E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

Data Discovery

A rough guide to microdata in Brazil, China, India and South Africa

Contents

This publication is based upon four reports commissioned by the UK Economic and Social Research Council, to inform an initiative to develop and promote knowledge about the wide variety of data available in different countries which can facilitate research in the social sciences and related disciplines.

The commissioning process and the development of this initiative has been overseen by a small panel of experts:

Mr Keith Cole, Director of the Economic and Social Data Service (International), MIMAS, University of Manchester
Dr Louise Corti, Associate Director, UK Data Archive, University of Essex
Professor Peter Elias, ESRC Strategic Advisor for Data Resources, University of Warwick
Professor Phil Rees, Department of Geography, University of Leeds

This summary of the four reports has been prepared by Iain Stewart, IIASA, stewart@iiaasa.ac.at

For further details please contact the report authors directly or:

Professor Peter Elias, ESRC Strategic Advisor for Data Resources, peter.elias@warwick.ac.uk
Mike Bright, ESRC Associate Director for Research Resources, michael.bright@esrc.ac.uk

Introduction:
Research without borders

It's 2015 and a team of Indian social scientists are advising their government on how best to adapt India for climate change. Thanks to open access to the data resources for a wide range of countries, the social scientists can analyse how factors outside India's borders will affect its climate. They can also learn how other countries are tackling climate change and apply successful policies to India. As a result, effective and targeted climate policies can be implemented and more accurate predictions made about the social and economic impacts of climate change on India.

Yet today in 2007, social scientists all over the world struggle to easily access microdata about, say, health or migration in every continent. The worldwide sharing of the invaluable research resource of data, that is so common in the physical sciences, is not widespread in the social sciences.

Many of the problems facing countries are not unique – poverty, AIDS, immigration, to mention a few. Much can be learned from comparing how other countries have tackled their challenges. And in our increasingly interconnected world, some of the most threatening problems are global. Problems such as climate change, global terrorism and an influenza pandemic. All affect people, society and the economy and necessitate first-class social science. All ignore borders between countries, and so must the researchers if they are to truly understand the issues at the global level.

Microdata and macrodata
Microdata are data about individual objects such as persons, companies, events and transactions. Macrodata are data aggregated to a country or regional level.

If part of making borders invisible to social scientists is providing easy access to international microdata, why can't so many researchers access the data they need? Sometimes the good quality social science data simply don't exist. Sometimes researchers don't know what datasets are available. And sometimes legal, cultural, financial and technical factors make valuable data hard to access.

Making datasets visible

Within countries there are often vast resources of data – surveys conducted by national governments, funded by national research organisations or performed by individual academic entrepreneurs. Yet, many especially foreign researchers don't know such data are available let alone who controls the datasets and how they can be accessed. Nor do researchers know how useful these data really are.

In order to increase social scientists' knowledge about data available in other countries, the Economic and Social Research Council (ESRC) funded four groups of social scientists to explore the scope of microdata in Brazil, China, India and South Africa. This publication summarises the data explorers' detailed reports. Through easy-to-read tables of major social and economic datasets, this booklet is a rough guide for researchers to the accessible data in the four countries. Call-outs from tables give readers an indication of the wealth of information available in the full reports. While explorers made every effort to cover the major microdata sets accessible in each country, there may be some datasets not listed in this summary or in the full reports.

Making datasets accessible

The rapidly developing economies of Brazil, South Africa, China and India, with the latter two predicted to become the world's dominant economies by 2050, make the four countries excellent case studies for this report. The data explorations also highlighted the access issues that can face international researchers wishing to use data from other countries (see conclusion). Social science data deal with people and often their well-being. National governments collect much of this data but are understandably sensitive about releasing some data widely because of security and privacy issues.

Global research for global problems: Pandemic Influenza
The lethal influenza pandemic immediately after the First World War killed more people than those who died fighting. This was not unique: a pandemic in 1830-32 was as deadly, relatively speaking, as the one in 1918-20. Milder pandemics occurred in 1957-58 and 1968-69. Influenza pandemics are an evolutionary certainty, and they recognise no borders.

When one hits, what will happen to our economies? How will societies react? How should governments respond? These are all questions where scenarios can be posed and research undertaken on 'case studies' of lesser outbreaks of highly contagious diseases (eg AIDS & HIV) to see what lessons can be learned. But only with reliable microdata to analyse, can social scientists most effectively help countries become less vulnerable to a pandemic.

However, the benefits from sharing microdata with bona fide researchers outweigh the concerns. Countless lives and societies have been improved because of social science research using microdata. In our increasingly complex societies, access to microdata allows researchers to pose and analyse complex questions that cannot be accurately answered with aggregate statistics. With microdata, social scientists can understand down to the individual level the implications of, say, government migration policy. Access to microdata also means researchers can replicate, and so validate, others work. It all means that more quality research is produced from statistical collections, getting the best value out of the data.

But the difficulties of data sharing persist. The United Kingdom makes much of the government-collected microdata available for research. But a recent report by the UK Academy of Medical Sciences[1] argues there is still much to be done.

At an international level, the barriers are even more daunting for individual social scientists. Fortunately, international organisations have been working with countries to overcome many obstacles to build major international datasets. For example, the International Household Survey Network (IHSN www.surveynetwork.org) was set up in 2004 to better coordinate and manage socioeconomic data collection and analysis, and to mobilise support for more efficient and effective approaches to conducting surveys in developing countries. Today, IHSN holds over 2,600 surveys.

The International Data Forum

International cooperation to promote data sharing works. This publication is launched at the foundation conference of the International Data Forum in Beijing in June 2007. The conference takes stock of the current data resources that facilitate national and cross-national research on issues of global importance. It also identifies the barriers towards greater use and sharing of these resources. By facilitating discussion and generating ideas from the international community, we hope the conference will kick start the coordination of efforts by national research funding agencies and statistical authorities to make data more widely available for research purposes.

[1] The Academy of Medical Sciences (2006) Personal data for pubic good: using health information in medical research. Available at: www.acmedsci.ac.uk/images/publication/Personal.pdf

Brazilian data

Brazil is the largest, most populous country in Latin America, the fifth largest in the world as regards area and population, and the ninth biggest economy based on GDP. Rich in commodities that are coveted by the rising economies of Asia, from soya to iron ore, Brazil is predicted to dominate the world economy by 2050 along with China, India and Russia.

There is great diversity in wealth and welfare among the regions, as well as widespread poverty and crime. Income inequality, like many developing countries, is large, but in Brazil it is beginning to shrink. As Brazil faces such challenges, social scientists can learn much and contribute much to its progress.

Easily accessible datasets

Brazil has produced national statistics regularly since its first population census in 1872 and today has a wealth of high-quality social, political and economic time-series data.

Health

Two big 'systems' produce and disseminate most of Brazil's health data: the National Statistics System (SEN), coordinated by the Brazilian Institute of Geography and Statistics (IBGE); and the Unified Health System (SUS), founded in 1990, which incorporates a host of public health providers belonging to federal, state and local governments.

Though both work to similar standards, IBGE is attempting to overcome fragmentation and redundancy and assure temporal and geographic compatibility, within and among SEN datasets, particularly across different health sectors. Greater integration is required.
See table A1: Health data.

