharmla Rama	Research Specialist, Child, Youth, Family and Social Development	Human Sciences Research Council
Prof. Thomas Rehle	Director, Social Aspects of HIV/AIDS and Health	Human Sciences Research Council
Ms Antoniette Louw	Senior Research Fellow, Crime and Justice Programme	Institute for Security Studies
Dr Debbie Bradshaw	Director, Burden of Disease Research Unit	Medical Research Council
Dr Annatjie Moore	Director, Provincial Performance Monitoring, Intergovernmental Relations	National Treasury
Ms Julia de Bruyn		National Treasury
Mr Lawrence Matemba	[Social Sector]	Office of the Presidency
Ms Mastoera Sadan	[Social Sector]	Office of the Presidency
Mr Sibusiso Masuka	[Justice, Crime Prevention and Security Sector]	Office of the Presidency
Ms Shirley Robinson	Chief Director, Resource Management	Provincial Government Western Cape - Treasury
Dr Catherine MacPhail	Senior Researcher	Reproductive Health Research Unit
Ms Caroline Skinner		School of Development Studies, University of KwaZulu-Natal
Dr Myriam Velia		School of Development Studies, University of KwaZulu-Natal
Dr Gavin Reagon	Senior Lecturer	School of Public Health, University of the Western Cape
Ms Fiona Lister	Administrator	South African Advertising Research Foundation
Mr Paul Haupt	Chief Executive Officer	South African Advertising Research Foundation
Ms Phindiwe Tsebe	[Data Processing, Data Orders]	South African Data Archive
Asst Comm. (Dr) Chris de Kock	Head, Information Management	South African Police Service
Ms Kathy Nicolaou		South African Social Security Agency
Mr Marius Cronje	[Data Management and Information Delivery]	Statistics South Africa
Mr Heston Phillips	[Demographic Analysis]	Statistics South Africa
Mr Kevin Parry	Dissemination Officer, Marketing	Statistics South Africa
Mr Michael Manamela	[Industry and Trade Statistics]	Statistics South Africa
Ms Patricia Koka	Principal Survey Statistician, Employment Statistics	Statistics South Africa
Dr Patrick Naidoo	Executive Manager, Financial Statistics	Statistics South Africa
Mr Peter Buwembo	[Employment and Price Statistics]	Statistics South Africa
Mr Piet Alberts	[Data Integration, Statistical Information Services]	Statistics South Africa
Ms Susanna Ubomba-Jaswa	[Health and Vital Statistics]	Statistics South Africa
Ms Sally Wheeler	Senior Survey Statistician, Financial Statistics	Statistics South Africa

A2.2 Government department websites

Department	Website
Agriculture	http://www.nda.agric.za/
Arts and Culture	http://www.dac.gov.za/
Education	http://www.education.gov.za/
Environmental Affairs and Tourism	http://www.environment.gov.za/
Health	http://www.doh.gov.za/
Home Affairs	http://www.home-affairs.gov.za/
Housing	http://www.housing.gov.za/
Labour	http://www.labour.gov.za/
Land Affairs	http://land.pwv.gov.za/
Minerals and Energy	http://www.dme.gov.za/
National Treasury	http://www.treasury.gov.za/
Provincial and Local Government	http://www.dplg.gov.za/
Public Enterprises	http://www.dpe.gov.za/
Public Service and Administration	http://www.dpsa.gov.za/
Public Works	http://www.publicworks.gov.za/
Science and Technology	http://www.dst.gov.za/
SA Police Service	http://www.saps.gov.za/
SA Revenue Service	http://www.sars.gov.za/
Social Development	http://www.welfare.gov.za/
Sport and Recreation South Africa	http://www.srsa.gov.za/
Statistics South Africa	http://www.statssa.gov.za/
The Presidency	http://www.thepresidency.gov.za/
Trade and Industry	http://www.thedti.gov.za/
Transport	http://www.transport.gov.za/
Water Affairs and Forestry	http://www.dwaf.gov.za/

