

PLANNING FOR THE 2001 CENSUS OF INDIA BASED ON THE 1991 CENSUS

Dr. M. VIJAYANUNNI
Registrar General & Census Commissioner, India

The next decennial census of India due in 2001 will be the 14th in the series of uninterrupted decennial censuses taken in India ever since 1872. Thus it has been serving as the richest single data source for time series data on the people of India for over a century and a quarter now and completed its quasiquicentenary in 1997. While each census is an important data collection exercise on its own, the 2001 census coinciding with the first year of the next century and millennium will be even more so since it will throw up unique and vital benchmark data on the state of the nation's society, economy and demography at this important milestone in time when one century and one millennium yield place to the next.

The census of India is conducted under the Census of India Act of 1948. The Act makes it obligatory for the public to answer all the questions faithfully while it also guarantees the confidentiality of the information relating to the individuals.

The census of India used to be a one-night count during the first 70 years and this placed a very heavy demand on the number of enumerators required. The number of enumerators engaged was 2 million in 1931 for what was then undivided India with one-fifth of the then world population spread on an area of 5 million square kilometres. With the abolition of the one-night theory replaced by enumeration over a staggered period, the number of enumerators was brought down to one million in 1941. The first census in post-partition India taken in 1951 covered a population of 340 million on a land area of 3 million square kilometres and engaged 600,000 enumerators. Since then the population and proportionately the number of enumerators have been steadily going up in each census. The population was 840 million in 1991 and the enumerator strength was 1.6 million. For the 2001 census the population would exceed 1000 million and the number of enumerators is expected to be 2 million.

Of the two well-known methods of enumeration viz. canvasser method and householder method, the Indian census has followed the former. Under this method, the enumerator visits every household and fills in the questionnaire based on the answers given by the respondent. Because of the high level of illiteracy among the public, the householder method is not suited to India and the 2001 census will continue with the canvasser method.

In India, the earlier censuses based on a one night enumeration were taken on the de facto approach viz. where the person is found, as against the de jure approach or at his usual residence. The de facto approach will appear simpler but requires a very large enumerator strength which may be difficult to mobilize. On the other hand, the de jure approach requires careful probing about the person's usual residence. The de jure data provide a better base for the purpose of planning and

"The administrator of Bastar State (in central India), when inspecting census work on the night itself, was attacked by a tiger, which sprang onto the bonnet of his car, but finding the pace and the radiator too hot for him failed to make an end either of the inspector or his inspection".

— From the Census of India Report of 1931

development. Considering the pros and cons of both the methods, the Indian census has made a compromise between the two and has been following what is known as the extended de facto approach. Under this method, an enumeration period of two to three weeks is fixed and the data are brought up to date with reference to the census date by a revisional round after the enumeration period is over. The houseless population is counted on the last night before the census date. In the 2001 census also the extended de facto approach will be adopted.

Houselisting

A houselisting will be done six to ten months in advance of the actual enumeration. The main purpose of houselisting is to locate and identify all places which are occupied or used or which are likely to be occupied or used by people. The houselisting operation will also give an advance estimate of the population and provide a frame for carving out the census enumeration blocks. Houselisting is preceded by house numbering which involves affixing of the number on the door of each house and structure in a systematic manner. Although a housing census as such is not taken, some basic information on housing is collected during the process of houselisting. In the housing statistics, therefore, the house and household would be the enumeration units.

Enumeration

At the 1991 census three schedules were canvassed viz. (i) houselist (ii) household schedule and (iii) individual slip. The houselist was canvassed during the houselisting operation while the other two schedules were canvassed during the actual enumeration. At the 2001 census, it is proposed to canvass the houselist during the houselisting as earlier, but only one schedule the household schedule during the actual enumeration. The main purpose of canvassing the two schedules at the enumeration stage in 1991 was that the household schedule would be used for manual tabulation while the individual slip would be used for direct data entry and generation of computer tables. The enumerator was required to copy out the information relating to each individual from the individual slip to the household schedule. However, during the scrutiny and editing of the household schedule, it was noticed that there were copying mistakes and therefore the entries in the individual slip and household schedule did not fully match. In order to avoid such discrepancies at the 2001 census it is proposed to canvass only one schedule, the household schedule, which will be used for generating the manual as well as computer tables.

