

Micro-data scoping study – China

Report submitted to

Economic and Social Research Council, U.K.

Prepared by

Dr. Sarah Cook
Institute of Development Studies
University of Sussex
(S.Cook@ids.ac.uk)

and

James Keeley
International Institute for Environment and Development
London
(James.Keeley@iied.org)

January 2007

Contents

Abbreviations

- 1. Introduction**
- 2. China's statistical system: Mapping the institutions**
 - i. The government statistical system
 - ii. Other government agencies
 - iii. Academic and research institutions
 - iv. Market research organisations
- 3. Data and resources**
 - i. General overview
 - ii. Areas of concern to ESRC
 - a. Population
 - b. Labour
 - c. Health
 - d. Education
 - e. Crime and public security
 - iii. Other data resources
 - a. Government economic and social data
 - b. Other non-governmental data and resources
 - c. International databanks
 - d. International agencies and funders
- 4. Issues of access: Opportunities and constraints**
 - i. The legal environment
 - ii. Making data public: Attitudes and cost
 - iii. What the ESRC has to offer China
- 5. What does this mean for ESRC? Conclusions and recommendations**
 - i. Data availability and access
 - ii. What can be done?

Appendices

- A.1 List of data sets**
- A.2 List of individuals and organisations contacted**
- A.3 Social Policy Analysis and Information Center, ISDPP**

Abbreviations

ACFTU	All China Federation of Trade Unions
ACWF	All China Women's Federation
CAS	Chinese Academy of Sciences
CASS	Chinese Academy of Social Sciences
CCER	China Center for Economic Research
CDC	Center for Disease Control
CDC, Michigan	China Data Center, Michigan
CEDAW	Convention on the Elimination of Discrimination Against Women
CHIP	China Household Income Project
CIDA	Canadian International Development Agency
CPDRC	China Population and Development Research Center
CPIRC	China Population and Information Research Center
CSDN	China Survey Data Network
CULS	China Urban Labour Survey
DFID	Department for International Development
ESRC	Economic and Social Research Council
FSRS (RCRE)	Fixed Site Rural Survey
IPLE	Institute of Population and Labour Economics
ILO	International Labour Organisation
ISDPP	Institute of Social Development and Public Policy
LFS	Labour Force Survey
MDG	Millenium Development Goals
MOCA	Ministry of Civil Affairs
MOE	Ministry of Education
MOH	Ministry of Health
MOLSS	Ministry of Labour and Social Security
NBS	National Bureau of Statistics
NSSF	National Social Science Foundation
PFPC	Population and Family Planning Commission
RCCC	Research Center for Contemporary China
RCRE	Research Center for Rural Economy
USC	University Service Center

1. Introduction

This report presents the results of a study of available micro-data resources in China. Following recommendations made by the 'Review of International Data Resources and Needs' commissioned by the ESRC in 2005, the main purposes of this study were to catalogue more systematically relevant sources of economic and social micro-data existing in China, locate their institutional host, identify their main characteristics, and establish the conditions for future availability to UK researchers. This report aims to allow ESRC and other research bodies in the United Kingdom to assess the value of different data sources, while also appreciating the restrictions on access to existing social science data in China.

In writing this report, the authors conducted a comprehensive survey of relevant sources of statistical data across a range of social and economic sectors. Such data are collected through official government channels as well as universities, research institutions and other relevant organisations. The following key questions were raised in interviews undertaken with a wide range of respondents, including government officials, academics, data collectors and users (see Appendix 2 for list of individuals and organisations interviewed):

- What are the main national social science data sources currently available through government agencies at national and sub-national levels?
- What other social science data sources are available (e.g. through sample surveys undertaken by academic institutions, commercial survey groups, etc.)?
- What are the current restrictions on access to and use of social science data by overseas researchers?
- What are the options for negotiating access and use of such data – whether for public access or on a case by case basis for specific projects?

In addition to providing a review of existing resources, including relevant documentation where available, the report discusses Chinese government regulations and limits on data access and use particularly by foreigners. It is informed by the experiences of both Chinese and overseas researchers in gaining access to data for research purposes. This information should help in identifying successful modes of partnership or agreements that could lead to reliable longer term access.

China's current statistical system has been developed and expanded since the opening of the economy from the late 1970s. A Statistics Law was passed in 1983 (and revised in 1996), since which time China has been engaged in an ongoing process of modernising the statistical system. The survey instruments and collection methods continue to be improved, with regular changes aimed at making national data more comparable with that of other countries. The government is frank about existing limitations in terms of quality and international comparability. Nonetheless, the quantity and quality of data collected at the national level, despite its limitations, provides a valuable resource for researchers.

In the Chinese context, the fundamental challenge in gaining access to data relates to the sensitivity surrounding various statistics and particularly micro-level data. Access to micro-level statistical data (and particularly to national sample surveys) is difficult for both Chinese and foreign researchers, but is particularly problematic for the latter. A series of regulations in recent years has made it increasingly difficult for foreigners to access data particularly in the social sciences. Restrictions apply to any micro-level survey data, whether produced by the National Bureau of Statistics (NBS), by other government agencies or by research institutions; and to foreigner involvement in the collection of social science survey data. According to both domestic and international scholars, it appears to have become significantly harder to obtain the permissions required to undertake surveys or participate in the collection of such data when foreign researchers are involved.

Some of the restrictions relate to the perceived sensitivity of social science data. There is a lack of clarity about whether data is a public good or a private resource, and considerable uncertainty about what kind of information might be considered confidential or 'secret' at different points in time. This makes institutions and individuals extremely risk averse about sharing data, when they believe they might in the future be held responsible for how those data are used or published by others. Such difficulties exist internally, between different institutions and individuals within China, as well as – more acutely – where foreigners (or even Chinese students studying overseas) are involved in the collection or analysis of data.

Another major issue concerns the use of data as a revenue raising resource. Many under-funded government departments or other institutions may see access fees or the sale of data and other information as a relatively easy way of raising revenues. In many cases (particularly for data from NBS), data has become prohibitively expensive for prospective users. Thus, while understanding what data is available and through what channels is a first step to improving access, the barriers to acquisition or use are likely to remain significant, both in financial and political terms.

Given these conditions, there are serious limitations on the micro-data that is readily available or could be easily accessible to the international research community through open access channels. There is also relatively limited detailed documentation available (particularly in English) of the kind requested by the ESRC and desirable to prospective users. Where more detailed documentation exists, it is usually in Chinese. This report provides a brief description of key data sources which might be more readily accessible, where possible indicating the main categories of variable and other documentation available. However, given the context of restrictions and sensitivity about social science data, described in further detail below, it has generally not been possible to discuss potential acquisition of specific data sets in detail. Instead, the report's recommendations focus principally on activities through which ESRC might contribute to improving conditions for access to data by the UK research community, given the restrictions currently operating in the Chinese context.

2. China's Statistical System: Mapping the Institutions

i. The government statistical system

China's overall statistical system can be divided into two main components: official government-collected statistics, and other 'non-government' or unofficial statistics. Government statistics include 'comprehensive' statistics collected by the National Bureau of Statistics (NBS) and survey teams directly under NBS, and by local statistical departments at each level of government; and departmental (or administrative) statistics, that is, data reported by government agencies under the State Council and corresponding agencies or personnel at sub-national levels. (See Figure 1)

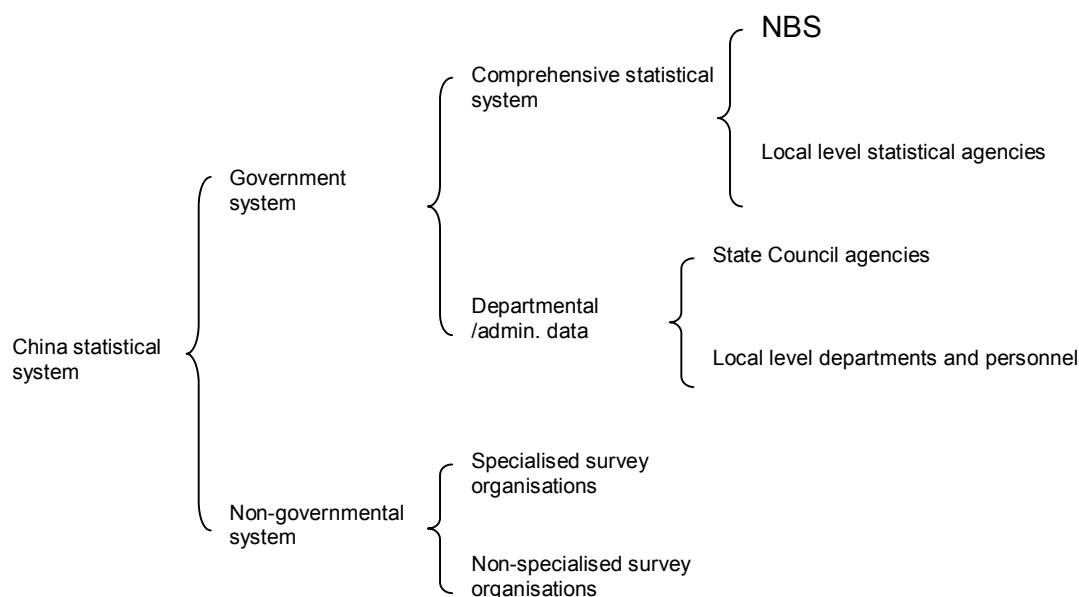


Figure 1: China's statistical system

The unofficial or non-governmental statistics system covers other organisations involved in data collection and analysis, including enterprises and public service units; university or academic institutions, and private commercial organisations. Data from these sources can vary in quality while their usefulness to researchers also varies based on the purpose for which they are being collected.

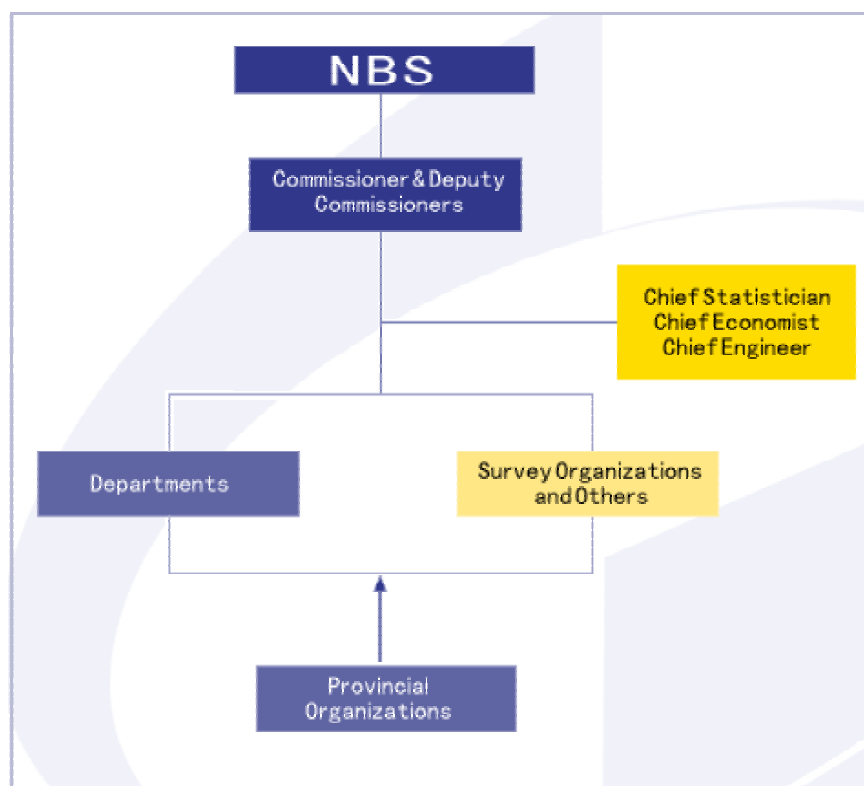
National Bureau of Statistics (NBS)

The National Bureau of Statistics is a government agency directly under the State Council (China's Cabinet-level body), and has overall responsibility for national statistics and national accounts work. It plays the leading regulatory and supervisory role for the Chinese data collection system, including over 'non-governmental' data collection agencies. It is responsible for the implementation of the National Statistics Law and supervises or grants permission for other institutions, such as research organisations, to collect data. All social science surveys in which foreign researchers or funders are involved, or where foreigners will have access to the data, should obtain approval from NBS.

In keeping with China's size and centralized system of government, the NBS structure is replicated at lower levels of government. At the provincial level, the Provincial Statistical Bureau is directly responsible to the Provincial Government. However, Provincial (and below them County) level statistics departments are responsible for reporting certain statistics directly up the governmental or departmental system to the Central level, and ultimately to

the NBS (figure 2). These data are publicly reported in summary form, for example in annual publications such as the National Statistical Yearbook.

Figure 2: Organisation of National Bureau of Statistics



NBS Departments and survey organisations

The NBS and local level statistics departments have statutory responsibility for organising, coordinating and regulating (at the relevant administrative level) the collection of official statistics. Responsibilities include coordinating overall data collection efforts, drafting laws and regulations, and planning surveys and data reporting systems. They are also responsible for national accounting, implementing surveys on issues of national significance, supervision and coordination of other survey activities, and management and supervision of departmental reporting systems.

The NBS has 16 departments, of which six are most directly involved with micro-data collection. These are: industry and transport; fixed asset investment, foreign trade; population and employment; rural socio-economic and urban socio-economic survey departments. Central NBS survey organizations (such as the urban and rural survey teams) fall directly under the jurisdiction of the NBS, and establish their own offices at the local level. Such comprehensive survey teams (which undertake the implementation of rural, urban and enterprise surveys) are established in 15 provinces, 318 municipalities, and 887 counties and are funded directly by the NBS. Temporary offices are set up at lower levels of government for major survey efforts such as the census.

Key categories of micro-data collected by NBS include:

- *Census*: Population, economic (including previous industrial) and agricultural censuses, overseen by the Census Department of the National Statistical Bureau.
- *Population and Employment data*: Collected through surveys of demographic change and labour force surveys (described further in section 3).

- *Rural and urban social and economic data*: Collected by the rural and urban household survey teams under the NBS, through surveys and diaries.
- *Industry and transport*: Surveys are organised and implemented on issues such as industrial enterprise production, management and energy resource use. The administrative reporting system focuses on national industrial economic development, and collects data on large scale enterprises. Indicators include productivity, economic, energy consumption, costs, employment and wages.
- *Fixed asset investment data*: This includes the construction and real estate sectors. Administrative reporting includes fixed asset investments and real estate development by all units, private investments in sectors such as urban industry, mining and construction

ii. Other Government Agencies - national and sub-national levels

Each government line agency is responsible for collecting administrative data, as well as for undertaking specific surveys relevant to its role. For example, the Ministry of Civil Affairs collects extensive administrative data on civil society organisations (through its registration process), on social assistance recipients, and on village elections. The Research Center for Rural Economy (RCRE) under the Ministry of Agriculture undertakes annually a longitudinal 320 village sample survey on rural and agricultural development. Other examples are described further below in relation to specific sectors of interest.

Each ministry has a Statistics Department responsible for receiving and publicising relevant data reported through the line agency, and for reporting aggregate data to NBS. Most ministries also have associated research institutes responsible for policy related research and survey work, such as the Labour Research Institute under MOLSS.

The central level statistical system is essentially replicated at sub-national levels, and data collected by provincial and other sub-national government agencies is fed up to the central agencies. Additionally, provincial and local governments may have their own data collection efforts that are separate from national surveys and reporting systems.

In addition to line agencies, other government organisations also collect information and undertake surveys on issues of particular policy concern. These include for example the Development Research Center under the State Council; the National Development and Reform Commission, and the Leading Group on Poverty Alleviation. A further group of institutions which collect data include mass organisations such as the All China Federation of Trade Unions (AFCTU), the All China Women's Federation (ACWF); Federations of the elderly and disabled; and Chambers of Commerce or Trade Associations. However, these tend to be infrequent and have not been covered in this report. Further work would be needed to assess the quality and content, or possibility of using such data.

iii. Academic and research institutions

A third set of data sources to explore are sample surveys implemented by research or policy institutions. These may be funded by Chinese government or international funds. Larger sample surveys which aim to be national in scope are often collected in collaboration with the NBS in order to benefit from its networks of trained enumerators. An example is a longitudinal study of income inequality implemented by CASS and supported by the Ford Foundation since the mid-1990s. Others are collected independently of the NBS though may require approval.

Government institutions include the Chinese Academy of Social Sciences (CASS) and other Academies (for example, agricultural, geography, the natural sciences); and parallel academies supported by each level of government. Sub-national (provincial or municipal)

academies tend to address issues of local concern; there is no direct hierarchical relationship with CASS, and they tend to have limited resources for large scale surveys.

Universities and research institutes While a large number of such institutes exist, only a few have the capacity or financial resources to collect good quality survey data. The main Chinese funding body for the social sciences, the National Social Science Foundation (NSSF) provides relatively small grants (on average approximately 60,000 yuan or £4000) – insufficient to undertake a reasonable sized survey. Thus good quality micro-survey data is collected principally by a small number of key institutions or research teams, most of which have access to international funding sources either directly or through international collaboration. Some of these key institutions are discussed in relation to specific data sets in section 3. They tend to be concentrated in Beijing: examples include Peking University, Tsinghua University, People’s University, Beijing Normal University, CASS. Outside Beijing, major cities such as Shanghai also have excellent capacity and funding and may engage in data collection activities, although often the scope is more limited in geographic scope to the local region.

iv. Market research organisations

While market organisations were not a primary focus of this study, some are of interest in that they engage with researchers (both domestic and international) to assist in the implementation of social surveys, and the collection of social science survey data on behalf of others. Some collect regular ‘public opinion’ type information that might also be useful to researchers. From the perspective of social science researchers, one of the most interesting such groups is HORIZON (*lingdian diaocha*) (See Box 1 for further details).

Another group of commercial institutions are involved in the marketing of Chinese data sets. Data tends to be more macro with the primary audience being the business rather than academic research community. Information might include for example, macroeconomic, trade, and investment data. Older census data or other survey data is sometimes also available through such agencies. These companies may be Chinese, or overseas companies with offices in China. They include:

- China Data Online, which offers access to census data including county level data for 2000 (<http://141.211.142.26/>);
- CEIC which holds Chinese macroeconomic, financial and industrial data (www.ceicdata.com);
- China Knowledge (www.chinaknowledge.com)
- China Statistical Data (industry, investment, prices, trade, import and export data) (www.china.org.cn)
- All China Marketing Research (known in Chinese as Huatong), a company set up by NBS and now fully privatised (www.acmr.com.cn/english/). Huatong carries out mainly business research and analysis, making use of its special access to NBS data.