Education

The two main sources of education data in Brazil are: the regular population and labour market surveys/censuses, collected by IBGE; and the quality of education surveys (SAEB and ENAD) and censuses of basic and higher educational institutions, collected and organised by the Education Ministry's Institute of Research on Education (INEP).

The Basic School Census and SAEB data sources can be used to estimate contextual and individual level effects on student test scores and evaluate the quality of basic education. Access to data from the Higher Education Census and, since 2004, the National Exam of Student Development (ENADE), is for academic purposes only.
See table A2: Basic and higher education data.

Economic data

The main source of economic data is IGBE which, in the mid-1990s replaced its quinquennial economic censuses with stratified annual sample surveys of firms with 20+ employees (30+ in manufacturing). Monthly economic surveys are available online for basic computations, but access to sensitive microdata is restricted.

In the 1990s continuous updating began of the Central Registry of Firms (CEMPRE), built from the Annual Relation of Social Information (RAIS), an administrative register dataset compiled by the Labour Ministry (MTE), from which IBGE draws all its economic sample surveys. Variables can be tracked to the municipality level, making RAIS the most important source of information on formal labour market dynamics in Brazil.
See table A3: Survey on manufacturing, trade, services and the regions.

Labour market

The labour market is fairly well measured by different sources and methodologies and can be drawn from several sources. In 2002 the methodology and data-gathering process of the Monthly Employment Survey (PME) underwent a profound revision to bring it to international standards. This resulted in broader, higher-quality data, with easier accessibility. However, a national PME that portrays the current (formal and informal) employment move from greater metropolitan regions to smaller cities would greatly enhance Brazil's analytical capacity.
See table A4: Labour market data.

Housing conditions

To address Brazil's housing deficit of some 7 million homes, there has been ongoing research by many institutions countrywide. The small datasets generated make it difficult to link research results, policymaking and effective policy actions.

The main source of information is IBGE with three datasets. The National Household Sample Survey (PNAD) covers in-depth dimensions of the demographic dynamics that census data cannot address. The third source concerns the housing deficit in the municipalities, particularly the slums or *favelas*, and can be disaggregated down to intra-urban sub-areas and family income levels. Local authorities and energy providers also provide housing data.
See table A5: Housing data.

Transport

Comprehensive statistics about transport in Brazil are lacking. Data usually come from federal, state, and municipal agencies and associations of transport-related companies. Some reliable data can be obtained through the Annual Survey of Services, produced by IBGE, but the Survey is not specific to transport.

The Yearbook of the National Agency of Terrestrial Transport (www.antt.gov.br) and the Brazilian Automotive Industry Yearbook, produced annually in Portuguese and English by the Brazilian Automotive Industry Association (www.anfavea.com.br/carta.html), are good information sources.

Crime and Violence

There are two major sources of data on crime and violent deaths in Brazil: the Mortality Information System Dataset (SIM), compiled by the National Health Foundation (FUNASA) and restricted to lethal crimes; and the Unified Public Safety System (SUSP) compiled by the National Secretariat of Public Safety (SENASP). Quality, reliability and validity of crime data vary greatly, with shortcomings in coverage and under-reporting problems, particularly in the North and Northeast.

SENASP has recently upgraded its operations and is now the major source of crime data in Brazil. However, it will be several years until a comprehensive and reliable national crime database is available.
See table A6: Crime data.

Demography

IBGE's demographic census is the main source of demographic data. The information can be disaggregated by gender, race (1980, 1991 and 2000), age and other personal and family characteristics. Information on race/colour was not included in the 1970 census.

Census sample design makes it a more representative survey than PNAD, as information can be tracked down to the intra-urban census areas. The most important administrative demographic data are the statistics of the Civil Register. No microdata are available for this research, for confidential reasons.

Data sources

Many federal, state, municipal and private institutions in Brazil generate statistics. The main provider is the Brazilian Institute of Geography and Statistics (IBGE), part of the Federal Ministry of Planning, Budget and Management. The decennial demographic census is its nucleus.

Brazil has many other institutions that gather, archive or produce surveys. The most important are the Brazilian Institute of Public Opinion (IBOPE), Datafolha Institute, Vox Populi and the Centre of Public Opinion Studies (CESOP), which hosts most Datafolha and IBOPE polls, providing free access to more than 2,300 surveys from 1986 to date.

Other archiving institutions include: the Social Information Consortium (CIS) (www.cis.org.br), which hosts datasets from hundreds of sources, classified under 58 subject denominations; the João Pinheiro Foundation of the Government of Minas Gerais, responsible for the production of the UNDP Human Development Index in Brazil; and the State System of Data Analysis (SEADE) Foundation, linked to the São Paulo state government, responsible for the important Employment and Unemployment Survey (PED).

Improving data availability and access

Most data are readily available to researchers, academics and policymakers, either free or at low cost, the only drawback for international researchers being that the material is in Portuguese. This summary includes acronyms in Portuguese for ease of reference.

Basic computations even using the biggest datasets can be performed online, and CD-ROMs of most of IBGE's data can be obtained at its virtual store (www.ibge.gov.br/lojavirtual). Data can be translated into English for a fee, and arrangements for data use are generally via the relevant institution. Some sensitive microdata, relating to personal or economic issues, can only be accessed in person.

About the Scoping Study

This study was produced by a team of seven renowned Brazilian scholars from four research institutions in Rio de Janeiro, chosen for their acknowledged expertise in each research topic and experienced in research projects involving microdata manipulation: Adalberto Cardoso (Coord, IUPERJ); Alberto Najar, Miguel Murat Vasconcellos, Jacques Levin, and Silvia Rangel (FIOCRUZ); Carlos Antônio Costa Ribeiro (IUPERJ); Glaucio Ary Dillon Soares (IUPERJ); José Ricardo Ramalho (IFCS/UFRJ); Luiz Cesar Queiros Ribeiro, (IPPUR/UFRJ); and Celi Scalon (CFCH/UFRJ).

The full report Brazil, Microdata Scoping Study is available at www.esrcsocietytoday.ac.uk/idfpapers

Brazilian datasets

A1: Health data

Dataset	Website or contact details
Administrative records	
Evaluation System for the Immunization Program (API)	www.datasus.gov.br/svs
Information System on Compulsory Notification of Health Events (SINAN)	www.saude.gov.br/sinanweb
Information System on Live Births (SINASC)	www.datasus.gov.br/svs
Mortality Information System (SIM)	www.datasus.gov.br
SUS Ambulatory Care Information System (SIA/SUS)	www.datasus.gov.br
SUS Hospital Information System (SIH/SUS)	www.datasus.gov.br
National Directory of Health Care Establishments (CNES)	www.datasus.gov.br/sas
Survey data	
Health Care Survey (AMS)	www.ibge.gov.br
Consumer Expenditure Survey (POF),	www.ibge.gov.br
National Household Sample Survey (PNAD), Health Appendix	www.ibge.gov.br
Demography and Health Survey (PNDS)	www.bemfam.org.br www.saude.gov.br/dab
Household Survey on Risk Behaviour and Reported Morbidity of Non-transmissible Diseases and Adverse Events	www.inca.gov.br

Investigates smoking, diet, alcohol consumption and other activities that affect a person's health.

