

DPDSYS: An Integrated Data Entry System for Sample Surveys (Indian experience)

1. Introduction:

1.1 In India Socio-economic surveys are conducted by National Sample Survey Organisation(NSSO) under the Ministry of Statistics and Programme Implementation. Each year there could be two or more surveys run concurrently on two or more than two subjects and is termed as a NSS round. (There were occasions a NSS round was conducted for a period of six months e.g, 47, 49, 54, 58 and 60 NSS rounds).

1.2 The NSSO data processing work is done through a decentralised system through Eight Data Processing Centres (DPCs) located at various parts of the country. The processed data based uniform data validation and cleaning rules are processed at the main Data Processing Division (DPD) where DP Centre-wise datasets are amalgamated for each subject and processed for making tabulations.

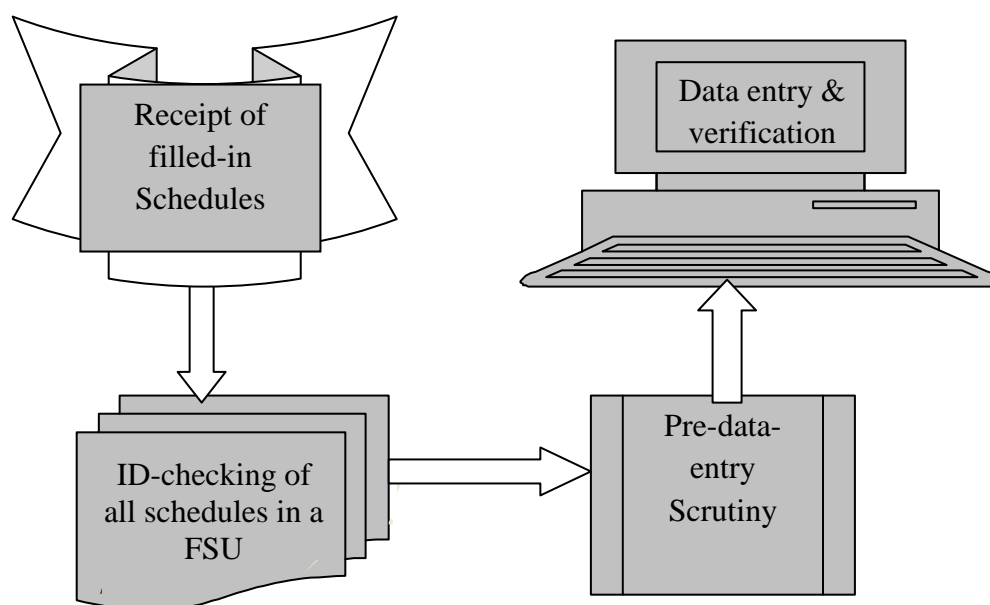
1.3 The Data Entry package used till 1999 was not based on integrated approach and as a result, there used to be mis-match among data entered from schedules/ questionnaires under same subject of a NSS round. It may be mentioned that for each specific survey, there are a minimum of two schedules – namely, listing schedule and a main survey schedule/questionnaire. Thus, within a selected FSU there could be many filled-in schedules along with a single filled in “Listing schedule”. Without having an Integrated Data entry module, there used to be lots of data mis-match resulting in considerable time over run for validation of survey data.

1.4 In mid-1999, the Data processing Division took an exercise to develop an Integrated Data entry and data validation software using GUI based interface. The result is DPDSYS. This is a unique kind of software but before it is introduced the same was tested under various conditions and various DP Centres. Based on extensive testing this inhouse developed software was finally used for NSS surveys from July2001-Jun02 round . In the foregoing sections various features of the DPDSYS has been described.

2. Features of DPDSYS:

2.1 Data entry is one of the basic steps of data processing of large scale sample survey that NSS does. Data entry of filled in schedule of enquiry received from Field Operations Division (FOD) of NSSO is done by screen-based data entry **software called DPDSYS developed in-house by DPD**. The software is fully menu-driven and user-friendly. Use of this software has greatly reduced and simplified the steps of data processing.

The following chart shows basic steps of data entry in NSS Round:



2.2 The work of Data Entry is done through eight(8) Data processing Centres. Each DPC has three(3) basic Units- Document Control Section (DCS), Scrutiny Section (SS) and Data Entry Section (DES). Schedules of enquiry are received at DCS from Regional/ Sub-regional offices of FOD spread over different parts of the country. A list indicating which FOD office should send canvassed schedules of enquiry (FSUs) to which DPC is sent before the survey work starts. Accordingly, FOD offices send all the canvassed FSUs to concerned DPC. In the DPC, DCS receives all the FSUs. For this in each DPC, 'Receipt-Register' is kept by DCS to record and monitor receipt position of different types of schedules. The following checks are exercised at DCS at the time of making entry in Receipt Register on receipt of schedules of an FSU.

