# Experience of NSSO with CAPI in Periodic Labour Force Survey

**FAO-UN CAPI Training**

**IASRI, PUSA, New Delhi, 11.07.2016**

# Introduction

For estimation of level of various labour force indicators and changes thereof at quarterly interval in urban areas, NSSO has designed Periodic Labour Force Survey(PLFS).

❖ It is based on rotational panel design, with 25% replacement of FSUs in each quarter. The selected households in an FSU will be surveyed for four successive quarters.

❖ During 2011-13, PLFS was piloted during 8 successive quarters in 3 states namely HP, Odisha and Gujarat.  In total 288 urban FSUs were surveyed with 8 households per FSU. The survey used paper schedules (PAPI).

❖ In PLFS(P), a 2-stage stratified sampling design was adopted. Complete listing of the HHS was done in the sample FSUs using a Listing Schedule (0.0).  In the second stage, listed HHS were stratified and 8 HHS were sampled. In the 1$^{st}$ visit to the HHS, paper Schedule 10.3v1 was canvassed. In the remaining 3 revisits, paper Schedule 10.3v2 was canvassed.

# Trying out CAPI-(1)

In order to evaluate the feasibility of WB CAPI solution for PLFS survey, a fresh pilot survey was conducted using CAPI covering some of the HHS of 7th quarter, with revisit in 8th Quarter of PLFS(P). That is, only one revisit was tried out in CAPI.

- In order to compare the PAPI data and CAPI data, 40 FSUs (i.e. 240 HHS)were initially considered.

- Field Offices to supply address of these 240 households along with other identification particulars from paper Listing Schedule of 7th Quarter of the PLFS(P), so that interviews could be generated in CAPI.

- NSSO obtained **NLTA (Non Lending Technical Assistance) from World Bank for a period of 1 year with effect from 02.12.2014** for conducting the pilot testing of WB CAPI solution for PLFS.

# Trying out CAPI-(2)

- World Bank provided all sort of technical support in designing the Questionnaire for testing, creating server domains in their server for conducting the pilot survey.

- During the development process they were in constant interaction with NSSO officers involved in designing the CAPI solution. Both World Bank Team and World Bank CAPI support team namely Mr. Michael M. Lokshin, Mr. Sergiy Radyakin, Mr. Zurab Sajaia, Ms. Rinku Murgai and Ms. Urmila Chatterjee provided support on logistic and technical issues.

- After developing the CAPI questionnaire by DPD, it was shared with WB Team for their review and suggestions for improvements and also for providing necessary remedial solutions to problems faced during designing stage. World Bank CAPI Team was regularly interacted through online email support and Skype, several meetings and VCs.

# Design-related issues

**(a) Second stage sampling** :

- CAPI does not allow online processing of collected data. It does not allow multiple schedules (e.g. 0.0, 10.3v1, 10.3v2) in a single survey.

- For 2nd stage sampling, the 200-500 listed HHS need to be grouped into 3 strata, and 6 sample HHS need to be selected using random numbers.  This was not possible in the field using CAPI.

- With roster size is limited to 40, it required a varying number of rosters, besides a separate survey for Listing Schedule.

- So, it was decided to use PAPI for Listing and 2nd stage sampling, as usual.

- However, for quick availability of the multiplier-related (i.e. design weights) information, a separate template was created on CAPI.

- Besides, two separate templates were created for v1 and v2 schedules.

# Design-related issues

**(b) Assignment of interviews**:

- In NSSO surveys, HHS (interviews) are assigned by the 49 regional offices independently, after the listing and 2$^{nd}$ stage sampling are completed.

- In CAPI interviews need to be assigned right at the beginning by the "Headquarter".

- This problem was solved by creating multiple HQ users in a single NSSO domain. **In the revised CAPI solution, provision of multiple HQ was developed as per demand of NSSO**.

- Names of Supervisors and Investigators were obtained from the Regional offices. Using respective Headquarter privileges, roles were created using batch upload.

- The created roles along with login/password were sent to respective R.O.s

# Design-related issues

**(b) Assignment of interviews – contd.**:

▸ As number of HHS per FSU is fixed (6), and IDs of the sampled FSUs are known,  separate .tab files were prepared with complete FSU ID and serial No. 1 to 6 as dummy HHS ID.

▸ Cases were generated at NSSO(Hq), segregated R.O.-wise and distributed to the R.O.s for assignment of cases to interviewers.

▸ The Interviewers linked the generated HHS-ID with the sample HHS-serial No. maintained  in the Listing Schedule (0.0) of the relevant FSU. The detailed addresses of sample HHS were available there.

# Design-related issues

**(b) Assignment of interviews – contd.**:

- Though under multiple HQ support in the revised server solution, user resources like Supervisors/ Investigators teams, interviews created by one HQ can be accessed / utilised by another HQ and act upon the interviews in terms HQ privilege, no serious threat was felt in actual survey operation.

- Only supervisor assigned and/or interviewer assigned questions (blank questionnaires/interviews) created by one HQ can be deleted by another HQ. Completed interviews cannot be deleted by any HQ user. Administrative instructions was given to HQ users not to act upon questionnaire created by another HQ.