A2: Basic and higher education data

Dataset	Website or contact details
Higher Education Census	CD-ROM. Tabular data: www.inep.gov.br/basica/saeb/anos_anteriores.htm
School Census (CE)	CD-ROM. E-mail: ines.pestana@inep.gov.br
The National Exam of Student Development (ENADE)	CD-ROM. E-mail: linda.goulart@inep.gov.br
The National System of Basic Education Evaluations (SAEB)	CD-ROM. Tabular data: www.edudatabrasil.inep.gov.br

A3: Surveys on manufacturing, trade, services and the regions

Dataset	Website or contact details
CAGED and RAIS	www.mte.gov.br/EstudiososPesquisadores/PDET/Acesso/Conteudo/TermoResponsabilidade.asp
Annual Manufacturing Survey, Enterprises (PIA Empresa)	www.ibge.gov.br
Annual Manufacturing Survey, Products (PIA Produto)	www.ibge.gov.br
Monthly Manufacturing Survey, Physical Productivity (PIM-PF)	www.ibge.com.br/home/estatistica/indicadores/industria/pimes/default.shtm
Monthly Manufacturing Survey, Employment and Wages (PIM-ES)	www.ibge.gov.br
Technological Innovation in Manufacturing Survey (PINTEC)	www.pintec.ibge.gov.br
Annual Trade Survey (PAC)	www.ibge.gov.br/home/estatistica/economia/comercioeservico/pac/2002/notatecnica.pdf
Monthly Trade Survey (PMC)	www.ibge.gov.br/home/estatistica/indicadores/comercio/pmc/default.shtm
Annual Services Survey (PAS)	www.ibge.gov.br/english/estatistica/economia/comercioeservico/pas/pas2003/default.shtm
Annual Survey of Products and Services 2002–2003	www.ibge.gov.br/home/estatistica/indicadores/comercio/pmc/default.shtm
Paulista Economic Activity Survey (PAEP) (SEADE)	www.seade.gov.br/produtos/paep/index.php?opt=apr

Survey of investment, R&D, development of products in 11,000 production units.

A4: Labour market data

Dataset	Website or contact details
Administrative microdata	
Annual Relation of Social Information (RAIS)	www.mte.gov.br/EstudiososPesquisadores/PDET/Acesso/Conteudo/TermoResponsabilidade.asp
General File of Employment and Unemployment (CAGED)	www.mte.gov.br/EstudiososPesquisadores/PDET/Acesso/Conteudo/TermoResponsabilidade.asp
Survey data	
Employment and Unemployment Survey (PED)	www.scielo.br/scielo.php?pid=S0102-88392003000300013&script=sci_arttext&tlng=en
The Monthly Employment Survey (PME)	www.ibge.gov.br
The National Household Survey (PNAD)	PNAD 2001–2005 at: www.sidra.ibge.gov.br/bda/pesquisas/pnad/default.asp

National survey of 80,000 households from living conditions to education to employment.

A5: Housing data

Dataset	Website or contact details
Survey data	
Decennial Census	www.ibge.gov.br
Consumer Expenditure Survey (POF)	www.ibge.gov.br
The National Household Survey (PNAD)	www.ibge.gov.br
Secondary data based on IBGE surveys	
Index of Basic Habitation Services (ICH)	www.ippur.ufrj.br/observatorio
The Atlas of Human Development in Brazil	www.fjp.gov.br

Atlas generates maps and figures for 125 social and economic pointers for over 5000 Brazilian cities.

A6: Crime data

Dataset	Website or contact details
Administrative records	
Mortality Information System (SIM)	www.datasus.gov.br
Unified Public Safety System (SP)	www.justica.gov.br
Survey data	
Center for Criminal and Public Safety Studies (CRISP, at Minais Gerais Federal University (UFMG)	www.crisp.ufmg.br
National Household Sample Survey (PNAD), supplement on Political Participation and Victimization	www.ibge.gov.br
United Nations Latin American Institute for Crime Prevention and Delinquent Treatment (ILANUD), Victimization Surveys	www.ilanud.org.br www.ilanud.org.br/?cat_id=35
Selected surveys at Social Information Consortium (CIS)	
Attitudes, Cultural Norms and Values Concerning Violence in 10 Brazilian Capital Cities, 1999	www.cis.org.br
Rebellions in the São Paulo Prison System, 1981–1998	www.cis.org.br
Violence against Women in Rio de Janeiro	www.cis.org.br
Violent Crime in Minas Gerais, 1986-1997	www.cis.org.br
Youth Vulnerability Index (São Paulo and districts)	www.cis.org.br

Chinese data

Over the last decade China has emerged as a world power. Since 2000, China's contribution to global GDP growth (in purchasing-power-parity terms) has been bigger than that of the United States, and more than half as big again as the combined contribution of India, Brazil and Russia, the three next-largest emerging economies.

China's low-priced manufacturers give Western consumers more buying power. Its entry to the World Trade Organisation in 2001 has speeded up the opening of the world's biggest market. Whereas a few years ago it might not have mattered much to the West if China's growth had faltered, today it would be a very different story. And it is not just economists that can learn much from and contribute much to China.

Accessible datasets

On the whole, microdata collected in government surveys and for administrative purposes are hard to access in China. Even data collected by academics are generally not easily accessible. However, there are exceptions. Surveys with international funding often make the data accessible because the international funder stipulates it. Foreign researchers with good relationships with Chinese researchers who in turn have good relationships with the data collectors can gain access to some microdata.

Demography

The China Population and Development Research Center (www.cpirc.org.cn/en/eindex.htm) provides datasets on areas ranging from migration to disabled people to children's situation. However, the surveys are dated, for example, the 1% Population Sample Survey is available for 1982 and 1990 but not for 2005.

The Chinese Academy of Social Sciences (CASS) Institute of Population and Labour Economics (http://iple.cass.cn/e/e.htm) holds population-related datasets. Researchers need to negotiate access to the data with CASS.
See table B1: Population datasets held by the CASS Institute of Polulation and Labour Economics.

Labour

The CASS Institute of Economics (http://ie.cass.cn/en/about/index.htm) in collaboration with the National Bureau of Statistics conducted a nationally representative survey of income inequality in 1988, 1995 and 2002. Labour economists, both Chinese and international, have used the data to study China's income distribution, changing inequality and labour market. The 1988 Chinese Household Income Project Survey is publicly available and the later surveys are available through Luo Chuliang (luocl2002@163.com).

The CASS Institute of Population and Labour Economics also conducted the China Urban Labour Survey in 2001 and 2002 to help understand urban poverty.

Health

Two longitudinal surveys about health in China offer overseas and Chinese researchers high-quality and easily accessible microdata.
See table B2: Easily accessible microdata on health.

Education

Researchers from a range of academic institutions and other organisations based both inside and outside China hold education datasets. Foreign researchers can either easily access the datasets or will need to negotiate a special agreement.
See table B3: Major education datasets collected by organisations outside the government.

Rural economy

International researchers have used the Fixed Site Rural Survey from the Ministry of Agriculture through collaborative research projects with a Chinese partner. The annual survey has followed the social and economic conditions of 22,000 farm households in 320 villages since 1984. The collaborative research projects have often funded additional questions in the survey.

