UNITED NATIONS SECRETARIAT ESA/STAT/AC.115/17
Department of Economic and Social Affairs May 2007
Statistics Division **English only**

United Nations Expert Group Meeting on
Contemporary Practices in Census Mapping and
Use of Geographical Information Systems
29 May-1 June 2007
United Nations, New York

Development of Census-based Geographic Database Applications
- India experience[*]

Prepared by

Chinhoy Chakravorty
Joint Director
Office of the Registrar General, India

**Development of Census-based Geographic Database Applications
- India experience**

1. *Introduction to Indian Census*

1.1.   A systematic and modern population census in India, in its present form conducted synchronously covering the entire territory was first held in 1881. Since then, censuses have been undertaken uninterruptedly once in every ten years. 2001 Census was the fourteenth census in this continuous series.

1.2.   The Census process in India is based on extended *de facto* method, i.e., counting persons present in an Enumeration Block during the time of enumeration. It is divided into two phases, Houselisting Operations and the Population Enumeration, the first phase preceding the second phase by about six to ten months. Duration of the actual population enumeration extends usually for 21 days followed by a revision round of 5 days. The houseless population is counted on the last night of census enumeration to avoid multiple counting. The 2001 Census was held from 9th to 28th February 2001 with reference date as 1st March 2001. India's population was enumerated as 1.029 billion, second only to China.

1.3.   For the purpose of census, the entire geographical area of the country comprising of villages, towns and other areas, is divided into Enumeration Blocks (EBs) covering about 120 to 150 households. One enumerator is assigned to each EB, who is responsible for complete enumeration of the area. At the 2001 Census, more than 2 million enumerators were deployed to collect information by visiting every household. The Indian Census is one of the largest administrative exercises undertaken in the world.

2. Preparing Rural-Urban frame for Census

2.1.   One of the important pre-requisites of census enumeration is the freezing of administrative boundaries of the constituent administrative divisions to enable preparation of rural-urban frame for enumeration. Unless there is a firm list of villages and towns it is not possible to assign codes and plan for enumeration. For 2001 Census, orders were issued against creation of new administrative divisions with effect from 1st January 2000, about a year preceding the census, thus freezing the list of villages and towns allowing finalizing the frame for census enumeration.

2.2. India is divided into 28 states and 7 union territories. Each state or union territory is then divided into districts and each district into sub-districts. Villages and towns are the lowest administrative units, which together constitute a sub-district, though there are a few exceptions. Towns are also divided into Wards. Whereas the towns, wards-in-towns and villages have all fixed boundary, Enumeration Blocks are curved out for each census on the basis of population size (about 120 to 150 households or 750 to 1000 persons). At the 2001 Census, there were 593 districts, divided into 5,463 sub-districts which in turn comprised of about 0.64 million villages and 5,161 towns. In this chain, clear and unambiguous hierarchy is established avoiding duplication or overlapping.

2.3. Coding scheme: A unique code is assigned to each of these administrative divisions, which makes it possible to identify each unit distinctly and allows presenting aggregated data at these levels. The coding scheme follows a serpentine route covering each unit beginning north-west and ending south-east of each administrative division. The code scheme for each units are as follows:

State/Union Territory	...	2 digits (within the country)
District	...	2 digits (within the state/UT)
Sub-district	...	4 digits (within the district)
Village	...	8 digits (within the state/UT)
Town	...	8 digits (within the district)
Ward	...	4 digits (within the town)

2.4. One of the major initiatives taken in the Census 2001 was the allotment of Permanent Location Code Number (PLCN) to each and every village within the State and not within a district/tahsil as in the earlier censuses. PLCN was thus assigned as one continuous number from the first village in the first district to the last village in the last district. PLCN is an eight digit unique location code number with the first six digits representing the code number of the village and the last two digits depicting two zeros '00'. These zeros are reserved as buffer to be used for coding any new village(s) that may come up between two villages in future. For example, if a new village comes up between two villages with PLCNs 01254600 and 01254700, the new village will be allotted PLCN 01254601 and so on.

2.5. The location code number for a town is also an eight digit number starting with the digit 4 situated at the extreme left acting as the unique identifier. The next two digits depict the code number of the district in which the town falls followed by two digits representing the town serial number in

the district. There are three zeros at the end as buffer mainly to meet the requirement for bringing the number of digits to eight to match the number of digits in the PLCN for the villages. Thus a town location code number 40305000 represents the town serial number 5 of the district number 03 in a State. The town codes are not permanent in nature in view of the fact that the boundary of the towns is more likely to change than the villages. As a result the name of the town may be same but the entity represented may be distinctly different in two different censuses.