Box 1: HORIZON Research Consultancy Group 零点调查

HORIZON is China's largest independent survey organisation. Established in 1992, it is headquartered in Beijing but has additional offices in Shanghai, Guangzhou and Wuhan. It has both commercial and non-profit, social activities. Approximately 90% of its work is commercial, while about 10% is related to social or social science issues, development, and public policy. The latter include several types of data collection activity:

1. *Surveys initiated and implemented by Horizon*: The data from such surveys is open access on a fee basis. They include for example public attitudes and concerns around issues such as SARS. An annual survey of quality of life has been undertaken for 8 years. Data is provided in a ready for use format at relatively low cost. Typically access to a data set would cost about \$1500.
2. *'Omnibus' or syndicate studies*: National representative samples to which modules / questions can be added as requested by different scholars who pay additional costs depending on the additional information to be collected. With a basic sample size of around 5000, the total cost for a survey is around \$100,000. This service is used by many Japanese scholars.
3. Responses to requests for specific assistance: e.g. research, literature review or background information on a particular research topic. Horizon can provide assistance in doing a background review of available data sources, making connections and introductions, assessment of quality of existing data; etc. This may be done for free, but depending on requirements costs may be up to \$2000.

Websites: www.horizon-china.com and www.horizon-key.com

3. Data and resources

i. Overview

Through the survey and administrative data collection systems described above, substantial quantities of data are collected in China, much of it of reasonably good quality. Most large scale nationally representative data sets are collected through China's National Bureau of Statistics. Major categories of data collected include:

- *censuses* (e.g. population, economic, and agricultural censuses)
- *national sample surveys* including 1% and 1⁰/₁₀₀ sample surveys of demographic change; labour force; rural and urban households; etc.
- *administrative data* reported by government line agencies, and surveys undertaken by government agencies. For example, Ministry of Agriculture (RCRE) longitudinal fixed point survey.
- *small sample surveys* undertaken by academic and other institutions. Many smaller data sets are collected, but only a few have nationally representative samples or are of sufficiently good quality to be worth pursuing further.

Examples of such data sets are described further below and details of specific data sets provided in Appendix 1.

Published statistics: Aggregate statistics and summaries of many of the data collected are available to the public in print and (in some cases) electronic format, through annual statistical yearbooks. These include the China Statistical Yearbook produced by the NBS, and other Yearbooks in specific areas published by relevant departments or Ministries – for example, Labour and Social Security, Agriculture, Population, Civil Affairs, and by different levels of government. The NBS now has an Information Center, with a public reading room containing all the Statistical Yearbooks, including many from sub-national levels of government; as well as a Consulting Center which can facilitate access to publicly available data.

ii. Data areas of particular concern to ESRC

This section provides information on data resources in the areas of population, labour, health, education and on a number of related social and economic issues. Further details for selected data sets are provided in Appendix 1.

a. POPULATION

Main sources of population data

The National Statistical Bureau is the main government agency responsible for collecting population data. The main instruments are the Population Census, the 1% Population Sample, and the Annual Population Sample Survey (see Appendix 1).

The Population Census (every ten years; 1982, 1990 and 2000). The census covers population by gender, ethnicity, age, cultural level, employment and professions, population movement, marriage situation, reproduction, birth-rates and death-rates.

A 1% Population Sample Survey is carried out between the main population censuses. It covers population size, births and deaths, migration, age, gender, social structure of different areas, citizen employment, income, expenditure, housing conditions, and energy use, and other basic economic information. It also covers basic information on health status, educational attainment, and marriage, as well as employment and social security data. The

population census and the 1% population sample survey are organised and carried out by statistical bureaux at different administrative levels.

The Annual Population Sample Survey. This is collected in three ways. It is filled in by individuals, households and community representatives. The individual provides personal information such as employment, unemployment, marital situation, reproductive situation, birth and deaths. Households supply information on household composition including gender, income for past year, births and deaths and migration. Finally, community data includes land type, number of households, births and deaths, transport, health, education, communications, land use and income. The sample is nationwide, with 1.2 million people surveyed. The survey uses a stratified, multi-stage, weighted approach.

The Population and Family Planning Commission (PFPC) collects administrative data through its representatives on the village committee (for rural China) and the neighbourhood committee (for urban China). PFPC operate a system for registering births and marriages, and require that a Population and Family Planning Basic Work Handbook is filled in at the local level. This data is reported up through different levels of government and aggregated. The PFPC publish some aggregated data on births, family planning and ageing. The key survey carried out by the PFPC is the National Family Planning and Reproductive Health Survey, 2001 which covers female reproductive status, contraceptive conditions and reproductive health needs (see appendix).

The Public Security Bureau also collects some population data, primarily related to household registration. This covers basic information about household composition, migration, employment and rural/urban change and social structure.

The China Population and Development Research Center (CPDRC, formerly CPIRC) was set up with support from UN Population Fund, it reports to the Population and Family Planning Commission (PFPC). It has the most comprehensive collection of population datasets, as well as an information centre, a website, and a Data User Service. The CPDRC has micro-data for a range for the following datasets, including the 1% Population Sample Survey:

Table 1: Datasets held by CPDRC

The China Elderly Health and Longevity Survey	1987	Beijing University, Population Research Institute, China Centre for Ageing Research
China Children's Situation Sample Survey	1992	PFPC
Survey of Migration in 74 Urban Centres	1986	CASS Institute of Population and Labour Economics
China Disabled People Sample Survey	1987	Ministry of Civil Affairs
Reproductive Health Sample Survey	1997	PFPC
Population Census, 1% sample of data	1982, 1990	NBS
China Birth Rate Sample Survey	1992	PFPC
Survey of Support Systems for the Elderly	1992	China Ageing Research Institute

Other research institutes and datasets:

The following demographic surveys were carried out by the CASS Institute of Population and Labour Economics (IPLE). Micro-data are potentially accessible, but details would need to be negotiated. Where there has been overseas funding, or sufficient time has elapsed, access is likely to be easier.

Table 2: Population datasets held by CASS Institute of Population and Labour Economics (IPLE)

Survey of Migration in 74 Urban Centres	1986
Sample Survey Of Elderly People Over 60	1987
Survey of Family Economy and Reproductive Situation in 10 provinces	1992
Family Trends Social Survey	2002, 2004
Survey on the Impacts of Migration on Rural Women	2005
Survey of Changing Marriage Practices in Rural China	2005

Other population related micro-data has been collected by the Institute of Sociology at People's University, the Research Centre for Contemporary China, Peking University, and Duke University (see appendix).

b. LABOUR

Main sources of labour force data

There are four major sources of labour related data in China:

- The labour force surveys of the NBS;
- The administrative data collected through relevant government line agencies: these include principally the Ministry of Labour and Social Security (MOLSS), but additionally the Industrial and Commercial bureau and Ministry of Personnel;
- Occasional sample surveys undertaken by research departments under MOLSS or other government agencies;
- Surveys undertaken by researchers in academic institutions.

In terms of *government statistics*, a relatively clear division of labour exists between the NBS and MOLSS in the types of data collected. NBS collects overall information about the structure of the labour force, while MOLSS reports data related to unemployment, social security and training, primarily through its administrative reporting system. Other government departments report employment and wage statistics for specific population groups (ICB for employees reported by registered private and small enterprises, while government and public sector employees are reported by departments under the Ministry of Personnel).

NBS survey data relating to labour includes the following:

Labour force surveys: The NBS Labour Force Survey was started in 1996. This was implemented only once a year, so was unable to capture most interim fluctuations. In 2006 new State Council regulations on labour force surveys were introduced with the aim of regularising labour force data collection, and a revised LFS was implemented in 2005-6. Although planned as a seasonal survey, it is currently being implemented twice a year due to insufficient resources and personnel. Variables collected include basic demographic information, education, and employment. The detailed occupational indicators used for the 5th Census in 2000 are used. (Detailed documentation of these categories are available in Chinese). From 2004/5 indicators of social insurance and security were added. Detailed data on hours worked are not currently included.

While the data is improving, problems were noted with the survey, including an inadequate sampling frame (currently the population survey sample frame is being used), and the lack of a sufficient data series for working out suitable weights. The sample cluster is regarded as too large. Unlike rural and urban surveys, the LFS is not fixed; that is, the sample sites change and there are no permanent enumerators, which raises the costs of training while reducing the quality of data collected.

Population surveys including the census, 1% population survey (every 10 years), and annual survey of demographic change. These surveys all include similar labour force statistics but vary principally in terms of sample size. Information collected includes the detailed employment categories introduced in the 5th census in 2000.

Other censuses, including the economic (incorporating the former industrial census), and the agricultural census. The *Economic Census* was first conducted in 2004 and includes among other information the number and distribution of employed persons across enterprises, their education and employment status. The second *Agricultural Census* is being carried out in 2006, and includes all units engaged in primary industry and rural household production, and information on employment and rural labour force mobility.

Rural and urban household surveys are implemented annually, and include basic data on labour force participation and employment information, as well as detailed household income and expenditure data (collected through household diaries). While there are problems with data, this remains the best source of nationally representative household level information (on a wide range of issues in addition to employment and income). However, access is generally limited.

Government reporting systems

The MOLSS reporting system at each level of government covers employment, unemployment, and training information for the urban population under the jurisdiction of the relevant reporting level of government. The basic annual reporting system includes the main economic indicators of enterprises, including unit employee data.

The Industrial and Commercial Bureau at each level of government provides several sources of data coming from private (*siying*) and small-scale or individual (*geti*) enterprise registration information. This includes information on employment in such enterprises. The Personnel Department at each level under the Ministry of Personnel reports information relating to government employees; wage data for government organs and public service units; and management and technical personnel in such units. These data therefore provide information on employment in government administrative and public service units at each level.

Unemployment and employment training data

Currently the main source of unemployment data comes from the MOLSS reporting system mentioned above. Indicators of urban registered unemployment are calculated from official unemployment registration data; this does not include state and collective 'laid-off' workers (not in receipt of formal unemployment benefits). Nor does it include rural migrants seeking work in urban areas. Alternative sources of unemployment data come from surveys which provide information on unemployment rates at a particular point in time among surveyed households. In general, estimates for such population or household surveys provide more accurate estimates of the true unemployment rate.

Data on labour costs

Basic data on labour costs are very limited: but some information can be obtained from existing surveys. The most common source is the 5 year investment and production survey, which includes detailed indicators of labour force investment costs based on a survey of costs and expenditures of large and medium size enterprises; in addition, urban and rural household survey data includes detailed household living expenses; from which labour force costs can be estimated.

Published statistics

Aggregate data (national and provincial level summary statistics) are available publicly as part of reporting systems: for example, on government websites (e.g. www.molss.gov.cn), through monthly or quarterly reports, and annual Yearbooks (all in Chinese). In some cases, sub-provincial data may be available in particular provinces through similar reporting systems and yearbooks. The Labour Studies Institute of the MOLSS produces an annual Employment Report (in Chinese) which includes some aggregate statistics, in addition to the Labour and Social Security Yearbook of the MOLSS.

A major new government initiative is underway to computerise administrative data systems in eight areas, including labour and social security. This would involve putting administrative data in a standardized electronic format, so that national standard indicators are readily available and information is consistent across all localities. The indicators to be made available however are primarily those used for government policy and administrative purposes (e.g. managing social security funds), rather than for research. Restrictions on micro-data mean that individual records cannot be released, and access to any part of such data would need to be negotiated on a case by case basis.

Government collected survey data

Additional surveys may be undertaken by government agencies on an *ad hoc* basis. In many cases, however, these activities are driven by short term policy concerns, are implemented quickly, and may not use good sampling methods. Some examples include:

- In 2002 MOLSS implemented a large scale survey on 'Labour employment and social security' covering 190,000 people in 66 cities in 2002. A key purpose was to understand the unemployment situation. However, the data have not been made public, few outside researchers have had access to it, and little analysis has been undertaken.
- The Labour Research Institute under MOLSS is currently undertaking a survey of youth employment with support from ILO. A recent survey of worker training was undertaken in 15 cities.
- From 2000 – 2003, the MOLSS collaborated with the NBS in surveys of rural migrants. The sample covered rural household labour with a sample size of 68,000.
- In 2005 MOLSS undertook a research project on migrant labour. However, this suffered from an unrepresentative sample, and collected only basic data, much of which is not useful to researchers.
- The World Bank collaborated with MOLSS in a small (two city) survey on training for laid off workers in 2002.

Other surveys implemented by the government, or in collaboration with other international agencies, could be identified, some of which (such as the 66 city survey above) would be valuable for research purposes. However, a range of concerns make MOLSS and similar

government institutions sensitive about wider utilisation of the data, and access is likely to remain extremely restricted. Additional sources of information may be worth pursuing further, including that obtained through industrial and enterprise surveys, or collected through the All China Federation of Trade Unions and Trade Associations.

Surveys undertaken by academic and research institutions

CASS, Institute of Economics

In collaboration with NBS, a nationally representative household survey on income inequality (CHIPS) was undertaken in 1988, 1995, 2002. This data has been widely used by scholars, particularly labour economists, in China and internationally. The first round is publicly available, while later rounds are being prepared for public use (see Appendix 1). Numerous publications describe the data in detail.¹ The sample is based on the NBS rural and urban household sample, and implemented by the NBS but with close supervision by the principal researchers, insuring good data quality.

CASS, Institute of Population and Labour Economics

IPLLE has undertaken several large surveys on labour and population related issues, including on migration and the elderly, dating back to the mid-1980s (see above and Appendix 1). Some of the earlier data sets were prepared for open access and housed in the University Service Center in Hong Kong². Others would need further work to be made available. Basic information such as the list of basic variables and questionnaires are available.

The key recent survey on labour force issues is the China Urban Labour Survey (CULS) first undertaken in 2001 and repeated in 2002. It covered 3500 households in 5 large cities and 3000 migrants. In 2005 the sample was expanded to cover 12 cities and 6000 households, plus 3500 migrants. The focus was on understanding urban poverty.

Other recent surveys include those on unemployment, migration, and social security. Some of the older data sets from IPLLE are publicly available through USC. Others may be available through direct negotiation with IPLLE. Newer data sets tend to be accessible only through direct collaboration with researchers at the Institute. While smaller in sample size than the CHIPS or NBS data, these data are generally regarded as of high quality.

Over the next few years several new sources of data on labour markets and migration will be collected. A large panel survey of migrants is also in the preparatory phase, led by Professor Li Shi (Beijing Normal University). While the plan is to make this public as soon as possible, in the first stages it will be used only by research collaborators.

Quality and accessibility of data

While it has limitations, the NBS survey data is still among the best available with advantages of large, nationally representative samples. Unfortunately it remains hard and costly to obtain, and is usually only made accessible through specific forms of collaboration. In general, all official administrative data is also of limited public accessibility apart from the published aggregate statistics. It may be possible to negotiate access or use of data, either with the NBS or other agencies, but this would need to be done on a case by case basis, and generally involves high transaction and/or financial costs. The best likelihood of being

¹ These include Griffin, Keith and Zhao, Renwei, Eds. (1993) *The Distribution of Income in China*, Macmillan Press; Riskin, Carl, Zhao Renwei and Li Shi, Eds. (2001) *China's Retreat from Inequality*, M.E.Sharpe; Knight, John and Song Lina (2005) *Towards a Labour market in China* Oxford University Press, among others.

² The basic agreement with USC included the provision of some support to IPLLE to clean and document the data, and a share of user fees.

able to use such data is probably through specific collaborative research projects, including with NBS and relevant departments.

The most readily available data are therefore smaller sample surveys undertaken by academic institutions: many international researchers have accessed such data, principally through participation in collaborative research with the relevant Chinese institutions. Other data sets (particularly where funded by international organisations) may become public access within a specified time frame. Currently a number of institutions are considering creating data resource centres (described in more detail below): one important such initiative will be the collaboration between the China Center for Economic Research (CCER) at Peking University with Michigan to house a number of data sets at CCER. Initially these would be older data sets, provided voluntarily by researchers, but the organisers hope that this initiative will lead to a stronger recognition of the benefits of sharing data and placing it in the public domain.

c. HEALTH

Major sources of data

There are two major sources of government data: administrative data compiled and reported in the Health Yearbook, and the National Health survey which is carried out every 5 years. The micro-data associated with the health survey is the most comprehensive and useful to researchers. However, even Chinese researchers claim it is very hard to get access to this data.

Administrative data

Administrative data is collected in standard reporting fashion. Facilities at county level and above report upwards, while at the higher level data is aggregated and passed upwards again. Reports are monthly, quarterly or annually depending on the type of indicator. National aggregated data is compiled in the publicly available Health Yearbook. A selection of key indicators are also reported in the NBS Statistical Yearbook.

The NBS Yearbook contains the following data: number of health and medical institutions, number of personnel, number of beds, patients treated and in-patients, major diseases as a cause of death, proportion of total deaths, incidence and cause of death by infectious disease and estimates of health costs. The Health Yearbook contains many more indicators; these are available online at www.moh.gov.cn/12.htm.

As with other sectors the yearbooks provide aggregated data, and give an indication as to what micro-data exist. The micro-data, however, is not in the public domain, and is very difficult to access; although parts may occasionally be made available through special project arrangements. Data below the county level may not be centralised in any case as this data is reported upwards and aggregated: lower levels do not necessarily pass on all the original data to higher levels.

Data quality is variable. Some researchers claim that utilisation data is good, but that financial data is more likely to be problematic.

The Chinese Centre for Disease Control (CDC) is a unit under the Ministry of Health which has its own administrative and reporting systems. It collects surveillance data for 37 diseases reported from township and county level. Again assessments of quality vary: for diseases with high morbidity the coverage is very good; for other illnesses reporting may be more sporadic, and the quality is less good, (for example, influenza). Many national experts collaborate with CDC to analyse this data. It now appears that CDC is more open than in the past to arrangements where students or researchers base themselves in CDC and

collaboratively work on data. A range of publications are available including Disease Surveillance Monthly (available by subscription on the internet).

Government surveys

Health Sample Survey (1993, 1998, 2003).

As noted the Health Survey is the most useful broad coverage dataset in the health sector. It is managed by the Health Information Centre under the Ministry of Health. On balance researchers are positive about the quality of the data, and the level of description available (for more details see appendix 1).

National Mortality Survey (1973, 1990, 2005) is carried out by MOH and the Ministry of Science and Technology (see appendix 1).

Data collected by research institutes

Several significant academic surveys have been carried out in the health sector. The most promising in terms of access are those where there has been overseas funding and therefore some requirement to make data open access (when a specified period has elapsed). Many smaller surveys also exist, but these remain with the original researchers and no central record or database exist. In some cases researchers will share their micro-data, often additional resources will be necessary to put into a format where it will be useful to other researchers.