"From the point of view of geology India has been described as the land of paradoxes. The peninsula is one of the oldest formations in the world and the Himalayas one of the most recent... The flora of India is more varied than that of any other area of the same extent in the Eastern Hemisphere, if not in the world; and the species of animals far surpass in number those found in Europe. The climate is equally diversified. Nowhere is the complex character of Indian conditions more clearly exemplified than in the physical type of its inhabitants... The linguistic survey has distinguished in India about a hundred and thirty indigenous dialects belonging to six distinct families of speech... The heterogeneity of political conditions is equally great."

-- From the Census of India Report on 1911

"As with the ethnography so also in the case of the languages of India much of the pioneer work has been done in connection with the decennial census; and the interest in the subject, which eventually led to its complete and systematic treatment under expert direction, is largely due to the contribution made by Census officers in their reports."

-- From the Census of India Report on 1921

Economic questions in the census

The census of India collects information on the economic activity of the people and the census data provide an inventory of human resources of the country showing their numbers, characteristics, occupation and distribution among the various sectors of the economy. The reference period is one year preceding the date of enumeration for both seasonal and regular activities. Those who had worked for less than 6 months in the preceding year are termed as marginal workers. In respect of non-workers, those seeking work were also netted so as to ascertain the number of new entrants to the labour force.

The Primary Census Abstract contained the economic data down to the village-level and the ward-level for towns by presenting information on main workers, marginal workers and non-workers. This also showed the distribution of main workers into the following nine industrial categories: I Cultivators; II Agricultural labourers; III Livestock, forestry, fishing, hunting and plantations, orchards and allied activities; IV Mining and quarrying; V Manufacturing, processing, servicing and repairs (a) in household industry; (b) in other than household industry; VI Construction; VII Trade and commerce; VIII Transport, storage and communications; and IX Other services. The data collected on the economic activities of the population cross classified with other characteristics like age, educational level etc. are generated on the computer.

In 1991, 34.18 per cent of the population were main workers, 3.32 per cent were marginal workers and the remaining 62.50 per cent were non-workers. Out of the main workers, 67.37 per cent were in the primary sector (I+II+III+IV), 12.13 per cent in the secondary sector (Va+Vb+VI) and 20.50 per cent in the tertiary sector (VII+VIII+IX).

A similar definition of work or economic activity and similar classification of workers is proposed to be followed at the 2001 census so as to generate comparable data for time series studies and analyses.

Migration

The details of migration are collected by place of birth and by place of last residence along with the reason for migration and the duration of residence at the place of enumeration. The reasons for migration are noted as employment, education, marriage, family moved, natural calamities and others. Migration is classified in all the four streams viz. rural to urban, urban to rural, rural to rural and urban to urban. It is interesting to note that in India the biggest reason for migration is marriage, accounting for over 56 per cent. This is explained by the social custom of women migrating to the husband's place on marriage. Work and business requirements account for 11 per cent of migrations while education accounts for about two per cent.

About the toll the census work took on the census officers.

"Mr. Drysdale, the Deputy Superintendent of the Central Provinces, laboured untiringly at his work, and was I regret to say, reduced to such a state of health that he was obliged to take short leave home..... Mr. Stokes (who) had charge of the Madras Census, too devoted himself so entirely to his duties that his health suffered, and he was obliged to take leave to England..... In my own report I feel there are many shortcomings and defects, which I hope will be judged leniently, as continued observation of heavy-figured statements has so weakened my sight....."

-- W.C. Plowden, Census Commissioner, India for the Census of India 1881

Fertility

Questions on fertility are asked in the form of age at marriage for all ever-married women, number of children ever born alive, number of children surviving at present and children born alive during the last one year.

