System Design

System Design for data processing

In accordance with the requirements of large-scale data processing, the total data processing work is structured over a number of functional steps. As the data processing work is decentralized and a number of people are engaged in the data processing, each step of it is planned from beforehand, documented in detail, the underlying concepts and definitions are explained, training workshops are organised both centrally and locally, and mid-course discussions are also held to resolve unforeseen data problems. Hence, it is a formal system of data processing which involves the following stages:

NSSO has been shifted from PAPI to CAPI which comes under e-SIGMA, (e-Survey Instrument and Generalised Multimodal Application) project of MoSPI's ongoing IT initiatives. It has been conceived to transform the data collection and data processing system of NSSO into a more simplified and efficient one. The endeavour is to convert from the current Paper-schedule mode to Computer Assisted Personal Interview (CAPI) mode with in-built data validation to enable a faster processing of quality data through a synchronized setup.

Major tasks of DPD starts after the finalization of questionnaire and survey instruments, Computer Scrutiny Software programme (CSP) and determining the sample size.

Flow of survey process in the field and data finalization at DPD is given below:

- Frame finalization: The rural and urban frame are finalized based on the inputs received from FOD in consultation with SDRD.. Sampling is done from the finalised frame in order to draw samples in an unbiased manner.

- Sample selection & preparation of sample list for Central & State samples: The rural/urban samples are selected on the basis of some pre-fixed conditions involving classification into different groups. In this regard separate list of state sample and central sample is prepared.

- Data Layout Preparation: Organization of data according to a pre-set manner for computer processing.

- Computer Assisted Personal Interview (CAPI)) preparation and testing: Software application (both windows version and android version) are developed by the outsourced Vendor under the guidance of DPD and other stakeholders as per the finalized survey instrument for collection of data with in-built data validation to enable a faster processing of quality data in a synchronized manner.

- During the process of development by the Vendor, the integrated CAPI with eSIGMA digital platform, thorough testing is done for proper implementation of the CAPI as per survey instruments.

- The CAPI/Web-portal is developed afresh for each new NSS survey. The primary

purpose of the entire system is to facilitate faster processing of quality data through comprehensive digitization of major functions of survey administration and monitoring.

- Sharing CAPI to field for large scale testing before AIWOT: The software for individual NSS surveys is then shared to field for large scale testing to perform user acceptance test. Changes are incorporated on the basis of the feedback received.

- Finalization of CAPI up to pre-AIWOT level and demonstration thereof: All India training is organized prior to launch of surveys for field staffs wherein the ~~finalized version of~~ CAPI is demonstrated.

- Preparation of other modules of eSIGMA.: Other peripheral Modules of eSIGMA like Survey Setup Module, User Management Module, Self-Learning Module, Query Module, Dashboard Module, Data Download & Data Backup Module etc. are also prepared and tested.

- Testing of Final version of CAPI for User Acceptance Test(UAT) and other preparatory works before roll-out: Final version of CAPI is tested with incorporated feedback received from AIWOT by FOD for UAT along with other relevant modules. Once the UAT is done, final Go-Live is done for survey.

- Checking the Data by Data Supervisor (DS Users) of FSUs as received from supervisor of FOD and referring back to Field in case of discrepancies: The data which is received form the supervisor of FOD are then checked and verified by DS Users based on the concepts and definitions of the survey as well as they check the remarks by the field enumerators and field supervisors to ensure the quality of data. In case of any anomalies/doubtful cases, the data is referred back to the field for further checking/ correction/ justification.. At this stage:

  I. coverage of data vis-a-vis listing done by field enumerators; remarks/ clarification provided by enumerators are examined.
  II. Additional warnings are generated by DS users after thorough checking of all the items **of each and every survey schedule** based on the concepts and definitions for referring back to FOD for further clarification/ correction.
  ~~III.~~ Check for already action taken by enumerators/ Supervisors **in each and every survey schedule**.
  IV. person-wise or item-wise information furnished in different blocks are checked for consistency with respect to various codes
  V. Unit value checks for quantity and value fields

- Additional validation checks: Data is subjected to additional validation checks which are developed on the basis of previous experiences of scrutiny in an aggregate manner
- Development of multiplier generation and other necessary software for tabulation: Multiplier generation software and other software are prepared during the course of the survey. At this stage multiplier files are prepared, i.e. calculation of weighting factors for each Ultimate stage Units as per sample design. During the survey multipliers are generated based on a sufficiently large data set for finalisation of the multiplier generation software and some key tables are also generated to see whether any intervention is required in data collection process. And once the survey is complete and the complete data is received from the field and all validation has been checked, multiplier program is run to generate the weights to be used in tabulation.
- Development of Tabulation software: Tabulation software are prepared during the survey period based on the tabulation plan finalized by the working group of the survey. The tables are usually generated sector x state x sex x other socio-economic category-wise. Trail tables are generated based on sufficiently large data sets for finalisation of table generation software during the survey period. Once data collection is complete and multiplier is generated, tabulation software is run on the multiplier posted data to generate draft tables. Based on the tables generated by DPD, the SDRD prepares subject wise reports which after due approval are subsequently published.

  Finalization of data and tables and dispatch of unit Data: The draft tables are then checked and sent to SDRD for final checking. In case of no discrepancies, the tables are sent to SDRD for preparation of report/keynotes and the unit level data are sent to Computer Centre for dissemination along with meta data after the release of report.