Two datasets are particularly noteworthy, both because of their longitudinal nature, and because the data are in the public domain. The China Health and Nutrition Survey is one of the most widely used micro-data sets on China. (For details, see appendix):

Table 3: Longitudinal Population datasets

The China Health and Nutrition Survey	1989, 1993, 1997, 2000, 2004	UNC Population Center/Ministry of Health
The China Elderly Health and Longevity Survey	1998, 2000, 2002, 2004	The Center for Healthy Ageing and Family Studies, CCER, Peking University

d. EDUCATION

Government education statistics

Education statistics are collected by the Ministry of Education in coordination with the National Bureau of Statistics. Within the MOE the Statistics Work and Management Reporting Group (under the Development Planning Division) oversees management and publication of statistics. Permission to use MOE data would have to be sought from the Development and Planning Division (for Chinese scholars), and the International Division for foreign organisations. Micro-data cannot officially be shared with researchers and others (domestic or overseas). However, as with other ministries it may be possible to negotiate access to small parts of micro-data sets with particular parts of the MOE (acting in partnership with a Chinese institution).

Published data

The MOE yearbook and the NBS yearbook present aggregated provincial data. This data gives an indication of what micro-data datasets exist. The reporting system is based on schools, colleges and other educational units report to townships (for basic education) who report to county, and so on to the provincial level. Provincial educational yearbooks contain

data aggregated at county level. It is not possible to get all county level data aggregated in one place. There are 20 different reporting lines: for example, the primary school reporting system and higher educational statistical reporting systems.

Data collected through the main MOE reporting system include number of schools, enrolled students, new students enrolled, number of graduates, number of staff and workers, full-time teachers; school construction and conditions; sources of and outlay of education funds; education expenditure from the state budget. This data is compiled in the Education Yearbook.

There are also 19 additional reporting lines which go to particular divisions within the Ministry of Education. Some report every six months, some annually. These include education expenditure reporting to the Finance Division of MOE, basic construction to the Development and Planning Division, higher education graduation and employment statistics to the Higher Education Student Division. These divisions are important for access as it is usually necessary to negotiate with the specific part of the ministry that holds the data; discussions with an umbrella body are not likely to be so fruitful.

Basic education data include categorisation of schools, numbers of students in different years, numbers by age of entrants, enrolled and graduating students, changes in student numbers, data on teachers and other staff (age, gender, educational achievement, teachers by subject, changes in numbers). Educational facilities data, education expenditure statistics and higher education statistics are also collected.

The ministry has other data where they have commissioned one-off surveys from researchers. There is no centralised list available of which surveys have been carried out and what data has been collected.

Data quality and usefulness to researchers

Interviews with Chinese researchers suggest that administrative data is likely to be of variable quality. Some basic information is likely to be fairly accurate, but where particular data are linked to allocations of resources there are more likely to be distortions. As with other areas researchers comment that the indicators are often so general and poorly broken down that they are of limited use for many research purposes.

Some Chinese researchers at key institutes like Beijing Normal University Education College have access to parts of the data, particularly where they are working with the Ministry to help with analysis.

Data collected by Chinese universities and research institutes

There are a range of datasets collected by Chinese researchers working in the education field (see appendix). These are generally held by researchers and not available in databanks. Many would need formatting and cleaning prior to use by others. The degree to which researchers would be willing to share datasets varies. In the case of Beijing University, Education College, Educational Economics and Finance Department there are strict regulations about sharing micro-data – special agreements would need to be arranged. Generally there is some interest in the idea of a databank, but researchers also recognise the difficulties in establishing such an arrangement.

Some researchers have close links to the MOE, as noted above. Researchers at the Education College of Beijing Normal University have added additional questions (or a full questionnaire) on education to the Rural Household Survey. There is no central resource listing which other institutes have done this and on which subjects.

China National Institute for Education Research is a research institute under MoE. They are responsible for a 66 county survey, which includes data on drop-out rates and children not attending school. The lead researcher is Ma Xiaochang. This is a very sensitive area, and so it is difficult to access this micro-data.

Table 4: Major education data sets collected by academic institutions

Urban Residents Education And Employment Situation Survey	2005-2007	Education College, Beijing University
National Higher Education Graduates Employment Situation Survey	2003 and 2005	Education College, Beijing University
Rural Compulsory Education Baseline Survey	2006	Education College, Beijing University
National Higher Education Students	2002	Education College, Beijing Normal University
National Higher Education Political Thought And Public Curriculum Implementation Situation Student Survey	2005	Education College, Beijing Normal University
Compulsory Education Period Student Family Education Expenditure Survey	2005-6	Education College, Beijing Normal University
National Ordinary Higher Education College Special Teacher Survey	2006	Education College, Beijing Normal University

Beijing Normal University Education College also have a plan to develop a databank of various educational datasets, and discussions are currently underway with organisations in Macao and Taiwan. Researchers commented that support is needed with formatting and preparing data for sharing.

Other institutes with education micro-data sets include Shanghai Education Science Research Institute who have been responsible for some data cleaning for the Ministry of Education. This institute has some education expenditure statistics based on administrative data. Huadong Normal University have also carried out a survey on curriculum development.

Data collected by researchers and organisations based outside China

Gansu Survey of Children and Families (GSCF) is a 6 year longitudinal, multi-level data set, with a sample of 2000 children in 20 rural counties, collected through a research project led by Emily Hannum, University of Pennsylvania and Jennifer Adams, GSCF. Good documentation on sampling and weighting and the original questionnaires are available, as is some of the early micro-data is publicly accessible; data for 2000 is available via the China Survey Data Network. (See Appendix).

Development agencies and UN organizations also support the collection of data; some of this is quite small in scale but not all. UNESCO is assisting China (via the Chinese National Commission for UNESCO) with its Education For All monitoring exercise, which currently covers five provinces (Guizhou, Gansu, Henan, Yunnan and Hebei). Micro-data is being collected, and reports based on this data will be released in 2007. It is unclear at present whether it will be possible to access micro-data and on what terms. It is likely that only aggregated data will be available.

UNICEF China is working with the Planning and Statistics Department of the MOE on Education Management Information Systems. The Gansu Basic Education Project is another potential source of micro-data. GBEP was a DFID funded project managed by Cambridge Education Consultants and Gansu Provincial Education Bureau (for more details see

appendix 1). The Basic Education Western Areas Project is a World Bank loan project for which some micro-data is available.

e. CRIME AND PUBLIC SECURITY

Very few statistics on crime and related issues are publicly available. Selected data on crime and public security issues collected by the Ministry of Public Security and other judicial or investigative organs are reported to the NBS (as with other administrative data). Selected data available are collated and summarised in the Yearbook. The main categories presented are summarised in the following table:

BOX 2:
Main categories of public security information presented in China Statistical Year book
Basic statistics on lawyers, notarisation and mediation (number of personnel, offices, cases handled, etc.)
Number of civil disputes mediated by type (e.g. family, housing, business related)
Labour disputes accepted and handled by labour dispute arbitration committees
Criminal cases registered in public security organs (e.g. homicide, robbery, etc.)
Offences against public order handled by public security organs
Information on traffic accidents and fires
Cases under investigation and appeals handled by procurator's offices
Arrests of criminal suspects and defendants under public prosecution
Trials accepted and settled by courts (including criminal, civil, economic and administrative cases).

Due to the sensitivity of crime and security data, the NBS has limited power to require the relevant government organs to provide information. Furthermore, there is little validation of these data. Such information needs to be interpreted within an institutional context where there is restricted openness about the proceedings of the security and judicial organs. Any researchers wishing to use such data for research purposes would need to find their own ways to cross-check data. From conversations with experts in the field, it is also clear that there is very little quantitative work by researchers in the legal field within China using such data. Given the sensitivity of these kind of data, and thus the very low probability of access to more detailed information, we did not pursue this aspect in depth.

What does increasingly exist however are surveys relating to public perceptions e.g. of rule of law, public security and stability, access to and use of legal channels of redress, as well as some documentation of use of channels such as the system of 'Letters and Petitions' to the government and collective incidents of social unrest.

Researchers wanting to work on these issues would however need to invest significant time and energy in understanding the issues, their sensitivities, the institutional environment, and the sources of data and possible biases. Researchers with the required expertise and connections to pursue such investigations would be unlikely to rely on publicly accessible data. We therefore recommend that there is little value to ESRC in pursuing this avenue of enquiry further at this stage.

iii. Other data resources

a. Government economic and social data

Social welfare and assistance: A number of data sets related to social welfare and related issues are being collected by the relevant government departments, particularly the Ministry of Civil Affairs (MOCA), and by the Institute of Social Development and Public Policy (ISDPP) at Beijing Normal University. Some of these data sets are collected directly by ISDPP; others are in collaboration with MOCA, either through administrative reporting systems, or surveys designed for a particular purpose, and with other agencies (including the Asian Development Bank and the World Bank). The ISDPP has a plan, supported by Beijing Normal University, to establish a 'Social Analysis Information Center' which would provide limited access to the data resources housed there. (See Appendix 4 for further details and a list of available data; selected data sets are presented in more detail in Appendix 1).

Gender: The aggregation of data in the reporting process, and the format in which survey data are made public, means that there is relatively little easily accessible information that is systematically disaggregated by sex. This limits the data available for undertaking gender analysis across a range of social and economic issues. Without access to micro-data, there are serious limitations on the capacity of academic or government and policy researchers to analyse a range of key issues from a gender perspective (including for example, labour force participation and employment, school drop outs, and health care).

The ACWF is leading an effort to require that more statistics are reported in a sex-disaggregated form. Currently their efforts are focused on key variables that would be useful for monitoring China's commitments to addressing discrimination against women through CEDAW, as well as progress towards the MDGs. The Women's Studies Institute of ACWF in particular is spear-heading this work by identifying key indicators linked to MDG reporting obligations in order to build a stronger awareness of the need to provide gender disaggregated data.

Many statistics by sex are however sensitive domestically, or serve to reinforce sensitivities around other issues: these include for example, school drop out rates, and infant and under-5 mortality.

Social statistics for monitoring MDGs: Monitoring progress towards the MDGs is currently a significant issue for many development agencies, both UN and bi-laterals. UN agencies are collectively trying to work with NBS to ensure the collection of relevant data for monitoring, but obtaining data with a sufficient level of disaggregation (at sub-national levels, as well as by gender or income groups) remains challenging.

Overall, while many public statements are being made relating to improving social statistics, funding remains very unbalanced: within the NBS the departments responsible for social statistics are relatively poorly funded compared to those responsible for the collection of economic data.

Fixed Site Rural Survey of Ministry of Agriculture (FSRS): This survey, implemented annually in 320 villages by the Research Center for Rural Economy under the Ministry of Agriculture is worth particular mention. Although relatively small in size, its major value lies in being a longitudinal study, with a panel collected since 1984. It covers 22,000 farm households; data is collected for over 1000 social and economic indicators on rural households, villages and village enterprises. While it has limitations as a panel (due to the rotation of households), it is widely used by researchers often as a base with modules added on specific topics. Access needs to

be negotiated on a case by case basis, and usually involves a collaborative research project and some payment for use of data.

b. Other non-governmental data and resources

The Chinese General Social Survey is an annual survey based on the International Social Survey Programme and the European Social Survey model. The aim is that it should be made publicly available. The survey was organised by Hong Kong University of Science and Technology, led by Bian Yanjie (2003), and Institute of Sociology, People's University, led by Li Lulu. It involved a sample of 10,000 (6000 urban and 4000 rural residents). (see www.chinagss.org)

The 2004 survey contains additional data on migration. The 2005 survey adds variables on community organisation, government functions and elections. The 2006 survey includes data on class structure and identity. The data for 2003 should be publicly available in 2007. Researchers at People's University have approached the National Social Science Foundation (China) to seek core funding for the survey. This is currently being discussed.

The Research Center for Contemporary China (RCCC) at Beijing University, under its founding director, Shen Mingming, has carried out a number of social science surveys, and are centrally involved in building social science survey research capacity in China. They have investigated building a databank of social survey datasets. Their flagship survey is the Beijing Area Study, an annual survey (since 1995) of 1200 urban residents. RCCC also carry out the National Value Survey, Environmental Awareness Surveys, and the China component of the World Mental Health Survey. (see www.rccc.pku.edu.cn/en/intro_1.php).

c. International databanks

The China Data Center (CDC) at the University of Michigan houses the China Data Archive, a databank developed through a partnership with a company under the NBS. The databank includes 3000 electronic books of aggregated macro-data down to township, county, province levels, including some neighbourhood data. CDC have a membership facility which allows access to census and survey data (including from the 2000 census). Leading US institutions including the University of California at Berkeley, Stanford and Columbia all have agreements to access this data. According to CDC administrators, despite good relationships with NBS (and the link via the NBS company) they have been unable to access micro-data for the censuses and NBS surveys. The researcher responsible for managing the CDC, Mr. Shuming Bao, raised the possibility of exploring options for a partnership arrangement between ESRC and CDC, which might perhaps result in data being made more widely available for UK/European researchers

The China Survey Data Network (CSDN) is a new initiative, also with the involvement of CDC, University of Michigan, in partnership with Beijing University. CSDN aims to bring together original small-scale survey data. The principle is that researchers donate data and in return have complete access to other datasets held on the network. Beijing University have the responsibility of sourcing datasets from researchers based in China; Michigan is responsible for data collected outside of China. So far getting data from Chinese institutes and other units has been difficult. The strategy of the organisers is to start by making older or already public data more

easily available, in the hope of encouraging greater willingness to share more recent data over the longer term.

This initiative has just been launched (late 2006), and is partially supported with a small initial grant from the Ford Foundation. It is one of a number of new activities in this area, aimed in part at making data more widely accessible to researchers, but also at changing the culture and attitudes to sharing data.

Table 5: Data sets held by CSDN

Gansu Survey of Children and Families	2000-5
Delinquency in a Birth Cohort in Wuchang District, Wuhan, China	1973-2000
China Housing Survey,	1993
Survey on the Support System for the Elderly in China	1992
National Sample Survey on the Situation of Children	1992
Fertility Sampling Survey	1992
Fertility and Contraception Survey	1988
Survey on China's Aged Population	1987
In-depth Fertility Sample Survey	1987
Survey on the Disabled	1987
Sampling Survey on Migration in 74 Towns and Cities	1986
One Per Thousand Sampling Survey on Fertility	1982

University Service Centre for Chinese Studies, Chinese University of Hong Kong (USC) has one of the most extensive and accessible collections of a great variety of materials on contemporary China. This includes: 1300 constantly updated regional and national statistical yearbooks; a large collection of provincial, city, village annals, including many volumes on special topics; a fast growing collection of carefully classified and readily searchable electronic data sources which are available on its website.

Table 6: Data sets held by USC

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

Various metadata are available online, including for example, the questionnaire for the Rural Household Survey 2000.

The China Archive at Texas A & M University has a relatively small archive though it is looking to expand. At present the collection is fairly limited. The following datasets are available to registered users: 1% of the 1990 Population Census; 1993 Survey on Social Change and Social Mobility; a survey of children in Changchun (1987); a Survey of Political Participation in Beijing (1988). The China Archive also has several datasets relating to land use/GIS/landsat data.

d. International agencies and other funding agencies

Organisations such as the World Bank collect many different types of data in China. The Department for International Development, UK, has also supported the collection of data linked to its projects. For example, the Poor Rural Communities Development Project supported and managed by the Government of China, World Bank and DFID has involved the collection of comprehensive poverty assessment data in project counties in Sichuan, Guangxi and Yunnan. Other examples for basic education were cited above. Such data might be of interest to researchers, but access would require discussions with the relevant funders or government counterparts. The World Bank is starting to take steps to make some of the data it has collected more readily available to a wider group of users. There is however a large amount of 'grey literature' relating to many development issues in China, much of which has value to researchers, which could potentially be made more accessible to a wider audience.

It is worth noting here that organisations such as international development agencies are users as well as funders of data. In fact, they support data collection largely because relevant information does not exist, or (as is often the case in China) exists but is not accessible. Even for major agencies such as the World Bank, access to NBS data must be negotiated on a case by case basis, with specific departments, involving very high transaction costs. These are compounded by terms of use, which often involve getting access to a small part (one province, or a sub-sample) of the data to analyse and produce routines, which the NBS then runs on the whole data set. While the WB aims to negotiate a more favourable agreement for collaboration, including providing technical support for anonymising individual records, such an arrangement seems far from being achieved.

Similar difficulties arise in the case of a long-standing collaboration between Statistics Canada and the NBS, with the involvement of CIDA (Canadian International Development Agency). The first phase of the Statistical Information Management Program ran from 1996-2003 and was followed by a second phase which aims to assist the NBS to improve its economics statistics program. Through the project, CIDA hoped for some reciprocity in terms of greater access to data but this appears on the whole not to have been achieved to date.

iv. Concluding remarks

As we have discussed in this section, administrative data and most NBS micro-survey data are very difficult to access, except as published tables or data series in highly aggregate form. While micro-data is sometimes more easily available for specific localities at the sub-national government level, obtaining the data would require significant efforts to negotiate, and would be prohibitively costly in terms of time if the objective was to construct a nationally representative sample.

The best chance for making progress to facilitate access to some good quality micro-data is probably through focusing on key areas where such data already exists largely outside the formal government statistical system. The areas of micro-data reviewed above include those where the richest resources exist: population and demographic change, labour and related issues, and some aspects of health. Such areas have received significant support from international agencies and donors, with both government and independent academic or research institutions collecting good quality survey data. In some of these areas (notably Population) efforts have already led to the establishment of data archives, which provide initial evidence of what is possible as well as the limits of such systems in China. In other areas (notably Labour) initiatives are currently being undertaken to place more data in the public domain. Many institutions now recognise the importance and benefits of making data more accessible, though significant obstacles remain to the realisation of these goals.