Social and cultural subjects

The census collects the mother tongue of every individual and two other languages known so as to present data on mother tongue strengths and bilingualism and trilingualism. India has over 400 languages out of the 6700 identified languages of the world. But as the census presents only those mother tongues with a minimum strength of 10,000 persons, the number of mother tongues presented in the census tables work out to 216. In the census the instructions are that the respondent is free to return the name of his mother tongue and that is to be faithfully recorded by the enumerator. This naturally leads to the recording of a very large number of mother tongue names. In 1991 the number of such raw returns came to 10,400. These were subjected to the thorough linguistic scrutiny, editing and rationalization which resulted in 1576 rationalised mother tongues. These 1576 rationalised mother tongues were further classified following the usual linguistic methods for rational grouping. Thus was arrived at the 216 identifiable mother tongues which returned 10,000 or more speakers each. These 216 mother tongues are further grouped into 114 languages whose details are presented in the census tables. The rate of bilingualism was 20 per cent and that of trilingualism was 8 per cent which means that one out of five Indians knew at least two languages and one out of twelve Indians knew at least three languages.

The census of India collected information on the number of retired defence personnel in the country at the 1991 census at the special request of the Defence Ministry, though it was pointed out that this is not an item of information which is amenable to accurate collection from a vast operation like census. The data netted turned out to be an underestimate as expected and hence it is proposed to drop this question at the 2001 census.

Though the census of India used to collect data on the physically handicapped during the first few censuses, the results were invariably found to be gross underestimates. This question was therefore dropped. In 1981, keeping in view the fact that 1981 was declared as the International Year of the Disabled and at the insistence of the Ministry of Welfare, the question on the disabled was re-introduced. The results again proved to be very unsatisfactory since the numbers turned out to be far short of what was expected. Accordingly the question was dropped in 1991.

"Some curious marriage customs may be mentioned. At the marriage of a Mukkuvan woman the consent of all persons present must be obtained. An Okkiliyan husband pays the bride price, not at marriage, but after the birth of a child. When a Toda girl is about to attain puberty, she is deflowered by a sturdy member of the tribe from another village. The Bhatiyas of Gujarat formerly allowed the priest to pass the first night after marriage with the bride. The sangarkhel Pathans who allow considerable freedom to an unmarried girl, permit her on the night of her marriage to slip away for an hour with some young man of her choice. A third marriage is regarded as unlucky; and when a man has lost two wives and contemplates a fresh matrimonial venture, he often goes through a mock marriage with a sheep, a pigeon or some plant, so that his next wife may be his fourth and not his third."

-- From the Census of India Report of 1911

The census of India collects and presents data on religion, the scheduled castes and tribes as also the marital status of the population.

Manual Processing

The manual processing of census schedules of the 1991 census was carried out in 163 Regional Tabulation Offices set up all over the country engaging 45,000 temporary personnel. These Tabulation Offices generated the Primary Census Abstract (PCA) which provided small area statistics for the lowest units viz. villages and wards of towns. The PCA presents the number of occupied residential houses, households, total population, population in the age group 0-6, literates, main workers by broad industrial categories, marginal workers and non-workers. Tables on mother tongue and religion are also generated manually from the household schedules. These tables were ready by the end of 1992.

Data entry

The data entry work of the houselist in 1991 involved 45.5 million records comprising 20 per cent sample of the census houses in major states and 100 per cent in smaller states. With a view to complete the work in time a portion of the data entry job i.e. 4.9 million records was got done by external agencies and the in-house data entry work amounted to 40.6 million records. The data entry work started in the second half of December 1990 and was completed by October 1992. The housing and household amenities data generated from the houselist included the following : predominant construction material of the wall, floor and roof, purpose for which the house or building is used, whether the house is owned or rented, number of living rooms, number of persons normally residing in the household, availability of the facilities of drinking water, electricity and toilet, type of cooking fuel used etc. It is proposed to collect similar data at the 2001 census also. It is also being examined as to whether additional data on amenities and facilities and the quality of housing can be collected at the time of houselisting based on the requests received from various data using agencies and ministries for new items of information.