Social survey

The Chinese General Social Survey is an annual survey based on the European Social Survey model which asks the same questions in different countries to help researchers compare countries. The Chinese survey started in 2003 and collects demographic, economic, social and attitudinal information on 10,000 individuals. The data from 2003 should be made publicly available in 2007. For more information contact Li Lulu from the Institute of Sociology at People's University on chinagss2003@yahoo.com.

Restricted datasets

Most large-scale nationally representative data sets are collected through China's National Bureau of Statistics and aggregate statistics are published through annual statistical yearbooks such as the China Statistical Yearbook.

However, accessing the data at a micro level is difficult especially for foreign researchers who would need to negotiate on a case-by-case basis with the relevant government body. Sensitivities about confidentiality and national security, along with fear of criticism, mean most government departments are reluctant to grant access. The following table gives some examples of these datasets which could provide valuable information to researchers.
See table B4: Major demographic, labour and health surveys collected by the Chinese government.

Data sources

National Bureau of Statistics

The National Bureau of Statistics (NBS) has overall responsibility for national statistics in China and surveys population, employment, rural and urban households, industry, transport and fixed asset investment. All social science surveys which involve foreign researchers or funders, or where foreigners will have access to the data, should obtain approval from NBS (www.stats.gov.cn/english).

International databanks

The following three organisations provide researchers with access to microdata, sometimes for a fee.

The China Data Center (http://chinadatacenter.org/newcdc/) at the University of Michigan holds China population data series, China economic census data series, and China GIS data series. It has a large range of macrodata. With Beijing University, it launched the China Survey Data Network in 2006 to bring together small-scale survey data. Researchers who donate data to the network receive complete access to other data on the network in return.

The University Service Centre for Chinese Studies at the Chinese University of Hong Kong (www.usc.cuhk.edu.hk/uscen.asp) contains a great variety of materials on contemporary China including datasets.
See table B5: Datasets held by the Chinese University of Hong Kong.

The China Archive at Texas A & M University (http://chinaarchive.tamu.edu/portal/site/chinaarchive) offers the following datasets: 1990 1% National Population Sample Survey, 1993 Survey on Social Change and Social Mobility, a 1987 survey of children in Changchun, and a 1988 survey of political participation in Beijing.

Improving data availability and access

Various legal restrictions make it hard for researchers, both Chinese and foreign, to access microdata. These include rules on data protection and protection of state secrets. Add to this a cultural attitude that places little value on sharing data, and not surprisingly it requires considerable effort and expense (eg translation) to access data. Even international organisations such as the World Bank need to negotiate access on a case-by-case basis. While the overall environment is difficult, researchers can obtain some access by building relationships and partnerships with research organisations in China.

There are also promising developments. For example, the Institute of Social Development and Public Policy at Beijing Normal University is setting up a Social Policy Analysis Information Center. Its major resources already include over 15 datasets ranging from surveys of orphans to HIV/AIDS families and children. International researchers will be able to access some of the data online.

About the scoping study

This summary is based on the work by Sarah Cook from the University of Sussex's Institute of Development Studies and James Keeley from the International Institute for Environment and Development. The data explorers conducted a comprehensive survey of sources of statistical data across a range of social and economic sectors. They interviewed 59 individuals and organisations including Chinese government officials, academics, data collectors and users.

The full report Micro-data scoping study – China is available at www.esrcsocietytoday.ac.uk/idfpapers

Chinese datasets

B1: Population datasets held by the CASS Institute of Population and Labour Economics

Dataset	Year
Family Trends Social Survey	2002, 2004
Sample Survey of Elderly People over 60	1987
Sampling Survey on Migration in 74 Towns and Cities	1986
Survey of Changing Marriage Practices in Rural China	2005
Survey of Family Economy and Reproductive Situation in 10 provinces	1992
Survey on the Impacts of Migration on Rural Women	2005

Survey of 100,000 people, living in 74 urban areas, from mega cities to townships, about migration.

B2: Easily accessible microdata on health

Dataset	Description	Year	Principle investigator
China Elderly Health and Longevity Survey	Survey that follows the health status and healthcare of over 9,000 elderly people across China	1998, 2000, 2002, 2004	The Center for Healthy Ageing and Family Studies, Peking University (www.pku.edu.cn/academic/ageing)
China Health and Nutrition Survey	Surveys 16,000 individuals in 4,400 households about health and nutrition, and how these are influenced by changes to socio-economic circumstances of household and community	1989,1993, 1997, 2000, 2004	University of North Carolina Population Center (www.cpc.unc.edu/china)

B3: Major education datasets collected by organisations outside the government

Dataset	Description	Year	Principle investigator
Compulsory Education Period Student Family Education Expenditure Survey	Survey of 17,400 students and their families, and their educational expenditure	2005, 2006	Beijing Normal University Contact: Prof Liu Huizhen on bnuhz@163.com
National Higher Education Political Thought and Public Curriculum Implementation Situation Student Survey	Survey of 2,500 students and their attitudes to political and moral education, the curricula and curricula reform	2005-06	Beijing Normal University Contact: Prof Liu Huizhen on bnuhz@163.com
National Higher Education Students Needs Situation Survey	Survey of 3,000 students and their attitudes to teaching content, methods and assessment	2002	Beijing Normal University Contact: Prof Liu Huizhen on bnuhz@163.com
National Ordinary Higher Education College Special Teacher Survey	Survey of 10,000 teachers and their work conditions, responsibilities and job satisfaction	2006	Beijing Normal University Contact: Prof Liu Huizhen on bnuhz@163.com
Gansu Survey of Children and Families	Longitudinal survey of 2000 children and their schooling experience and attitudes in 20 rural counties	2000-05	University of Pennsylvania and Stanford University Contact: Emily Hannum on hannumem@ssc.upenn.edu
National Higher Education Graduates Employment Situation Survey	Survey of employment among higher education graduates from 34 higher education institutions	2003, 2005	Beijing University Contact: Prof Yue Changjun on cjyue@gse.pku.edu.cn
Urban Residents Education and Employment Situation Survey	Survey into education and employment and the relationship between education and income of 10,000 households	2005	Beijing University Contact: Prof Yue Changjun on cjyue@gse.pku.edu.cn

B4: Major demographic, labour and health surveys collected by the Chinese government

Research interest	Dataset	Year	Principal investigator
Demography	National Population Census	1953, 1964, 1982, 1990, 2000	National Bureau of Statistics
	1% National Population Sample Survey	1982, 1990, 2005	National Bureau of Statistics
	Annual Population Sample Survey	Annual	National Bureau of Statistics
Labour	Labour Force Survey	Annually from 1996, twice yearly from 1997	National Bureau of Statistics
	National Census of Agriculture	1997, 2006	National Bureau of Statistics
	National Economic Census of China	2004	National Bureau of Statistics
Health	National Health Sample Surveys	1993, 1998, 2003	Ministry of Health
	National Mortality Survey	1973, 1990, 2005	Ministry of Health

B5: Datasets held by the Chinese University of Hong Kong

Dataset	Year
National Population Census	1982
Rural Permanent Observation Sites Village Level Surveys	1986-91, 1993, 1995-2000
Sampling Survey on Migration in 74 Towns and Cities	1986
Second China In-depth Fertility Sample Survey	1987
Survey on China's Aged Population	1982, 1986, 1987
Survey of Privately-Owned Enterprises	1991, 1993, 1995, 1997, 2000, 2002, 2004
Survey on the Support System for the Elderly in China	1992
The Rural Household Survey	1986-2000
The Urban Household Survey	1986-97

National census of population includes gender and nationality composition, educational level.