4. ISSUES OF ACCESS: OPPORTUNITIES AND CONSTRAINTS

The difficulties in access to and use of micro-data resources in China have been highlighted in the preceding sections. This applies particularly to government statistics and surveys, but also to non-governmental sources of data. Several factors contribute to this situation, including: legal restrictions; a fear that others may use data in inappropriate ways; a culture among researchers of not sharing datasets; the cost of acquiring data and of making it accessible to the public; the absence of good metadata; and (for international access) language barriers and the cost of translation.

i. The legal environment

The legal environment will remain a critical issue for those seeking to access to micro-data or to place data in the public domain. There are four main types of regulation which place limitations on access to original data:

- rules on data protection and the anonymity of data subjects
- rules on state secrets and endangering national security
- rules on who can carry out surveys
- rules on releasing data to foreigners.

a) *Rules on data protection* specify that rights to anonymity have to be respected; this is often interpreted to mean that micro-data cannot be released, and their use are closely supervised and restricted. The NBS is also responsible for interpreting the Statistics Law. Article 15 of the Statistics Law states: ‘Single item investigation data concerning any individual or his/her family shall not be divulged without the consent of the said person’³. This reason is commonly invoked to explain why micro-data – particularly government micro-data – cannot be shared. Greater use of techniques for anonymising data may offer more possibilities for use of micro-data without falling foul of data protection rules. The World Bank is exploring options with Chinese partners in this area. It is also the case that parts of the NBS household surveys and the census (with identifying variables omitted) have been made available to outside organisations which suggests that the protection of individual identities is not an insuperable problem. However, in the short term it seems likely that these rules will continue to be applied as justifying a default position of no release of government micro-data; critically, it is the NBS which interprets the provisions of the Statistics Law.

b) *Rules on protection of state secrets* are covered by the Law on Protection of State Secrets (1988) and the subsequent Implementation Guidelines (1990). Again infringement of these rules can be invoked as a reason for not sharing micro-data. Certain areas of information such as some economic and social data, and some crime and public order data, are inherently sensitive. However the broad way that national security legislation is framed means that apparently less controversial areas could also be included. The Implementation Guidelines specify that it is illegal to: ‘impact on national unity, popular solidarity, and social peace and stability’ or to ‘undermine national economic, scientific or technological strength’. Exactly how state secrets or national security might be undermined is often a grey area. In practice the lack of specificity, and fear of unfavourable interpretations of broad categories like ‘undermining social peace and stability’ can lead to risk-averse behaviour both by government officials and researchers.

c) A third significant area of regulation concerns who has a right to carry out surveys. The relevant rules make clear that foreigners cannot carry out surveys in China unless they are working with a Chinese partner, and that the Chinese counterpart needs to be given

³ Similar stipulations are also found in the Regulations for the 5th Population Census, Article 24.

official approval for this function⁴. Any proposed surveys (including the full questionnaires) need to be approved by the NBS. Further to this, the NBS has an oversight role for all survey and statistical work carried out in China. This means that Chinese organisations should also seek approval from the NBS for surveys they carry out. In practice many do not, viewing it as a formality that is unlikely to be enforced. This is generally the case, but such researchers are also aware that if they share their data or make it more widely available (particularly to foreigners), they are potentially putting themselves at risk, since many of the original surveys were not approved.

d) There are also regulations requiring that release of micro-data to foreigners needs formal approval, including for academic survey data⁵. Such approvals can be difficult to obtain, which acts as a disincentive even to make the effort. Nevertheless there is also a gap between what is specified in regulations and actual practice. Researchers will release micro-data where they think that it is unlikely to lead to problems, but they are also aware of the risks. Any process for widespread sharing of data, such as construction of databanks, would need to address this issue directly to protect the researchers and providers, as well as the users, of data.

These then are some of the key legal dimensions. It should be noted that there is quite a lot of uncertainty around these issues. It is often repeated that the NBS and administrative micro-data cannot be shared as if this is a rule in itself, rather than a consequence of data protection or national security regulations. Getting a precise answer on whether this is the case is difficult.

Finally, different academic institutions also have their own rules on sharing data. Beijing University College of Education College for example has its own internal regulations which specify the terms and conditions of access for researchers outside the college to original data collected through College projects.

ii. Making data public: Attitudes and cost

Data as a public good: There are a lot of good datasets collected by Chinese researchers. However, unless they are part of an international project where data sharing is built into the project design they are generally not in the public domain. There is no equivalent of the UK Data Archive in China. Even government collected data is not seen as a public good. Different departments within different ministries guard their datasets fiercely (including from each other) and within certain constraints use them as a resource to generate income for their particular section. Research institutes are also likely to view datasets in a similar way. This means that efforts to encourage sharing of datasets have tended to be unsuccessful: researchers and institutes are unlikely to freely contribute their data unless they see very clear benefits for themselves.

Attitudes to information: As noted in the section on legal issues there are many grey areas in relation to what counts as sensitive information. This creates a broader problem, namely, that government officials and researchers tend to be risk-averse in sharing information, for fear that it will be used by others in ways that they cannot control and which might cause

4 Implementation Guidelines, Article 35.

5 Statistical Information Services Management Temporary Regulation (2005). According to Article 19 'The supply of statistical information products or services to foreign organisations or individuals should be reported to departments at province level or above responsible for the administration of statistical information and approval obtained'. Article 26 specifies that any organisation breaking these rules will be fined and the responsible person in the organisation will be subject to administrative disciplinary measures.

political embarrassment. If analysis of government micro-data were to contravene some aspect of official policy, for example, this would be viewed as a serious issue. Some data are politically sensitive: for instance, school dropout rates or unemployment data. Researchers can to some extent present their primary data in academic and government policy communities quite freely, but if data were to be presented in the media and appeared to contradict government policy, or to be highlighting unwelcome social trends, then there could be negative repercussions for the original researcher, even if it were a secondary user who directly caused the problem.

Researcher attitudes: As noted above, in China the academic culture appears to place little value on the sharing of datasets to facilitate research by others. There are exceptions, and many researchers bemoan the current situation, but attempts to develop voluntary databanks have to date not flourished. There are several reasons for this:

- Data sets are often seen as a private resource.
- Researchers often lack the time and resources to make data available to others. Informants expressed concern that if they made their data available, they might end up spending excessive amounts of time answering queries.
- Researchers are also concerned that others will use data sets in ways that are politically sensitive, and which may cause problems for the original researcher or their organisation.

Lack of good meta-data: Given that there is not in general a culture of sharing data, and that this is not mandated by Chinese funding sources, there is often a lack of good meta-data pertaining to different datasets. There are exceptions, but generally little information is available beyond, for example, a brief description of the focus of a survey and the sample size. This lack of information makes it difficult to form a clear idea about the strengths and weaknesses of many datasets. Any meta-data that exists is also likely to be in Chinese.

Translation and preparation of data: Data sets and any associated descriptive materials are almost certainly only in Chinese. The costs of translating into English are likely to be high. Given practices around sharing data it is also likely that time would need to be invested in preparing data sets to make them useful to others. Chinese institutes may be willing to do this, and many see this as a useful activity, but there will be costs.

The cost of obtaining data: For government collected micro-data costs of acquisition of even small parts of a data set are likely to be very high. Since this access needs to be negotiated on a case-by-case basis, and usually as part of a collaborative project, it is not possible to be precise about the costs of acquisition in advance of concrete discussions about specific data.

iii. What the ESRC can offer China

Another set of factors relevant to UK attempts to acquire Chinese micro-data relate to Chinese perceptions of what the ESRC might bring to future partnerships. Finance is one consideration, but there are others:

- Accessing data for a specific project is viewed differently from accessing data with a view to making it available to others, as ESRC would want to do.
- There is interest among Chinese research institutes in ESRC expertise in data management, data preparation and a range of other services offered by ESDS. Training courses and sharing of training materials for example could be well received by some Chinese organisations.
- Chinese government officials and researchers expressed limited interest in UK specific datasets. This is significant in that it means that any offers of reciprocal

arrangements limited to information about the UK are not likely to be of great appeal. Reciprocal arrangements with the EU may, however, be of more interest, and it may be worth the ESRC thinking about opportunities in this regard.

To summarise, this section has suggested that accessing and using micro-data in China presents several challenges which should not be underestimated. However, while the overall environment for obtaining micro-data is difficult, it is still possible to move towards meeting strategic objectives if certain conditions are met. These include building the right relationships, identifying appropriate entry points, behaving sensitively, and offering something that provides benefits to both sides. Individuals and organisations in China are adept at finding ways of negotiating their way through a mass of complex and sometimes ambiguous regulations if they see adequate benefits. Constructive partnerships can also develop which make it possible to do things that, at the early stages of collaboration, might have been dismissed as too difficult. The final section draws together reflections on what opportunities exist to move forward constructively and makes some suggestions for possible next steps.

5. What does this mean for ESRC? Conclusions and recommendations

i. Data availability and access

- Almost all micro-data in China is difficult to access – statistical data, administrative data, and even data collected by researchers. The major exception is research survey data involving Chinese academics where there has been international funding and where open access to the data has been stipulated by funders and implemented. There are a range of useful datasets that fall into this category.
- The current environment however is not conducive to direct access to significant data sets or archives. This is especially the case for nationally representative micro-data sets collected by the National Statistical Bureau (NBS) – such as censuses and annual household or labour force surveys. But it also applies to most other micro-data.
- There are always exceptions – researchers access some NBS data, or administrative data and micro-data collected by Chinese researchers. However there is no single formula that explains how access is obtained. Discussions for this study suggest that in the absence of a concrete project proposal it is hard to get a firm indication of what might be available on the Chinese side. Where government data is accessed it is almost always only a small part of the whole dataset, and emerges from a specific collaborative project. Buying copies of any micro-data outright is generally unrealistic (especially complete datasets).
- The availability of data varies by subject area, but there is quite a lot of information collected, through multiple channels, much of which is of reasonably good quality. The quality of data tends to be better where funding has come from international sources or where collaboration or support from international organisations have been involved; such data often tends to be more readily accessible, often due to the requirements of international funding agencies. One of the earliest such fields is population, family planning and reproductive health issues; others include labour markets and health. For domestically funded social science projects, funds are often insufficient to undertake significant data collection, and when this is undertaken samples tend to be small or of poor quality.

NBS data

- The Population Census, the 1% Population Sample Survey, the Labour Survey and the Rural and Urban Household Surveys are the most useful data sets collected by NBS. Officially it is not possible to access the micro-data for these datasets. It may, however, be possible to negotiate access to subsections of the data, this would require discussions with the *specific sections* of NBS responsible, and may involve collaboration with a Chinese third party with good relationships with NBS.
- The fundamental question is whether there a cost and time effective way to access NBS data. To date, the experiences of major funders (including the World Bank and CIDA) or researchers is not positive. While some access has been provided, this is essentially negotiated on a case by case basis, with specific departments within NBS. Data tends to be expensive, and to involve high transaction costs. Alternatively, it depends on individual relationships and collaboration, usually developed over a long period of time. There is no obvious way to negotiate a general agreement or MOU that would cover multiple data sets, departments and users, and would provide access to more than the already accessible aggregate data. This does not mean that ESRC should not hold discussions with NBS, expressing their interest in data access,

and exploring whether ESRC has resources and capacities which would provide a basis for a relationship with NBS. This could provide a starting point to other discussions. However, this would be an investment in building a long term relationship; short term pay offs in terms of access to data would be highly unlikely.

Administrative data:

- The general consensus is that most administrative data collected through government reporting systems is of limited value to researchers. Access is extremely limited, especially to micro-data. The exceptions within the government system are some sample surveys, and particularly repeat surveys or longitudinal panel studies.
- One of the most widely used is RCRE's fixed point survey of 320 villages. A number of researchers have used this data, usually in conjunction with collaborative research projects which have funded additional modules or questions on specific topics. RCRE may be willing to come to some kind of data sharing agreements, but would probably require significant resources to cover costs of managing the data; as well as financial income for the Center. The Ministry of Health commissioned Health Survey is also a valuable resource. It is difficult to access, but joint research projects allowing access to some of the data might be possible.
- Apart from these specific opportunities, which could be a case study for exploring the options in greater depth, in general the restrictions on government data, and the financial incentives for institutions involved in its collection, mean that these are unlikely to be fruitful avenues to pursue.

ii. What can be done?

At a general level, the key challenges that currently exist in China to improving access to good quality micro-data also point to areas where there is scope for collaborative ventures with Chinese institutions:

- Attitudes to making data a 'public good': even among scholars, the challenge will be to foster an environment where data is seen as a public good, where the benefits of greater openness and sharing are recognised, and where collectors of data have adequate incentives to make data public. Longer term benefits may come in increasing quality of data, and greater recognition of researchers and institutions that collect good quality data.
- Strengthening the capacity to clean, document and manage data sets so that they can more easily be made available; and investing in the institutional infrastructure such as data archives for easy accessibility.
- Reducing the risks and providing incentives for open access data and more collaborative research projects: one possible mechanism is of course through funders who generally have most leverage in terms of specifying terms and conditions for making data public.

Addressing some of these challenges will involve long term efforts. However, some more concrete and immediate steps can be initiated that should promote greater access over time. These might include:

- Negotiate arrangements where data is not taken away but analysed on site with Chinese assistance; or where part of the data is obtained to test and develop a model which is then run by Chinese researchers on the entire data set. CPRDC

for population data, the Health Survey and the RCRE fixed point survey are useful examples. It may be possible to negotiate agreements with such organisations where a certain number of activities are allowed over a specified period of time (say five years).

- In terms of academic surveys – with the exception of internationally-funded surveys – there is an urgent need for work on the step prior to the construction of large databanks. Researchers and institutes could benefit from support to document what micro-data they have with supplementary descriptive information (in a consistent format). This kind of information is currently woefully lacking. Several key institutes have expressed an interest in undertaking such work. There could also be support for preparing data in formats that are useful to other researchers. Some of the older (and perhaps smaller) datasets might then be accessed at low cost, or offered to ESRC. Chinese researchers express concerns that sharing data may lead to additional work for them in answering queries. Arrangements would need to be made to manage this pressure.
- Make better use of the category of datasets that are already public access. Look at some of these which are subscription access etc (often via the websites of international organisations/universities). Some of these could potentially be linked to ESDS.
- Make better use of development agency funded data collection, some of this data could be made public access and is potentially useful. World Bank have useful datasets, but not all are available to date.
- Explore collaboration/reciprocal arrangements with others who have already started to build archives of Chinese micro-data. Michigan who manage the China Data Archive, and Chinese Survey Data Network are the most promising interlocutors.

Work with Chinese institutions:

- Create an easily accessible document listing very basic information and contact details about more easily available data, including any current or previous users. This would provide scholars with basic information and a first point of contact when trying to access and evaluate data for specific research topics.
- Identify and work with institutions and individuals (principally in the academic research community) who are concerned with making data more open. A first step to establish relations could be a basic introductory training course on the technical issues for making data available. (Possible institutions include: CCER, RCCC, BNU, CASS)
- Continue to support the relationship with CASS International Office. Provide support to CASS as they construct their own databank of datasets. Be aware that these are likely to be limited to those within the CASS system and not comprehensive.
- This could then be followed by provision of more targeted technical support to organisations which are creating (or already have established) data archives for public access.
- Identify one (or more) such organisation to create a collaborative data archive or centre, through which more specific or open data sharing arrangements might be negotiated.

Work with funders:

- The requirements of funding agencies appear to be one of the key mechanisms for ensuring that data collected by researchers becomes widely available. These may be both Chinese or international funding agencies.
- The closest equivalent organisation to the ESRC in China is the National Social Science Foundation. At present the NSSF places few demands on researchers in relation to sharing data. It is also not the most significant provider of funding, as ministry or overseas funds can be much larger than NSSF grants. While we were unable to get access to NSSF, it may be possible for ESRC to convene an exploratory meeting to assess the level of interest in a potential future collaborative arrangement. This might have broad objectives of improving the quality of projects funded and data collected. Possible options for consideration could include: joint activities and sharing of information about ESRC work as a funding agency; how to assess and evaluate proposals; how to ensure the quality of data collection; technical assistance around terms and conditions, and skills needed to make data public. It may be possible to arrange meetings to discuss possible options through CASS in association with the 2007 Conference.
- As an exploratory initiative, it may be worth considering collaboration with NSSF on a specific project. Discussions around the Chinese General Social Survey (currently managed by the Institute of Sociology at People's University, but with a proposal under discussion for core NSSF support) is one way in, but would need to be handled sensitively. Some kind of partnership around the CGSS (if NSSF do agree to support it) may be a viable modality for building a longer-lasting relationship.
- In discussion or collaboration with other funders, it may be possible to consider joint activities to support the development of a data center or archives; or activities to change the culture around data. This might link for example to the China Data Survey Network at the University of Michigan (mentioned above and in Section 3).

Making better use of existing data

- Currently much data is underutilised or badly used due to lack of understanding of its limitations, or of what can be done with it. To some extent, all data is defective and has limitations. However, it may be possible to encourage better use of existing data by having more detailed information and documentation about the data, including drawing on 'local knowledge' of many of those involved in its collection and analysis, who can provide insights into its strengths and limitations.
- One option may be for ESRC to support a more detailed study on how to use existing data, and what can be done with it, despite its limitations. This should start by building on areas of existing strengths, e.g. in fields of population and labour, where reasonable amounts of data of good quality exist. Through a more in-depth understanding of data in narrower areas, it may be possible to identify specific areas where efforts to improve the quality of data or fill key gaps could pay off in the short to medium term.
- Other sources of existing data may also be worth exploring: DFID for example supports many data intensive activities, and collected background resources related to its programmes. Many of these may be useful to researchers. However, most of this 'grey' literature and related data is not easily accessible. A data

archive of background materials and data which placed such information in the public domain may be of value to researchers.

