©2013, ISSN 2180-2491

Development of Census-GIS methodology for Pakistan

Syed Nawaz ul Huda[1], Farkhunda Burke[2], Ghazi Salahuddin[1], Imran Ahmad[1], Mazhar Jamal[1]

[1]Population Census Organization, Statistics Division, Karachi, Government of Pakistan, [2]Department of Geography, University of Karachi, Karachi, Pakistan

Correspondence: Syed Nawas ul Huda (email: nawaz_huda@hotmail.com)

Abstract

The importance of a reliable and efficient methodology for data collection at the national level cannot be overemphasized. Conducted at decennial intervals, census plays an essential role in the collection of a wide range of data in Pakistan. It is of prime importance for national planning as it is the main source of information for the study of various aspects of the population who are the beneficiaries or otherwise of any governmental policies. Census exercises thus need to be efficiently performed to safeguard data accuracy. In this context, censuses conducted by manual methods have become obsolete and are being increasingly replaced by automated methods such as the Census GIS. This paper presents a model for the implementation of GIs for census enumeration in Pakistan so as to overcome the inordinate delays often encountered in the country's previous census exercises. It is believed that this new mode of census management would enable the streamlining of not only national, local and regional planning but also that of fool-proof disaster management.

Keywords: Census GIS, census exercise, census modeling, data collection, disaster management, Google Earth

Introduction

Population census plays an essential role for a wide range of a country's data collection as it serves as the main data source for the study of various population issues. Spatial distribution of population both rural and urban, temporal variations, social and economic characteristics ranging from housing to employment, migration, perspectives of ethnicity, epidemiology and health risk assessments, inequalities and poverty mapping, etc., are some of the aspects highlighted by censuses (Raymer et al., 2009; Norman et al., 2008; Finney & Simpson, 2008; Stillwell et al., 2008; Briggs et al; 2007, Norman & Bambra, 2007; Huda et al., 2011; Shen, 2005, Minot & Baulch, 2005). In-fact, it is the most basic and essential source of data for any subsequent policy research and implementation (Schroeder, 2007).

The manual method of census enumeration, estimation and analysis is not only cumbersome and erroneous but quite obsolete and outdated. The post modern era is witnessing the implementation and continuous development of Census-GIS without which the handling of massive and complex data seems well nigh impossible. Maps are symbolic representations of real space and form the basis of spatial data collection, especially census data ((Martin, 2000). The gap between reality and its symbolic representation has been shortened in recent years with digital multimedia information systems (Camara & Dias, 2004). With reference to census data collection, different administrative units are followed in different countries e.g., in United Kingdom and United States the Enumeration Districts, in Canada Enumeration Areas, while in Japan the Basic Unit Blocks are the smallest census data collection units (Martin, 2004; Côté, 2005; Ragains, 1995; IHARA, 1998).

Geographic Information System (GIS) is one of the applications most in demand for the development of census activities (Martin, 2002; 1998a; 1998b; Martin et al., 2001; Wu et al., 2008; Wang et al., 2001; Duke-Williams & Rees, 1998). In the past several years, various developed as well as developing countries have been crucial in GIS application in census (Schlossberg, 2003; Mobbs, 1998; Openshaw, 1977; Xiru, 2009; Tye, 2009; Wall, 2007; Verhoef & Grobbelaar, 2005, Ralphs &

Ang, 2009; Prasad, 2006; Owiti, 2008; Laldaparsad, 2007; Gregory & Ell, 2005; Eze, 2009). The growth of GIS through the 1990's led to a dramatic increase in the use of digital geographic data (Mokhele, 2011).

The GIS technology used in census becomes more sophisticated through open source high resolution satellite imageries of Google Earth including GPS built-in Trimple, Tablet and PDAs. The role of UN for development support in terms of established GIS lab. with modern and latest software and equipments in developing countries has been one of the basic causes for promotion of Census GIS in the contemporary world. In addition, the UN Statistics Division succeeded in ensuring the NSOs of most developing countries with no-cost versions of commercial GIS software (UN, 2000; 2004; 2009). In Pakistan, UN based NGOs UNFPA with collaboration of Statistics Division, Government of Pakistan rendered massive support in terms of funds for the development of GIS based census. The present study has been designed from available and funded sources of Population Census Organization of Pakistan for the development of Census GIS. Census operation in Pakistan gained legal support in 1959, through a presidential ordinance, in which the definition of census along with other relevant matters were promulgated. The census not only involves coverage of area and population, but all topics of information regarding information on various social and economic parameters of the population (GoP, 2010).

Materials and method

Census history in Pakistan

The first census launched in 1881 under the British rule, covered the areas now comprising Pakistan (GoP). The first census of independent Pakistan was conducted in 1951. The census tiers were Census District, Charges, Circles and Blocks. Each of the administrative districts with the exception of cities like Karachi, Lahore, Multan, Rawalpindi and Sargodha were subdivided into Census Districts. Each Tehsil/Taluka and in some areas Qanoongo Halqa (QH)/ Supervisory Tapadar Circle (STC), Patwar Circle (PC)/Tapedar Circle (TC) and Blocks that consisted of around 800 persons and a revenue estate or its part were formed. In the 1961 census, further changes were made. Census Districts were arranged as:

(a) Sub-Division or portions of a district
(b) Cities consisted of one or more Census Districts, and
(c) Big towns (above 0.1 million population) were considered as separate Census Districts.

