Socio-economic Databases as a Support System for Inter-disciplinary Research : Indian Scenario

P.R. Goswami, Director, National Social Science Documentation Centre, Indian Council of Social Science Research, New Delhi, India

International Federation of Library Associations, Social Science Libraries Section, Satellite Conference

Disappearing disciplinary borders in the social science library - global studies or sea change?

University of Toronto, Toronto, Canada
6-7 August 2008

Sponsored by:

CQ Press
Worldwide Universities Network (WUN)
Center for Global Studies (CGS) at the University of Illinois
University of Illinois Library
University of Toronto Libraries
Faculty Information Studies, University of Toronto

Introduction

Globalization (i.e. increasing degree of economic interaction among countries) and liberalization (i.e. less government intervention in market) are important themes for research in most of the developing countries. Both have contributed to debate ranging from the very optimistic to very skeptical. In fact, globalization is a multidimensional phenomenon and it extends much beyond the economy to polity and society (World Commission on the Social Dimension of Globalization, 2004). Researchers from non-economic disciplines are trying to measure how these two factors have affected inequality (and also poverty) at the national and regional level. In India, a country with many languages, castes, religions as well as many customs, rituals and traditions, development is judged from different angles using different sets of quantitative and qualitative data. The pluralist Indian society has also given rise to interdisciplinary research transcending the so-called boundaries of social science disciplines. One can also say that economic globalization, the Internet and World Wide Web with easy and real time access to a wide variety of information, research funding by international agencies and a system of evaluation that encourages publishing in high impact journals are also the factors working towards the homogenization of social sciences.

> Apart from socio-economic issues, the Indian government and its policy making and administrative machinery are required to deal with problems of ethnic, religions and other differences that often promote militancy and result in law and order problems. In order to develop a better understanding of these differences and their social and political implications, ideas are drawn from many sources particularly those belonging to hyphenated social science subjects such as political sociology, political psychology etc. It is often said that networks of cross-disciplinary influences are such that they are obliterating the old classification of the social sciences. In fact, there is a movement from the old formal academic disciplines towards new hybrid social sciences.

Moreover, World Conference on Science emphasizes the necessary enhancement of dialogue between science and society. International Forum on the Social Sciences, a body supported by the UNESCO, endorses the viewpoint that addressing hunger and poverty, lack of education, poor health and environmental degradations – the five areas to which the eight Millennium Development Goals relate - is crucial for human welfare, social and economic development, the achievement of social cohesion and the strengthening of democratic governance. One can say that

none of these five areas is solely within the purview of social science but without social sciences, none is fully understood or capable of being properly investigated. All these factors work in favour of interdisciplinary research in social sciences.

The purpose of this paper is to familiarize the LIS professionals with the statistical databases that act as knowledge support system for interdisciplinary research in social sciences. Recent changes made in the Indian statistical system to meet the growing needs of research have also been discussed.

1. Social Science Research: Recent Trends

In India, two kinds of institutional structures are prominent today for social science research. They are i) universities wherein research is dispersed across individuals; and ii) research institutes where the model of quantitative research based on empirical evidence has worked successfully for satisfying the criteria of output and accountability [John 2008]. The Chairman of the Indian Council of Social Science Research (ICSSR) has recently said that the government have policy requirements which are pressing requirements and which therefore, require research which will yield results in the short run…….. it will be a great pity if immediate return research (i.e. sponsored research mostly done in the institutes) is allowed to displace delayed return research (i.e. research done by the individuals in universities) [IEG 2007].

In fact, the social sciences changed significantly in the years following independence (i.e. 1947). The new Indian states' commitment to economic development through centralized planning and measures to protect socially and economically weaker sections were responsible for creating new areas of social science research. However, prior to independence, British officials, travelers and missionaries collected a considerable quantity of basic information on the cultural diversity and resources of India, its people, ecology, agriculture, land revenue and regional problems. More organized and regular data collection began with the starting of Population

Census in 1871, which became more elaborate and regular in the following decades. The Ethnographic Survey of India led to the compilation of Imperial Gazetteer of India and many volumes of Indian castes and tribes living in different regions. There was also the Linguistic Survey of India, which provided data on languages and dialects of the country [Srinivas 1987].