Appendix 3 Organisational structure: Stats SA and HSRC

A3.1 Stats SA

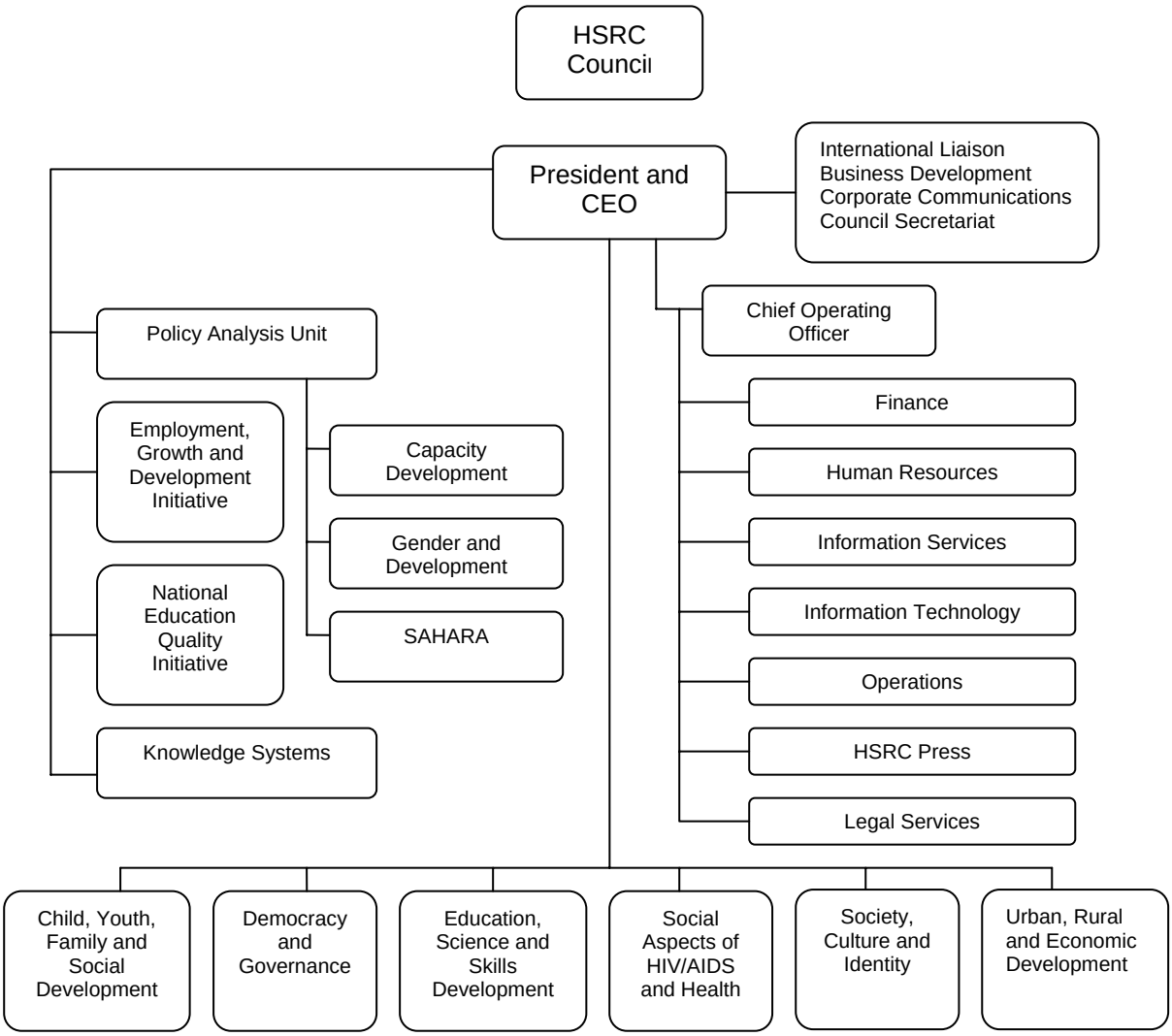


In terms of economic and social micro data, the relevant sections are Economic Statistics and Population and Social Statistics.