At Charge level, the Divisions comprised Tehsil/Taluka or QH/STC, while urban areas were composed of Wards or parts of Wards. Census Circles were formed of small QH/STCs or PC/TC, while Census Blocks consisted of Mouza/Deh/Village or its part. For urban areas, Blocks have covered the specific ranges of housing units.

The third census which was conducted in 1972, envisaged more changes in methodology in comparison to the previous census. Major urban areas with population of 0.1 million or more were treated as one or more independent Census District. Census Charge is followed by each Tehsil/Taluka or QH/STC. In the urban areas they were treated as a Census District. A Census Charge was formed of Ward/ Union Committee or a part there of Medium size urban areas, constituted independent Charges, while the Census Circles were composed of QH/STC or their part. In the latter case, a group of PC/TC's combined to form a Census Circle. In urban areas (Census District or Charges) a Circle was formed of a Union Committee or its part. Small urban areas constituted independent Census Circles. A Census Circle was generally formed of about 1000 households. In the un-settled areas of NWFP (Tribal Area) a Census Circle was formed of a Union Council or its part. Census Blocks which was the smallest census area consisted of approximately 200-300 households or 1000-1500 population. In the rural setup a Mouza/Deh comprised of at least one Census Block.

There was a felt need in the third Census for maps in urban areas with large population concentrations. It was realized that delimitation of census areas without maps was a difficult task. Three years prior to the conducting of Census, maps for field use were collected from the local councils i.e., Municipal Committees (MC)/ Town Committees (TC)/ Cantonment Boards. Some maps were collected from the Survey of Pakistan and Town Planning Departments. These maps were

reviewed and where necessary, they were enlarged or reduced to suitable scales for the purpose of delimitation. Copies of tracings which were prepared were dispatched to the Local Councils for updating and for the purpose of Census and Blocks and Circles were marked. In the case of rural areas, Taluka maps with boundaries of Census Charges and Circles marked on them were used.

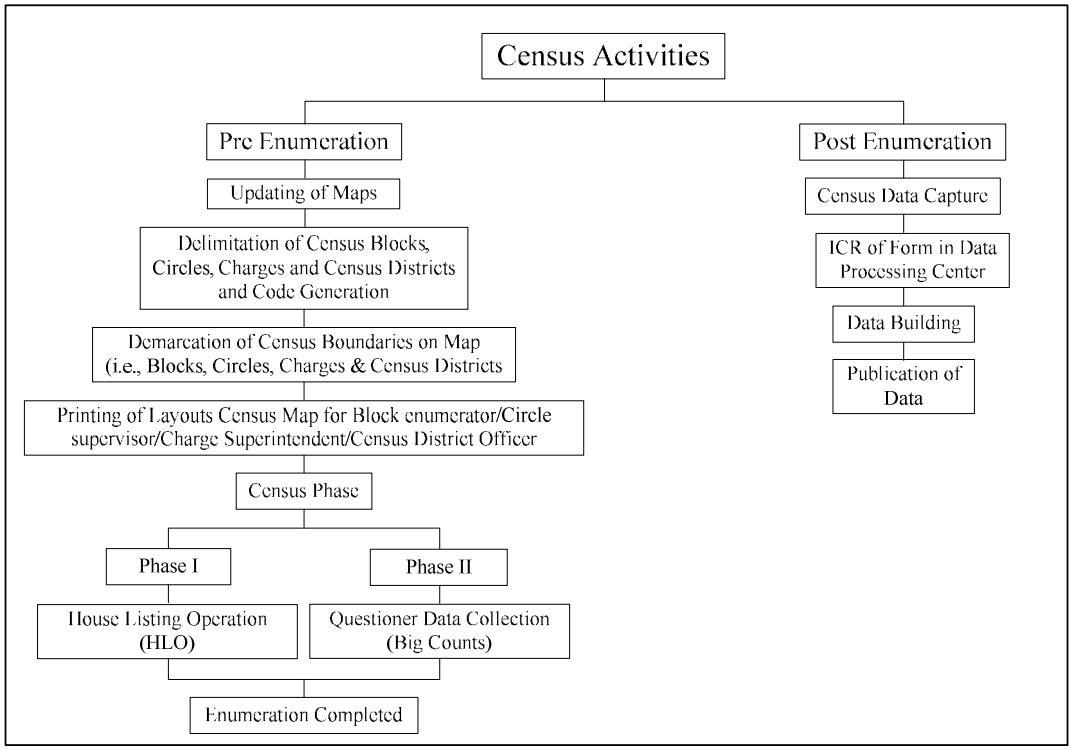

Figure.1 *Ongoing census processes in Pakistan*

The four census tiers i.e. Census Districts, Charges, Circles and Blocks prevailed in the 1981 Census. Due to increase in the population of the administrative district and some undefined reasons each Sub-division of a district in Sindh, Punjab & NWFP were treated as Census Districts. In tribal areas and Balochistan, each administrative district remained as Census District. However, each Municipal Corporation (M.C)/Cantonments (Cantt.) having estimated population of 50,000 were created as separate Census Districts. Each Census Charge comprised of QH/STC Charge in Sindh, Punjab and settled areas of NWFP while in other areas each Tehsil/ Sub-Tehsil was formed as a Census Charge. Each urban Census District was delimited into two or more Charges. In addition, all other urban areas were treated as separate Census Charges. In the case of Census Circles, rural areas mostly PC/TC's were treated as Census Circles along with the un-settled areas of each Union Council of NWFP. In urban areas a Ward or a part of Ward or two or more Ward groups were treated as a Census Circle, on the basis of population. Each urban area (MC/TC/Cantt.) treated as Census Charge was further subdivided into two or more Circles. Census Blocks comprised of 200-300 households.