Similarly, Indian states' policy to promote social and economic development and carry forward social welfare to the weaker sections of the society mostly living in rural areas created demand for information. Newly established agencies like the Planning Commission, Indian Statistical Institute (ISI) and other autonomous institutes such as National Council for Applied Economic Research (NCAER) etc. started collecting data through surveys on different socio-economic variables. And a few years later, the Indian government laid the foundation for statistical information system to assist in planning and policy making. Now India has a well developed statistical system which is deemed as an integral part of knowledge base of the state.

2. Statistical Databases: Structure and Contents

Researchers in social sciences are avid users of statistical databases. And for them, statistics is not viewed as a purely mathematical discipline isolated from life. It is a practical science whose main concern particularly in developing countries is promotion of human welfare through collection of data and information for planning and implementing various development programmes.

Needless to say, socio-economic data provide vital raw material for research. Statistical literacy, i.e. ability to understand and critically evaluate statistical results is now considered an `essential qualification' for a researcher. Raw data, graphs, charts, rates, percentages, probabilities, averages, forecasts and trend lines are now considered part of research processes. Time series data on different socio-economic

variables and their analysis using various econometric models are utilized for the purpose of resource allocation, for planning in areas like health, agriculture, industry, employment and many other matters. Statistical data on Indian society and economy available at national, state and district level on different variables have their origin in government through three major processes [Goswami 2001]:

- Collection on repetitive basis by official (i.e. government established) statistical agencies through censuses and sample surveys. The examples are Population Census (done every ten years) and Sample Surveys on subjects like employment, health, education, housing conditions, household spending/expenditure etc.
- Collection on continuing basis by government regulatory agencies for administrative purposes by using statistical returns. The examples are Income Tax Statistics, Foreign Trade Statistics, etc. This process is also known as routine collection of data as a by-product of administrative activities of the government.
- Ad-hoc collection of data through surveys for specific purposes by specialized agencies and research institutes as part of their research projects. The examples are surveys conducted by National Council of Applied Economic Research (NCAER), International Institute for Population Science (IIPS) and other autonomous agencies.

There are many statistical agencies in the country, which collect, process and disseminate data on a wide variety of subjects. The statistics wing in the Ministry of Statistics and Programme Implementation (MOSPI) is the apex body in the system. The vision statement of this organization is

> "To be recognised as an efficient apex organization for timely dissemination of reliable and credible statistics consistent with international standards to all users"[India 2007].

One of the major responsibilities of this apex federal body is to coordinate statistical work in the country with the view to identifying gaps in data availability or duplication of statistical work and to suggest necessary remedial measures.

Following is a list of certain important federal statistical Agencies in India and their databases [Goswami 2001]:

a. Central Statistical Organization (CSO)

a.1 The CSO is one of the major statistical agencies which function under the direct control of the MOSPI. The CSO coordinates statistical activities of central and state agencies and keeps a close liaison with international bodies like UN, IMF etc.

a.2. Major databases developed by the CSO include i) National Accounts Statistics (NAS), ii) Annual Survey of Industries (ASI), iii) Index of Industrial Production (IIP), iv) Economic Census (EC), v) Consumer Price Index for Urban Non-Manual Employees, vi) Energy Statistics, vii) Social statistics.

b. National Sample Survey Organization (NSSO)

b.1. NSSO is the primary data collection wing of the MOSPI. Data is collected regularly through nation-wide household surveys on various socio-economic subjects such as household consumption, employment and unemployment, agriculture, social sectors (including education, mortality, maternity and child, medical services, public distribution), Housing condition, prices (including data on rural credit).

b.2. In addition, NSSO also collects data for Annual Survey of Industries (ASI) under collection of Statistics Act and Enterprise Surveys as a follow up of the Economic Census.

b.3. NSSO collects data on rural and urban prices and also plays an important role in the improvement of crop statistics.