- **Industry and Trade Statistics** provides information on turnover and volumes in various sectors of the economy.
- **Employment and Price Statistics** provides information on employment in the formal non-agricultural sectors, and on changes in the prices of consumer and producer goods.
- **Financial Statistics** tracks public sector spending and the financial performance of private sector organisations.
- **Population Census** provides the most comprehensive, detailed small area data on the number, distribution and characteristics of individuals and households in South Africa.
- **Social Statistics** conducts detailed investigations on the state of the labour market and the living conditions of the population.
- **Health and Vital Statistics** publishes statistics on births, deaths, marriages, divorces, tourism and migration, based on administrative records.

Source: Statistics South Africa Annual Report 2006 and StatsOnline Newsletter Issue No.14/2004 (see http://www.statssa.gov.za/newsletters/StatsOnline1620Aug2004.pdf)

A3.2 HSRC



Source: Adapted from http://www.hsrc.ac.za/about/structure/index.html

Appendix 4 Sample records

A4.1 South African Data Archive

STUDY NUMBER: 0137
Labour Force Survey (LFS), March 2004
PRINCIPAL INVESTIGATOR: *Statistics South Africa (Stats SA)*
ABSTRACT: The Labour Force Survey is a twice-yearly rotating panel household survey, specifically designed to measure the dynamics of employment and unemployment in the country. It plans to measure a variety of issues related to the labour market, including unemployment rates (official and expanded), according to standard definitions of the International Labour Organisation (ILO).
KEYWORDS: Employment; Unemployment; Household surveys; Housing; Informal sector; Labour supply.
FOCUS AREA: Household Survey
DEMOGRAPHIC VARIABLES: Age, gender, level of education, marital status, migration, use of health services, economic activity, unemployment, employment and self-employment.
IMPORTANT VARIABLES: Employment, unemployment, informal sector.
UNIVERSE: Households in the nine provinces of South Africa
METHOD OF DATA COLLECTION: Survey Questionnaire
SAMPLING DESIGN: A multi-stage stratified sample was drawn using probability proportional to size principles. The sample was drawn from the master sample, which Statistics South Africa uses to draw samples for its regular household surveys. The master sample is drawn from the database of enumeration areas (EAs) established during the demarcation phase of Census 1996. As part of the master sample, small EAs consisting of fewer than 100 households are combined with adjacent EAs to form primary sampling units (PSUs) of at least 100 households, to allow for repeated sampling of dwelling units within each PSU. The sampling procedure for the master sample involves explicit stratification by province and within each province, by urban and non-urban areas. Within each stratum, the sample was allocated disproportionately. A PPS sample of PSUs was drawn in each stratum, with the measure of size being the number of households in the PSU. Altogether approximately 3 000 PSUs were selected. In each selected PSU a systematic sample of ten dwelling units was drawn, thus, resulting in approximately 30 000 dwelling units. All households in the sampled dwelling units were enumerated. The LFS is based on a rotating panel design to collect labour force statistics from households, to ultimately obtain a better picture of movements into and out of the labour market over time. A rotating panel sample involves visiting the same dwelling units on a number of occasions (in this instance, five at most). After the panel is established, a proportion of these dwelling units is replaced each round (in this instance, 20%). New dwelling units are added to the sample to replace those that are taken out. The advantage of this type of design is that it provides the basis for monitoring changes in the work situation of members of the same households over time, while retaining the larger picture of the overall employment situation in the country. It also allows for both longitudinal and cross-sectional analysis. A subset of 1 000 PSUs was drawn from the master sample for the pilot LFS of February 2000. For the subsequent rounds, the sample was increased to 3 000 PSUs for September 2000 and February 2001, in which the same 30 000 dwelling units were visited. In September 2001 a new sample of 30 000 dwelling units was drawn. In February 2002, 80% of the dwelling