The five tiers of delimitation of census areas remained the same. But once again the Census District changed in the 1998 census. Each Cantonment, Municipal Corporation and each administrative district with the exception of the two i.e., (Karachi and Lahore) were treated as Census Districts. Similarly, in rural areas each QH/STC was treated as a Census Charge while in urban areas each PC/TC and MC and part of big cities with population ranging from 350,00 to 500,00 were treated as Census Charges. In rural areas each PC/TC were treated as Census Circle. In urban areas each (smaller town) committee was treated as a Circle. Generally, a group of 5 to 7 Blocks formed a Circle. A group of 150-200 households in rural areas and 175-225 households in urban areas were treated as Census Blocks.

Methodology

GIS applications can provide multi dimensional solution regarding census activities. In Pakistan, census activities designed can be according to the following steps i.e., geographical works (preparation of maps and delimitation of blocks) and data collection, updating of area list, in the case of bifurcation of urban/rural administration recommendation based on population and housing data must be made, and Geo-coding of areas in accordance to the manual patterns (GoP, 2010). The present study has proposed Census GIS Model shown in Fig.1.

The proposed model anticipates need of the following supports:

1. All existing maps regarding census purpose must be available at micro-level i.e., layout maps, sketch maps, slum maps and maps of smallest revenue estates (Mouza/Deh) of administrative district.
2. A Google Earth satellite image for the base map while Google Earth Professional for image layout printing at A° size for field map.
3. Mapinfo Professional/ArcGIS, Global Mapper and Microsoft database packages for layering, database and conversion of format for ongoing activities.
4. A3 and Plotter for printing job and 42 inch Scanner for manual map scanning.
5. High speed internet connection (minimum 1mb bandwidth).
6. All supportive equipments for GIS and Data Processing Labs. (e.g. Computers, Window 7, Office Packages, SQL server, Networking, GPS etc).

For application of the suggested model (Fig. 2) the applied methodology consists of several parts. Most of the stages are designed for laboratory work, while others are related to field data. After the setting of equipments in the GIS laboratory the prime concern is the organized running of the system. The general notion is that technological applications run automatically or under a single command. However, contrary to this, applications designed for any work varies. It is characterized not only by temporal variation, but spatial as well as environmental variations. In the post modern age, team work has been recognized as capable of producing the most exemplary results. The aim of the present study is the establishment of Census GIS with minimum expenditure and time lapse, where in every employed personnel will contribute their optimum efficiency. Fig.2 shows the placement of personnel with reference to the prime requirement of geographic data collection in the GIS Laboratory.

Fig.3 shows classification of works in the GIS Laboratory. This model consists of three major desks on the basis of experience and expertise. Significance of this model is that requirement of geographic data for census have been classified where low experience perform their work simple vectorization e.g., roads, rivers, canals, drainage, shape of rural settlements, etc., while highly experienced personnel who have stronger commands in manual cartographic and geographic work but with lesser experience in computer applications will do the job, e.g., demarcation of Census Blocks in terms of rural/urban, labeling of road networks and other line data, etc. The concept of layer development will be accomplished by personnel who have the knowledge of GIS software applications like Mapinfo/ArcGIS etc. GIS analysts perform the final task of development of Census GIS. Fig. 4 shows channelized mechanism in GIS Lab. where all the different stages from initiation to completion of census work is performed.