ESRC support for UK researchers:

- There is a need to recognise the degree to which research in China, particularly at the micro-level, still depends heavily on building personal relationships and contacts. It is also important to find ways to minimise competition between scholars in accessing scarce data resources. In addition to creating a list of data resources and contacts, noted above, other possible starting points could include:
- In connection with support for UK centres of China Studies, focus on individuals and institutions that already have strong links with Chinese scholars / institutions as a base for creating a link program around micro-data resources.
- Create a resource base of individuals with good connections in particular areas who could work in an advisory capacity to assist particularly younger scholars in gaining access to relevant data.
- Support a seminar of senior UK academics working on China, to discuss experiences to date, to work out a list of priorities for ESRC, and relationships that need supporting in different sectors

Appendix 1 Information on key data sets

A1.1 Census data and national sample surveys:

Name	National Population Census (1) 全国人口普查
Principal Investigator	NBS
Year(s)	1953, 1964, 1982, 1990, 2000
Area (s) of interest	General Population and demographic data, labour force, location Specific Age, gender, marital status, ethnicity, education, employment, residence etc., birth, deaths
Brief description	Summary National census now carried out every 10 years Geography National coverage: 34 provinces (including capital cities)
How useful to researchers?	Very useful
Availability of data descriptions	Additional information is available in Chinese from various sources: First National Population Census: implemented 30 th , June, 1953 (only reported publicly the total population) <i>Source:</i> http://www.bbtpress.com/homepagebook/345/zq6.htm Second: implemented 30 th , June, 1964. Information includes total population, provincial population, nationality composition, age composition, and educational level. <i>Source:</i> http://www.stats.gov.cn/tjgb/rkpcqb/qgrkpcqb/t20020404_16768.htm Third implemented in 1982. Information includes total population, gender composition, nationality composition, educational level, birth rate, mortality rate, provincial population, data-check information, like error rate, etc. <i>Source:</i> http://tj.tengzhou.gov.cn/ZYLM/tjgb/rkpc/qj/1982.htm Fourth implemented in 1990. Information includes total population, provincial population, population density, population natural growth rate, urban-rural composition, nationality composition, educational level, industry distribution, marital situation, birth situation, mortality situation, household population, and migration situation, etc. <i>Source:</i> http://www.cpirc.org.cn/tjsj/tjsj_cd_detail.asp?id=284 Fifth implemented in 2000. Information includes total population, provincial population, gender composition, population natural growth rate, urban-rural composition, nationality composition, educational level, labor force, the aged population, household population, etc. <i>Source:</i> http://www.popinfo.gov.cn/popinfo/popdocrkxx.nsf/v_rkblnew/8B97851859B2CCFB48256E59003291D7
Conditions	Access to recent data (apart from aggregate published statistics) is difficult and costly, and needs to be negotiated on a case by case basis.
Tabled outputs	Published summary tables. Aggregate or older data is available in published form and on the web (see above), through a number of databanks or through NBS. Publications and reports using the data are also available from various sources.
Contact	Census Department, NBS

Name	National Census of Agriculture (2) 全国农业普查
Principal Investigator	NBS with other relevant government agencies (e.g. Ministry of Agriculture)
Year(s)	1997, 2006
Area (s) of interest	General Agricultural population, labour force, agricultural production, enterprises, rural socio-economic conditions Specific Rural household size and composition, labour and productive activities, number and location of rural employed, cultivated area, agricultural production, township and village enterprises, agricultural machinery, number of technicians, size of towns and villages, infrastructure and facilities, etc.
Brief description	Summary The most recent Agricultural Census was being prepared as this report was underway. Financial support was provided by the Government of Italy and technical assistance was provided by the Food and Agriculture Organization of the United Nations (FAO) through a series of projects spanning a period of thirteen years. The census scope covered 38 items with 687 variables. The time reference point for the previous 1997 census was 1 January 1997. 2006 census to cover six aspects: a) Status of establishments in agriculture and rural household production activities, b) Circumstances related to township and towns, village committee and community environment, c) Utilization of agricultural land, d) Investment in fixed assets of agriculture and rural area, e) Employment and movement of rural labour force and f) Living conditions of farmers.
How useful to researchers?	Potentially very useful but limited availability. Currently (apart from NBS) FAO and Government of Italy are understood to have access to earlier census data.
Availability of data description	Information sources include: First census: http://www.stats.gov.cn/english/specialtopics/agcensus/t20020520_19971.htm Second census: http://www.stats.gov.cn/was40/gjtjj_en_detail.jsp
Tabled outputs	Published aggregate and summary data available.
Contact	NBS

Name	National Economic Census of China (3) 全国经济普查
Principal Investigator	NBS
Year(s)	2003
Area (s) of interest	General Economic and productive activities, enterprises, employment Specific (see below)
Brief description	Summary China conducted its first economic census in 2004 with the aim to keeping abreast of the development of the secondary and tertiary industries in terms of size, structure and economic results, to establish a sound system for registration and a database of basic economic units, and to provide the

	<p>foundation for preparing plans for national socio-economic development and improving policy-making and management.</p> <p>Dataset</p> <p>Several aspects of information were included as below:</p> <p><i>Basic condition of Units</i> (number of units and self-employed individuals, number of corporations by registered type of ownership, regional distribution of unites and self-employed individuals, sector distribution of establishments, distribution of self-employed individuals by sector)</p> <p><i>Number of employed persons</i> (distribution of employed persons in units, education and technical titles of employed persons in units)</p> <p><i>Capital investments</i> in enterprises and its composition</p> <p><i>Industry</i> (number of enterprises and employees, output of main industrial products, energy consumption, assets, liabilities and equity, business revenue and gross profits from principal activity, corporate science and technology activities)</p> <p><i>Construction</i> (number of enterprises and employees, Total output value of construction, Floor space and value of completed construction projects, Assets, liabilities and equity, Business revenue and gross profits from construction project settlement)</p> <p><i>Transport, Storage and Postal Service</i> (number of corporations and employees, Assets, Liabilities and Equity, Business Revenue and Gross Profits from Principal Activities)</p> <p><i>Real Estate Sector</i> (number of corporations and employees, Assets, Liabilities and Equity, Business revenue and Gross Profits from Principal Activities, Production and Management Situation of Real Estate Sector)</p> <p><i>Wholesale and retail trade</i> (number of corporations, employees and business volume, Assets, Liabilities and Equity, Business revenue and Gross Profits from Principal Activities)</p> <p><i>Hotel and Catering Services</i> (number of corporations, employees and business volume, Assets, Liabilities and Equity, Business revenue and Gross Profits from Principal Activities)</p> <p><i>Other Sectors</i> in Tertiary Industry (number of units and employees, Liabilities and Equity, Business revenue and Gross Profits from Principal Activities, Assets, Revenue and Expenditure of Government Agencies, Institutions and other Non-corporation Legal Person Units)</p> <p>Data Format</p>
How useful to researchers?	No information but potentially high value
Availability of data descriptions	<p>Several volumes of background information and some preliminary summary statistics are now available in Chinese; minimal summary information available in English</p> <p>Source http://www.stats.gov.cn/was40/</p>
Conditions	Access restricted: by negotiation
Tabled outputs	As above: some aggregated tabled outputs published by NBS
Contact	NBS

Name	National Census of Basic Units (4) 全国单位普查
Principal Investigator	NBS and other relevant government ministries (including e.g. Taxation, Industry and Commerce, and Quality Control, Inspection and Quarantine)
Year(s)	2002
Area (s) of interest	General Basic economic and other units Specific Economic activity and organisation; employment
Brief description	<p>Summary</p> <p>Objective of the census was to understand the number of different kinds of units in China, their organization, economic composition, scale and structure, vocational distribution and the use of factors of production; and to establish a directory of systems of basic units covering all departments and sectors; to share information and create a database that can be continuously updated; to provide basic information for standardizing markets, strengthening social supervision and management, adjusting economic structure, optimizing industrial policy, and normalizing urban and rural construction, and lay a foundation for other census and sample surveys.</p> <p>Dataset</p> <p>The data covered the following categories of information: Number and Proportion of Corporate Units and Establishments by Type, Distribution of Corporate Units and Establishments in Eastern, Central and Western Regions, Distribution of Basic Units Among Sectors, The Size of Corporations, Corporation Ownership Structure, Number of Corporations by Employed Persons Number of Corporations by Size of Income from Operation Quality of the Census Data. Information available (in Chinese) includes the quality of the Census Data.</p> <p>Scope: Entities and industry activity units of PRC. Self-employed businesses covered as determined by local governments.</p> <p>Content: Census forms are divided into grass-root forms and grouping forms. Grass-root forms include "Background of Impersonal Entities" and "Background of Industrial Activity units"; group forms consist of five forms, "Number of Impersonal Entities and Industrial Activity Units", "Number of Impersonal Entities Grouped by Type of Unit", "Number of Industrial Activity Units by Unit Type", "Background of Impersonal Entities", and "Backgrounds of Industrial Activity Units". Content in the forms include basic identification of units, such as unit code, title, legal representative, unit location, and contact number ; main attributes of units, such as vocation category, registration type and membership; basis status of units, such as the time of going into operation and operation state; master data of units, such as number of employees, paid-up capital of enterprises, fixed assets for production operation and operating income.</p>
How useful to researchers?	Uncertain.
Availability of data descriptions	In Chinese.
Tabled outputs	Published tables and aggregate data available in Chinese
Contact	NBS

Name	National Industrial Census (5) 全国工业普查
Principal Investigator	NBS
Year(s)	1950 1985 1995
Area (s) of interest	General Economic activity, production, employment Specific (see details below)
Brief description	Dataset The third NIC (1995) covered 16 aspects: (1) background of industrial enterprises; (2) financial status; (3) employment situation; (4) the total value of industrial production, distribution, and storage; (5) production, distribution, and storage of principal industrial products; (6) production status of attached industrial units; (7) raw and processed materials, energy consumption and stock of industrial enterprises; (8) energy processing and transformation; (9) scientific and technical activities of industrial enterprises; (10) major industrial technical and economic indices; (11) productivity of principal industrial products; (12) major industry production units and their service state (installed equipment); (13) technical status of major industry production units (installed equipment); (14) quality of principal industrial products; (15) incomings and outgoing of principal industrial products distribution; (16) basic operation status of jointly-funded part of industrial enterprises. Now integrated into Economic Census.
How useful to researchers?	This is now dated but may be useful to compare with current economic census information.
Availability of descriptions	Information on first census: http://www.stats.gov.cn/50znjn/tjds/t20020617_22714.htm Second census: http://www.mtg.bjstats.gov.cn/gjchjk/gongyepucha2.htm Third census: http://www.stats.gov.cn/english/surveyprograms/censuses/t20021025_39810.htm
Conditions	May be more readily accessible due to age
Contact	NBS

Name	National Tertiary Industry Census (6) 全国第三产业普查
Principal Investigator	National Committee of Reform and Development
Year(s)	1993
Area (s) of interest	General Service sector activities and employment Specific
Brief description	Summary Dataset Includes information on: Number of third industry corporations and employees by type of ownership, Number of third industry corporations and employees by sectors, the growth rate of service sector, income and expenditure of the tertiary sector corporation, assets of the third industry

	corporation, etc.
How useful to researchers?	While potentially an interesting source of information on the development of the service industry, this is now dated. The new economic census includes more recent information. Possibly useful for comparison.
Availability of data descriptions	Source: http://www.stats.gov.cn/tjgb/scpcgb/qgscpcgb/t20020331_15507.htm

Name	1% National Population Sample Survey (7) 全国 1%人口抽样调查
Principal Investigator	NBS
Year(s)	2005
Area (s) of interest	General Demographic, population, labour Specific Total population, migrant population, urban-rural composition, sex, age, ethnicity, educational level, household population, unemployment, etc
Brief description	Summary China carried out a 1% National Population Sample Survey at the end of 2005. With the whole country as the population, and provinces, autonomous regions and municipalities as sub-populations, a sampling method of stratified, multi-stage, cluster, and probability proportioned to size was adopted. The final sample was enumeration district. The sample size was 17.05 million persons, accounting for 1.31% of the total population of China. The time reference of the survey was 0:00 a.m. November 1, 2005. The survey targets were the population having permanent residence, possessing Chinese nationality and living in the mainland of the People's Republic of China. Sampling stratified, multi-stage, cluster, and probability proportioned to size Weighting Geography 31 provinces Data Format
How useful to researchers?	Extremely valuable but limited accessibility to researchers.
Availability of data descriptions	Further background information available in Chinese
Tabled outputs	Published tables and aggregate data from NBS
Contact	NBS

Name	National Rural Household Survey (8) 全国农村家庭抽样调查
Principal Investigator	NBS
Year(s)	Annual
Area (s) of interest	General Rural economic conditions, employment

	Specific Rural household demographic information, income, expenditure, consumption, education, employment, etc.
Brief description	<p>Summary</p> <p>The survey covers sampled rural household in every province, municipality and autonomous region, providing information about production, income, consumption, accumulation, education, social activity etc. for rural households.</p> <p>Dataset</p> <p>Basic information of rural households' community, family member in each household in sample units, fixed asset of households, living situation of households, durable products in each households in sample units, production of each household, sales on products of households, total income and net income of each household, total expenditure of rural household, consumption expenditure of each household, the need of commodity, the food consumption and nutrition of each rural household, etc.</p> <p>Cash income and expenditure of rural households, as well as sales on main agricultural products are collected covering certain interval of under a year. Other data is collected annually. Detailed expenditure and consumption diaries are kept throughout the year by households.</p> <p>Geography All the provinces, municipalities and autonomous regions in China</p>
How useful to researchers?	Extremely valuable
Availability of data descriptions	Survey instruments and other background information available through Rural Survey Team
Conditions	Negotiable: micro-data will not be released; may be accessed through collaborative projects
Costs	Parts of data sets have been acquired by research organisations but at high cost and under restrictive conditions
Tabled outputs	Published aggregate and summary statistics available; many publications available containing analysis based on parts of data
Contact	NBS

NAME	NATIONAL URBAN HOUSEHOLD SURVEY (9) 全国城市家庭抽样调查
	Similar to rural household survey (above) but covering urban households

A1.2 Other data sets

Labour

Name	Labor Force Survey (10) 劳动力调查
Principal Investigator	NBS
Year(s)	Every year from 1996; twice yearly since 1997
Area (s) of interest	General Employment structure Specific Employment, unemployment, occupation, demographic information
Brief description	<p>Summary</p> <p>The survey covers the population aged over 16 in sampled households. Sample size: 950,000 people in November round; 420,000 people in May (urban population only). It provides the employment and unemployment situation of respondents in the one week preceding the survey.</p> <p>Sampling</p> <p>Multi-stage, stratified, clustered, pps</p> <p>Weighting</p> <p>Sample frame is census; weighting by provincial population</p> <p>Geography</p> <p>Rural and urban areas covered in November round; urban areas only in May</p> <p>Data Format</p> <p>Text files</p>
How useful to researchers?	Very useful but limited accessibility
Availability of data descriptions	Questionnaire is available in English Background information available from NBS in Chinese, including information on variable descriptions, sampling etc.
Conditions	By negotiation with NBS
Costs	Negotiable
Tabled outputs	Reported in NBS publications
Contact	NBS

Name	Population and Labor Force DataBASE (11) 人口与劳动力数据库
Principal Investigator	Chinese Academy of Sciences Geographic Science and Resources Institute (中国科学院地理科学与资源研究所)
Year(s)	1986 – 2002
Area (s) of interest	General Labour and population
Brief description	Summary The dataset is a compilation of publicly available (aggregate) data from the NBS Census and population data sets. It includes variables such as number and changes of population, basic info of population, age distribution, households, gender, educational level, job, marital situation, birth, mortality, migration (all these info are at provincial, county, city and township level). For further information see NBS surveys above.
How useful to researchers?	Uncertain, but may be easy way to access some NBS data.
Availability of data descriptions	Yes. Information at http://www.resip.ac.cn/SPT--FullRecord.php?ResourceId=118581 http://support.csdb.cn/datasharing/suburiread.jsp?subURI=cn.csdb.naturalresources.population
Conditions	Open access
Costs	Unknown
Contact	Li Zhehui (李泽辉) CAS Geography and Resources Institute 中科院地理与资源所
	Tel: 010-64889070 E-Mail to lizh@igsnr.ac.cn

Name	Labour employment and Social Security (12) 2002 年全国城镇劳动力就业和社会保障状况调查
Principal Investigator	MOLSS
Year(s)	2002
Area (s) of interest	General: Employment, poverty and social security Specific: Unemployment, laid-off workers; social security benefits, demographic information; occupation; working time, wages etc.
Brief description	Summary In 2002 MOLSS implemented a large scale survey on 'Labour employment and social security' covering 190,000 people in 66 cities in 2002. A key purpose was to understand the unemployment situation. However, the data have not been made public, few outside researchers have had access to it, and little analysis has been undertaken. Sampling: multi-stage, stratified, clustered, pps

	<p>Weighting: sample observations and the population in 66 cities in 2002.</p> <p>Geography: National coverage</p> <p>Data Format: Stata</p>
How useful to researchers?	Potentially the richest source of employment data (especially for looking at informal employment) and unemployment information with a large nationally representative sample
Availability of data descriptions	Only one researcher is known to have access to this data (Wu Yaowu, IPLE, CASS) to undertake analysis for MOLSS. The data on unemployment made it sensitive to the government.
Conditions	Generally unavailable
Tabled outputs	Some studies based on the survey (in Chinese), by Wu Yaowu including: <i>Informal employment in urban China, size and characteristics.</i> Published in China Labor Economics. 2006.4 <i>The evaluation of labor market policy: coverage, targeting and effect.</i> Working paper. IPLE, CASS.
Contact	MOLSS

Name	CHINESE HOUSEHOLD INCOME PROJECT SURVEY (CHIP) (13) 中国居民收入分配调查														
Principal Investigator	CASS, Institute of Economics with NBS and international collaborators														
Year(s)	1988, 1995, 2002														
Area (s) of interest	<p>General Employment, income distribution and inequality</p> <p>Specific Poverty, labour, education</p>														
Brief description	<p>Summary In collaboration with NBS, a nationally representative household survey on income inequality was undertaken in 1988, 1995, 2002. This data has been widely used by scholars, particularly labour economists, in China and internationally. The sample is based on the NBS rural and urban household sample, and implemented by the NBS but with closely supervision by the principal researchers, insuring good quality of data.</p> <p>Sampling: all the samples were draw from the large sample used by the NBS in its annual household survey. A detailed description of the sampling method is found in Li et al. (2007) cited below.</p> <p>Geography:</p> <table border="1"> <thead> <tr> <th></th> <th>CHIP1988</th> <th>CHIP1995</th> <th>CHIP2002</th> </tr> </thead> <tbody> <tr> <td>Urban</td> <td>Beijing, Shanxi, Liaoning, Jiangsu, Anhui, Henan, Hubei, Guangdong, Yunnan, and Gansu</td> <td>All the provinces covered in CHIP 1988 urban survey, and Sichuan province added.</td> <td>The same as CHIP 1995 urban survey, except that the Chongqing was separated from Sichuan as a province-level city</td> </tr> <tr> <td>Rural</td> <td>All the provinces in China except for Xinjiang, Tibet, and Taiwan</td> <td>CHIP 1995 urban survey + Hebei, Jilin, Zhejiang, Jiangxi, Shandong, Hunan, Guizhou,</td> <td>CHIP 1995 rural survey + Xinjiang, Guangxi, Chongqing</td> </tr> </tbody> </table>				CHIP1988	CHIP1995	CHIP2002	Urban	Beijing, Shanxi, Liaoning, Jiangsu, Anhui, Henan, Hubei, Guangdong, Yunnan, and Gansu	All the provinces covered in CHIP 1988 urban survey, and Sichuan province added.	The same as CHIP 1995 urban survey, except that the Chongqing was separated from Sichuan as a province-level city	Rural	All the provinces in China except for Xinjiang, Tibet, and Taiwan	CHIP 1995 urban survey + Hebei, Jilin, Zhejiang, Jiangxi, Shandong, Hunan, Guizhou,	CHIP 1995 rural survey + Xinjiang, Guangxi, Chongqing
	CHIP1988	CHIP1995	CHIP2002												
Urban	Beijing, Shanxi, Liaoning, Jiangsu, Anhui, Henan, Hubei, Guangdong, Yunnan, and Gansu	All the provinces covered in CHIP 1988 urban survey, and Sichuan province added.	The same as CHIP 1995 urban survey, except that the Chongqing was separated from Sichuan as a province-level city												
Rural	All the provinces in China except for Xinjiang, Tibet, and Taiwan	CHIP 1995 urban survey + Hebei, Jilin, Zhejiang, Jiangxi, Shandong, Hunan, Guizhou,	CHIP 1995 rural survey + Xinjiang, Guangxi, Chongqing												

		Shannxi	
	Data Format: SPSS, Stata		
How useful to researchers?	The most useful and widely used source of data on many aspects of China's income distribution, changing inequality, labour market and employment, education etc.		
Availability of data descriptions	<p>Two papers describes the basic information of the dataset:</p> <p>(1) Eichen, M., and M. Zhang, (1993), <i>Annex: The 1988 household sample survey - Data description and availability</i>, In Griffin, K., and Zhao, R., eds., <i>The Distribution of Income in China</i>, New York, St. Martin's Press.</p> <p>(2) Li Shi, Luo Chuliang, Wei Zhong, and Yue Ximing, 2007, "Appendix: the 1995 and 2002 household survey: sampling methods and data description", in Björn A. Gustafsson, Li Shi, and Terry Sicular, eds (2007), <i>Inequality and Public Policy in China</i>, Cambridge University Press, forthcoming.</p>		
Conditions	1988 public availability; 1995 and 2002 surveys are available through contact with Luo Chuliang (luocl2002@163.com).		
Costs	<p>CHIPs 1988: 120000 USD</p> <p>CHIPs 1995: 160000 USD</p> <p>CHIPs 2002: 200000 USD</p>		
Tabled outputs	<p>Numerous sources and publications based on the datasets.</p> <p>There are three edited volumes of papers produced during the primary stages in each survey round:</p> <p>Griffin, K., and R. Zhao, eds., (1993), <i>The Distribution of Income in China</i>, New York, St. Martin's Press.</p> <p>Riskin, C., R. Zhao and S. Li, eds., (2001), <i>China's Retreat from Equality: Income Distribution and Economic Transition</i>, Armonk, N.Y., M.E. Sharpe.</p> <p>Björn A. Gustafsson, Li Shi, and Terry Sicular, eds (2007), <i>Inequality and Public Policy in China</i>, Cambridge University Press, forthcoming.</p>		
Contact	Luo Chuliang: luocl2002@163.com		
	The Research Team is preparing for the publication of all the CHIP surveys. Until then, researchers can contact Luo Chuliang for questions or information.		