Example:

State: Jammu & Kashmir (code 01)

State Code	District Code	Sub-District Code	Village Code 2001	Name of state, district, sub-district and village
01	**00**	**0000**	00000000	**JAMMU & KASHMIR STATE**
01	**01**	**0000**	00000000	**Kupwara District**
01	01	**0001**	00000000	**Kupwara Tahsil**
01	01	0001	00000100	Bore Village
01	01	0001	00000200	Keran Village
01	01	0001	00000300	Bugna Village
01	01	0001	00000400	Bichwal Village
01	01	0001	40101000	Kupwara (NAC) Town
01	**13**	**0000**	00000000	**Jammu District**
01	13	**0001**	00000000	**Akhnoor Tahsil**
01	13	0001	00491600	Khara Village
01	13	0001	00491700	Seri Palai Village
01	13	0001	00491800	Mala Village
01	**14**	**0000**	00000000	**Kathua District**
01	14	**0001**	00000000	**Billawar Tahsil**
01	14	0001	00607800	Bhattwal Village
01	14	0001	00607900	Sadrota Village

2.6. For uniquely identifying any village in the country, state code is required to be prefixed. The PLCN, not only allows easy location, but also allows time series analysis of the characteristics of a village over different censuses. Another operational convenience relates to reorganization of villages/towns into new state, district or sub-district, when a new administrative unit is curved out. Location Codes being permanent in nature,

allow time-series analysis of data, even after relocation to new administrative

Village Boundary Map
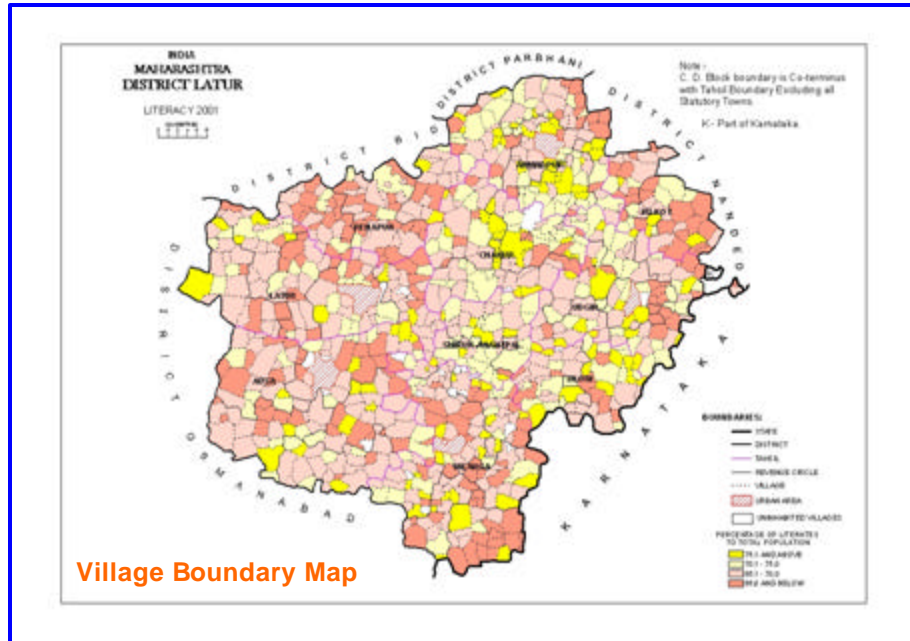
units. For a country with more than 0.6 million villages, this is a welcome improvement. Most of the other government organisations engaged in statistical and survey work in India have adopted the census coding scheme.

3. Geographic Information System database design

3.1 Maps showing jurisdictions at different levels of administrative hierarchy are extensively used for various pre and post enumeration census activities as well as to ensure complete geographical coverage. Different types of maps for both rural as well as urban areas are prepared. Before the actual census count is undertaken, it becomes quite essential to prepare maps showing latest jurisdiction of the area because numerous changes occur during the span of ten years, i.e. from one census to another. Updating the maps manually before each census is a Herculean task involving considerable time, manpower and energy.