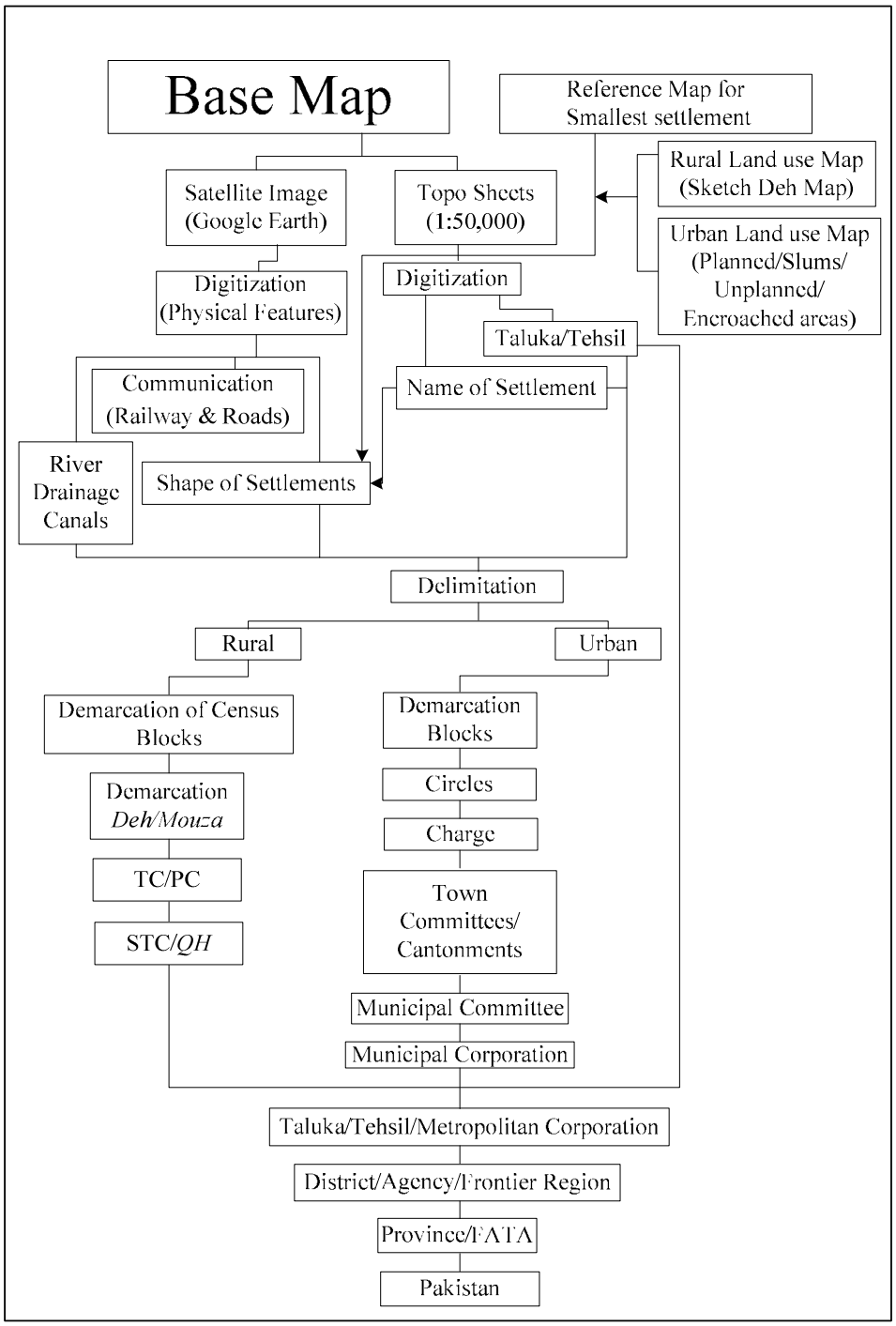

Figure.2: *Proposed model for the development of census GIS*

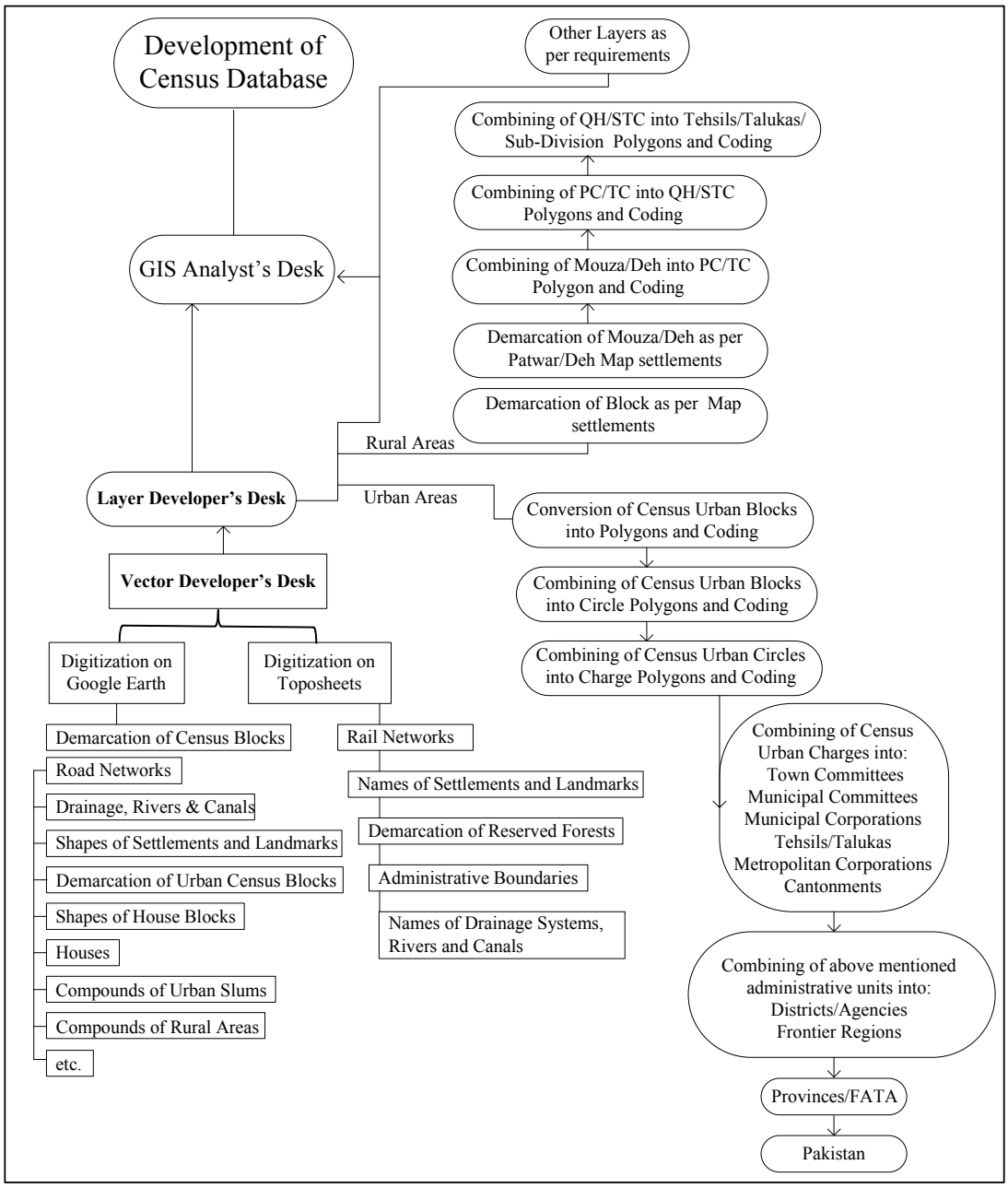

Figure.3. *Desks in GIS laboratory*

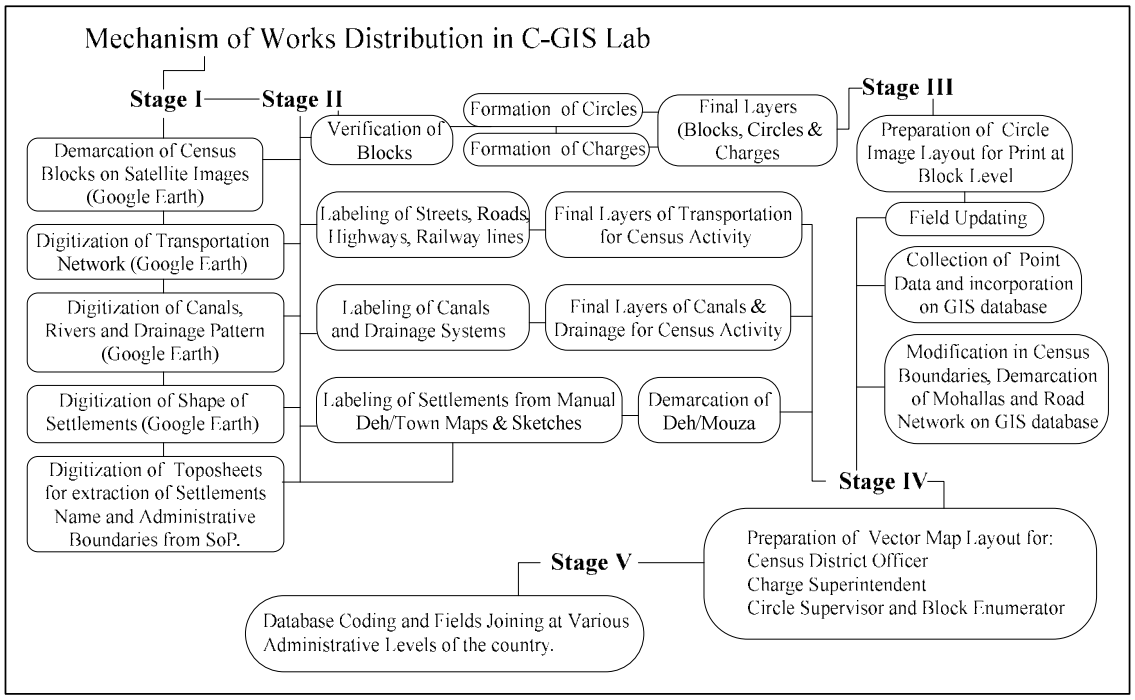

Figure.4. *Channelization of work in GIS Laboratory*

Results and discussion

GIS is a convenient method for collecting census data and providing the underlying geographical framework. It serves as a fundamental resource for social, demographic, economic applications as well as for regional planning (Barr 2007, Cai et al. 2006, Martin et al. 2002, Harris & Longley 2000). However, there are some challenges faced during geographic data collection regarding Census GIS :

- Google Earth images do not cover entire Pakistan at high resolution.
- Mosaics of images show geo-reference/ rectification errors.
- Images pertain to various dates, some being recent while others couple of years old which causes digitization problems.
- There is temporal variation of settlements over the same space as a result of which administrative and census boundaries keep changing in each census.

These problems lead to cost escalation of census GIS exercises in developing countries. However, enhancement of Google Earth images along with maintaining intact boundaries of census and administrative areas can help control/reduce the cost incurred.

