

Form Processing Technology used in Census of India 2001

Census data processing

Application of Image based form processing technology has been the salient feature of the Information technology plans for the 2001 census. In census data processing one of the major constraints has ever been handling of the large scale data entry from paper based census schedules. The new technology has been a boon as it helped in creating data base of more than one billion individuals in record time of about 2 years.

Technology

Selection of technology was a complex decision making process given the size of Indian census and uncertainties associated with the technology. After interacting with various vendors and going through the demonstrations of their products Automatic Form Processing Systems (AFPS/ E_flow) of Top Imaging Systems, Israel, High end scanners from Kodak and software integrator M/s CMC Ltd were short listed. The results were encouraging and we could process the Houselist and Household project of more than 200 million forms covering more than a billion population in a record time frame of two years.

Business Model

Objective was to convert the information recorded in census schedule to digital form through ICR technology besides maintaining the quality within a span of two years. Since the output is dependent on the quality of images and performance of the ICR software, it was felt appropriate to have single point responsibility on this count. Accordingly the special hardware (scanners) and ICR software were tied up with the front end integrator i.e. M/s CMC Ltd. with Top Imaging System (TIS) and M/s Kodak India Ltd as partners. Single point control with the integrator and productivity linked payment plan ensured that goods were delivered in time besides ensuring that Integrator plays a crucial role in assuring the quality of Images and ICR performance. Initially only 20% of the project amount was paid

at the time of installation and subsequently proportionate payments were made quarterly based on performance i.e. actual number of forms converted to digital data. The payment was subject to ensuring more than 90 percent recognition and less than 2 percent false positive. The project management including operational management was within the organization while the integrator was to extend support services for ensuring smooth operations utilizing organizational manpower. Quality controls were introduced for batches picked up through selection process. Comprehensive maintenance arrangements were made for scanners including stocking of spares etc.

Implementation details

The Image based form processing technology (using ICR/OCR) was introduced for the first time in India for processing census 2001 data. Also 100% census data processing was taken up for the first time in India. About 9 million “House listing” schedules and 203 million “Household” schedules were processed through this technology. The technology was implemented at 15 ORGI offices located at 15 major state capitals.

Data was captured using 25 high end Kodak Scanners. In all about 45 NT servers (P-III 733 MHz, 512 MB RAM, 216 GB HD etc), 1060 PC (P-III 800 MHz, 128 MB, 20 GB HD etc.), 25 high end duplex scanners (40 sheets of A3 size in one minute) were installed under Local Area Network along with others supporting peripherals like printers (Laser & Line), DLT drives, CD writers, ZIP drive and SLR tape drives.

The processing of 9 million (100%) “House listing” schedules was targeted and achieved in 6 months time period. Similarly 203 million (100%) “Household” schedules was targeted and achieved in 18 months time period. Besides Computer Assisted Coding (CAC) for the descriptive fields like Languages spoken, Education level, Migration particulars have been completed. The CAC for National Industrial Classification, National Classification of Occupations etc have also been taken up.

The both side filled-in A-3 size schedules were scanned by using the High-volume-production scanners and the electronic images were created.

Various data checks like range checks, field checks and inter consistency checks between the fields were applied on the ASCII data file besides full fledged editing with Auto correction/Imputations to enhance the quality of the data.

Impact

Impact of the Information technology in Indian census could be highlighted as follows:

1. Compilation of basic population data (Primary Census Abstract) which was hitherto done on manual basis is totally computerized which resulted in not setting up of about 200 regional tabulation offices resulting in savings of 6.5 billion rupees (USD 150 million) to the Government.
2. Entire census data belonging to more than one billion population has been computerized using ICR (Intelligent Character Recognition) based form processing technology.
3. Almost all the census tables can now be prepared on 100 percent basis.
4. Images of all the census schedules are available electronically and can be called back for other processing work at any point of time in future.
5. The technology is being assimilated within the organization. The infrastructure and the software is being put to use for other projects like Economic Census, Sample Registration System and Verbal Autopsy etc.