Population

Name	National Family Planning and Reproductive Health Survey (14) 国家计划生育与生殖健康调查
Principal Investigator	National Population and Family Planning Commission of China, and CPRDC
Year(s)	2001
Area (s) of interest	General Population Specific Reproductive Health
Brief description	Summary This survey reviews female reproductive status, contraceptive conditions and reproductive health needs. This survey contains basic information on the communities; reproductive histories; contraceptive use and reproductive health of people at reproductive age; and, services and technology for reproductive health at township and community level. Dataset 1000 communities were sampled, including all the people in these communities. About 180,000 individuals were interviewed, among which 25,000 were women of reproductive age. Sampling The survey used stratified random sampling.
Contact	CPRDC (Zhuang Ya'er - yaerzh@vip.sina.com) www.cpirc.org.cn/tjsj/tjsj_cd_detail.asp?id=5081

Name	The China Elderly Health and Longevity Survey (15) 中国高龄老人健康长寿影响因素项目
Principal Investigator	Beijing University, Population Research Institute, China Centre for Ageing Research
Year(s)	1987
Area (s) of interest	General Population; Ageing Specific
Contact	CPRDC (Zhuang Ya'er - yaerzh@vip.sina.com); also CSDN, University of Michigan; also University Service Centre, Chinese University of Hong Kong

Name	China Children's Situation Sample Survey (16) 中国儿童状况抽样调查数据汇总表
Principal Investigator	PFPC
Year(s)	1992
Area (s) of interest	General Children
Brief description	Format Excel

Contact	CPRDC (Zhuang Ya'er - yaerzh@vip.sina.com); also CSDN, University of Michigan
----------------	---

Name	Survey of Migration in 74 Urban Centres (17) 中国 74 城镇迁移抽样调查
Principal Investigator	Institute of Population and Labour Economics, Chinese Academy of Social Sciences
Year(s)	1986
Area (s) of interest	General Population; migration
Brief description	Summary Survey of 74 urban centres including 15 mega-cities, 6 large cities, 12 medium-size cities, 10 small cities and 31 townships from 16 provinces. Dataset 25000 questionnaires filled in for 23895 households and 1643 collectives. 100267 people were sampled in total. The total population of the area covered was 3.5 m people. The survey took place over 6 months between July and the end of the year.
How useful to researchers?	Useful
Availability of data descriptions	Yes
Contact	CASS; CPRDC also have this dataset (contact Zhuang Ya'er - yaerzh@vip.sina.com); also CSDN, University of Michigan; also University Service Centre, Chinese University of Hong Kong

Name	China Disabled People Sample Survey (18) 中国残疾人抽样调查
Principal Investigator	Ministry of Civil Affairs
Year(s)	1987
Area (s) of interest	General Population; disability Specific
Contact	CPRDC (Zhuang Ya'er - yaerzh@vip.sina.com);

Name	Reproductive Health Sample Survey (19)
Principal Investigator	PFPC
Year(s)	1997
Contact	CPRDC (Zhuang Ya'er - yaerzh@vip.sina.com); CSDN, University of Michigan

Name	China Birth Rate Sample Survey (20) 中国生育率抽样调查
-------------	---

Principal Investigator	PFPC
Year(s)	1992
Area (s) of interest	General Population
Contact	CPRDC (Zhuang Ya'er - yaerzh@vip.sina.com); CSDN, University of Michigan

Name	Survey of Support Systems for the Elderly (21) 中国老年人供养体系调查
Principal Investigator	China Ageing Research Institute
Year(s)	1992
Area (s) of interest	General Elderly Specific
Availability of data descriptions	Available from CPRDC
Tabled outputs	Available from CPRDC
Contact	CPRDC (Zhuang Ya'er - yaerzh@vip.sina.com); CSDN, University of Michigan; also University Service Centre, Chinese University of Hong Kong

Name	Sample survey of elderly people over 60 (22)
Principal Investigator	Institute of Population and Labour Economics, Chinese Academy of Social Sciences
Year(s)	1987
Area (s) of interest	General Population Specific Ageing
Brief description	Dataset 36755 people from 28 provinces were surveyed; of these 17819 were urban-registered, and 18936 rural (the latter from 223 counties). Sampling A stratified, multi-stage, random cluster process was used.
Contact	Institute of Population and Labour Economics, CASS; also University Service Centre, Chinese University of Hong Kong

Name	Survey of Family Economy and Reproductive Situation in 10 provinces (23)
Principal Investigator	Institute of Population and Labour Economics, Chinese Academy of Social Sciences
Year(s)	1992
Area (s) of interest	General Population
Brief description	Summary 1400 households were sampled from each province. This sample was split into 500 urban households, 300 households from townships, 300

	households from developed counties, and 300 from less-developed counties.
Contact	Institute of Population and Labour Economics, CASS

Name	Family Trends Social Survey (24)
Principal Investigator	Institute of Population and Labour Economics, Chinese Academy of Social Sciences
Year(s)	2002 and 2004
Area (s) of interest	General Specific
Brief description	Summary 2002: This covered 9 provinces. 4477 elderly and their adult children were questioned on family trends. Themes include looking after the elderly and the generation gap. 2004: 4700 households and 9400 people were sampled. It focussed on 3 provinces in the Yangtse River delta.
Contact	Institute of Population and Labour Economics, CASS

Name	Survey on the Impacts of Migration on Rural Women (25)
Principal Investigator	Institute of Population and Labour Economics, Chinese Academy of Social Sciences
Year(s)	2005
Area (s) of interest	General Population; migration; women; gender
Brief description	Dataset The sample for this was 1000 women.
Contact	Institute of Population and Labour Economics, CASS

Name	Survey of Changing Marriage Practices in Rural China (26)
Principal Investigator	Institute of Population and Labour Economics, Chinese Academy of Social Sciences
Year(s)	2005
Area (s) of interest	General Population
Brief description	Summary. This single province survey looked at changing attitudes towards marriage, and effects on the status of women and care for the elderly. Dataset 600 Geography Hebei
Contact	Institute of Population and Labour Economics, CASS

Health

Name	National Health Sample Survey (27) 国家卫生服务调查
Principal Investigator	Ministry of Health
Year(s)	1993, 1998, 2003
Area (s) of interest	General Health Specific Health status; health resource structure
Brief description	<p>Summary The Health Survey is the most useful broad coverage dataset in the health sector. It is managed by the Health Information Centre under the Ministry of Health.</p> <p>The Third National Health Survey focused on: health status rural and urban populations; needs and demand for healthcare; utilisation of healthcare services and costs; impacts of healthcare reform on different groups.</p> <p>The survey includes: demographic and social characteristics of the household members, household financial status, housing, modification of water supply systems and latrines in rural areas, health insurances of family members, self-reported illnesses in the two weeks prior to the survey, chronic disease in the previous 6 months, health behaviours and knowledge of the population aged at 15 years old and above, medical consultations in the previous two weeks, reasons for not seeking medical care, self-care, hospitalization in the past year, reasons for not being hospitalized when patients were referred by their doctors, health care for married women of child-bearing age and children under 5 years old and the responses of patients to healthcare services.</p> <p>Survey data is collected through a Nationwide Household Interview Survey, using household interviews, carried out by local health care workers and supervised by medical personnel from township health centres.</p> <p>There is also a supplementary qualitative Localized Social Assessment. In 1993 and 1998 the Health Survey included a Health Services Survey; this is no longer carried out, as administrative data is now perceived as adequately covering its contents.</p> <p>Dataset 95 counties were randomly selected from a nationwide frame, from these sample townships (rural) or administrative streets (urban) were chosen. Altogether 475 sample townships (or streets) were chosen. 2 villages or neighbourhood committees (from each township or street) were then chosen, giving a total of 950 sample villages (or neighbourhood committees). From each sample village or neighbourhood committee, 60 sample households chosen. Overall a total of 57,000 sample households were selected from across the whole country. The First and Second National Health Surveys were of comparable size (in terms of number of sample counties, and number of sample households).</p> <p>Sampling The survey used multistage stratified random cluster sampling.</p>
How useful to	Researchers are generally positive about the quality of the data, and the level of description available. The main problem is that this dataset

researchers?	is very difficult to access. Parts might be made available linked to a specific collaborative research project. Various quality control measures were applied during the last survey. The consistency rate between the original and the countercheck was 95%. The Myers Index, Test for Goodness Fit gave good results; and the Gini Concentration Ratio of households, and Delta Dissimilarity Index gave good matches when compared with the general population.
Tabled outputs	《1993 年国家卫生服务调查分析报告》, by the National Ministry of Health, the catalog is available at: http://www.moh.gov.cn/newshtml/8720.htm 《1998 年国家卫生服务调查分析报告》 by National Ministry of Health, related catalog at : http://www.moh.gov.cn/newshtml/8739.htm 《2003 年国家卫生服务调查分析报告》 by National Ministry of Health, main result is available at: http://www.moh.gov.cn/newshtml/8981.htm
Contact	Rao Keqing, Ministry of Health Information Centre

Name	National Mortality Survey (28) 全国死因回顾抽样调查
Principal Investigator	Ministry of Health; Ministry of Science and Technology
Year(s)	1973, 1990, 2005
Area (s) of interest	General Health Specific Mortality; focus on death from cancer
Brief description	Summary It covers total number of deaths, reasons for death, death rate and death locations; as well as birth rates in the sample area and basic socio-economic information. Dataset The sample size is 6% of the population Sampling This is a multi-stage, stratified, clustered survey, using random sampling methodology.
Availability of data descriptions	The third survey has not been completed, related information can be found at www.moh.gov.cn/moh/fs/uploadfile/200609/200698144026228.doc

Name	CHina Health and Nutrition Survey (29) 中国健康与营养调查
Principal Investigator	UNC Population Centre and the Ministry of Health.
Year(s)	1989,1993, 1997, 2000, 2004
Area (s) of interest	General Health Specific
Brief description	Summary This survey is a longitudinal survey conducted to understand the nutritional and health status of the Chinese population, and how

	<p>socioeconomic influences relate to outcomes. Detailed community data are also collected in surveys of food markets, health facilities, family planning officials, and other social services and community leaders.</p> <p>Dataset 9 provinces, with 4400 households, and about 16,000 individuals interviewed</p> <p>Sampling The sampling method was a multistage, random cluster process</p>
How useful to researchers?	Very useful and widely used
Availability of data descriptions	Available at www.cpc.unc.edu/china
Contact	The data is available for public use at www.cpc.unc.edu/china .

Name	<p>China Residents Nutrition and Health Survey (30) 中国居民营养健康调查</p>
Principal Investigator	The Ministry of Health and National Bureau of Statistics and The Ministry of Science and Technology
Year(s)	2002
Area (s) of interest	<p>General Public Health</p> <p>Specific Nutrition and health status for the residents</p>
Brief description	<p>Summary This survey is the first national representative survey on nutrition and health status. This survey contains four parts:</p> <ol style="list-style-type: none"> 1. General information 2. physical examination (221044), 3. lab inspection, and 4. food investigation (69205) <p>153259 participants had blood pressure tests, 94996 had blood fat tests taken, and 211726 haemoglobin tests, 98509 blood sugar tests were taken, and 13870 Vitamin A plasma tests.</p> <p>Dataset</p> <p>71971 households (24034 urban and 47937 rural) and 243479 individuals (68656 urban and 174823 rural). With additional pregnant and children observations, this survey totals 272023 participants.</p> <p>Sampling Multi-stage, stratified, clustered, random sampling. It divided 6 types of area according to the economic development: big city, middle and small city, type one rural area, type 2 rural area, type 3 rural area and type 4 rural area.</p>
Availability of data descriptions	<p>No publication of this survey is available now, while the main result is available at : http://news.xinhuanet.com/video/2004-10/12/content_2080855.htm</p> <p>http://www.chinacdc.net.cn/n272442/n272530/n3246177/11031.html</p>
Contact	NBS

Name	<p>China Healthy longevity Survey (31) 中国高龄老人健康长寿调查</p>
Principal Investigator	Prof. Zeng Yi, The Center for Healthy Ageing and Family Studies, Peking University.

	and China Research Center on Aging
Year(s)	1998, 2000, 2002, 2004
Area (s) of interest	General Public health Specific Health status of most elderly in China
Brief description	<p>Summary This survey is one of the best examples of a well-designed longitudinal survey where the micro-data is now in the public domain. Funding came from the National Institute for Ageing, US; Beijing University; UN Population Fund; National Social Science Foundation (China) and Hong Kong Research Grant Office (RGC). The survey focuses on the health status and health care of the most elderly in China. The first wave of the Chinese longitudinal survey on healthy longevity was conducted in 1998. It randomly selected 50 percent of the counties and cities in 22 provinces with a total population of 985 million. The contents of the individual questionnaire included questions concerning family households, activities of daily living (ADL), lifestyle, diet, psychological characteristics (disposition), economic resources, family support and medical care services. Some demographic, socioeconomic, and environmental data on the 666 sampled counties and cities were also collected. A total of 9,073 valid individual questionnaires were completed in 1998 survey.</p> <p>Dataset The procedure of sampling for the 2000 survey was similar to 1998, as a longitudinal survey, it interviewed 11216 elderly (over 80), of which 4844 had been interviewed in 1998 and survived to 2000. In 2002, the third survey had been expanded to the 65+ elderly, thus the sample size became 15000, and about 2800 interviewees had survived and received all three interviews. In 2004, a fourth survey was conducted, the data has been cleaned, but little has been published so far.</p> <p>Sampling The survey randomly selected 50 percent of the counties and cities in 22 provinces. The 22 surveyed provinces are: Liaoning, Jilin, Heilongjiang, Hebei, Beijing, Tianjing, Shanxi, Shaanxi, Shanghai, Jiangsu, Zhejiang, Anhui, Fujian, Jiangxi, Shandong, Henan, Hubei, Hunan, Guangdong, Guangxi, Sichuan, and Chongqing, which have a total population of 985 million.</p> <p>Geography 22 provinces Data Format SPSS</p>
How useful to researchers?	Very useful
Availability of data descriptions	The general information and the aggregated data of 1998 is published and is available at: http://www.pku.edu.cn/academic/ageing/
Contact	Professor Zeng Yi

Name	Health and Retirement Survey (32)
Principal Investigator	John Strauss, Zhao Yaohui CCER
Year(s)	2007
Area (s) of interest	General Health; ageing
Brief description	<p>Dataset 6-8000 households Plan for longitudinal survey modelled on internationally comparable studies. Currently at design stage.</p>

Availability of data descriptions	NIA funded, should be public access after a certain period of time has elapsed
--	--

Name	INVEstigation into Psychological Reactions during the SARS OuTBREAK (33) 非典期间人们的社会心理调查
Principal Investigator	Social Psychology Research Institute, People's University
Year(s)	2003
Area (s) of interest	General: Social Psychology
Brief description	Dataset : 1602 Sampling : random sampling on Internet Data Format: SPSS
Contact	Wang Weidong 电邮: wwd@ruc.edu.cn 地址: 中国人民大学社会心理学研究所 邮编: 100872

Name	World Mental Health Survey (34) 世界精神健康调查
Principal Investigator	Research Centre for Contemporary China
Year(s)	2001
Area (s) of interest	General: Social Psychology Specific
Brief description	Summary To investigate the incidence of certain mental health problems including depression and anxiety Dataset : 5,201 Sampling: Probabilities proportional to size (PPS) Geography: Beijing, Shanghai Data Format: SPSS
Contact	Tel: 010-62755443 Email: rccc@rccc.pku.org Peking University 100871

Name	Economic Cost of Adverse Drug Reaction (35)
Principal Investigator	Gordon Liu, Centre for Pharmacoeconomics and Outcomes Research, Guanghua School of Management, Beijing University
Year(s)	2006-7

Area (s) of interest	General Health Specific Pharmaceuticals
Brief description	Summary Data from Beijing hospitals.
Contact	Professor Gordon Liu

Education

Name	Urban Residents Education And Employment Situation Survey (36) 城镇居民教育与就业情况调查
Principal Investigator	Beijing University, (Education College, Educational Economics And Finance Department and Education Finance Scientific Research Institute)
Year(s)	2005
Area (s) of interest	General Education and employment Specific Relation between income and education expenditure; parental education and children educational experience
Brief description	Summary 10000 households were surveyed on education and employment. Data was collected using household interviews for 752 variables including educational attainment, education expenditure, employment and income. Families were asked about family members, and parental circumstances, educational levels of children, work histories, family numbers, family income and expenditure, housing conditions, assets, computer and car ownership, residential status, migration, job satisfaction, changes of employment and training courses attended. Altogether information was collected for about 67,000 people. The Educational Economics and Finance Department of the Education College also have data from the NBS Urban Household Survey, including data from 68,000 informants on income, employment and education. Dataset 12 provinces; 10000 households Geography 12 provinces: Beijing, Shanxi, Liaoning, Heilongjiang, Zhejiang, Anhui, Hubei, Gaungdong, Sichuan, Guizhou, Shanxi, Gansu.
How useful to researchers?	Can carry out research on educational expenditure, employment, income, training etc
Availability of data descriptions	Available
Contact	Professor Yue Changjun; cjyue@gse.pku.edu.cn

Name	Rural Compulsory Education Baseline Survey (37) 农村义务教育基线调查
Principal Investigator	Beijing University, (Education College, Educational Economics And Finance Department and Education Finance Scientific Research