In the case of settlement expansions, areas of new cities/towns maybe demarcated. With the help of GIS techniques trend analysis and comparative studies can be made, which will prove beneficial for regional, urban, environmental, demographic, administrative, disaster planning, etc. Fig.5. indicates the overall Census GIS activities suggested. The model covers both manual as well as modernized census conduction in Pakistan. In the post enumeration phase of census activities editing of Intelligent Character Recognition (ICR) census data is an essential stage in which the editing program should be developed by personnel in accordance to edit specification. Edit packages which have been developed for particular geographical socio-economic and demographic environments are also available.

Prior to the actual editing operation, the editing package/program must be tested. For this purpose dummy data file can be created or the data collected through pretest/pilot census may be used to check all steps of editing. Data from ICR census form is converted into single record for each person within a household and households within a block. Each block contains 9 digit geo identification numbers. The computer assumes an important role in cleaning, editing and imputing of missing values,

checking consistency, completeness and accuracy of data. The success in achieving these objectives depends on the editing program/package used for editing of data and the quality of manpower deployed for this purpose. Edit specification are the basic tool for computer editing which should be developed by the subject specialist in collaboration with the computer personnel. Edit instructions should describe the action to be taken on each data item. The instruction should be clear, concise and unambiguous which are clearly understandable for the program. Sorting of data prior to the actual edit, a sort program is run to arrange the records in sequence by blocks and households within blocks.

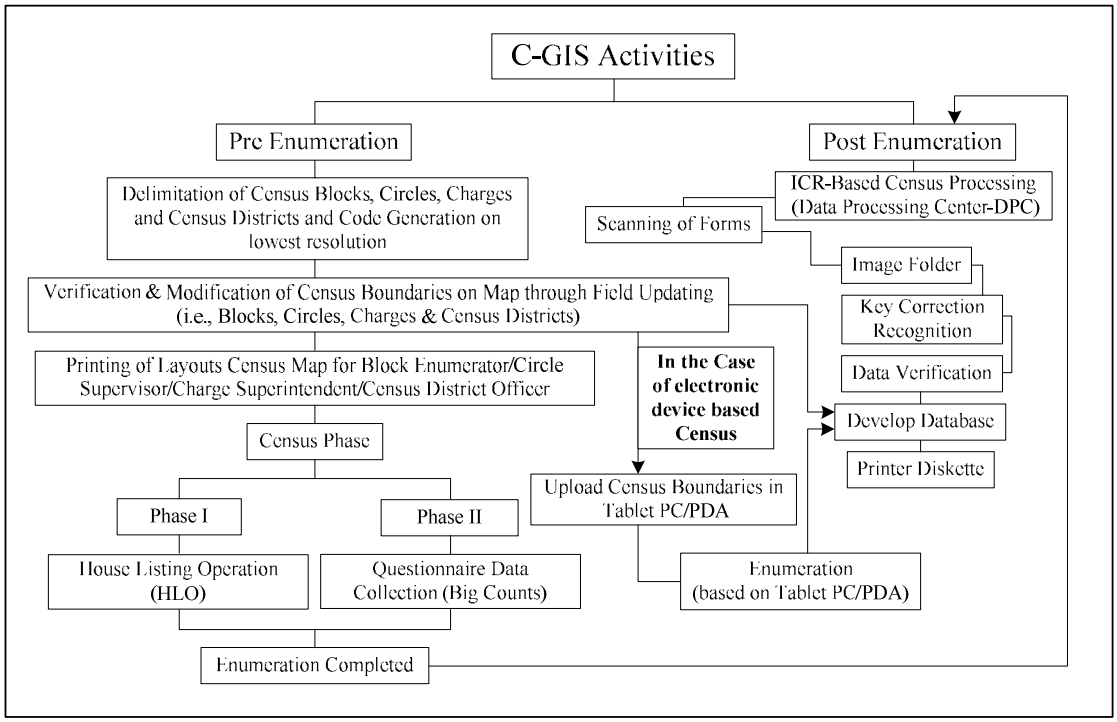

Figure.5 *Census-GIS activities*

A control edits checks the duplication and missing records and produces a summary total. After these operations, a master file is created. The edit program is executed to edit/correct invalid data of all census forms according to their edit specification. Edit statistics for each data item is generated. In case, numbers of corrections for any item are more than acceptable limit, data file will be reviewed. After cleaning the data files of a district, common information of short and sample forms will be merged to give 100% population and housing information.

The first step of computer editing will be the range edit which checks each response to determine whether or not they are valid codes. The range edit also includes a check for blanks/omissions. This edit should produce error listing and edit diary. The consistency edit checks whether a predetermined relationship exists between two or more data items within a record as well as between records. This edit check will also produce error listing and edit diary. The edit program is used to incorporate the corrections in the file. All updated records are reprocessed through the edit program till the data are cleaned. Raising factors will be developed at district level separately for urban, rural and by sex and age groups for preparation of sample tables. Imputation is the process of assigning reasonable entries in the data field with missing or invalid entries. For this purpose the following methods are commonly used:

(a) Hot Deck method is the method where by missing/invalid entries are imputed using a valid response from the last record in the file with similar characteristics. As the data are processed the stored values are constantly updated by valid response.

(b) Cold Deck method provides help to the missing/ invalid entries which are imputed using the fixed values from the auxiliary data e.g., data of previous census or pilot census.

(c) Frequency Distribution technique provides support to the missing or invalid entries which are imputed on a proportional basis, from a distribution of valid responses of the current census/survey.