c. Registrar General of India (RGI)

c.1. RGI is responsible for conducting population census and registration of births and deaths in the country.

c.2. Census of India is the largest repository of data. It contains data in various forms, which include classification of person by sex and age, marital status, rural/urban residence, educational standard, economic activities etc.

d. Labour Bureau (LB)

d.1. LB is responsible for developing databases on labour and employment. Data are collected from different regulatory agencies and statistical organizations like NSSO, RGI etc.

d.2. Most of the employment data are collected by LB as by product of administration of various labour laws operating in various sectors.

d.3. Data on employment and unemployment of agricultural Labour are collected through Rural Labour Enquiry Survey in collaboration with NSSO.

e. Reserve Bank of India (RBI)

e.1. The RBI is responsible for developing database on banking and finance. The source of data for RBI is statistical returns submitted by banks according to Banking Regulation Act, 1949.

e.2. The RBI also conducts ad-hoc surveys on banking business particularly in rural areas.

f. Directorate General of Commercial Intelligence and Statistics (DGCIS)

f.1. The DGCIS has created a foreign trade database. The database developed by the DGCIS is basically a by-product of regulatory mechanism of the government.

f.2. Statistics of India's foreign trade are compiled according to general system, i.e. all goods entering the customs' area by land, air or sea are called imports; and

similarly all goods which are products of the country and which leave the customs frontiers previously imported and subsequently exported are called `re-export'.

g. Directorate of Economics and Statistics (DES), Ministry of Agriculture

g.1. DES is the apex body for agriculture statistics in India, working as the coordinating agency for data collection and processing of results.

g.2. The collection of agriculture statistics is the responsibility of the states. Database pertaining to area and yield and `crop estimation' surveys are conducted by the states.

g.3. Agricultural census is conducted every five years. Livestock census is also coordinated by the DES.

Apart from certain major federal statistical agencies, there are state statistical offices and in many states, there are district statistical offices. On the whole one can say that India has a vast network of statistical agencies and databases pertaining to different socio-economic variables are produced in a distributed manner.

3. Statistical Databases: Towards Interdisciplinary Research

In the present era of globalised market economy, release of timely and reliable data on variables that related to I) poverty estimation, ii) human development and particularly iii) impact of globalization and liberalization on the weaker or poorer sections

of the society has become imperative. On the demand side, national statistical system is getting a new group of data users whose varied needs are to be identified and met. And on the supply side, technology is making it possible to provide data in the form of electronic products (CD-ROM, on-line databases etc.) Recent innovations in the filed of information and communication technology (ICT) are offering new opportunities to developing countries to utilize their knowledge base for research in diverse areas.

In India, several new initiatives have been started by the statistical agencies to support interdisciplinary research. They relate to I) expansion in data collection, tabulation and dissemination programmes; ii) new subjects for conducting sample surveys and iii) special surveys to ascertain the present state of education/literacy, health, and other related variables; iv) policy measures, pertaining to release of `unpublished' `validated' unit level data to the researchers.

As regards sample surveys, NSSO has expanded its data collection programmes keeping in view the recommendations of the National Statistical Commission. It says, "Under continuing globalization and liberalization of the Indian Economy, to monitor rapid socio-economic changes, and to meet international spatial data dissemination standards, Government of India would need reliable and timely statistical information on a much large scale than before. NSSO, as the only organization in the government with requisite technical capability and experience would have to take up this responsibility" [Vidwans, 2002].

NSSO has recently released several reports based on Survey on Household Consumer Expenditure, Employment and unemployment (2004-2005). Following are the titles of some of the reports.

- NSS Report No. 508: Level and Pattern of Consumer Expenditure, 2004-2005.
- NSS Report No. 510: Public Distribution System and Other Sources of Household Consumption, 2004-2005.

- NSS Report No. 511: Energy Used by Indian Households, 2004-2005.
- NSS Report No. 515: Employment and Unemployment situation in India, 2004-2005.
- NSS Report No. 518: Participation of Women in Specific Activities, 2004-2005.

