

Prospects and Problems In The Use of Government Generated Data in India

by N.K. Nijhawan¹
Director, Data Archives,
Indian Council of Social Science Research
(ICSSR),
New Delhi

This paper attempts to discuss the nature and extent of government effort, in India, in generating various kinds of socio-economic data and also to examine the problems faced by researchers in having access to such data. India is one of the few countries in the world which has such an elaborate, comprehensive and multilevel data-base. But it must be admitted that very few of these data are put to effective use. The government departments and its other organizations have a limited purpose and accordingly analyse the data only marginally. The researchers in the institutions and university departments who have interest, necessary skills and competence to use data in greater depth

and dimensions do not have access to these data. Thus despite a 'data revolution' in the country, both policy makers and researchers are unable to make effective use of a valuable data base for various reasons. An attempt has been made to analyse some of the problems underlying this process and also to explore possibilities that exist for remedying the situation.

Historical background

The present statistical system in the country has evolved through centuries. It dates back at least to the sixteenth century, when the Mughal rulers had data collected and tabulated systematically for about four thousand geographical units of the Mughal Empire. A comparable system for that century may not be found in any other part of the world. The nature of tabulated data also indicates that the Mughals may have borrowed some of its important features from the system prevalent in an earlier period. It seems even the British rulers, in India, leaned heavily on the data system developed by the Mughals during the early period of their rule. The British, however, transformed, and in many ways enriched, the system as is evident from the fact that the first gazetteer was published in 1866, the statistical abstract of British India in 1868, and the first all-India population census conducted in 1881. In 1895, a statistical bureau was established at the Centre to coordinate agricultural and foreign trade statistics and to collect data on prices, wages and industry. A number of Committees and Commissions were set up from time to time to examine in detail the available statistical material with a view to improving its quality.

¹Presented at IASSIST/IFDO International Conference May 1985, Amsterdam. The views expressed in this paper are those of the author and not those of the ICSSR.

Post Independence Developments

A real breakthrough in the official statistical system, however, came only after independence. A nucleic statistical unit called Central Statistical Organization (CSO) was set up in 1951. Subsequently, State Statistical Bureaux (SSB's) were set up in almost all the States. Following this statistical offices were set up, each in the charge of a district statistical officer. Like the CSO at the centre, the SSB's are entrusted with the task of coordination and publication of statistics collected by different government departments and district statistical offices.

At the instigation of, then, Prime Minister Jawaharlal Nehru, a National Sample Survey (NSS) was started in 1950 by the government to expeditiously collect data relating to all aspects of the national economy, on a continuing basis, using sampling methods. Beginning with the first survey (called a Round) in 1950-51, the National Sample Survey Organization (NSSO) is about to complete data collection for the 40th Round by July, 1985. Currently, NSSO has a large staff. Its field staff alone consists of about 5000 full time research investigators and field supervisors engaged in data collection.

The need for collection of relevant data for the purposes of National Five-Year Plans, as well as for meeting the demands of various international organizations, has been increasingly felt since independence. A large body of these data are collected by the government in addition to those that are usually collected for its routine and normal functions. It is appropriate to say that the country is witnessing data explosion since 1951. A brief mention of the nature and extent of socioeconomic data generated by the government during the past three decades or so is in order.

Nature of Data

Four decennial all India population censuses have been completed. Basic population characteristics for all the districts (approximately 365) are available in published form for all four censuses. Another extremely important tabulation of the Indian census which goes right up to the village and urban blocks called the primary census abstract (PCA) are available for each village/town in the district census handbooks. The census data also serve as the major source of basic economic information about the population. Besides data on population composition and distribution, data on fertility and mortality are collected under the vital registration system.

Considering the significance of agriculture in Indian life, information on agricultural holdings, their characteristics such as: size, form of tenancy, land utilization, production of crops, irrigation, livestock, prices, wages, etc., are being collected systematically by various government agencies. Data pertaining to land utilization, areas under various crops and area irrigated are almost continuously available in published form since 1884. Apart from the data collected from village records, NSSO has also collected data on land holdings, land use, agricultural production, cost of cultivation, etc., in its various Rounds. Besides, the NSSO carried out a special survey on land holdings, at the instigation of FAO, in its 16th and 17th Rounds to meet the requirements for the 1960 World Agricultural Census.