The last step in ICR is quality control of computer edits. The purpose of the quality control is to prevent distortion, duplication and a loss of data. Excess editing should be avoided. A diary should be prepared to show the results of computer editing. This diary should contain the number of errors, types of errors and percentage of errors by area. These diaries are used to document the results of an edit and information about changes made in the data.

Fig.6 shows the choice of topics to be covered in a census which depends on national needs and which is tremendously increasing, while census has its own limits. Thus, a balance has to be maintained in what census can produce with desired accuracy against the demands of most of data users (GoP 2010).

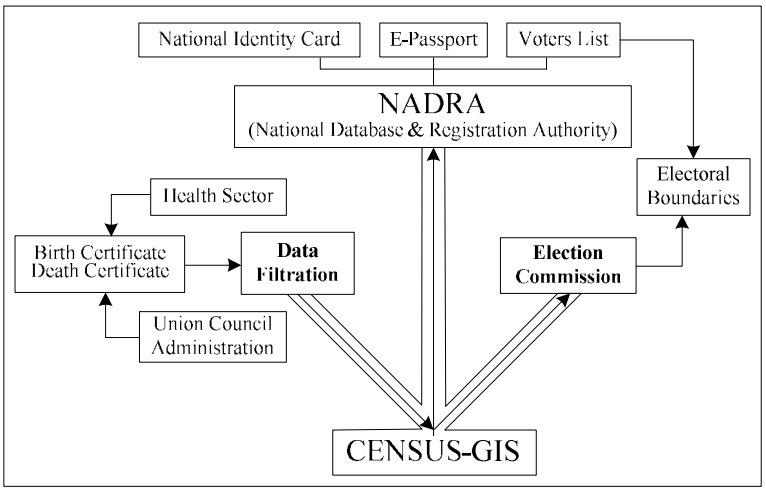

Figure 6. *Collaboration with other departments*

Conclusion

In developing countries like Pakistan the GIS is a new technology which utilization has increasingly enhanced the work of government organizations, especially in terms of data base development. This technology growth is promising not only in all sectors of planning at the governmental level but also at the nongovernmental level. The Statistics Division of the United Nations has played a very constructive role in promoting the development of GIS based census technology especially with reference to software equipments and manpower training. Qualitative as well as quantitative benefits of handheld devices (Tablet, Mobiles etc.) have been proved in the field in many countries (e.g. Australia, Canada, Brazil, Malaysia, New Zealand, etc.) and there are various options available for selecting those devices. Clear objective identification is necessary in the selection of the best devices to be used.

Census GIS provides immense help not only in regional planning but also in disaster management. There is great need, therefore, for detailed research in applications of GIS as well as other GIS technologies with regard to census data to inform comprehensive policy formulations and decision making.

References

Barr R (2007) A comparison of aspects of the US and UK censuses of population. *Transaction in GIS* **1** (1), 49-60.

Briggs D, Fecht D, Hoogh-de K (2007) Census data issues for epidemiology and health risk assessment: Experiences from the Small Area Health Statistics Unit. *J. R. Statist. Soc.* **170** (2), 355-378.

Cai Q, Rushton G, Bhaduri B, Bright E, Coleman P (2006) Estimating small-area populations by age and sex using spatial interpolation and statistical inference methods. *Transaction in GIS* **10** (4), 577-598.

Camara SA, Dias EA (2004) Location-based services for WAP phone users in a shopping centre. In: Stillwell J, Clarke G (eds) *Applied GIS and spatial analysis*, pp. 55-70. Wiley.

Côté JM (2005) A note on 'Bed allocation techniques based on census data'. *Socio-Economic Planning Sciences* **39** (2), 183-192.

Duke-Williams O, Rees P (1998) Can offices publish statistics for more than one small area geography? An analysis of the differencing problem in statistical disclosure. *International Journal of Geographical Information Science* **12** (6), 579-605.

Eze CG (2009) The role of satellite remote sensing data and GIS in population census and management in Nigeria: A case study of an enumeration area in Enugu, Nigeria. *Scientific Research and Essay* **4** (8), 763-772.

Gregory IN, Ell PS (2005) Analyzing spatiotemporal change by use of National Historical Geographical Information Systems; population change during and after the Great Irish Famine. *Historical Methods* **38** (4), 149-167.

Harris JR, Longley AP (2000) New data and approaches for urban analysis: Modelling residential densities. *Transaction in GIS* **4** (3), 217-234.

Huda SN, Burke F, Azam M (2011) Socio-economic disparities in Balochistan, Pakistan–A multivariate analysis. *Geografia - Malaysian Journal of Society and Space* **7** (4), 38-50.

IHARA H (1998) *The outline of the CMS in the population census of Japan*. Statistics Bureau Management and Coordination Agency, Government of Japan.

Laldaparsad S (2007) Census mapping and the use of geo-spatial technologies (A case of South Africa). United Nations Expert Group Meeting on Contemporary Practices in Census Mapping and Use of Geographical Information Systems, New York. , 29 May – 1 June.

Martin D (1998a) 2001 Census output areas: From concept to prototype, *Population Trends* **94**, 19-24.