Consumer expenditure data are used for the preparation of estimates of poverty in India. The economists have repeatedly raised the issue of deficiency in NSSO data. It has also been said that there is a need to adjust poverty estimates keeping in view the increased consumption of health and education services by the household [Mahendra Dev 2008].

On the other hand NSSO has released databases on backward castes and religion making it possible to disaggregate across a range of interlocking social categories based on caste and religion. One can say that the Indian Statistical System is witnessing the extension of economic frames and quantifiable techniques into other areas and classes of people whose identities and histories remained out of the definitions of development.

Similarly, Population Census of India 2001 has published literacy and work participation data by religion for the first time after independence (i.e. 1947 A.D.). The census data also provides an opportunity to analyse and understand nearly 1700 religious groups that constitute India's religious and cultural diversity. However, for a detailed analysis, unpublished `cross-religion' data are to be made available to the researchers [Bhagat 2004].

National Family Health Surveys (NFHS) conducted in the years 1992-93, 1998-99 and the third phase in 2005-2006 provides valuable demographic and health information on India. One of the special features of NFHS is the application of quantitative methods and survey techniques to a large number of family health related questions and certain socio-cultural issues, which are normally described in subjective

terms. These includes questions pertaining to a) fertility, b) family planning and contraceptive practices, c) maternal and child health and mortality, d) nutrition and anaemia, e) sexual behaviour, f) women's empowerment, g) domestic violence [Irudaya Rajan 2007].

One can critically examine the questionnaire used in NFHS to appreciate the extent to which the lives of the people are probed in order to get detailed data on the above socio-cultural and health issues. One of the features of NFHS is its timely publication and wide dissemination of results of state and national level. NFHS datasets are squarely in the public domain and they are not in-house documents meant only for the sponsoring bodies. Moreover, these reports can be taken as a model for `best practices' in such type of survey methodology are available to the interested scholars free of charge. Also, there are special volumes providing information on the sampling design as well as questionnaires used [John 2008]. Easy availability of NFHS reports has led several researchers in India as well as abroad to work extensively on India's demographic diversity and human development issues.

Human development indicators (HDI) is now one of the major concern of the fast growing developing nations. In India, many states such as Madhya Pradesh (1995, 1998 and 2002), Karnataka (2000), West Bengal (2004), Kerala (2007) have brought out Human Development Report (HDR); and in several other states HDR is under preparation. A national HDR has been brought out by the Planning Commission in collaboration with the United Nations Development Fund (UNDP). However, the experts say that HDI approach reveals that the kind of growth the countries like India and China are experiencing does not necessarily come with better income distribution. The growth allowed by the global economy in these days is a growth that does not reach the weaker and poorer section of the society. One cannot deny the fact that no progress is feasible without fast GDP growth. But GDP growth alone cannot ameliorate the conditions of poor. There have to be social policies, inclusion policies and agricultural policies that try to help poor and weaker sections.

In India, new indicators get invented as new social problems are discovered. Sometimes, it becomes the other way, i.e. the indicators can also accidentally reveal the presence of a social problem. For instance, population census results have shown an alarmingly declining sex-ratio in some parts of India indicating presence of an evil like female infanticide.

Another notable feature of India's socio-economic data system is the National Policy on Dissemination of Statistical Data. According to this policy, the apex body i.e. MOSPI provides validated unit level data to the researchers for their analytical studies. Now a large volume of data generated through various socio-economic surveys, enterprise surveys, economic censuses (i.e. census of unorganized sector), annual survey of industries and other statistical operations are preserved for the users on CD-ROM and other digital media. These data are disseminated to a large number of national and international users on special requests. Technical guidance for the use of basic data and their processing is also given to the users. Indian government has now decided to create a data warehouse of official statistics to enable the data users to have easy access to published as well as unpublished validated data from one source [India 2007].