3.2 To overcome this problem in 1992, GIS technology was introduced to create digital database of maps at India, State, District, Sub-district and C.D. Block levels showing jurisdictions at these administrative levels reaching finally up to the village level. GIS software, such as, ArcInfo, ArcView and ArcGIS have been extensively used for creating geographical database. The digital database structure has been designed and stored as per the following hierarchy in different layers:

- India: Showing States/Union Territories

- India: Showing Districts

5

- India: Showing Sub-Districts

- States: Showing Sub-Districts

- District: Showing Sub-Districts

- Sub-District: Showing Village locations or Village boundaries

- District: Showing Village locations or Village boundaries

3.3   The spatial database created has also been used to check and analyze census data collected during Census before release. While checking draft output tables for consistency, maps showing spatial distribution of indicators at different levels were compared with previous census data and changes examined.

3.4   As Census of India is a treasure mine of attribute as well as spatial data, a variety of products are published integrating the two in both print and digital media for dissemination. A few of such products are:

- Census Atlas - National Volume

- Population Atlas - National Volume

- Census Atlases of each State and Union Territory

- District Census Hand Book maps

- Tribal Atlas of India.

- Administrative Atlas of India

- Administrative Atlases of States and Union territories

- Language Atlas of India

- Housing Atlas of India

- Map Profile 2001 – India, State and Union Territories

- Preparation of 593 District maps of the country showing villages

- Thematic maps for use in various analytical reports

- Thematic maps for use in booklets, pamphlets, data sheets brought out on various census subjects

- Thematic static maps on census themes hosted at the Census of India website on the Internet.

4. Digital geographic database for dissemination

4.1.   Indian Census has been a major supplier of map products in the country. After every decennial census atlases are published in print on

various subjects, such as, administrative boundary and various census themes, which are used by different data users for their authenticity. With the introduction of digital technology, it has become possible to produce a large number of maps on a variety of census subjects for dissemination. Whenever any analytical report is published, it invariably contains a few maps showing spatial distribution of important indicators covered in the report. Each District Census Hand Book (DCHB), one of the popular publications used by the District Administration providing basic census and amenity data on each village and town contains a collection of maps at sub-district level showing location and other important physical features, road network, etc.

4.2. In the past few years, demand for digital database of various administrative boundaries has increased. With the recent revision in government policy it has now become possible to disseminate geographical digital database for use by other government and non-government agencies. There are also various agencies which are taking up development of GIS modules for use by medium and large business houses. Be it in the field of communication for laying optical fibre network, or setting up new retail outlets for marketing fast moving consumer goods, use of maps, especially in digital format is becoming crucial. Census Maps are fast becoming an unavoidable component in the decision making process.

4.3. The requirement for digital database is confined to the following three groups in general:

a) Users who want to procure geographic database for developing their own applications

b) Users who would like to use a pre-packaged database product developed by the Census Organisation and

c) Users who would like to view thematic maps in the form of static pages

4.4 The first category of users above usually comprise of various government departments, large business houses and private entrepreneurs developing GIS products for sale. The second category includes those general users who like to view spatial distribution of census data for quick analysis and making simple queries. Products brought out by the Census Organisation on various census themes are popular among this group. The third group usually comprise of new users who like to view static maps and use them in their reports and presentations. As the spread of technology for developing applications is becoming popular in India, the demand for supplying digital database is increasing in exponential terms.

5. Development of GIS applications

5.1. Before any GIS application is developed, it is important to keep in mind the target users and their need. Usually it may not be advisable to pack too much information in one application and make it focused and user-friendly. Need of the development would usually arise by regular interactions with the data users over a period of time. For instance, a general user may only be interested in basic census data and maps for use in their study or presentation. Those engaged in marketing on the other hand, may like to analyze village level data with the help of suitable thematic maps to develop their strategy. Planning departments in government or an NGO working on a specific issue may like to have combination of both to first understand the spatial distribution at both macro and micro level and then plan interventions.

5.2. With the growing popularity of computer in India there is a strong demand for new products based on the above models. Indian Census publishes a large number of preformatted tables as decided in advance. To extract tabular data and to prepare thematic maps is not very easy for a new user. The GIS modules prepared for various target groups not only allows preparing thematic maps, but also in many case, enable extraction of tabular data from the backend database. Thus these modules serve as complete data source for easy and quick use.

5.3. <u>Application 1- CensusInfo India:</u> One of the first applications developed by Indian Census for use is entitled 'CensusInfo India' based on the ChildInfo software developed by UNICEF. CensusInfo available in CD, allows extraction of 2001 Census data along with facility for creating various theme based maps in an interactive manner. The quality of data tables and maps prepared are excellent. As a result CensusInfo has become quite popular with a variety of data users.