units sampled in September 2001 were visited again. The remaining 20% comprised new dwelling units. The same rotation procedure has been implemented for all subsequent rounds. A new master sample will be drawn in September 2004 benchmarked to Census 2001 for the LFS for the next five years.
WEIGHTING: A two-stage weighting procedure was carried out on LFS March 2003. The first stage weighted the results to separate estimates of the population size, based on the population census of October 2001, as adjusted by a post-enumeration survey (PES). The second stage used post-stratification by province, gender, population group and five-year interval age groups
UNITS OF OBSERVATION: Households
DATE OF RESEARCH: March 2004
EXTENT OF COLLECTION: 4 data files in ASCII and machine-readable documentation.
PUBLICATIONS: Statistics South Africa. 2003. Labour force survey: round 9: March 2004. Pretoria: Statistics South Africa.
RESTRICTIONS: Only available to bona fide researchers.
NOTES: Most questions in the Labour Force Survey questionnaire are pre-coded, i.e. there is a set number of choices from which one or more must be selected. For open-ended 'write-in' questions, the description will note that post-coding occurred and explain how this was done. For most variables the coding is apparent from the questionnaire (available elsewhere in the documentation) and is not repeated in the variable description. Where the coding is not apparent, the description either provides the codes or indicates where code lists are to be found.
DATE STUDY ADDED: 07 December 2004
MS Word Codebook (uncompressed) (580 KB)

A4.2 HRD Data Warehouse

Database Information			
Name	Annual School Survey	Acronym	ASS
Year	1997 to 2003	Month	April
Format	MS Access	Size	350 MB per year
Description	1997 was the first year that the 9 provinces submitted data, collected in March/April in the Annual School Survey, on learners and educators for ordinary public and independent schools to the national Department of Education. The national Department of Education collates the provincial data of school level information into one national database for each year and cleaning and clarifying data if possible. The Annual School Survey database contains information on each school in the country, its learner profile in terms of grade, age, home language and race. It also contains numbers of secondary learners enrolled by subject in each school as well as detailed information on educators.		
Organisation which owns the data			
Name	Department of Education		
Acronym	DOE		
WebSite	**http://education.pwv.gov.za**		
Related Documents			
Survey	**Annual School Survey Form 2003**		
How to gain access			
Contact	Department of Education: Mr CJ Lombaard	Tel	012-312 5242
Email	lombaard.c@doe.gov.za	Fax	012-328 3089
Description	Requests for information must please be e-mailed to: Mr Christo Lombaard (lombaard.c@doe.gov.za) The request must specify the type of information required.		
Access Conditions	The public does not have access to the database. Selected data is available to the public on request. All enquiries should be directed to the Department of Education.		
Rights	Any data received from the Department of Education (DoE) may be used as long as the DoE is acknowledged as the source of the data, and the data is used for non-profit purposes only.		

Glossary

ABET	Adult Basic Education and Training
ACDIS	Africa Centre Demographic Information System
AHDSS	Agincourt Health and Demographic Surveillance System
AMPS	All Media Products Survey
ART	Anti-Retroviral Therapy
BMR	Bureau of Market Research
CAPS	Cape Area Panel Study
CASASP	Centre for the Analysis of South African Social Policy
CASE	Community Agency for Social Enquiry
CENPOPS	Centre for Population Studies (University of Pretoria)
CeSTII	Centre for Science, Technology and Innovation Indicators (unit within HSRC)
CIET	Community Information and Epidemiological Technologies
CJCP	Centre for Justice and Crime Prevention
CPI	Consumer Price Index
CPS	Current Population Surveys
DC	District Council
DHIS	District Health Information System
DMID	Data Management and Information Delivery (Statistics South Africa)
DoE	Department of Education
DoH	Department of Health
DoT	Department of Transport
EA	Enumeration Area
ECD	Early Childhood Development
EMIS	Education Management Information System
ESD	Enumerator Sub-Districts
ESDMF	End to End Statistical Data Management Facility
ESDS	Economic and Social Data Service
ESRC	Economic and Social Research Council
FET	Further Education and Training
FTP	File Transfer Protocol
GDMA	Greater Durban Metropolitan Area
GHS	General Household Survey
GIS	Geographic Information Systems
GJMA	Greater Johannesburg Metropolitan Area
HANIS	Home Affairs National Identification System
HE	Higher Education
HH	Households
HRD	Human Resources Development
HSRC	Human Sciences Research Council
HSS	Housing Subsidy System
IDASA	Institute for Democracy in South Africa
IES	Income and Expenditure Survey
ILO	International Labour Organisation
ISS	Institute of Security Studies
KIDS	KwaZulu-Natal Income Dynamics Study
KMP	Khayelitsha/Mitchell's Plain Survey
KS	Knowledge Systems (unit within HSRC)