	Institute)
Year(s)	To be carried out in 2007
Area (s) of interest	General Education Specific Rural Compulsory Education Baseline Survey
Brief description	Summary A Rural Compulsory Education Baseline Survey in three parts looking at schools, teachers and students and their family situation. School survey: Basic information about schools, student enrolment, basic information about students, school facilities, school finances, examinations; school management and decision-making. The teacher survey component includes basic data on each teacher, teacher motivation, curriculum reform and teacher-training, teacher recruitment and movement, salaries. The family survey component includes basic information about the location of families, information about family members, income, education expenditure, social relationships, family perspectives on education, and relations between family heads and schools Geography Probably 15 provinces
How useful to researchers?	Very useful
Availability of data descriptions	Available
Contact	Professor Wang Rong: rwang@gse.pku.edu.cn

Name	National Higher Education Graduates Employment Situation Survey (38) 年全国高校学生课程需求情况调查数据
Principal Investigator	Beijing University, (Education College, Educational Economics And Finance Department and Education Finance Scientific Research Institute)
Year(s)	2003 and 2005
Area (s) of interest	General Education and employment Specific Looks at employment among higher education graduates
Brief description	Summary This survey covers 16 provinces and 34 higher education institutions Dataset Includes 9 key colleges and universities participating in the 211 programme, 20 ordinary institutions and 5 vocational institutions. Each institution issued 1000 questionnaires, 21220 were returned. Geography 16 provinces and 34 higher education institutions East: Beijing, Tianjin, Liaoning, Hebei, Shandong, Jiangsu; Middle: Jilin, Heilongjiang, Anhui, Heinan, Hubei, Hunan; West: Sichuan, Guizhou, Yunnan, Shanxi.
How useful to researchers?	Very useful
Availability of data	Available

descriptions	
Contact	Professor Yue Changjun: cjyue@gse.pku.edu.cn

Name	National Higher Education Students needs situation survey. (39) 全国高校学生课程需求情况调查数据
Principal Investigator	Liu Huizhen, College of Education, Beijing Normal University
Year(s)	2002
Area (s) of interest	General Education Specific Higher education student needs
Brief description	Summary <ul style="list-style-type: none"> Information on students Student attitudes to a range of issues including teaching content, teaching methods, assessment; towards educational reform, including tuition fees etc 25 variables in all Dataset 3002 students Sampling Randomised multilevel accounting for: a) 6 regions of China b) college management structure c) MOE designation of college Geography Nationwide Data Format SPSS
How useful to researchers?	Useful
Availability of data descriptions	Available
Contact	Professor Liu Huizhen ; bnuhz@163.com

Name	National Higher Education Political Thought And Public Curriculum Implementation Situation Student Survey (40) 全国高校政治思想课程和公共课程实施情况的学生调查数据
Principal Investigator	Liu Huizhen, College of Education, Beijing Normal University
Year(s)	2005-2006
Area (s) of interest	General Education Specific Moral and political education
Brief description	Summary <ul style="list-style-type: none"> Information on individual students: age, sex, subject studied, political identity, membership of societies Information on the college: administration etc. Students family background Students provided information on: <ol style="list-style-type: none"> Attitudes to political and moral education Effectiveness of political and moral education classes (what have students learnt?) Student views on curricula and curricula reform,

	<p>teaching methodologies, public classes etc</p> <p>Dataset 2560 students were surveyed</p> <p>Sampling Stratified random sampling using 3 indicators: region and economic level of area; college management system; special designation of college. On this basis 19 colleges selected.</p> <p>Geography Nationwide</p> <p>Data Format SPSS</p>
How useful to researchers?	Useful
Availability of data descriptions	Available
Contact	Professor Liu Huizhen ; bnuhz@163.com

Name	<p>Compulsory Education Period Student Family Education Expenditure Survey (41)</p> <p>义务教育阶段学生家庭教育支出调查数据</p>
Principal Investigator	Liu Huizhen, College of Education, Beijing Normal University
Year(s)	2005 and 2006
Area (s) of interest	<p>General Education</p> <p>Specific Educational expenditure</p>
Brief description	<p>Summary</p> <ol style="list-style-type: none"> 1. Basic information about students and family 2. Range of educational fees charged by schools (including exams, heating etc) 3. Other family spending on education <p>Dataset 17400</p> <p>Sampling Stratified random sampling accounting for representativeness of different areas, and levels of economic development, types of school, age and gender.</p> <p>Geography Nationwide</p> <p>Data Format SPSS</p>
How useful to researchers?	Useful
Availability of data descriptions	From Liu Huizhen
Contact	Professor Liu Huizhen ; bnuhz@163.com

Name	<p>National Ordinary Higher Education College Special Teacher (42)</p> <p>全国普通高校专任教师工作情况调查数据</p>
Principal Investigator	Liu Huizhen, College of Education, Beijing Normal University
Year(s)	2006
Area (s) of interest	<p>General Education</p> <p>Specific This survey looks into the work situation of special teachers in ordinary higher education colleges</p>

Brief description	<p>Summary</p> <p>Basic information about teachers: sex, educational background and achievements, income, background of parents</p> <p>Teachers work situation: working conditions, work responsibilities, job satisfaction, work difficulties</p> <p>Dataset 10000</p> <p>Sampling Based on 5 indicators (administrative area, economic level, type of college management, college level, school speciality) 22 areas were selected, from these 34 colleges were randomly chosen</p> <p>Geography Nationwide</p> <p>Data Format SPSS</p>
How useful to researchers?	Useful
Availability of data descriptions	Available
Costs	Contact Liu Huizhen
Contact	Professor Liu Huizhen ; bnuhz@163.com

Name	Gansu Survey of Children and Families (GSCF) (43)
Principal Investigator	Emily Hannum, University of Pennsylvania and Jennifer Adams, Stanford University
Year(s)	2000-5
Area (s) of interest	<p>General Education</p> <p>Specific The focus of the survey is on: boys and girls development and schooling experiences; family/children attitudes to schooling; mechanisms linking poverty and education; education and social mobility.</p>
Brief description	<p>Summary</p> <p>GSCF is a longitudinal, multi-level data collection project, with a sample of 2000 children in 20 rural counties, and data collected for 6 years. Some of the early micro-data is publicly accessible.</p>
How useful to researchers?	Very useful
Availability of data descriptions	Good documentation on sampling and weighting is available, and the original questionnaires (www.ssc.upenn.edu/china/Gansu/intro.htm)
Conditions	Data for 2000 is available via the China Survey Data Network. Data for later years should be publicly available.
Contact	Emily Hannum hannumem@ssc.upenn.edu

Name	Basic Education Western Areas Project (BEWAP) impact assessment survey (44)
Principal Investigator	Du Yuhong, College of Education Management, Beijing Normal University
Year(s)	2006

Area (s) of interest	General Education Specific Data collected in connection with a World Bank education loan project
Brief description	Summary BEWAP is a World Bank loan project, with DFID grant funding blended with the loan. It covers 112 poor counties in five provinces in western China. Impact assessment surveys are currently being carried out looking at civil works, participatory teacher training, and school development planning. Dataset Impact assessment used a sample of 15 counties, 270 primary and junior schools, 270 headteachers, 4000 teachers and 12000 students achievement and family information records. Sampling The Sampling method is proportional random sampling. Geography Sichuan, Gansu, Ningxia, Yunnan and Guangxi
How useful to researchers?	Possibly limited as this is project relevant data
Contact	Ministry of Education; Du Yuhong - du_yuhong@hotmail.com

Name	Gansu Basic Education Project End of Project survey (45)
Principal Investigator	Cambridge Education Consultants and Gansu Provincial Education Bureau
Year(s)	2005
Area (s) of interest	General Education
Brief description	Summary Gansu Basic Education Project is a source of in-depth micro-data for a small area. GBEP was a DFID funded project managed by Cambridge Education Consultants and Gansu Provincial Education Bureau, the project ended in 2005. Baseline survey and end of project survey using complete sample covering all four project counties (all schools, teachers, and head-teachers in the participating counties). Qualitative data is also available.
How useful to researchers?	Very small scale, project oriented, but could go together with other datasets on Gansu. See Gansu Survey of Children and Families
Availability of data descriptions	Should be available as funded by DFID
Costs	Low?
Contact	This data could probably be made available through application to Gansu Provincial Education Bureau (contact can be made via Cambridge Education Consultants). Andy Brock, Cambridge Education Consultants; andy.brock@camb-ed.com

Name	Education For All monitoring exercise (46)
-------------	---

Principal Investigator	Chinese National Commission for UNESCO
Year(s)	2007
Area (s) of interest	General Education Specific Basic education
Brief description	Summary UNESCO is assisting China (via the Chinese National Commission for UNESCO) with its Education For All monitoring exercise, which currently covers five provinces (Guizhou, Gansu, Henan, Yunnan and Hebei). Micro-data is being collected, and reports based on this data will be released in 2007. Geography Guizhou, Gansu, Henan, Yunnan and Hebei
Availability of data descriptions	It is unclear at present whether it will be possible to access micro-data and on what terms. It is likely that only aggregated data will be available.
Contact	Sun Lei, l.sun@unesco.org ; s.yano@unesco.org

Name	Employment and education survey covering rural and urban areas in Chifeng, Inner Mongolia (47) 人力资本与欠发达地区可持续发展调查
Principal Investigator	Du Yuhong, College of Education Management, Beijing Normal University
Year(s)	2006
Area (s) of interest	General Education and employment Specific Employment and education survey covering rural and urban areas in Chifeng, Inner Mongolia
Brief description	Summary <ul style="list-style-type: none"> • Basic information about informants: age, educational background, occupation, salary and place of work. • Employment history and experience. It contains some indicators on employment and training, including sources of training. • Family income and consumption, education expenditure. • Geographic and economic information about living conditions. The survey also contains data on ad hoc fees, tax and other levies. Dataset Urban sample size 1000, including 78 variables. The rural part includes 340 informants, with about 100 variables. Sampling The survey uses a random selecting method. Geography Chifeng, Inner Mongolia
Contact	Professor Du Yuhong ; du_yuhong@hotmail.com

Social welfare and assistance

Name	2004 Urban Heating Survey (48)
Principal Investigator	ISDPP under a contract with the Ministry of Construction.
Year(s)	2004
Area (s) of interest	General: Urban housing, social assistance Specific: Household demographic information, social assistance, heating
Brief description	<p>Summary To measure the impact of the current heat reform on urban low-income and poor residents, and to evaluate the effectiveness of the social mitigation policies adopted by local governments for poor households</p> <p>Dataset</p> <p>The questionnaire includes variables on basic social demographic characteristics of household members, household income and expenses, social welfare entitlement (qualification for minimum living standard benefits), housing condition and insulation improvement, heating bill payment, heating quality, heat supply methods, and social assistance measures.</p> <p>The survey was done between November and December 2004. In addition, the researchers also conducted in-depth interviews with residents and key informants about policy implementation and their attitudes towards the situation.</p> <p>Sampling Methods unknown. 1920 Households in 4 cities. 500 residents from each city (2000) in all were interviewed.</p> <p>Weighting Unknown</p> <p>Geography 4 Cities: Changchun, Harbin, Taiyuan and Tianjin</p> <p>Data Format SAS Format</p>
How useful to researchers?	Narrow topic and small sample, useful for addressing specific questions.
Availability of data descriptions	Instruments and data available through ISDPP
Conditions	Access negotiable
Tabled outputs	Draft Report by Xiulan Zhang and Yuebin Xu: Impact and Mitigating Policies of Urban Heat reform on Low-Income and Poor People, 2005
Contact	Professor Zhang Xiulan, ISDPP

Name	2004 Urban and Rural Social Protection Survey (49)
Principal Investigator	ISDPP, Beijing Normal University
Year(s)	2004
Area (s) of interest	General: Social assistance Specific: Rural and urban household demographic information, social assistance benefits, needs, household difficulties.
Brief description	Summary: Purpose of data was to support evidence-based policy making in China on social protection. Two separate questionnaires (one for the urban

	<p>families and the other for rural families) were developed.</p> <p>Dataset Originally 43 urban and rural communities with a size of 3,850 households were planned for the study. The researchers adjusted the localities with a focus on the northwest poor areas in China, and increased both the rural and urban samples. The eight localities were selected to be broadly representative of the different social and economic situation of China.</p> <p>Geography: 8 provinces: Beijing, Guangzhou, Liaoning, Shanxi, Chongqing, Hubei, Jiangxi, Gansu</p> <p>Data Format SPSS/SAS</p>
How useful to researchers?	Useful for answering a narrow range of questions related to social protection schemes in China; limited by sampling and design issues.
Availability of data descriptions	Through contact with research team
Conditions	Use may be negotiable
Tabled outputs	Some preliminary reports using data available from ISDPP
Contact	Professor Zhang Xiulan

Name	2005 Rural Social Protection Survey (50)
Principal Investigator	ISDPP, Beijing Normal University
Year(s)	2005
Area (s) of interest	<p>General: Rural household conditions, social assistance</p> <p>Specific: Household characteristics, incomes, living conditions; the financing and implementation of social protection programs, informal coping mechanism of poor households.</p>
Brief description	<p>Summary The purpose of the survey was to provide information on the general livelihood of rural households, with particular attention given to the current rural social protection recipients such as Five-Guarantee households, the extreme poor (<i>tekun</i>), and recipients of Minimum Living Standard Guarantee (<i>dibao</i>) programme benefits. The survey instrument was jointly designed and developed by researchers of ISDPP and experts from the World Bank. The resulting questionnaire includes ten broad categories of information: 1) basic characteristics of households; 2) education and training; 3) employment and migrant labor; 4) household economy; 5) household income and assets; 6) consumption and expenditure; 7) informal support and coping; 8) health, and utilization of health care services; 9) the implementation of the formal social protection programs, including those such as Five-guarantee, <i>tekun</i>, <i>dibao</i>, and Medical Financial Assistance; and 10) respondents' attitudes on programs, and their assessment of its performance.</p> <p>Dataset Sampling This survey used a multistage sampling method and combined random sampling and purposive sampling. To ensure an adequate sampling of the rural poor, these categories (<i>Five Guarantees</i>, <i>dibao</i> and <i>tekun</i>) were oversampled. Four provinces - Zhejiang, Fujian, Guangxi and Gansu, were selected based mainly on the difference in the development levels of rural social protection and their geographic locations. The selection of townships and villages was done at the county level.</p> <p>Geography: Townships and villages in 4 provinces: Zhejiang, Fujian, Guangxi</p>

	and Gansu Data Format SPSS/SAS
How useful to researchers?	Useful in understanding specific set of questions about China's social protection schemes and welfare of rural households.
Availability of data descriptions	Instruments available; otherwise limited background documentation.
Conditions	May be available by negotiation
Tabled outputs	Report on the Implementation of Rural Social Protection Survey, by ISDPP, Beijing Normal University, 2005
Contact	Professor Zhang Xiulan

Name	Urban Dibao 2005 (51)
Principal Investigator	ISDPP, Beijing Normal University
Year(s)	2005
Area (s) of interest	General: Urban household social and economic conditions, social assistance Specific: Urban household demographic structure, income and expenditure, social support beneficiaries, and standard of benefits
Brief description	Summary: Designed to provide a better understanding of Urban Minimum Living Standard Scheme (<i>Dibao</i>) and to assist the government in determining appropriate standards of benefits. This survey provides information about urban <i>Dibao</i> beneficiary families including their demographic characteristics and expenditures on food, clothing, health, heating, children education and entertainment. Dataset Sampling: 2794 <i>Dibao</i> beneficiary households, including 7111 individual beneficiaries, randomly selected from 6 cities. Geography: 6 cities: Jinan, Changchun, Chengdu, Xi'an, Hefei, Changsha Data Format
How useful to researchers?	Useful for understanding consumption patterns of urban poor households
Availability of data descriptions	Instruments available; otherwise limited background documentation.
Conditions	May be available by negotiation
Tabled outputs	Report on the Adjustment Mechanism of <i>Dibao</i> Benefit Level, by ISDPP, Beijing Normal University, 2006
Contact	Professor Zhang Xiulan

Other surveys

Name	China General SociAL Survey (CGSS) (52) 中国综合社会调查
Principal Investigator	Sociology Department, People's University in collaboration with Sociology Department, Hong Kong University of Science and Technology
Year(s)	2003 2004 2005
Area (s) of interest	General General social survey
Brief description	<p>Summary</p> <p>A social survey similar to the US and European model, may continue funded by the National Social Science Foundation</p> <p>Survey conducted in 2003</p> <p>Sampling size: 5894 individuals Respondents: adults from 18 to 69</p> <p>Survey conducted in 2004</p> <p>Sampling size: 11000 individuals Respondents: adults from 18 to 69</p> <p>Survey conducted in 2005</p> <p>Sampling size: 10000 individuals Respondents: adults from 18 to 69</p> <p>Sampling methods in these three surveys: multi-stage stratified sampling</p> <p>Main information: Demographic information of respondents, their parents, and spouse, economic situation of respondents, their parents, and spouse, political affiliation, social network, information about their working units, job satisfaction, changes of Registered Permanent Residence of respondents, their parents, and spouse, migration frequency and reason, housing and living situation, educational experience, career experience, the current job information, evaluation of themselves, their life and job, attitudes to some social problem</p> <p>Sampling multi-stage stratified sampling</p>
How useful to researchers?	Useful
Availability of data descriptions	The original database of the 2003 survey is available at http://www.cssod.org/
Contact	Li Lulu chinagss2003@yahoo.com.cn tel: 62511447

Name	Survey on Social Mobility and Social Change in China (SSMSC) (53) 社会变迁与社会意识调查
Principal Investigator	Sociology Department, People's University and Social Survey Centre, People's University; Shi Tianjian, Duke University
Year(s)	1993
Area (s) of interest	General Social mobility and social change

Brief description	Dataset 3287 surveyed Sampling Random sampling Geography 26 provinces and municipalities
Availability of data descriptions	Dataset available at: http://chinaarchive.tamu.edu/portal/site/chinaarchive/menuitem.cfa20b56c0b77e69b5c923d7f00011ca/
Contact	Professor Shi Tianjian, Duke University

Name	survey on nationalTY IDENTITY in CHINA (54) 中国民族性调查
Principal Investigator	Social Psychology Research Institute, People's University
Year(s)	1987
Area (s) of interest	General National identity
Brief description	Dataset 1838 surveyed people between the age of 20 and 60 Sampling PPS
Availability of data descriptions	A book (《中国民族性》) has been published on the research based on the survey results in 1993
Contact	Sha Lianxiang sha@263.net.cn tel: 62511949

Name	Environmental Awareness Survey (55) 环境意识调查
Principal Investigator	Sociology Department People's University
Year(s)	1998
Area (s) of interest	General Public awareness Specific Environment
Brief description	Dataset : 3662 Sampling : PPS Data Format : SPSS
Contact	Li Qiang, Sociology Department, People's University (http://social.ruc.edu.cn)

Name	National Public Environmental Awareness Survey (56) 年全国公众环境意识调查
Principal Investigator	Research Centre for Contemporary China