4. Statistical Databases : Other Issues

In India, there has been a quantum increase in the availability of statistical information after independence. However, the quality of data is often questioned by the researchers. There are wide gaps in data pertaining to unorganized sector and rural income. Moreover, there is a shortage of qualified researchers who can undertake in depth analysis of income inequality, its trends, its determinants, and impact of policy instruments. New sources of knowledge and more nuanced understanding of socio-economic processes in a trans-disciplinary mode are needed to study rural poverty and inequality. A mere quantitative approach is not valid because of its apolitical nature and general lack of explanatory capacity.

There are certain other issues that need to be addressed for providing better `knowledge base' to interdisciplinary researchers in India.

a) As a result of liberalization of Indian economy, there has been considerable erosion of the statistical system. The dismantling of several regulatory bodies and their procedures, which used to generate economic data on different subjects, has contributed to the decline of the system. The statistical agencies are required to take a fresh look on this matter.

b) Data for the construction of HDI is generally available at the national level though it may vary in quality, reliability and timeliness. One can encounter serious problem while trying to obtain disaggregated data on human development for women and men, for different ethnic groups, minority communities and others at the level of states and districts. It is only by monitoring trends one can trace progress. Average and macro-level data mark serious inequalities and low levels of achievements. One can also find that much of empirical social science is about the poor and underprivileged – social and economic strata, which is lower than that of the researcher. It is very difficult to get good data on the rich and powerful because they can deny access to researchers and otherwise exercise control over whether and what kind of information is available on them [Deshpande, 2006].

c) One of the major reasons which contribute to the improvement in the quality of statistical data and their timely release is extensive use of data sets by policy makers in the government and academic researchers. In India, users' feedback has resulted in substantial improvement in various datasets. However, use of data by policy makers is limited because Indian planners and policy makers are not inclined to base their decisions on statistical evidence. There is an absence of analytic tradition in Indian officialdom.

d) The students of economics and other social science disciplines are not exposed to the vast myriad of data sources produced by the statistical system. One of the reasons for this ignorance is the lack of directories and guidebooks on the sources and nature of both official and non-official socio-economic data. Two major publications brought out by the CSO i.e. Guide to Official Statistics (1987) and Statistical System in India (2002) are not revised regularly. More such `guide to literature' type of publications are needed for the users of data on the lines of Directory of International Statistics published by UN Statistical Office [Dasgupta 2002].

e) The users of datasets are required to familiarize themselves on the availability of both published and unpublished data. Regular workshop/seminars in subjects like research methodology are needed to discuss sources of generation, explanatory notes regarding concepts, coverage, time lag and level of aggregation and disaggregation of datasets. Directories and guidebooks are required to understand the content and structure of various socio-economic datasets and also to enhance the utilization of data. In addition, analysis, interpretation and integration of various types of data, facilitate their use by researchers. Of late, Indian statistical agencies have started releasing time-series data on important variables maintaining their consistency and comparability.

f) EPW Research Foundation, Mumbai also publishes special statistics series. Economic & Political Weekly, the most widely circulated social science periodical in India, also publishes data on various macro-economic and social indicators. Centre for Monitoring Indian Economy, Mumbai publishes socio-economic data in a repackaged form in its serial publications.

g) Another factor, which needs attention, is the indexing of rich data contents of research papers in primary journals. They include field level data collected by the authors and also repackaged data gleaned from primary sources often in a disaggregated form. In the domain of science and technology, the system of preparing data-descriptive research to indicate the data contents of primary scientific literature is known as `flagging and tagging of data'. Thus, an example of index to statistics is

different from an index to statistical publications. Index to International Statistics introduced by Congressional Information Service, USA is a distinct example as it provides a comprehensive and up to date index of statistical information published in serials, monographs and reports [Dasgupta 2002].

On the whole, it can be said that the Indian government has expanded its statistical system to meet the needs of researchers and policy makers. However, it has not paid adequate attention to wider and user-friendly dissemination of statistical data. An important policy decision taken by the government is to make available validated data through unpublished, including unit/household/establishment level data to the users' community. The unit level data is released after deleting their identification on particulars to maintain confidentiality. The impact of this new policy is good, as many agencies have started releasing their data in electronic form with a reduced time lag. Major agencies such as CSO, NSSO, RGI and others have started marketing their datasets announcing the availability of unit level data to the users against nominal payment.