Identification particulars of an FSU- Sample Serial No. (FSU no.), Sector, State code, Stratum, sub-Stratum etc.

Identification particulars of an SSU- FSU No., HG/SB No., Second Stage Stratum no. (SSS), Sample Household number.

2.3. **Schedules of enquiries of an FSU are not segregated. This is called LOT (= Batch).** Each LOT accompanies a FLY-SHEET indicating its composition. A “movement register” is maintained by DCS to keep an account of the movement of LOTs to and from different sections. DCS sends these lots (FSUs) to SS for ‘**Pre-data-entry scrutiny**’ and thereafter to the DES for data-entry & verification. A register is also maintained in the DES to record the movements of the LOTs. FLY-SHEET is updated at each stage of processing.

2.4. Once pre-data entry scrutiny job is over, the FSUs are ready for data entry and verification job. This job is done by DES of DPC using the data entry software called DPDSYS. It has two 'list of options' in the opening screen. First 'list of option' contains option for different types of operation i.e. 'Data Entry', 'Update', 'Delete', 'Verify', 'ID_Update_SSU' and 'Exit'. Second 'list of option' contains option for selecting schedule type for which data entry or update or verification etc. is to be done i.e. 'Sch 0.0', 'Sch 1.0' and 'Sch 10'.

Next option is the text box for entering the FSU serial no. Once this is entered and if it is a valid sample number given as per the sample list file, one data file with extension '.DAT' will be created automatically by the Data Entry Package in a specified path of the Computer '**C:\R68\DATA**' for **NSS 68th Round of NSSO**. Data of all the schedules of an FSU will be stored in one single '.DAT' file. This '.DAT' file is an encrypted file in text format. One separate software is used for conversion of encrypted '.DAT' file into text data files as per text data layout.

Data file naming convention:

The naming convention of a data file is structured as explained below:

- (i) The first two entries of the batch file will refer to a round number i.e. 68 for 68th round of NSSO.
- (ii) The third letter of the data file represents the Data Processing Centre.

A	—	Ahmedabad
B	—	Bangalore
C	—	Kolkata, HQU-III
D	—	Delhi
E	—	Kolkata, HQU-I
F	—	Kolkata, HQU-II
G	—	Giridih
N	—	Nagpur

One example of data file name created by DPDSYS is "**68D12345.DAT**", where

- '68' stands for 68th round (fixed for all data files of 68th Round),
- 'D' is the code for Data Processing Centre (viz. **Delhi**),
- '12345' represents the FSU No.,
- 'DAT' is the file extension name used by DPDSYS package,

5. Workload during NSS 68th round

Srl. No.	Sch. No.	Description of Subject	Unit for Canvassing Schedule	No. of FSUs		No. of SSUs per FSU
				Central Sample	State Sample	
1.	0.0	List of Households	Village & Urban Block	12784	14988	--
2.	1.0(Type-1)	Household Consumer Expenditure	Household	12784	14988	8
3.	1.0(Type-2)	Household Consumer Expenditure	Household	12784	14988	8
4.	10	Employment & Unemployment	Household	12784	14988	8

6. Scope of data entry

A particular Data Processing Centre will be allowed to enter data of only those FSUs which are allocated to them as per the **sample list**. The sample list file to be used in DPDSYS package is provided in the installation CD. The identification fields are checked online by the software against their respective values for their validity. Data entered in each field is as per text data layout. Some validation checks are also kept in the data entry package. If provision of bypassing validation checks is kept with any particular field, background colour of that field will **change to WHITE** as soon as the check fails. This is to distinguish it from other '**hard checks**' where the background colour will **change to RED** if the check is not satisfied. When the colour changes to white, one can either correct the entry or press Escape and Return to bypass the check, if no suitable correction is known at that moment. But for hard check, validation rule must be satisfied.

7. Order of data entry

In NSS 68th round, there are four schedules viz. Schedule 0.0, Schedule 1.0 (Type-1), Schedule 1.0 (Type-2), and Schedule 10. All the screens of all schedules are integrated in a single software. **At first, the data entry work for Sch.0.0 will be completed.** The data file created in this process will contain coverage information of all the schedules as given in the Block-6 of Sch. 0.0. Next, data entry and verification work of other schedules may be completed in any order.

All the data entry and verification works relating to the schedules of a single FSU have to be finished in a single go, before taking up another FSU.