Year(s)	1998
Area (s) of interest	General: Environment Specific: Public awareness
Brief description	Summary The project was proposed by State of Environmental Protection Administration of China and Ministry of Education of the PRC, designed and implemented by The Research Center for Contemporary China of Peking University. The method applied in this survey was a large-scale probability sampling. The national public individual data is based on face-to-face interview. The target population was defined to be the adults between 16 to 65 and teenagers between 10 to 15 living in the households in 31 provinces(Hong Kong, Macao, and Taiwan were not included), municipalities and Autonomous Region at the standard point of time of the survey, 0 P.M. on July 1 st , 1998. This survey applied the stratified, multi-level, probabilities proportional to size sampling method. The scheming sample size was 13,500 households, the national sampling ratio was 0.0042%. In the whole nation, 139 county-level sample unites, 11,630 sample households were chosen. Data Format: SPSS
Contact	Tel: 010-62755443 Email: rccc@rcccpku.org Peking University 100871

Name	Life history and social change survey (57) 社会变迁与社会现代化调查
Principal Investigator	Donald J. Treiman, Andrew Walder
Year(s)	2006
Area (s) of interest	Economy, social change
Brief description	Summary The survey covers 6090 adult samples Sampling PPS Geography 28 provinces, municipalities or autonomous regions in China
Contact	http://www.sscnet.ucla.edu/issr/da/lhscs/chinaweb.html

Name	Chinese Sexual Behavior Survey (58) 中国人性行为调查
Principal Investigator	Institute for Sexuality and Gender, People's University Pan Suiming

Year(s)	2000
Area (s) of interest	General Social development Specific Sexual behaviour
Brief description	Summary The survey covers 3824 informants aged from 20-64 in China. to make sense of the sexual behaviour of Chinese people. Sampling PPS Geography The whole country
Contact	Mail:book@sexstudy.org address: the sociology department in Renmin University postcode:100872 tel:62514498
	http://www.sexstudy.org/list.php?type=67&channel=2

Name	SURVEY OF CHINESE COLLEGE STUDENTS' SEXUAL VALUES AND SEXUAL BEHAVIOUR (59) 中国大学生性观念与性行为调查
Principal Investigator	Institute for Sexuality and Gender, People's University Pan Suiming
Year(s)	1997
Area (s) of interest	General Education, social development Specific The sexual values and behaviour of college students in China
Brief description	Summary The survey covers 1991 undergraduate students of Chinese college Geography Nationwide
Contact	Mail:book@sexstudy.org address: the sociology department in Renmin University postcode:100872 tel:62514498
	http://www.sexstudy.org/list.php?type=66&channel=2

Name	2006 Rural Social Capital Survey (60)
Principal Investigator	ISDPP, Beijing Normal University
Year(s)	2006
Area (s) of interest	Specific Social capital, community development and participatory poverty alleviation
Brief description	<p>Summary Survey to investigate how rural people's economic and social status influence their participation in local community affairs and their coping strategies for dealing with social risks.</p> <p>The questionnaire adapted the Social Capital Assessment Tool (SCAT) developed by Krishna and Shrader (1999), which measures social capital at the community level. Variables include basic social demographic characteristics of household members, household income and expenses, family coping strategies of social risks, and participation to organizations in rural areas.</p> <p>Dataset Sampling 2400 households from rural areas of 4 provinces</p> <p>Weighting</p> <p>Geography 1 village each from Shaanxi, Sichuan, Guangxi and Inner Mongolia</p> <p>Data Format Spss/sas</p>
How useful to researchers?	Uncertain, but probably limited except to small group of researchers due to narrow focus
Availability of data descriptions	Instruments available; otherwise limited background documentation.
Conditions	May be available by negotiation
Tabled outputs	Data set is undergoing cleaning
Contact	Professor Zhang Xiulan

Name	2003 Urban Social Capital Survey (61)
Principal Investigator	ISDPP, Beijing Normal University
Year(s)	2003
Area (s) of interest	<p>General Social capital, civil society participation</p> <p>Specific Urban community organisation; civil society, participation, social capital measures</p>
Brief description	<p>Summary To understand the changes in urban communities and the situation of residents' participation in civil society organizations</p> <p>The instrument used in the survey was adapted from the "Household Questionnaire" in the Social Capital Assessment Tool (SCAT) developed by Krishna and Shrader (1999), which measures social capital at the community level. Part of this instrument measures the quantity and quality of residents' memberships in civic society organizations. For the purpose of this study, only the part that measures residents' memberships in social organizations (organizational density) was adapted. A few modifications were made in the survey questions.</p>

	<p>One was in the list of organizations in the SCAT, which were replaced by those currently existing and in operation in China at the urban community level. Another one was in the category of occupations, which was also revised according to the classification commonly used in China. The interviews mainly explored the changes in urban communities and the operation of social organizations and groups at the community level particularly the Community Residents' Committee.</p> <p>Dataset Sampling 368 interviewees from 120 households from three urban communities in Beijing</p> <p>Weighting</p> <p>Geography Beijing</p> <p>Data Format SPSS/SAS</p>
How useful to researchers?	Limited due to narrow focus of topic and small size of sample
Availability of data descriptions	A paper published with more detailed description
Conditions	May be available by negotiation
Contact	Professor Zhang Xiulan

Appendix 2: List of individuals and organisations contacted

Name	Position / Title	Institution and address	Email	Phone
ATHAYDE, Chris	Assistant Representative	DFID, Beijing	c-athayde@dfid.gov.uk	
BAO Shuming	Director	China Data Center, Michigan	sbao@umich.edu	
BIKALES, Bill	Chief Economist	UNDP, Beijing	bill.bikales@undp.org	
CAI Fang	Professor	Institute of Population and Labour Economics	caifang@cass.org.cn	
DAVIN Delia	Professor, Retired	CASS		
		Leeds University, UK	D.Davin@leeds.ac.uk	
GU Baochang	Professor	Center for Population & Development Studies, Renmin University	bcgu@263.net	13910056983
HAO Daming	Doctoral Candidate	Capital University of Finance and Trade, Beijing		13811597072
		Department of Finance and Logistics, Ministry of Civil Affairs	hshsh@mca.gov.cn	
HE Shanshan	Deputy Director			58123308
HEATH, Douglas	Counsellor	CIDA	douglas.heath@international.gc.ca	
HODGES, Anthony		UNICEF, Beijing	ahodges@unicef.org	65323131
HOFFMAN, Bert	Chief Economist	World Bank, Beijing Office	bhoffman@worldbank.org	58617600
		The Asia Center, London School of Economics	a.hussain@lse.ac.uk	
HUSSAIN, Athar	Professor	Women's Studies Institute of China	jyongping@263.net	85116305
JIANG Yongping				
KRUGER, Mark	Head of Economics	CIDA, Beijing	mark.kruger@international.gc.ca	
LAI, Janet	Second Secretary	CIDA, Beijing	Janet.lai@international.gc.ca	65323472
LEE, Chin Kei	Epidemiologist	WHO	leec@chn.wpro.who.int	
		Department of Population & Employment NBS		68782666
LI Huimin	Director	Department of Sociology, Renmin University	lilulu@public.bta.net.cn	
Li Lu-Lu	Professor Chair	Department of Economics, Beijing Normal University	lishi89@263.net	
LI Shi	Professor	World Bank, Beijing Office	xliu1@worldbank.org	58617600
LI Xiaofan	Economist	Guanghua School of Management, Peking University	gordon@gsm.pku.edu.cn	
LIU Gordon	Professor, Director	Beijing Normal University, School of Education	bnuhz@163.com	13681203348
LIU Huizhen	Associate Professor	Center for Human Development, Peking University	minquanliu@hotmail.com	
LIU Minquan	Professor, Director	American & Oceania Division, CASS	liuqp@cass.org.cn	85195137
LIU Quanping	Project Officer			

MENG, Xin	Professor	Division of Economics Australian National University	Xin.Meng@anu.edu.au	
MO Rong	Deputy Director General	Institute for Labor Studies, Ministry of Labor and Social Security	morong@263.net	
PARK, Albert	Associate Professor	Dept. of Economics, University of Michigan	alpark@umich.edu	
PIEKE, Frank	Director	Center for Contemporary China, University of Oxford	Frank.pieke@chinese.ox.ac.uk	
SHEN Ming Ming	Director	Research Center for Contemporary China, Peking University	rccc@pku.edu	
SMITH, Kent	Head of Development	CIDA, Beijing	kent.smith@international.gc.ca	
SPOHR, Chris		ADB, Beijing	cspohr@adb.org	
TAN Lin	Professor, Director	Women's Studies Institute, All China Women's Federation	tanlin@wsic.ac.cn; tanlin@nankai.edu.cn	85116305
TANG Shenglan	Health & Poverty Advisor	WHO, BEIJING	tangs@chn.wpro.who.int	
WANG Fei	Foreign Affairs Office	Ministry of Labor and Social Security		13581506947
WANG Pingping	Statistician	National Bureau of Statistics, Rural Survey Team	wangpp@stats.gov.cn	
WANG Qian	Deputy Director	China Population Information and Research Center ??)	wqian1966@yahoo.com.cn	62131558
WANG Rong	Professor, Director	Peking University	rwang@gse.pku.edu.cn	
WANG Xiaomin		Research Center for Rural Economy, Ministry of Agriculture	wxm@rcre.org.cn; xiaomin11@hotmail.com	66115898
WANG Youjuan	Director	Urban Survey Team, NBS	wangyj@stats.gov.cn	13611061695
WANG Yubin	??	Ministry of Labor and Social Security	wangyubin@molss.gov.cn	
WU Zhigang	Deputy Division Director	RCRE, MoAg	wuzg@rcre.org.cn	
XU Jin	Chief, Division of Research Management	Institute of Population and Labour Economics, CASS	xujin@cass.org.cn	
XUE Haiping	Dr.	Graduate School of Education, Peking University	st462@gse.pku.edu.cn	13520552216
YANG Jianchun	Consultant Assistant	Department of Population & Employment NBS	yangjc@stats.gov.cn	
YANG Yang	Director General	Bureau of International Cooperation CASS	yyang@cass.org.cn	85196362
YAO Yang	Professor	CCER, Peking University	yyao@cser.pku.edu.cn	
YUAN Yue	Director	Horizon Research Group CCER, Peking	www.horizon-key.com	
ZENG Yi	Professor	University Institute of Population and Labour Economics	zengyi@cser.edu.cn	13522601381
ZHANG Juwei	Professor	CASS Institute for Labor Studies, Ministry of Labor and Social	Zhangjw@cass.org.cn	85195429

Security				
ZHANG Lihua	Deputy Director	American & Oceania Division, CASS	zhanglh@cass.org.cn	85195137 / 6142
ZHANG Linxiu	Professor, Director	China Center for Agricultural Policy, Chinese Academy of Sciences	lxzhang.ccap@igsrr.ac.cn	
ZHANG Shu	Head	Information department, NBS	stats@263.net ; xinxi@stats.gov.cn	63328230 / 6157
ZHANG Xiulan	Professor, Director	Institute of Social Development and Public Policy	isdpp@bnu.edu.cn	62200728
ZHAO Yaohui	Professor	CCER, Peking University	yhzhao@ccer.pku.edu.cn	
ZHAO Yutian	Director	Office for Fixed-Site Rural Survey RCRE, MoAg	yutianzhao@263.net	
ZHENG Xiaoying	Professor, Director	Institute of Population Research, Peking University	xzheng@pku.edu.cn	
ZHU Ling	Professor, Director	Institute of Economics, CASS	zhuling@cass.org.cn	64294016
ZHU Shouyin	Division Director	Research Center for Rural Economy, Ministry of Agriculture	(see Wang Xiaomin)	
ZHUANG Ya'er	Director	Information Service Department, CPDRC	yaerzh@vip.sina.com ; info@cpirc.org.cn	62173519

Appendix 3: Social Policy Analysis Information Center, ISDPP

In the past several years, ISDPP has prioritized its resources in building a social policy analysis information center to support its evidence based policy analysis, and teaching center. While hardware and software are critical, ISDPP will treat these as necessary but not sufficient conditions for the functioning of SPAIC. The key source of the competitive advantage of ISDPP's SPAIC will be its databases, of which there are four types. The first of these are data from large national and international databases relevant to social policy analysis in China and related comparative work (e.g., Chinese census data). The primary value added will be in assembling these in one place and providing ready access to these data. The second type of data relevant to social policy analysis in China are those released periodically or irregularly from sources such as Chinese provincial statistical department, for profit businesses, NGOs, and other academic institutions. Assembling such data obviously depends on the willingness to share them of the agency, individual, or company that collected them but ISDPP's credibility and academic/research mission will ensure it has an advantage in collecting and centralizing such data, to the extent that doing so is possible. The third type of dataset housed at SPAIC will be central and provincial government data collected and/or maintained by ISDPP under contract to the central government. As the repository of such data, ISDPP will obviously face certain constraints with respect to providing access, which will be handled through passcode-based access to both the network and datasets, and other standard security measures. The final type of dataset will be data collected by the ISDPP faculty and collaborators as part of their ongoing research.

The major resources in SPAIC:

1. China Charity Documentation Center: A Joint Center with the Ministry of Civil Affairs;
2. China Orphanages Information Center: the information on adoption and assistance for 577,000 orphans. It is jointly managed by the Ministry of Civil Affairs and ISDPP.
3. China Social Protection Survey Data (2004): 8 cities and 5 provinces with a sample size of 7,300 households and 29,000 individuals (The Ford Foundation). It is exclusively maintained by ISDPP.
4. China Rural Social Protection Survey (2005): 6,300 households in 4 provinces (The World Bank). It will be available for public use in one year when the World Bank finishes its final report.
5. China Urban Heating Survey (2005): 1,920 households in 4 cities (The World Bank). It is available for public use in September 2006.
6. Wuhan and Shenyang Laid off Worker Survey (2003): 7,800 households (The World Bank). It is obtained from the World Bank, and needs to get approval from the World Bank for public use.
7. Rural Medicaid Institution Survey (2005): 30 provinces and 2,000 counties (The World Bank). It will be available for public use in one year when the World Bank finishes its final report. It is anticipated the data will be available for public use in December 2006.
8. Urban Medicaid Institution Survey (2005): 200 cities. It will be available for public use in September 2006.
9. Urban Medicaid Household survey (2005): 4,000 households in 200 cities. It will be available for public use in December 2006.
10. Medicaid Provider Survey: Xining, Baiyin and Lanzhou cities. It will be available for public use in September 2006.
11. Village Governance Survey (2005): 120 villages nationwide. It is available for public use in December 2006.
12. Situation data of rural communities (2004): 120 villages nationwide. It is available for public use in December 2006.
13. Beijing and Shanghai Informal Long Term Care Survey (2003): 400 households. It is available for public use in September 2006.
14. Social Assistance to HIV/AIDS Families and Children (2005): 1,700 counties (ISDPP). It will be available for public use in December 2006.
15. MoCA 100 city community survey (2005): It is available for ISDPP but needs to negotiate a public usage agreement.
16. Linkage with MoCA's Archive Center (before 1949, and since 1949): Access is limited. Data Sharing with MoCA Statistical Center. Access is limited.
17. Data Sharing with Economic Census at NBS. Access is limited.

18. Social Policy Link: 10,000 websites linkages. It is available for public use in December 2006.
19. Other government statistics and World Bank Development Indicators. It needs more work.
20. Social Policy On-Line Documents; over 20,000 English and Chinese documents. It is available for public use in December 2006.
21. Book Sharing and Library Establishment: Faculty donated Chinese and English books.

On-going longitudinal surveys:

In addition, two large and on-going longitudinal surveys are underway:

Social Development Monitoring System

This three million RMB project is intended to collect extensive information on urban China's progress toward social development. It randomly surveys households in three cities in Northwest China's urban areas (when funding is available, rural areas will be included). The questionnaires contain modules concerning household composition, income and expenses, social program benefits, social assistance and social insurance, social and health services, family relations, education and training, etc. Each household will be followed every six months for three years, and more years if additional funding becomes available. Questionnaire modules in the follow-ups will contain additional information besides the core modules in the baseline. It is anticipated that the survey cities will be extended to central and eastern China, and funding for this system is vigorously sought. The baseline surveys were finished in the last quarter of 2005, and the primary analysis on the data indicated the very high quality of the data collected.

China Child Monitoring System

This three and half million RMB project will be partially funded by the Plan International to collect extensive information on children in both rural and urban areas. Its aim is to follow 8,000-10,000 households and 110,000-140,000 individuals, with an over sample of disadvantaged children, such as children with HIV/AIDs, orphans, children with disabilities, children living in broken families, children living in migrant families, etc. These children will be assessed every two years for three times over six years. The questionnaires include modules to the family situation, children's education, nutrition, psychological development, etc. Community and schools will be included to assess the living environment.

Continued Effort

In addition to the above mentioned projects, several new projects will allow SPAIC to obtain information on HIV/AIDs families and third sector organizations.

ISDPP has prioritized the usage of the above data and documentation for full time ISDPP faculty members and students. Most of the students at ISDPP wrote/are writing their theses based on these data. However, given the vast information collected and maintained by SPAIC, and given the demand for detailed data on the social development and social programs, SPAIC is positioned to offer its services in social surveys and limited government administrative data to academic and domestic academicians, social policy researchers and students, and also use the information to provide teaching courses in social policy analysis.

Amassing, housing, and providing access to the data center will not only improve the research capacity of ISDPP but will also enhance its attractiveness as a site for collaboration with other domestic and international scholars. Such partnerships are essential to academic enterprise and it is anticipated that ISDPP's data infrastructure will be a source of competitive advantage in attracting research funding and productive research collaboration. As one of the leading institutions in social policy in China, ISDPP aims to become a domestic and international focus of social policy exchange and SPAIC will serve as a linkage.

SPAIC Foci:

Amassing and housing of information: This will include standardizing the questionnaires, describing the database, providing a dictionary of the variables; providing both Chinese and English

versions of the questionnaires and data reports, securing the raw data, and periodically providing on-line reports on data availability and publications, as well as linking the micro-level databases with macro-level data files. A passcode management system will be installed and maintained by SPAIC.

Processing data usage requests: This will include reviewing data usage requests for both research proposal based usage and individual usage for publications and thesis writing; providing sample data and example programs in various programming languages to users to test out models; and final data analysis and result explanation and write-ups of the analytical results (when they are needed) to domestic users (Chinese versions) and international users (English versions).⁶ User fees will be used to finance this activity. Currently the user fees are set as follow: \$500 for international, \$100 for domestic and \$50 for BNU and free for ISDPP researchers.

Provide short term training courses: SPAIC will provide short term training courses for students and researchers in key database usage. The courses will cover the sampling framework, survey procedures, questionnaire design, hypothesis testing, and programming of models. The courses will be bilingual (Chinese and English).

⁶ China Statistical Law does not allow the raw data to be out of the county. This step is necessary to ensure the SPAIC follows this law.