Concluding Remarks

Globalization and liberalization of Indian economy has an impact on the social science research in India. Trans-disciplinary research in social sciences on subjects like i) human welfare measures and their impact on various social, religious and cultural groups; ii) good governance; iii) inclusive growth; iv) civil society movement etc. are gaining importance. Statistical agencies and research organizations are building databases and social indicators to support research on these areas. Regular statistical programmes such as decennial population census and National Sample Survey are expanding their data collection schedules and dissemination programmes to include new subjects.

Needless to say, the spread of markets as a result of globalization has an influence on research agenda of universities, as resources for research in life sciences, medicine, engineering, or economics are abundant while funding for research in philosophy, linguistics, history or literature is severely restricted. In fact, there is a premium on applied research (even in social sciences) and a discount on theoretical research. Moreover, globalization is also working towards harmonization of academic programmes. The professions like law, management, economics, etc. which have a close affinity to social sciences are becoming internationalized (Nayyar 2007]. As a result, their context is more global and less national.

In order to promote global interdisciplinary research, a great deal of data are how available at international level mostly compiled under the aegis of World Bank, United Nations Statistical Office, International Monetary Fund and others. The problem with the data provided by these bodies is that "they are often reconstructed rather than directly observed …… the reconstruction is often cruder with more guesswork, than analysis. There are enough arbitrariness in these figures to make the purists shrink at the thought of touching such data. Given the scarcity of reliable primary information the purist position can almost lead to abstinence from empirical work altogether" [Sen 2004]. Many international organizations publish a number of data serials and reports providing comparative estimates for different countries of the world. These data series are collected from respective national statistical offices.

One can say that LIS professionals working in social science libraries are required to understand and interpret various data series that are needed to support interdisciplinary research. A carefully designed training workshop to understand the nuances of statistical data collection, tabulation, processing and dissemination; and sources of grey literature (including results of ad-hoc social surveys) would be useful. Such programmes are to be organized in collaboration with stakeholders i.e. user groups and statistical agencies.

REFERENCES

Bhagat, R.B. (2004). Census 2001 and religion data : Fact and fiction. Economic and Political Weekly 39(39). p.4292-4294.

Dasgupta, A.K. (2002). Restructuring the statistical system. Economic and Political Weekly 37(15). p.1448-1449.

Deshpande, Satish (2006). From indicators to entitlements : The Politics of social statistics. Round Table on Depoliticizing Backwardness – Alternative Approaches, 29-30 August 2006. Delhi. Department of Geography : Delhi School of Economics.

Goswami, P.R. (2001). Statistical databases in India : Towards a better dissemination. International Information and Library Review. 33. p.181-196.

India. Ministry of Statistics and Programme Implementation (2007). Annual Report, 2006-2007. New Delhi.

Institute of Economic Growth (2007). Proceedings of the inaugural function of the Golden Jubilee Year. December 15, 2007. p.14.

Irudaya Rajan, S. and James K.S. (2007). Second National Family Health Survey : Emerging issues. Economic and Political Weekly. 39(7), p.647-651.

John, Mary E. (2008). New Agendas for social science research. Economic and Political Weekly. 43(5). P.36-38.

Mahendra Dev, S. and Ravi. C. (2008). Revising estimation of poverty. Economic and Political Weekly. 43(10). March 8, 2008. p.8-10.

Nayyar, Deepak. (2007). Globalisation : What does it mean for higher education? Economic and Political Weekly. 42(50). p.30-35.

Sen, Amartya (1994). Economic regress : Concepts and features. Proceedings of the World Bank Annual Conference on Development Economics, 1993. Washington DC : World Bank. P.318.

Srinivas, M.N. (1987). Development of Sociology in India : An overview. Economic and Political Weekly. 22(4). p.135-138.

Vidwans, S.M. (2002). Indian statistical system at crossroads II. Expansion of National Sample Survey. Economic and Political Weekly. 37(38). p.3943-3955.