# HW & SW resources

**(c) Server resources**:

➡ NSSO decided to host the CAPI solution on its own server, which is NIC cloud HW & SW resources, which is Govt.-owned.

➡ NIC requires auditing of CAPI software by empanelled vendors, before hosting. For this, two types of server resources are required – one for staging to facilitate SW audit, and the other for production.

➡ Keeping in view that around 100 Supervisors and 650 investigators likely to be created in the main PLFS survey, and the resultant peak data traffic, WB CAPI representative specified the resources required in the NIC Cloud Service.

➡ *Required SW resources  : MS Windows Server 2012R2, .Net Framework 4.5.2+, PostgreSQL 9.4.3. Microsoft C++ 2013 runtime.*

➡ For each of WEB server and DB server, CAPI requires  1 server instance with CPU: 32 cores; Memory (RAM):128GB; Storage: 2TB on SSD drives;

# HW & SW resources

## (d) Specification of Tablets used in Pilot :

- In the field, it is essential that the devices should have 3G service facilities for downloading interviewer application, assignments, uploading collected data, syncing with the server etc. Tablets with the following specifications were procured :

- Operating System -Android 4.2.

- Processor Speed -1.2 GHz or more

- Memory - 8 GB or more with provision of micro SD external memory slot. Internal RAM 1 GB or more

- Display - Screen size 10/10.1 inches , Resolution 1280X800 or above with capacitive multi touch

- Camera - Rear camera 3MP or more, front camera (1.3 MP or more)

- Connectivity- Bluetooth, WiFi, 3G either through 3G SIM (slot) or dongle ( USB)

- Battery - at least 8 hours in built power backup

- Power Bank - As per requirement in the field

# Issues in Questionnaire designing

**(e) grid layout - Navigation**:

- NSSO PLFS schedule 10.3 has 6 blocks, some of which has a grid layout, with 10-12 columns in each row. In absence of grid layout in CAPI, the blocks were converted to one-way layout.

- Initially it was apprehended that this will cause major problem in data quality and time taken to canvass. But the comparison of PAPI and CAPI data proved that it was quite convenient. Random access to chapters (i.e. Blocks) and scrolling facility in tablets made the navigation easy.

- Moreover, grid layout on tablet screen would have been very congested.

- Comparison between PAPI and CAPI on "Time taken to canvass" confirmed that there was no inconvenience faced.

# Questionnaire designing

## (e) grid layout – Navigation Contd:

- It is expected that the time would be reduced once the investigators get more and more used to the Tablet and the CAPI interface.

| State | Comparison of Average Time taken (in Minutes) | | | |
|---|---|---|---|---|
| | PLFS-Q7 (Visit-1) | CAPI Visit-1 | PLFS-Q8 (Visit-2) | CAPI Visit-2 |
| Gujarat | 63.00 | 75.57 | 46.81 | 45.00 |
| HP | 39.39 | 39.06 | 26.39 | 35.94 |
| Odisha | 50.38 | 76.30 | 36.11 | 63.81 |
| All | 55.19 | 67.60 | 38.70 | 49.31 |

# Questionnaire designing

**(f) Pre-filling data**:

▶ The re-visit schedule contains many data items copied from visit-1 (e.g. FSU and HHS ID particulars, Person IDs, etc.)

▶ After receiving all HQ approved data, they were exported, and suitable .tab file were prepared for distribution to R.O.s for batch uploading to pre-fill visit-2 questionnaire.

▶ For this purpose, all interviews of Visit-1 need to be completed, before any of the re-Visit interview could take place.  This is because FSU-wise/Investigator-wise export (filtered export) of data is not possible in CAPI at any level (HQ or Supervisor) and export of entire data is to be done at HQ level.

▶ After data is exported, if it is edited externally, it can not be pushed back into the CAPI system.

# Questionnaire designing

**(g) Mid-course revision in survey instruments**:

➤ In WB CAPI, once a Questionnaire is revised in the designer tool and imported in the server module, a separate template is generated with same questionnaire name padded with version number.

➤ Such revised template is as good as separate survey questionnaire and all subsequent survey assignments are to be made w.r.t. the revised template. The questionnaire once finalized including validation checks, should not be changed during the course of the survey.

➤ This poses a limitation for NSSO surveys. Even if questionnaire is finalised after thorough piloting, validation checks are often fine-tuned/modified in the mid course.

# Data editing issues

**(h) Issues of data editing** :

- In WB CAPI, once the HQ approves the interviews it cannot be corrected by the investigator as the same cannot be pushed back to investigator. Thus any mistakes detected through any elaborate validation of data by exporting the same from HQ approved interviews, cannot be corrected in the CAPI.

- In NSSO 'data-dependent validation checks' (called 'Howler checks') are performed on consolidated data. Doubtful cases are referred back to the investigators for correction. This is not possible in CAPI.

- Often supervisor may need to edit the data, errors are detected by him. Regarding giving provision in the Supervisor module to change/modify data, it was informed by CAPI team that the same cannot be given.