Unlike agricultural statistics, data on Indian industry is of relatively recent origin. Before independence, industrial statistics in the country, were extremely inadequate and of poor quality. After independence, legislation had to be passed to facilitate collection of statistical data relating to industries. A breakthrough came with the passing of the Collection of Statistics Act, 1953

and the Collection of Statistics (Central) Rules, 1959. An annual Survey of Industries (ASI) has been carried out since 1959, covering all the registered and licensed factories in the country. A number of government agencies periodically collect data on employment, output, earnings, etc., related to small scale and cottage industries. At present, the data relating to the industrial sector are available from at least twenty different sources of the government.

The National Sample Surveys are a rich treasure of data encompassing a large spectrum of socio-economic and cultural life of Indian society in the post-independence era. The NSS are multipurpose surveys covering several socio-economic and other subjects simultaneously. Comprehensive data have been collected, in various Rounds, on different aspects of agriculture including land utilization, cost of cultivation, rural indebtedness, rural labour etc. Likewise, a great deal of effort has gone into data collection in various surveys on consumer expenditure, demographic characteristics of sampled households, internal migration, employment and unemployment, housing conditions, trade, transport and rural electrification.

The quantity of data being collected through these surveys is quite voluminous. Just to give an idea, it may be mentioned that the sample size varies from Round to Round, but normally a survey covers about 5000 urban blocks including 200,000 sampled households. Sometime back, at the initiative of the Indian Council of Social Science Research, an estimate was made of the number of schedules/pages canvassed in various NSS Rounds. It was estimated that in the first nineteen Rounds alone (1950-51 to 1964-65) over two million schedules, comprising over nine million pages, had been canvassed. The large quantity of information collected during the course of all these years can thus be imagined.

Lack of perspective

The above account, which is purely illustrative in nature, clearly reveals the nature and extent of data being collected by the government. By any standards, India can claim to be one of the richest countries in the world in terms of data resources. But at the same time it cannot be denied that both policy makers and researchers are unable to make effective use of this resource, as a large part of the available data is put to a very limited use. The obvious question to ask is: why is that so? The single most important factor responsible for this sorry state of affairs is the government's lack of perspective in this regard. It may be worth noting that there is a disproportionate emphasis on data collection as opposed to data utilization. Emphasis continues to be on data collection rather than on its utilization and improvement of its quality. The main problem, therefore, is that the effort in data collection has multiplied manifold in the past 35 years, but the techniques of handling data continue to be quite rudimentary. Most of the data collected by the government is tabulated manually even today. This results in long delays, even in bringing out limited tabulated data. For instance, the results of the 31st NSS Round on rural electrification, conducted in 1976-77, were published as late as January, 1984. Likewise, although the summary results have been released, the detailed results of the 1978-79 Annual Survey of Industries are still being processed by the CSO. The last detailed results of ASI available for public use relate to the year 1973-74.

Notwithstanding unusually long delays in the publication of government-produced data, scholars can make little use of them owing to high levels of aggregation in the published data. Most government agencies collect data at a micro level, such as, village, household or industrial unit, but these are not made available to researchers. The analysis of micro data at

the village level would have been immensely useful to the researchers, but such information is not available to them. Only all-India and State-level results are normally published by the government.

Lack of Uniformity

Multiplicity of government agencies in the collection of similar data also adds to the difficulties of the researchers. For instance, as pointed out above, data relating to Indian industrial sector are available from at least twenty different sources. These alternate sources of data differ in their conceptualization, coverage and measurement procedures. For example, the Annual Survey of Industries (ASI) gives the gross output of industry in value terms whereas the Index of Industrial Production (IIP) gives the index in terms of quantities. Similarly, the ASI covers all companies registered under the Factories Act, whereas the IIP covers units registered with institutions, such as the Director General of Technical Development, Coal Commissioner and Textile Commissioner. Consequently one may arrive at different answers to the same question. For instance, to the question how much did Indian industry grow in a particular year, the estimates vary depending upon the source used for estimation. In one such exercise, the divergence among the estimates of growth rates using different sources of data for the year 1978-79 was found to vary between 7.2 percent and 14.1 percent, at the extremes.

Non Accessibility of Unpublished Data

It may be noted further that researchers usually do not have easy access to the unpublished data. Procedural complexities and red tape more often discourage them from making use of such data. Worst yet, potential users do not even have any means of knowing about the available data. If, through personal contacts or by sheer coincidence, they come to know about some unpublished data, it is extremely difficult for them to get access to such data. The dictum of confidentiality and privacy in respect of government held data is so rigidly adhered to by various government agencies, that even interdepartmental exchange of data within the government is difficult. The problem becomes all the more serious when the agency collects data under the provisions of the Collection of Statistics Act of 1953. In denying access to unpublished data they invariably invoke the sanctity of rule 7(1) of the Act which prohibits giving information on individual units. The said rule reads:

"No information, no individual return and no part of an individual return with respect to any particular industrial or commercial concern given for the purpose of this Act shall, without previous consent in writing of the owner for the time being of the industrial or commercial concern in relation to which the information or return was given or made or his authorized agent, be published in such a manner as would enable any particulars to be identified as referring to a particular concern."