Indian data

India is the most populous liberal democracy in the world, the seventh largest country geographically, and the fourth largest economy in purchasing power terms. A pluralistic, multi-lingual, multi-ethnic society, it became a modern nation-state in 1947.

The country faces serious challenges, such as a soaring population, environmental degradation, extensive poverty, and ethnic and religious strife. Yet its economy is growing rapidly. Predicted to dominate the world economy by 2050 along with China, Brazil and Russia, India will be a major producer and consumer. From welfare economists to anthropologists, social scientists can both discover much about and play an important part in India's progress.

Easily accessible datasets

Indian microdata offer enormous potential for international researchers. Data are usually available for a relatively modest charge, although their intended use may be checked. Virtually all documentation is in English and summaries of datasets are frequently available online.

Demography

After the decennial census, the key dataset for demographic information is the Sample Registration System (SRS), which provides more robust and accurate annual estimates of birth and death rates. The National Family Health Survey (NFHS) and the National Sample Survey (NSS) also collect information on gender, age and other health-related issues.
See table C1: Demographic data.

Economy

Most of the large volume of Indian financial data is published by Reserve Bank of India (RBI) (https://cdbmsi.reservebank.org.in). RBI's large database of bank transactions is not published; however, the annual accounts of private financial companies are released. Cooperative and credit organisation data are held by National Bank for Agriculture and Rural Development (www.nabard.org), and insurance data by the Life Insurance Corporation of India (www.licindia.com). The National Accounts Statistics are usually published annually by the Central Statistical Organization (CSO) and, inter alia, carry GNP, NNP and GDP data. In the finance sector, the All-India Debt and Investment Survey is perhaps the only truly micro database available (contact National Sample Survey Organisation).
See table C2: Economic data.

Three central agencies provide trade statistics for India: Director General of Foreign Trade (http://dgft.delhi.nic.in) (licensing); Director General of Commercial Intelligence and Statistics (www.dgciskol.nic.in) (balance of trade) and RBI (balance of payments).

Labour market

Detailed data on the formal employment sector is collected by various ministries and coordinated by the Ministry of Labour (MoL). There is a huge informal economy (accounting for perhaps 90% of the Indian work force), especially in the agricultural sector.

Several datasets are produced by the Directorate General of Employment and Training (DGE&T) of the MoL. Other important sources are the Small Scale Industry surveys, the CSO's Index of Industrial Production, the Annual Survey of Industry (http://mospi.nic.in/mospi_asi.htm), covering only some aspects of the service sector (which contributes 49% of India's GDP), the Economic Census, and the Decennial Population Census.
See table C3: Employment data.

Housing

Basic housing information is collected in the Decennial Population Census which reports information on access to basic services for the village as an aggregate. NSS surveys collect information on housing conditions and other amenities such as domestic fuel use. However, the best source for micro-level data for access to services is the National Council for Applied Economic Research (NCAER).
See table C4: Housing data.

Health and social welfare

The Central Bureau of Health Intelligence (CBHI) in the Ministry of Health & Family Welfare has dealt with health data at national level since 1961. Morbidity and mortality are covered by the NFHS, conducted by the International Institute for Population Sciences (IIPS), supplemented by much more disaggregated health data from the Reproductive and Child Health District Level Household Survey (DLHS-RCH).

The contribution of the private health-care sector is an indispensable aspect of the Indian health system and is covered by the National Institute of Medical Statistics, a nodal agency for 26 independent research centres on specific diseases.
See table C5: Health data.

Education

Educational data are compiled by the statistics division of the department of secondary and higher education of the Ministry of Human Resource Development (MHRD), with detailed information collected every 5–7 years by the National Council for Educational Research and Training (NCERT) through the All-India Educational Survey. For achievement levels, see the Public Report on Basic Education (PRBE) 1996 and recent NCERT results published by the Ministry of Human Resources Development (MHRD).
See table C6: Education data.

Infrastructure and Transport

Currently, there are no comprehensive statistics on infrastructure and transport. Some information is available through the Economic Census. The NSS and Census also have information on various infrastructure facilities. The village directory gives information for each village on the availability of various infrastructure services, but is updated only every ten years. Transport statistics are published by Ministry of Surface Transport (MST).

Crime

Crime data are either administrative data, usually based on recorded crimes, and survey data, based on victim reports. The National Crime Records Bureau (NCRB) publishes an annual statistical report Crime in India. It lists crimes under the Indian Penal Code (IPC) and Special and Local Laws (SLL – the IPC series shows rather more consistency than the SLL total). Data is presented by crime type and there is also data on police actions. Corruption is covered by a 2005 survey by Transparency India with international collaboration, *Corruption in India, 2005* (www.transparency.org/regional_pages/asia_pacific/newsroom/current_in_focus/india_study_2005).

Consumption Surveys

The most widely used datasets for consumer expenditure data are the annual NSS surveys. Other important sources are: the Family Living Surveys since 1958; the *Great Indian Market,* NCAER's Market Information Survey of Households (MISH). The only longitudinal dataset in this field appears to be the ICRISAT's 1975–84 study of 400 households.

Data sources

India has a long history of collecting social and economic statistics and, more recently, of conducting very large scale national- and state-level social surveys. After independence, development planning to build a modern India was heavily dependent on good national statistics. Thus, the Central Statistical Organisation (CSO) was established in 1951; the National Sample Survey (NSS) in 1950; the National Sample Survey Organisation (NSSO) in 1970 and the National Statistical Commission (NSC) in 2000 (http://mospi.nic.in/). A growing volume of administrative data forms the basis of many national and local statistics in India.

Improving data availability and access

Making data and its analysis more easily available, including in microdata form wherever possible, has been a high government priority. Major survey data have tended to remain unchanged in format despite India's now rapidly growing economy. Notwithstanding the scale of the data (with samples well into six figures) and relatively easy access, more complex analyses have been relatively limited until recently. While there are key central nodal points for certain datasets, there is as yet no overall central access point for researchers. The feasibility of developing a national archive is the subject of ongoing debate. There are also some commercial data sources; some non-governmental organisations restrict access to their privately collected data.

About the scoping study

This study was produced by the Department of Social Policy and Social Work, University of Oxford, by a research team comprising George Smith, Sony Pellissery, Sweta Rajan and Sylvie Dubuc. After establishing what Indian microdata was already in use in the academic and social research community, the team undertook brief visits to potential microdata sources and major data users in Delhi and Mumbai, following up contacts by e-mail and telephone. This study covers only the national level; there is an enormous body of data sources at state and district level studies.