5.4. While developing any GIS application, basically three steps are involved. These are:

      a. Finalizing geographic database

      b. Finalizing attribute database

      c. Integration of the two using GIS software.

5.5. All the above steps of developing GIS applications would use the same geographic coding. In CensusInfo, separate spreadsheet files were prepared on 65 selected indicators, showing the area ID, area name, sub population, value and the source as per format required. CensusInfo, being a general purpose application provided information on various demographic, economic

and education indicators as well as on housing, households amenities and assets from 2001 Census at country, state, district and sub-district levels. Validation programme was subsequently run to check data consistency and accuracy. Metadata, an integral part of any GIS application, was also prepared indicating the concept and definition involved. All these were then handed over to the agency appointed by UNICEF for integration. After the draft module has been prepared officials in the Census Organization checked it for functionality. After thorough checking and approval, the final CensusInfo module was published in CD for dissemination.

5.6.   Every Census Organization should have a regular channel available for sensitizing the data users about the availability of data, data products and the technique to use them. Without proper publicity and sensitization of the

(Map prepared using CensusInfo India)
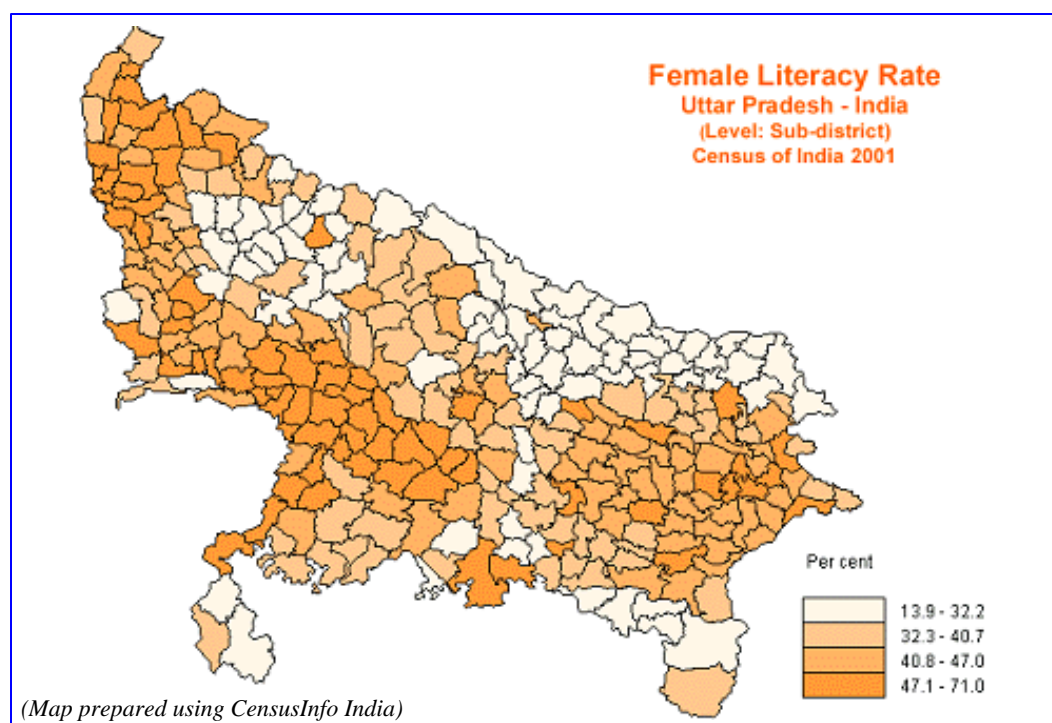
prospective users, any product, particularly special GIS application developed after investing substantial money and time, are likely to remain unused. In India, Data Dissemination Workshops are held across the country in a regular and continuous manner to apprise the users or the prospective users about new product releases on Census data along with data highlights. After the release of CensusInfo India a large number of workshops were held in universities and various government departments to sensitize them about the features available in CensusInfo and providing hands on training. Such efforts went a long way in popularizing CensusInfo and ultimately the results from the 2001 Census.

5.7.　Application 2 - Punjab GIS: The main objective of developing Punjab GIS software was to empower District Administration in identifying backward regions of the district. Selection of villages for implementing any government programme is controlled by various push and pull factors especially when the resources are limited. Be it in setting up primary or middle school, providing medical facility or supply of power etc, there are always regions which are less developed. As Punjab is one of the advanced states in terms of infrastructure development as well as providing facilities for education and economic development, identification of such regions or villages using Punjab GIS was aimed to direct development efforts to these areas.