Opening screen of data entry package:

8. Data verification

After completion of data entry, verification of each schedule is done by selecting **VERIFY** option from the Opening Screen in order to ensure correctness of data entry as per filled in schedule. The job of verification of an FSU is done independently. In case there is a verify mismatch in a particular field, the background colour of the field turns **blue**, and the cursor halts there. The user should **re-enter the correct figure of the field twice successively**. The entry will then be accepted as final, provided it does not fail any other validation rule imposed in the field. **Unverified schedules are not accepted for subsequent stages of processing.**

9. Data updation

While data entry is in progress in respect of any particular household of a schedule and it is not yet saved, one can go back to any field of any previous block by using **PAGEUP**

and UP/DOWN ARROW keys, and content of the field can be changed. **But once the data is saved, changes in the already entered data can be effected only in UPDATION mode.** The following figure shows the **Updation screen**, which comes up when **UPDATE** option is selected in the Opening Screen.

Updation Opening Form				
Please Enter Schedule Identification				
FSU	Hg./Sb. No.	SSS	HHS. No.(SSU).	
12345	1	1	01	O.K.
<p>F1=Help,F2/F7=Continue,F5=Quick Exit,F10=Save & Next Sch., F12=Save & Main Menu, PageUp/Dn=Last/Next Fo</p>				

10. Updation of key identification fields of a sample household (ssu)

For a sample household (SSU) of an FSU, Key Identifications fields are- Hg/ Sb, SSS and Household No. While data entry is in progress in respect of any particular household of a schedule and it is not yet saved, one can go back to the screen for Block 1 and any of these identification fields can be changed. **But once the data is saved, changes in the identification fields can be effected only in ID_UPDATE_SSU mode.** The following figure shows the **ID Updation form of an SSU**, which comes up when **ID_UPDATE_SSU** option is selected in the Opening Screen.

ID Updation form for a SSU				
Please Enter Schedule Identification you want to update				
FSU	Hg./Sb. No.	SSS	HHS. No.(SSU).	
15517	2	2	02	
Please Enter modified Schedule Identification				
Hg./Sb. No.	SSS	HHS. No.(SSU).		
1	2	02		O.K.
<p>F1=Help,F2/F7=Continue,F5=Quick Exit,F10=Save & Next Sch., F12=Save & Main Menu, PageUp/Dn=Last/Next Fo</p>				

11. Data deletion

Just as in the case of UPDATION procedure, one has to select DELETE menu from the Opening Screen and the **Deletion Form** will come up. It is exactly the **same as updation form**. By entering the Key Identification fields, one can **delete complete data of one household at a time**. By repeating this process, more than one households of any particular type can be deleted.

It may be noted that, deletion of schedules will automatically attract data integrity and coverage checks. For example, deletion of Schedule 0.0 is not permitted if there are data of other detailed schedules related to this FSU exist.

12. Despatch of data files to CPU of DPD (HQ), Kolkata

On completion of data entry, verification and various stages of processing, data files of all the FSUs allotted to DPCs are sent to the CPU of DPD (HQ) for further processing and tabulation.

Sample formats of various Register

RECEIPT REGISTER: NSS 68th ROUND

July 2011 – June 2012

(DCS: DOCUMENT CONTROL SECTION)

Page No.:

D.P.Centre :

State / U.T. :

FOD Region :

Sector : RURAL / URBAN

Sub Round :

FSU NO.	Number of Schedules received for Schedule Type				Date of despatch by FOD	Date of receipt at DPD	Remarks
	0.0	1.0(Type-1)	1.0(Type-2)	10			

MOVEMENT REGISTER: NSS 68th ROUND

July 2011 – June 2012

DOCUMENT CONTROL SECTION

Page No.:

Activity: PDES & DE/ DV

D.P.Centre :

State / U.T. :

FOD Region :

Sector : RURAL / URBAN

Sub Round :

FSU No. / Lot No.	Date of Despatch to SS	Date of Receipt from SS	Date of Despatch to DES	Date of Receipt from DES	Remarks

--	--	--	--	--	--

LOT REGISTER: NSS 68th ROUND

July 2011 – June 2012

DATA ENTRY SECTION

Page No.:

Activity: Data Entry /Data Validation

D.P.Centre :

State / U.T. :

FOD Region :

Sector : RURAL / URBAN

Sub Round :

FSU No. / Lot No.	Date of Receipt from DCS	Name of DPA for DE/DV	Date of Allotment for data entry	Date of DE	Date of DV	Date of despatch to DCS	Remarks

LOT REGISTER: NSS 68th ROUND
July 2011 – June 2012
DATA ENTRY SECTION

Page No.:

Activity: Updation

D.P.Centre :

State / U.T. :

FOD Region :

Sector : RURAL / URBAN

Sub Round :

FSU No. / Lot No.	Date of Receipt from DCS	Name of DPA for Updation	Date of Allotment	Date of despatch to DCS	Remarks