LFS	Labour Force Survey
LURITS	Learner Unit Record Information Tracking System
MD	Magisterial District
NESSTAR	Networked Social Science Tools and Resources
NGO	Non-Governmental Organisation
NIMSS	National Injury Mortality Surveillance System
NLD	Needy Learner Database
NSS	National Statistics System
NVCS	National Victims of Crime Survey
OHS	October Household Survey
OVC	Orphans and Vulnerable Children
PMTCT	Prevention of Mother to Child Transmission
PSLSD	Project for Statistics on Living Standards and Development
PSU	Primary Sampling Unit
RHRU	Reproductive Health Research Unit (University of the Witwatersrand)
SAARF	South African Advertising Research Foundation
SADA	South African Data Archive
SAPS	South African Police Service
SAL	Small Area Layer
SALDRU	South African Labour and Development Research Unit (University of Cape Town)
SAMS	School Administration Management System
SARB	South African Reserve Bank
SARPN	South African Regional Poverty Network
SARS	South African Revenue Services
SA-SAMS	South Africa School Administration Management System
SASAS	South African Social Attitudes Survey
SASSA	South African Social Security Agency
SAYP	Survey of Activities of Young People
SDC	Statistical Disclosure Control
SES	Socio-Economic Surveys (unit within HSRC)
SDS	School of Development Studies (University of KwaZulu-Natal)
SNE	Special Needs Education
SOCPEN	Social Pensions Database
Stats SA	Statistics South Africa
TB	Tuberculosis
TBVC	Transkei, Bophuthatswana, Venda and Ciskei states
TLC	Transitional Local Council
TRC	Transitional Rural Council
UCT	University of Cape Town
UIF	Unemployment Insurance Fund
UKZN	University of KwaZulu-Natal
UNICRI	United Nations Interregional Crime and Justice Research Institute

References

Akoojee, S. (2005) *Private Further Education and Training in South Africa: the changing landscape.* Cape Town: HSRC Press.

Ardington, C., Lam, D., Leibbrandt, M. and Welch, M. (2005) 'The sensitivity of estimates of post-apartheid changes in South African poverty and inequality to key data imputations'. Working Paper No. 106. Centre for Social Science Research, University of Cape Town.

Barnes, H. Gutierrez-Romero, R. and Noble, M. (2006) 'Multiple imputation of missing data in the 2001 South African Census'. CASASP Working Paper 4. Department of Social Policy and Social Work, University of Oxford.

Bhorat, H., Poswell, L. and Naidoo, P. (2004) *Dimensions of Poverty in Post-Apartheid South Africa 1996-2001*. Cape Town: Development Policy Research Unit, University of Cape Town, South Africa.

Burton, P. (2006) *Snapshot Results of the 2005 National Youth Victimisation Study*. Research Bulletin No. 1. Cape Town: Centre for Justice and Crime Prevention.

Burton, P., du Plessis, A., Leggett, T., Louw, A., Mistry, D. and van Vuuren, H. (2004) *National Victims of Crime Survey South Africa 2003*. ISS Mongraph No 101.

Chobokoane, N. and Budlender, D. (2002) 'Activities over time: Further analysis of the Time Use Survey'. Occasional paper 2002/04. Pretoria: Statistics South Africa.

Cosser, M.C. and Du Toit, J.L. (2002) *From School to Higher Education? Factors affecting the choices of grade 12 learners.* (The Student Choice Behaviour Project, Phase 1). Cape Town: Human Sciences Research Council.

Cronje, M. and Budlender, D. (2004) 'Comparing Census 1996 and Census 2001: An operational perspective'. *Southern African Journal of Demography*. 9(1): 67-89.

Desai, T. and Cowell, F. (2006) *ESRC Review of Data Resources and Needs*. Economic and Social Research Council.