5.8.　With the above objective village level data from 2001 Census was compiled on the following subjects for use in Punjab GIS:

a) Data on housing, household amenities and assets

b) Data on population, number of literates and literacy rates, type and category of workers, etc.

c) Data on amenities and infrastructural facilities available in villages

5.9.　For developing the GIS application, geographic database at district level showing village boundary has been used by creating separate files for each district. Various layers like, village and urban boundary, national highway, major roads, railway lines and rivers have been added. The application finally developed allows preparing thematic maps showing spatial distribution of any selected indicator. One of the most important features of the application is to make simple or complex queries and view the result on the map. It also generates a report showing the name of the villages that satisfies the condition(s) in the query in MS Word or MS Excel. There is also facility for generating two different maps in a single frame for comparison.

5.10. As in CensusInfo application, basic datasheets and geographic database for Punjab was made available to the agency which developed the GIS module free of cost. It has been the experience that the GIS software used for preparing such applications which allows independent map-viewing facility is quite costly. Therefore, the applications developed using such software have to be priced steeply to recover the cost either in full or in part. One of the easier options, to begin with, would be to make available the software at the Census website of the Internet.

5.11. The map on the next page prepared using Punjab GIS shows villages in Rupnagar district which has at least one primary school.
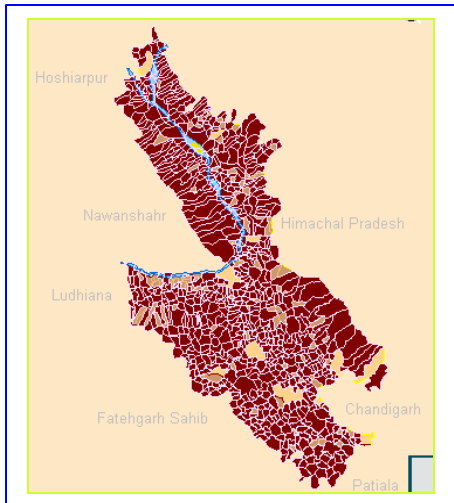
VILLAGES WITH PRIMARY SCHOOL
Rupnagar District - Punjab
Census of India 2001

(Map prepared using Punjab GIS)

Internet GIS applications

5.5. Census organization in India is one of the few who have made available Census GIS module at their website to depict census data on maps up to sub-district level. This free interactive facility (available at http://www.censusindia.gov.in) enables any user to depict 2001 Census data on map for analyzing spatial distribution and make queries. This facility on the Internet has helped tremendously in popularizing 2001 Census results and focusing on important issues.

Child Sex Ratio (0-6 years) : India by Districts
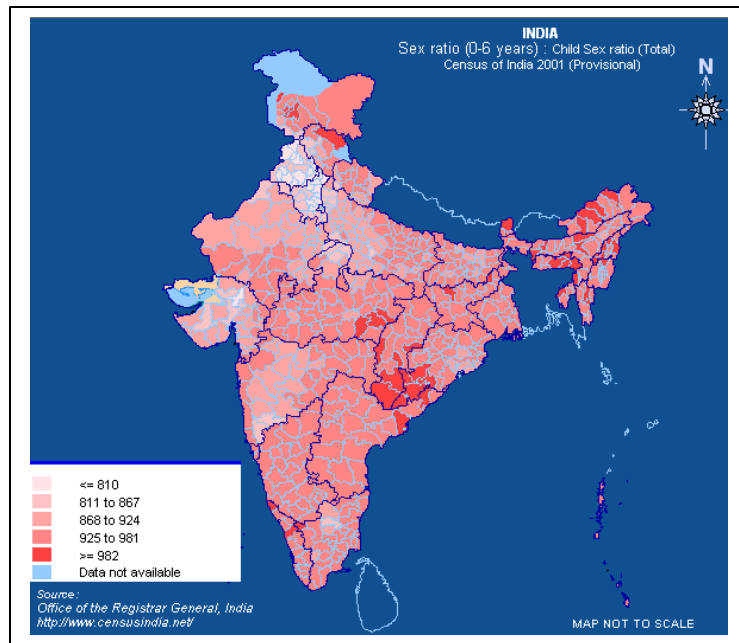
(Map prepared using Census India GIS available at Census of India website)

———————

11