The full report India, Microdata Scoping Study is available at www.esrcsocietytoday.ac.uk/idfpapers

Indian datasets

C1: Demographic data

Dataset	Year	Principal investigator
Census of India	1871–2001	Registrar General of India (RGI) www.censusindia.net
Civil Registration	1850s–2006	Registrar General of India (RGI) www.censusindia.net
SRS	1964–65 (pilot)	Registrar General of India (RGI) www.censusindia.net
	1970–2006	Registrar General of India (RGI) www.censusindia.net
SRS, Mortality and Fertility	1972, 1979	Registrar General of India (RGI) www.censusindia.net
NFHS	1992–93, 1998–99, 2006 (first results)	Indian Institute of Population Sciences (IIPS) www.nfhsindia.org or www.measredhs.com
Reproductive and Child Health Project–Rapid Household Survey (RCH–RHS)	1998–99, 2002–03	Indian Institute of Population Sciences (IIPS) www.nfhsindia.org or www.measredhs.com
Multiple Indicator Cluster Survey (MICS)	2000–01	UNICEF and Department of Women and Child Development (DWCD), Ministry of Human Resource Development (MHRD) www.childinfo.org/MICS2/m2reports/reports.htm
National Sample Survey (NSS)	1950-2005	National Sample Survey Organisation (NSSO) http://mospi.nic.in/nsso_test1.htm
Survey of Living Conditions	1997–98 (Bihar and Uttar Pradesh)	World Bank

Surveys the living conditions, health and educational position of women and children in households. Third round in progress.

C2: Economic data

Dataset	Year	Principal investigator
Annual Survey of Industries	1960–2006	Ministry of Labour http://labour.nic.in
Small Scale Industry Surveys	Annual	Ministry of Labour http://labour.nic.in
Economic Census	1977 to date (5 rounds)	CSO http://mospi.nic.in/cso_test1.htm
Economic Surveys	Annual	CSO http://mospi.nic.in/cso_test1.htm
Index of Industrial Production	Annual	CSO http://mospi.nic.in/cso_test1.htm
Family Living Standards Survey	Since 1958	Labour Bureau
Market Information and Household Survey (MIHS)	2002	National Council for Applied Economic Research (NCAER) www.ncaer.org
Market Information Survey	2005	CMIE www.cmie.com
NSS Household Consumption Expenditure Survey	Annually since 1950	NSSO http://mospi.nic.in/nsso_test1.htm

Important source on number of enterprises and their employment details with more detailed information in the Follow-up Enterprise Surveys.

C3: Labour Market data

Dataset	Year	Principal investigator
Employment and Unemployment Situation in India (NSS)	Since 1965	NSSO http://mospi.nic.in/nsso_test1.htm
Apprentice Training in India	Annually	DGE&T http://dget.nic.in
Bulletin of Job Opportunities in India	Annually	DGE&T http://dget.nic.in
Census of Central Government Employees	Annually	DGE&T http://dget.nic.in
Employment Exchange Statistics	Annually	DGE&T http://dget.nic.in
Employment in the Organized Sector	Quarterly	DGE&T http://dget.nic.in
Employment Review	Annually	DGE&T http://dget.nic.in
Occupational–Educational Patterns of Employees in India	Annually	DGE&T http://dget.nic.in
Quarterly Employment Review	Quarterly	DGE&T http://dget.nic.in
Quick Estimates of Employment in the Organized Sector	Annually	DGE&T http://dget.nic.in

C4: Housing data

Dataset	Year	Principal investigator
Census of India	1871–2001	RGI www.censusindia.net
India Human Development Report	1999	NCAER www.ncaer.org
Housing Conditions in India (NSS 49th Round)	1993	NSSO http://mospi.nic.in/nsso_test1.htm
Migration in India (NSS 49th Round)	1993	NSSO http://mospi.nic.in/nsso_test1.htm
Slums in India (NSS 49th Round)	1993	NSSO http://mospi.nic.in/nsso_test1.htm
Handbook of Urban Statistics	1993, 1995	National Institute of Urban Affairs, www.niua.org

C5: Health and social welfare data

Dataset	Year	Principal investigator
NSS Household Consumption Expenditure Survey	1993–94, 1995–96, 1998–99, 2002–03	NSSO
Human Development Profile Survey	1994	NCAER
NFHS	1992–93, 1998–99, 2006 (first results)	IIPS
Men and Women in India	2001	Department of Women and Child Development
Reproductive and Child Health–District Level Household Survey (RCH–DCL)	1998	IIPS
DLHS–RCH	2002–04	IIPS
Health Information of India	Since 1963	CBHI www.cbhidghs.nic.in
Rural Health Statistics	Since 1963	CBHI www.cbhidghs.nic.in
National Aids Survey	Annually	National Aids Control Organization (NACO) and National Institute of Medical Statistics

C6: Education data

Dataset	Year(s)	Principal investigator
All-India Education Survey	Since 1977 (every 5–7 years)	NCERT www.ncert.nic.in
District Information System of Education (DISE)	Since 1990s	National Institute of Education, Planning and Administration www.educationforallinindia.com
PROBE	1996	MHRD www.education.nic.in
Census of India	1871-2001	RGI www.censusindia.net
Attending Educational Institutions in India: Level, Nature and Cost (NSS 52nd Round)	1996	NSSO
Educational Achievement Surveys at Class IV, V, VII / VIII	2000	NCERT

Infrastructure, teacher-pupil ratios, educational attainment, rural-urban differences and school-level performance at district level.

South African data

The Republic of South Africa has a diverse population of 47 million with 11 official languages. It has Africa's largest and most developed economy, producing over one third of sub-Saharan Africa's GDP with 6% of its population.

Yet twenty years ago, South Africa was internationally isolated and ruled by apartheid. The transition to successful democracy astounded those predicting a violent demise to apartheid. Today, South Africa welcomes representatives from every troubled corner of the planet eager to find out how the country achieved such a peaceful transition.

And there is much more to learn from South Africa. It faces challenges common to many countries such as high unemployment, poverty, crime, HIV and Aids. Social scientists, working in collaboration with epidemiologists, medical scientists, environmental scientists and others can make a significant contribution to understanding and informing appropriate responses to these challenges.

Easily accessible data

There are over 40 key South African datasets that are easily accessible. Researchers from anywhere in the world can access the data usually for little or no costs and without overly restrictive conditions of use. Descriptions of the surveys and data, costs, conditions of use, contact details and an assessment of the utility of the data for researchers can be found in the data explorers' full report.
See table D1: Easily accessible datasets by area of research interest.

Data sources

The following four organisations provide researchers with access to the majority of easily accessible microdata in South Africa. They are also all trying to make data resources more widely available:

Statistics South Africa

Statistics South Africa (Stats SA) is the national statistics agency in South Africa and collects, processes and produces official statistics (www.statssa.gov.za). The main microdata sources are the Population Census and household surveys.

Publicly available data can be downloaded from Stats SA's website (mainly macrodata), or found in Stats SA publications (macrodata), or by contacting Stats SA directly (macro and microdata). Most datasets and publications are free of charge. Once researchers have a dataset, they are allowed to disseminate the data further provided no charge is made and Stats SA is acknowledged as the supplier and owner of the data and copyright.

Human Sciences Research Council

The Human Sciences Research Council (HSRC) is the premier social science research organisation in South Africa (www.hsrc.ac.za) and carries out various national and sub-national surveys.

Although the majority of HSRC's datasets are not officially in the public domain, there is a new drive for data to be released as soon as possible after collection, cleaning and preliminary analysis. In the meantime, researchers can find out about HSRC surveys on education, employment and skills at the Human Resources Development Data Warehouse (http://hrdwarehouse.hsrc.ac.za) along with details on how to gain access.