Devey, R. Valodia, I. and Velia, M. (2005) *Constraints to Growth and Employment: Evidence from the Greater Durban Metropolitan Area*. Research Report, 64.

Du Toit, J.L. (2004) *Independent Schooling in Post-Apartheid South Africa: A quantitative overview.* (Research Programme on Human Resources Development. Occasional paper, 1). Cape Town: HSRC Publishers.

Hirschowitz, R. (1996) *Living in South Africa: Selected findings of the 1995 October Household Survey*. Pretoria: Central Statistical Service.

Hirschowitz, R., Buwembo, P., Serwadda-Luwaga, J. and Hasholm, H. (1998) *Victims of Crime Survey*. Pretoria: Statistics South Africa.

Hirschowitz, R., Sekwati, W.M. and Budlender, D. (2001) *South Africa in Transition: Selected findings from the October Household Survey of 1999 and changes that have occurred between 1995 and 1999*. Pretoria: Statistics South Africa.

Kok, P., Gelderblom, D., Oucho, J.O. and van Zyl, J. (eds) (2006) *Migration in South and Southern Africa: Dynamics and Determinants*. Cape Town: HSRC Press.

Leoschut, L. and Burton, P. (2006) *How Rich the Rewards? Results of the 2005 National Youth Victimisation Study*. Cape Town: Centre for Justice and Crime Prevention.

Meth, C. (2006a) 'What was the poverty headcount in 2004 and how does it compare with recent estimates by van der Berg et al?' SALDRU Working Paper No. 06/01. Southern African Labour and Development Research Unit, University of Cape Town.

Meth, C. (2006b) 'Income poverty in 2004: A second encounter with the recent van der Berg et al figures'. Working Paper No. 47. School of Development Studies, University of KwaZulu-Natal.

Noble, M., Babita, M., Barnes, H., Dibben, C., Magasela, W., Noble, S., Ntshongwana, P., Phillips, H., Rama, S., Roberts, B., Wright, G. and Zungu, S. (2006) *The Provincial Indices of Multiple Deprivation for South Africa 2001*. Department of Social Policy and Social Work, University of Oxford.

Pillay, U., Roberts, B. and Rule, S. (eds) (2006) *South African Social Attitudes: Changing times, diverse voices*. Cape Town: HSRC Press.

Shisana, O., Rehle, T., Simbayi, L.C., Parker, W., Zuma, K., Bhana, A., Connoly, C., Jooste, S. and Piillay, V. (2005) *South African National HIV Prevalence, HIV Incidence, Behaviour and Communication Survey, 2005.* Cape Town: HSRC Press.

Simkins, C. (2003) *A Critical Assessment of the 1995 and 2000 Income and Expenditure Surveys as Sources of Information on Incomes.* University of the Witwatersrand. Unpublished.

Smith, G.A.N. and Noble, M., (2000) 'Developing the use of administrative data to study poverty'. In J. Bradshaw and R. Sainsbury (Eds) *Research Poverty*. Ashgate. pp.77-97.

Smith, G.A.N., Noble, M., Antilla, C., Gill, L., Zaidi, A., Wright, G., Dibben, C. and Barnes, H. (2004) *The Value of Linked Administrative Records for Longitudinal Analysis*. Report to the ESRC National Longitudinal Strategy Committee. Social Disadvantage Research Centre, University of Oxford.