It can be seen that the Act does not explicitly restrict the release of raw data to researchers. Government agencies, in fact, use the Act in self defense and to conceal inefficiency in handling data. It may also be appropriate to mention here that while the Act can be invoked by the competent authority to punish or impose a fine if any person "willfully refuses or

without lawful excuse neglects to furnish such information... as may be required under the Act", the Act is silent about the obligation of the 'statistics authority' in respect of utilization of the collected information. Consequently, large volumes of collected information continue to gather dust while genuine researchers are denied access to the required data on some pretext or other.

Prospects

A large majority of researchers are working in the universities and research institutions in various disciplines. The universities and deemed universities (institutions which have been given the status of a university), numbering over 150, are financially supported by the government via the University Grants Commission (UGC). The research institutions are financed or are in some way supported by national level organizations in various fields like agriculture, medicine, social sciences, education, science and technology and culture, etc. These national level organizations are, to name just a few: the Indian Council of Agricultural Research, Indian Council of Medical Research, Indian Council of Social Science Research, Indian Council for Scientific and Industrial Research, National Council of Educational Research and Training, etc. All such national level organizations as well as the UGC are financed by the government. As already pointed out, researchers based in these organizations, or institutions supported by them, have quite often little or no access to unpublished government generated data. The published data, on the other hand, serve only a very limited purpose for the scholar. Consequently, researchers have no option but to generate their own data. For want of adequate funds, most of the scholars both in the universities and research institutions end up taking only micro level studies. It is ironic that

at least some of the scholars in the universities and research institutions, keen, and qualified to handle large data for macro analysis, must content themselves with only micro studies for want of financial resources. On the other hand vast body of data generated by the government remain largely unutilized. Thus, research activity is seriously hampered not for want of resources, initiative, willingness, skills or expertise in the country, but for lack of perspective and policy for proper storage and sharing of data generated by different wings of the government.

Thus what is needed is not additional output of financial resources for the collection of new data, but a perspective and policy for utilizing data collected by various government organizations. It is necessary to improve communication channels between the producers and users of data so that interested scholars can know what type of data is available, where, and in what form. Various national level organizations may help to bridge this communication gap. Besides, these organizations may try to persuade government agencies to release raw data to genuine researchers. Computerization of government data may also greatly facilitate utilization of the existing data. In the absence of a computerized data base, it is extremely difficult for researchers to utilize available data even when these are released by the government. The volume of data is so large, and the structure so complex, that individual scholars with limited funds at their disposal, would find it difficult to put them in proper shape for meaningful analysis.

It should be mentioned that the Indian Council of Social Science Research is making efforts to persuade the CSC, NSSO and other major data producing agencies to, at least, release 'non-sensitive' and unclassified data to interested scholars. The ICSSR is also preparing an inventory of data available in various government agencies and research institutions which will, hopefully, go a long way towards

disseminating information to scholars about available data. The ICSSR has also recently financed a project in collaboration with the Bureau of Industrial Costs and Prices, the Planning Commission and the Industrial Development Bank, to develop an Information System on Industrial Data. The System would be developed by pooling industrial data collected by various government agencies and standardizing them as much as possible.

Currently, the government is following a more liberal policy in the use of computers in various government activities. A number of government agencies have already set up computers and transformed their databases in machine-readable form. The National Informatics centre, set up by the Department of Electronics, is also helping various government agencies to computerize their databases. With the help of the National Informatics centre, separate information systems are being developed, containing data on education and manpower,

energy, health and demography, industrial resources, agriculture, transport and industry. The ICSSR is part of the NIC network and, therefore, will be in a position to help scholars get access to these information systems, when developed.

Thus the potential of utilizing government generated data in India is enormous, provided proper perspective is developed and policy formulated. Hopefully, given recent developments in government, particularly, in the computerization of government data, it may be possible for researchers to access primary data. Proper utilization of existing data may in turn help government agencies to cut the cost of collection of unnecessary and irrelevant data. This will, however, require greater competence on the part of the researchers in handling complex data. Recent developments do seem to hold promise but time alone will bear testimony to this. ■