The main work of data entry of the individual slips in 1991 was taken up in two stages. The first stage data entry of the 10 per cent sample data for major states and 100 per cent for smaller states amounted to 108.4 million records. The second stage data entry involved the keying in of 100 per cent data of main workers in the non-agricultural sector, marginal workers and those seeking work, and the total volume was 121 million records. The original target dates for the completion of the first and second stage data entry were fixed as June 1994 and June 1995 respectively. In the light of this optimistic time frame for the generation of final tables,

"The penalties inflicted by caste panchayats usually take the form of feasts or fines, but never corporal punishment. In some castes, like to Chamars, the offender is put to some form of humiliation e.g. he has to collect the shoes of all his caste fellows and carry them on his head, or shave one side of his moustache. Korkus put a grinding stone round the neck of a women and make her go round the village with it on. Among the castes given to smoking, the offender is often subjected to the disgrace of preparing the smoking bowl (hukka) for the Chaudhris. Where fines are not imposed, the offender is required to feed the panchayat or sometimes the whole community. It is only for very grave offences that the person is excommunicated from the society. The control of the caste panchayat is very efficient. An offender usually finds himself unable to elude them as the caste can make his life a burden to him. By outcasting him, they stop not only all intercourse with his caste fellows, but can prevent him from enjoying the usual necessities and amenities of life. They can order the barber not to shave him, the dhobi not to wash his clothes."

-- From the Census of India report of 1911

a programme for advance tabulation of data based on a smaller sample was not envisaged in the 1991 census unlike in the previous censuses. However, this time frame turned out to be unrealistic later in view of the procedural delay in the sanctioning and procurement of data entry equipment and the sanction and creation of additional data entry staff. As a result the data entry first stage could be completed only by April 1996 and the second stage by September 1997. Though an effort was made initially to speed up the data entry by entrusting part of the work to private data entry agencies, this did not prove satisfactory in terms of quality. In the light of the above experience, the data capture, tabulation and processing at the 2001 census will have to be planned in a different way to yield quicker and more effective results. In addition to the generation of the provisional totals and the primary census abstracts through manual compilation methods, it is also proposed to do data entry of schedules on a suitable sample basis and generate advance tables on important characteristics for the country and states within a time frame of one to two years at the 2001 census. This will require careful and detailed planning of the various steps in advance.

Data entry centres were established for the first time in 1981 for computerizing the data so that it can be put directly on to magnetic tapes instead of the punch cards used earlier. There were 32 data entry systems with 16 terminals each (total 512 terminals) and these were located in 14 centres spread all over the country. By 1991 these data entry equipments had become largely outdated and had outlived their life and hence needed to be replaced. After considering the different options regarding equipment available at that time, it was decided that Intel 80386 chip-based systems with ten terminals each would be sufficient for data entry purposes. It was decided to replace the 32 systems (total of 512 terminals) of 1981 with 64 systems (total of 640 terminals) in 1991 in a total of 15 centres with one more centre being opened in addition to the 14 centres of 1981. These 64 new systems were procured in three phases to stagger the financial commitments. The first phase was installed in April 1992, the second phase in March 1993 and the third phase in January 1994.

The data entry equipment installed for the 1991 census would have outlived their life and utility by the 2001 census and will require to be replaced by state-of-the-art facility for data capture. This will have to ensure capability to handle the massive volumes of schedules involved, reliability, speed, accuracy and time-bound results. This is crucial to ensure prompt and timely dissemination of data avoiding bottlenecks and delays so as to commendably discharge the obligations of the census to the wide spectrum of data users.

Data Processing

One of the difficulties experienced at the 1981 census was the absence of in-house computing facility leading to dependence on other organizations for census data processing. An in-house computer centre was therefore established for

"It was another of the misfortunes of the 1931 census that it coincided with a great fall in revenue and a period of economic depression which made the most rigorous economy necessary and which left me no choice but to cut all expenditure as fine as possible and to goad my census superintendents unremittingly in an attempt to finish sooner and spend less."

"The census of India is not only by far the most extensive census operation in the world but, besides being one of the quickest, it is probably the cheapest."

-- From the Census of India Report of 1931.

the 1991 census. The computer medha 930/11 which is a version of the CDC Cyber 930 computer was installed in March 1991. Using this hardware and associated softwares, all the processing of 1991 could be done in-house.