Martin D (1998b) Optimizing census geography: The separation of collection and output geographies. *International Journal of Geographical Information Science* **12**, 673 – 685.

Martin D (2000) Towards an Integrated National Socioeconomic GIS: The geography of the 2001 census in England and Wales. 3rd AGILE Conference on Geographic Information Science, Finland. 25 – 27 May.

Martin D (2002) Geography for the 2001 census in England and Wales, *Population Trends* **108**, 7-15.

Martin D (2004) Neighbourhoods and area statistics in the Post 2001 Census Era. *Area* **36** (2), 136-145.

Martin D, Nolan A, Tranmer M (2001) The application of zone-design methodology in the 2001 UK census. *Environment and Planning A* **33**, 1949-1962.

Martin D, Tate JN, Langford M (2002) Refining population surface models: Experiments with Northern Ireland census data. *Transaction in GIS* **4** (4), 343-360.

Minot N, Baulch B (2005) Poverty mapping with aggregate census data: What is the loss in precision? *Review of Development Economics* **9** (1), 5–24.

Mobbs J D (1998) Australia comes to its census: The public sector mapping agencies and the 1996 Australian Census of Population and Housing. Public Sector Mapping Agencies, Australia. Available from: http://www.geom.unimelb.edu.au/fig7/Brighton98/Comm7Papers/SS34Mobbs.html.

Mokhele T (2011) The applications of GIS in census data. Human Science Research Council Seminar, South Africa. 9th May.

Norman P, Bambra C (2007) Incapacity or unemployment? The utility of an administrative data source as an updatable indicator of population health. *Population, Space Place* **13**, 333–352.

Norman P, Simpson L, Sabater A (2008) Estimating with confidence and hindsight: New UK small-area population estimates for 199. *Population, Space Place* **14**, 449–472.

Openshaw S (1977) A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling. *Transactions of the Institute of British Geographers, NS* **2** (4), 459-472.

Orford LS, RadcliVe J (2007) Modelling UK residential dwelling types using OS mastermap data: A comparison to the 2001 census. *Computers, Environment and Urban Systems* **31**, 206–227.

Owiti PO (2008) Embedded GIS for Census Mapping [cited 11 Aug. 2010]. Nairobi, Kenya. Available from: http://www.map-gis-rs.blogspot.com/2008/03/gis-for-census.html.

Prasad M (2006) Maps and census. *GIS Development: Asia Pacific. The Monthly Magazine on GIS,* **10** (12), 7-48.

Ragains P (1995) Promoting U.S. census data in an academic environment: Opportunities and challenges. *Journal of Government Information* **22** (4), 321-335.

Ralphs M, Ang L (2009) Optimized geographies for data reporting: Zone design tools for census output geographies. *Statistics New Zealand Working Paper No 09–01*. Statistics New Zealand, Wellington.

Raymer J P, Smith WFP, Giulietti C (2009) Combining census and registration data to analyse ethnic migration patterns in England from 1991 to 2007. *Population Space Place* **15**, 565.

Schlossberg M (2003) GIS, the US census and neighbourhood scale analysis. *Planning, Practice and Research* **18** (2-3), 213-217.

Schroeder PJ (2007) Target-density weighting interpolation and uncertainty evaluation for temporal analysis of census data. *Geographical Analysis* **39**, 311–335.

Shen J (2005) Counting urban population in Chinese censuses 1953-2000: Changing definitions, problems and solutions. *Population, Space Place* **11**, 381-400.

Tye A (2009) The emergence of GIS and GPS integrating with census data and its impact for the developing world. *The 57th Session of The International Statistical Institute,* South Africa. 16-22 August.

UN (2000) *Handbook on Geographic Information Systems and digital mapping. Studies in methods, Series F No. 79*. United Nations Publication, New York.

UN (2004) Integration of GPS, Digital Imagery and GIS with Census Mapping. United Nations Expert Group Meeting to Review Critical Issue Relevant to the Planning of the 2010 Round of Population and housing Censuses, 15-17.

UN (2009) Handbook on Geospatial Infrastructure in support of census activities. Studies in methods, Series F No. 103. United Nations Publication, New York.

Verhoef H, Grobbelaar N (2005) The development of a small area layer for South Africa for census data dissemination. Statistics South Africa. Available from: http://www.cartesia.org/geodoc/icc2005/pdf/poster/TEMA26/HELENE%20VERHOEF.pdf.

Wall HND (2007) GIS based dissemination of census data in Trinidad and Tobago: A Caribbean Experience. Central Statistical Office, Ministry of Planning and Development, Republic of Trinidad and Tobago. [Cited 11 Aug, 2010]. Available from: http://www.unstats.un.org/unsd/demographic/meetings/wshops/Trinidad_22Oct07/docs/countries_presentations/Trinidad .

Wang X, Liu Z, Chen S (2001) Some issues on urban census GIS designing. *Geo-spatial Information Science* **4** (4), 25-31.

Wu S, Wang L, Qiu X (2008) Incorporating GIS building data and census housing statistics for sub-block-level population estimation. *The Professional Geographer* **60** (1), 121-135.

Xiru Li (2009) Mapping and GIS application in population census of China. 24th Population Census Conference, Hong Kong, China. March 25-27.