South African Data Archive

The South African Data Archive (SADA) acts as a broker between various data providers and the research community by safeguarding datasets and making them available over the internet (www.nrf.ac.za/sada). SADA's data includes census and household surveys, demographic and health related studies, and surveys on crime, income and poverty. Researchers can order datasets online and if within South African working hours can receive the data within one hour.

DataFirst

DataFirst at the University of Cape Town also archives datasets from South African surveys. Due to copyright restrictions it can only make datasets of the university surveys available to researchers online (www.datafirst.uct.ac.za). For the other data, researchers must physically visit the university.

Administrative data

The South African government collects a wide range of data for administrative purposes such as information on all schools in the country and records on individuals receiving benefits. The data are normally aggregated to a national or regional level and published by the government. The data at an individual level are not routinely made available for researchers outside the government because of concerns about confidentiality. Therefore, in most cases researchers will have to negotiate access individually to the data by demonstrating the benefits of the research. Some government departments (eg Department of Education) are more open to making datasets available to bona fide researchers.
See table D2: Key administrative datasets for some areas of interest to social scientists.

Improving data availability and access

In South Africa easily accessible microdata are held by several different organisations. It would be extremely useful for researchers if all data were held in one repository, which would also reduce the duplication of effort that inevitably takes place between the organisations. The UK's Data Archive is a good example of a single repository that could share its expertise with South Africa. Researchers would also benefit from a microdata user group to share local knowledge and experience of using different datasets.

Improving access to administrative data is more complicated. Fortunately, a promising development in the Department of Social Development may encourage change. The department has commissioned a study into how its administrative data can be used by researchers to develop a comprehensive system for the care and support of orphans and vulnerable children. A positive outcome could encourage more accessibility to administrative data for researchers such as in the UK which has greatly benefited from this openness.

About the scoping study

This summary is based on the work by Helen Barnes, Michael Noble, Chris Dibben, Charles Meth, Gemma Wright and Lucie Cluver from the Centre for the Analysis of South African Social Policy (CASASP) at the University of Oxford. The study is based on CASASP's existing knowledge of microdata in South Africa along with interviews with 57 senior members of staff in South African organisations involved in the collection and use of data. The organisations ranged from government departments to universities to Stats SA and the HSRC.

The full report South Africa Microdata Scoping Study is available at www.esrcsocietytoday.ac.uk/idfpapers

South African datasets

D1: Easily accessible datasets by area of research interest

Research interest	Dataset	Website or contact details
Demography	Africa Centre Demographic Information System	www.africacentre.ac.za
	Agincourt Health and Demographic Surveillance System	http://hermes.wits.ac.za/www/Health/PublicHealth/Agincourt
	Birth to Twenty	http://sunsite.wits.ac.za/birthto20
	Census of Population (10% sample)	www.statssa.gov.za/census01/html
	Demographic and Health Survey	www.doh.gov.za/facts/1998
	General Household Survey	Contact Stats SA User Information Services by email: info@statssa.gov.za
	October Household Survey	Contact Stats SA User Information Services by email: info@statssa.gov.za
	South African Migration and Health Survey	www.nrf.ac.za/sada
	The 2001-2002 HSRC Migration Survey	www.nrf.ac.za/sada
Housing	Census of Population (10% sample)	www.statssa.gov.za/census01/html
	All Media Products Survey	www.sarf.co.za
	General Household Survey	Contact Stats SA User Information Services by email: info@statssa.gov.za
	KwaZulu-Natal Income Dynamics Study	http://sds.ukzn.ac.za/default.php?7,12,9,4,0
	October Household Survey	Contact Stats SA User Information Services by email: info@statssa.gov.za
	Rural Survey	Contact Stats SA User Information Services by email: info@statssa.gov.za
Social welfare	General Household Survey	Contact Stats SA User Information Services by email: info@statssa.gov.za
	KwaZulu-Natal Income Dynamics Study	http://sds.ukzn.ac.za/default.php?7,12,9,4,0
Economy	Census of Population (10% sample)	www.statssa.gov.za/census01/html
	Financial Diaries Project	www.financialdiaries.com
	Greater Durban Metropolitan Area Large Manufacturing Firms Survey	http://sds.ukzn.ac.za/default.php?11,0,0,0,0
	Income and Expenditure Survey	Contact Stats SA User Information Services by email: info@statssa.gov.za
	KwaZulu-Natal Income Dynamics Study	http://sds.ukzn.ac.za/default.php?7,12,9,4,0
Labour market	Cape Area Panel Study	www.caps.uct.ac.za
	Census of Population (10% sample)	www.statssa.gov.za/census01/html/
	General Household Survey	Contact Stats SA User Information Services by email: info@statssa.gov.za
	Khayelitsha/Mitchell's Plain Survey	www.datafirst.uct.ac.za/data_kmp.html
	KwaZulu-Natal Income Dynamics Study	http://sds.ukzn.ac.za/default.php?7,12,9,4,0
	Labour Force Survey	Contact Stats SA User Information Services by email: info@statssa.gov.za
	October Household Survey	Contact Stats SA User Information Services by email: info@statssa.gov.za
	Rural Survey	Contact Stats SA User Information Services by email: info@statssa.gov.za
	Survey of Activities of Young People	http://www.ilo.org/public/english/standards/ipec/simpoc/southafrica/index.htm
	The 2001-2002 HSRC Migration Survey	www.nrf.ac.za/sada
	Time Use Survey	Contact Stats SA User Information Services by email: info@statssa.gov.za

Longitudinal health and demographic survey since 1992 of 70,000 people living in rural area of Mpumalanga province.

A national annual survey of the living circumstances of 30,000 South African households across the country from 2002-05.

A national sample of over 4,000 individuals investigating the causes of migration in 2001-02.

D1: Easily accessible datasets by area of research interest (cont.)