The entered data were checked for completeness and the master files were prepared. To ensure quality, extensive edit checks were designed and a suitable edit package was developed. Detailed computerized checking of the inter-consistency between the responses to different questions was done and data corrected to the extent possible in this massive exercise. All the software for tabulation was prepared in-house. Some of these were developed using the tools provided by the U.S. Bureau of the Census (CONCOR for editing of data and CENTS for tabulating data). All other softwares were prepared using COBOL/FORTRAN within the organization. The tables were generated on computer in a controlled environment and then subjected to post-tabulation scrutiny and also extensively checked at the State Directorates and at the headquarters. As soon as they were ready the tables were released on floppies in a PC-compatible form to the data users. This eliminated the long time lags involved in bringing them out in printed form through the busy Government Presses and the data were available at once. This office has also acquired the facility to cut CD Roms and it will be possible to supply census data on CD Roms also on demand at the 2001 census. We are also preparing camera-ready copies which eliminates the errors at the composing and proof-reading stage and speeds up the process of printing of the publications. These processes will be continued at the 2001 census.

Tabulations

The following series of tables were brought out after the 1991 census and are planned to be continued at the 2001 census :

A - series	Population Tables
B - series	Economic Tables
C - series	Social and Cultural Tables
D - series	Migration Tables
F - series	Fertility Tables
H - series	Housing Tables
SC/ST - series	Special Tables on Schedules Castes and Tribes

The number of tables has increased from just 37 in 1961 to 120 in 1991. There are 8 tables in the A-series, 31 tables under the B-series, 15 tables under the C-series, 17 tables under the D-series, 32 tables under the F-series, 15 tables

"The organization was so thorough that the results for the whole of India were received complete on the 19th March i.e. within nine days of the census, and were issued in print next day... The record was broken by two Native States (Rampur and Sarangarh) where, by dint of working all night, the local officers were able, with the aid of mounted messengers and other means of conveyance, to get the figures for all parts of the State to headquarters in time for the telegram reporting the results to reach me in Calcutta by 8 a.m. on the following morning..."

"Apart from the extreme celerity with which this work was accomplished, which is not approached even in the smallest European states, the accuracy of the provisional totals is also noteworthy. The net difference in the whole of India between them and those arrived at after detailed tabulation was only 0.04 per cent."

-- From the Census of India Report of 1911

under the H-series and 2 tables under the SC/ST-series, working out to a total of 120 tables. The compilation and generation of so many tables is very time-consuming and an exercise is on to see how many basic tables are essentially required to be generated for common use so that other tables required by specialised users can be left to be prepared on demand.

Dissemination

Previously the census data were disseminated only in printed publications. A beginning was made in the dissemination through computer media in the 1980s when the Primary Census Abstract data was put on the National Informatics Centre Network (NICNET). In 1991 almost the entire census data is being made available on floppies. When the PCA was finalised in 1992 these data at the village and town level were put on the NICNET and simultaneously made available to users on floppies. Initially it was available on 480 floppies which was reduced to 51 using data compression techniques. The tables on housing and household amenities prepared from the houselist were also made available on floppies in Dbase format much before the printed publication was available. One of the tables of this series giving data on material of construction of floor, wall and roof of the house, generated for district-level, was so voluminous that it could not be brought out in printed volumes but could be conveniently made available in computer media.

The Dbase format which was used for presenting the PCA data and the housing tables as mentioned above was found a bit difficult to use by some data users. Considering this, it was decided to make available the subsequent tables under the B series (Economic Tables), C series (Socio-Cultural Tables), D series (Migration Tables) and F series (Fertility Tables) in spreadsheet format that can be used with a wide range of spreadsheet softwares.

The total output of the 1991 census consisted of over one hundred all-India volumes and over one thousand state volumes running into 600,000 printed pages and over 300 floppies.

The census of India has opened a web site on the Internet and the address is <http://www.censusindia.net>. The web page will provide an overview of the census and vital statistics publications and data put out by the Registrar General and Census Commissioner, India by giving information on what data are available where and how to get them. This will open up new vistas in data dissemination and user interaction.

"Few people outside realise that census is an administrative operation of great dimensions and, in addition, it is a scientific process. Indian census in particular covers the largest population in the world and it is also one of the most economical administrative operations. Census as an institution goes back to the remote past, but it is no longer a mere counting of the heads; it involves extraction of information which plays a vital role in the determination of many of our administrative policies. The facts elicited during the course of this operation yield valuable scientific data of sociological importance. In many matters it provides a useful guide for the effectiveness or otherwise of our economic policies. The theory of population is in itself an interesting part of economics. The census helps us to test and adapt that theory to facts."

-- Sardar Vallabhbhai Patel, Home Minister of India on the 1951 census