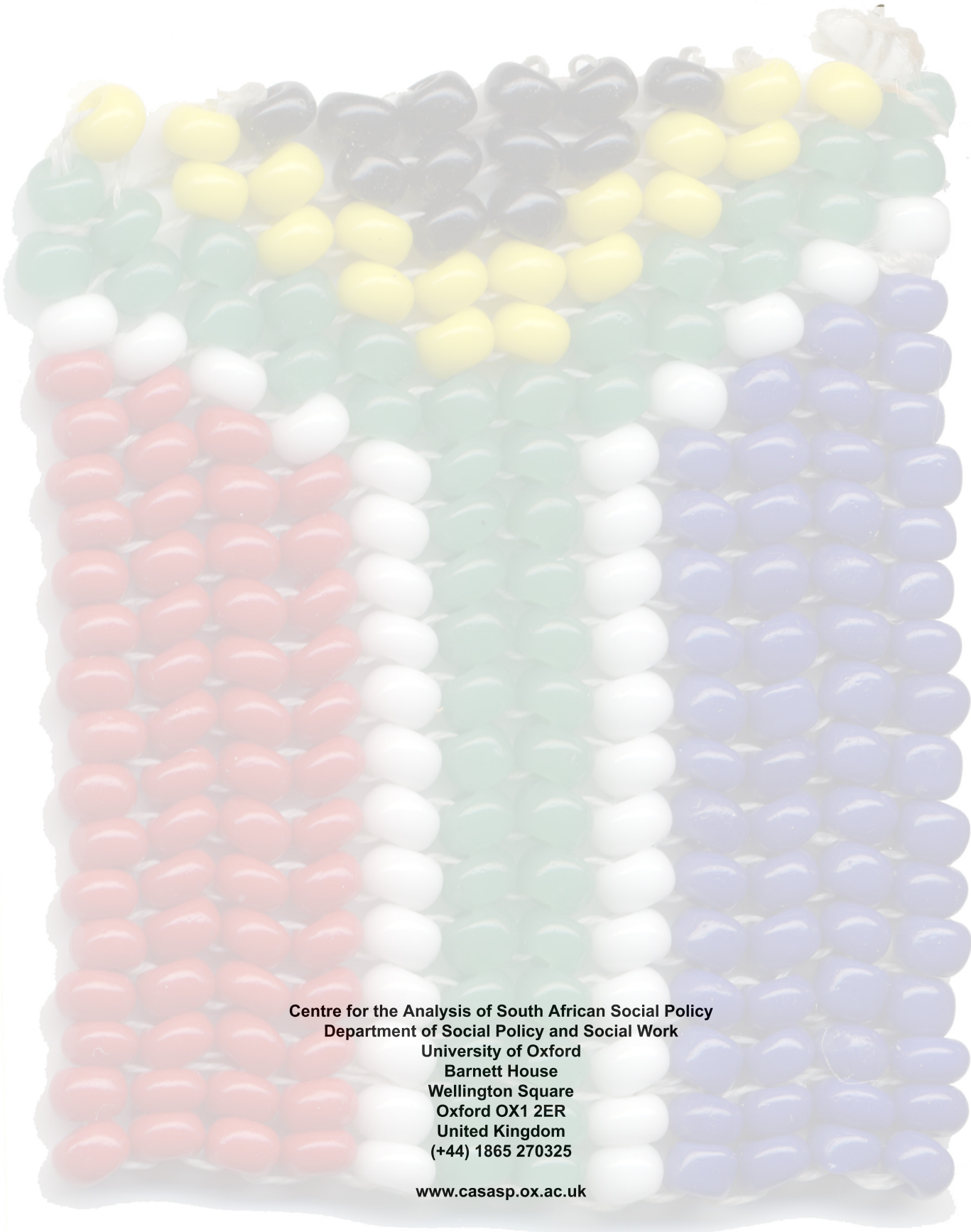
Statistics South Africa (2001a) *Survey of Activities of Young People in South Africa 1999: Country report on children's work-related activities*. Pretoria: Statistics South Africa.

Statistics South Africa (2001b) *Survey of Activities of Young People in South Africa 1999: Tables on children's work-related activities*. Pretoria: Statistics South Africa.

Statistics South Africa (2001c) *A Survey of Time Use: How South African women and men spend their time*. Pretoria: Statistics South Africa.

Statistics South Africa (2002) *Pricing Policy*. Pretoria: Statistics South Africa.

Statistics South Africa (2005) *Achieving a Better Life for All: Progress between Census '96 and Census 2001*. Pretoria: Statistics South Africa.



Centre for the Analysis of South African Social Policy
Department of Social Policy and Social Work
University of Oxford
Barnett House
Wellington Square
Oxford OX1 2ER
United Kingdom
(+44) 1865 270325

www.casasp.ox.ac.uk