Research interest	Dataset	Website or contact details
Education	Cape Area Panel Survey	www.caps.uct.ac.za
	Census of Population (10% sample)	www.statssa.gov.za/census01/html
	General Household Survey	Contact Stats SA User Information Services by email: info@statssa.gov.za
	KwaZulu-Natal Income Dynamics Study	http://sds.ukzn.ac.za/default.php?7,12,9,4,0
	October Household Survey	Contact Stats SA User Information Services by email: info@statssa.gov.za
	Student Choice Behaviour Project Phase 1	Contact Mr Michael Cosser, Chief Research Specialist, HSRC by email: mcosser@hsrc.ac.za
	Survey of Activities of Young People	http://www.ilo.org/public/english/standards/ipec/simpoc/southafrica/index.htm
	Survey of Independent Schools	Contact HRD Data Warehouse by email: hrdwarehouse@hsrc.ac.za
	Survey of Private and Further Education in South Africa	Contact HRD Data Warehouse by email: hrdwarehouse@hsrc.ac.za
	Time Use Survey	Contact Stats SA User Information Services by email: info@statssa.gov.za
Transport	All Media Products Survey	www.sarf.co.za
	Census of Population (10% sample)	www.statssa.gov.za/census01/html
	National Household Travel Survey	www.dot.gov.za/projects/nts/framesPage.htm
	Time Use Survey	Contact Stats SA User Information Services by email: info@statssa.gov.za
Crime	All Media Products Survey	www.sarf.co.za
	National Victims of Crime Survey	www.nrf.ac.za/sada
	National Youth Victimisation Survey	Contact Mr Patrick Burton Research Director, CJCP by email: Patrick@cjcp.org.za
Health	Africa Centre Demographic Information System	www.africacentre.ac.za
	Agincourt Health and Demographic Surveillance System	http://hermes.wits.ac.za/www/Health/PublicHealth/Agincourt
	Birth to Twenty	http://sunsite.wits.ac.za/birthto20
	Cape Area Panel Survey	www.caps.uct.ac.za
	Census of Population (10% sample)	www.statssa.gov.za/census01/html
	Demographic and Health Survey	www.doh.gov.za/facts/1998
	General Household Survey	Contact Stats SA User Information Services by email: info@statssa.gov.za
	HIV and Sexual Behaviour Among Young South Africans	www.rhru.co.za
	KwaZulu-Natal Income Dynamics Study	http://sds.ukzn.ac.za/default.php?7,12,9,4,0
	National Food Consumption Survey	www.sahealthinfo.org/nutrition/foodconsumption.htm
	National HIV and Syphilis Antenatal Sero Prevalence Survey	www.doh.gov.za/docs/reports/2005/hiv.pdf
	South African Migration and Health Survey	www.nrf.ac.za/sada
	South African National HIV Prevalence, HIV Incidence, Behaviour and Communication Survey	Contact Prof Thomas Rehle, Director, Social Aspects of HIV/AIDS and Health, HSRC by email: trehle@hsrc.ac.za
	Time Use Survey	Contact Stats SA User Information Services by email: info@statssa.gov.za

A longitudinal study of the living conditions of over 1,000 households in one of South Africa's poorest provinces in 1993, 1998 & 2004.

Survey of the extent, nature and consequences of work-related activities of children aged 5-17 years.

Longitudinal study of the health and development of over 3,000 children born in 1990 in Johannesburg-Soweto.

D2 Key administrative datasets for some areas of interest to social scientists

Research interest	Potential dataset	Description	Principal investigator
Demography	Population Register	Records on births, marriages, deaths, adoptions, naturalisation and resumption of citizenship of every South African citizen	Department of Home Affairs
Social welfare	SOCPEN	Records on all individuals receiving social security benefits	South African Social Security Agency
Labour market	Unemployment Insurance Fund	Records on employers (from business type to salaries paid) and employees in South Africa	Department of Labour
Education	Headcount Survey, Annual Survey, Learner Unit Record Information and Tracking System, Senior Certificate Examination Results	Various sources of data on pupils, their background, subjects studied, progression of learners and exam results	Department of Education

Conclusion:

Towards easily accessible microdata across the world

This *Rough Guide to Microdata in Brazil, China, India and South Africa* gives a flavour of the enormous wealth of data collected in four of the world's most important countries. It also argues that the rewards from sharing microdata are huge – easy access to microdata improves research, better research improves government policies and the lives of countless people.

In each of the four countries, organisations are working to increase access to their data. They already see the benefits of sharing microdata within their country. On the international level, our increasingly interwoven world means many pressing problems from poverty to climate change cannot be properly understood by confining research within the borders of a single country. So it is not surprising that representatives from research funding agencies and statistical authorities in over 30 !##check figure at proof##! countries are enthusiastically taking part in the Foundation Conference for the International Data Forum. The time is ripe for the international community to negotiate ways to share microdata effectively across national borders.

However, there are stumbling blocks to achieving this objective. The four scoping studies highlighted some of the obstacles and suggested ways forward.

Language

A social scientist wishing to analyse Brazil's housing datasets needs to understand Portuguese or pay for translation. The fee is often beyond the means of individual researchers but translation to, say, English could be performed centrally by relevant science funding agencies to benefit all English speaking researchers worldwide.

No central resource

Brazil, China, India and South Africa are vast with many different organisations collecting and holding data. It would be extremely useful for researchers if all the data were held in one repository in each country. Besides making data easier to find, a central data archive can improve survey documentation and begin efforts to make surveys compatible and comparable both for policymaking and for research purposes. Yet building and maintaining a central data archive is expensive. It also requires considerable technical skills. Some countries, such as the United Kingdom which runs a well-regarded national data archive for social scientists, could provide useful technical expertise. Other initiatives, such as CESSDA (see opposite) are being developed to use the internet and the web to bring together many different archives to create a virtual central resource.

Confidentiality

Governments are justifiably reluctant to release large amounts of private information about individuals because of privacy issues. Researchers can address this concern in a number of ways. Some data can be made anonymous removing the risk of disclosure. For other data, it can be a matter of managing the risk. Some data is naturally less sensitive than others; data in this category could be made available.

Legal and cultural

Finally in some countries, legal constraints and a culture of not sharing data are major obstacles. Clearly trust needs to be built to show benefits outweigh costs. The way forward needs to be to engage as much as possible with scholars in such countries in open and transparent efforts to build comparative research networks.

Overcoming these stumbling blocks will not be easy. But the rewards of better research leading to better lives are far greater than the costs.

All four full scoping reports are available from:
www.esrcsocietytoday.ac.uk/idfpapers

CESSDA
(Council of European Social Science Data Archives)

CESSDA is a distributed research infrastructure that has provided and facilitated Social Science researchers' access to high quality data and supported their use for over thirty years. The CESSDA network currently extends across 21 countries in Europe, holds around 15,000 data collections and provides access to over 20,000 researchers conducting comparative cross-national research, regardless of their location in the infrastructure.

The CESSDA data archives have a long-standing record for the acquisition, support and supply of data range across official government censuses and surveys, election studies, longitudinal and cohort studies, opinion polls and surveys, addressing most issues relating to society and human activity. Many of these types of International and European datasets, such as the European Social Survey, Eurobarometers, European Values Surveys and the International Social Survey Programme can be accessed through the CESSDA Portal. The CESSDA Portal is the principle means by which users can locate research data and metadata, as well as questions or variables within datasets, stored at CESSDA member archives throughout Europe. Additionally, CESSDA has data access agreements and arrangements with other data holding organisations worldwide.

CESSDA will continue to ease researchers' access to data across international and organisational boundaries and plans a major upgrade in the near future. Currently CESSDA works as a common platform for its national members, who share a mission statement and promote the development and implementation of metadata standards, Social Science thesauri, information security standards/user authentification procedures and the simplification and standardisation of the rights management framework. Over the coming years CESSDA plans to upgrade its Portal to provide a comprehensive and integrated 'One-Stop Shop' for data location, access, analysis and delivery across the social science community in Europe, at the same time extending and deepening the CESSDA network to include new and associated data-holding and creating organisations.

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

Economic and Social Research Council
Polaris House
North Star Avenue
Swindon
SN2 1UJ
United Kingdom

Telephone: +44 (0)1793 413000
Fax: +44 (0)1793 413001

The Economic and Social Research Council is the UK's leading research and training agency addressing economic and social concerns. We aim to provide high-quality research on issues of importance to business, the public sector and Government. The issues considered include economic competitiveness, the effectiveness of public services and policy, and our quality of life.

The ESRC is an independent organisation, established by Royal Charter in 1965, and funded mainly by the Government.