# Data Validation

**(i) Template design & Validation checks - Advantages of CAPI :**

- Easy to handle and develop simple questionnaires like PLFS (visit-1)

- Inclusion of validation expression helps avoiding inconsistent /inadmissible entry

- Inclusion of Instruction helps interviewer for better understanding at the time of interview

- Conditional expressions allow complex routing

- Interviewers cannot miss mandatory questions and ask wrong question due to provision of mandatory questions and provision of Navigation Panel in the interview application

- Dashboard and Report facilities provide automated survey management tools with real time information on the status of survey

# Data Validation

**(j) Validation checks – Limitations faced**:

- Display/use of count of records, derived total from constituent items, any kind of auto-generated field etc. not possible in CAPI.

- On-screen calculator not available. Calculator of the Tablet to be used

- It was noticed that once radio button is touched on categorical item of single selection type, it gets selected and one cannot de-select it even with disabling by preceding conditional fields.

- Numeric fields whether integer or decimal does not allow entry of value more than 9999999. Mobile/Landline numbers were to be stored as text fields.

- CAPI allows to modify even pre-filled columns, which are only for viewing.

- Display filters on Rosters based on a previous Roster does not work.

# Feedback from Field Staff

**(k) Feedback/Suggestions on the Interviewers Module**

**(Feedback was collected on a structured Feedback sheet)**

➡ **Framing of questions, instructions, its sequence :-** All the investigators of the Regional Offices felt "Satisfactory/Extremely Satisfactory". Same experienced by the supervisors also.

➡ **Ease of navigations between questions, sections, blocks etc :-** All the investigators of the Regional Offices felt "Very Quick/Moderately Quick". Same experienced by the supervisors also.

➡ **Accuracy and quality of recording responses using virtual keyboard and touch considering interviewing process and handling device simultaneously including surrounding environment :-** Most of the investigators felt "Average/Accurate" response in terms of accuracy and quality of recording responses and also few investigators from RO, Ahmedabad felt "Below Average". All the supervisors felt "Average" response in this regard.

# Feedback from Field Staff

**Feedback/Suggestions on the Interviewers Module (contd.)**

- **Ease of recording responses considering mode of collections such as numeric field, text field, list, radio button or filter combo box :-** All the investigators of the Regional Offices felt "Easy" in recording responses in this regard. Same experienced by Supervisors also.

- **How user-friendly is application interface? :-** Most of the investigators have experienced "User-friendly" whereas very few investigators from RO(Ahmedabad) have experienced "Not at all user-friendly". Same experienced by the supervisors also.

- **How successful would be this solution in performing its intended task in large scale PLFS survey? :-** All the investigators of the Regional Offices have felt this device would be "Successful / May be Successful". Same experienced by the supervisors also.

# Feedback from Field Staff

**(I) Feedback/Suggestions on the Supervisor Module**

- **Scheduling and Allotting questionnaire to Investigators :-** All the supervisors have felt "Easy" in scheduling and allotting questionnaire.

- **Checking and assessing/reviewing the completed interviews :-** Most of the supervisors have experienced "Easy" whereas very few supervisors from RO, Ahmedabad/Vadodara have experienced "Difficult".

- **Revert back the completed interviews to investigator and final acceptance :-** All the supervisors have felt "Easy".

- **Are you satisfied with the CAPI supervisory role compared to the usual paper-based scrutiny system ? :-** All the supervisors have experienced "satisfied to a great extent/ to some extent"

# Experience of Data quality with CAPI

**(m) Comparison CAPI & PAPI on important indicators based on 82 matched HHS covered in CAPI pilot:-**

➤ Distribution of HHS were compared.

| Household type | CAPI data | PAPI data |
|---|---:|---:|
| 1 Self-employed | 23 | 23 |
| 2 Regular wage/salary earning | 35 | 36 |
| 3 Casual labour | 6 | 5 |
| 9 Others | 18 | 18 |
| Total | 82 | 82 |

| Social Group | CAPI data | PAPI data |
|---|---:|---:|
| 1 Scheduled Tribe (ST) | 7 | 7 |
| 2 Scheduled Caste (SC) | 12 | 11 |
| 3 Other Backward Class (OBC) | 26 | 27 |
| 9 Others | 37 | 37 |
| All | 82 | 82 |

# Experience of Data quality with CAPI

| Religion | CAPI data | PAPI data |
|---|---|---|
| 1 Hinduism | 79 | 79 |
| 2 Islam | 3 | 3 |
| All | 82 | 82 |

| Marital Status of Persons | CAPI data | PAPI data |
|---|---|---|
| 1 . Never married | 106 | 114 |
| 2. Currently married | 161 | 163 |
| 3. Widowed | 20 | 20 |
| 4. Divorced/separated | 4 | 2 |
| All | 291 | 299 |

| Education Level of Persons | CAPI  data | PAPI  data |
|---|---|---|
| 2 Literate with formal schooling: below middle | 59 | 64 |
| 3 Middle | 59 | 51 |
| 4 Secondary | 27 | 33 |
| 5 Higher secondary | 34 | 28 |
| 6 Graduate | 30 | 34 |
| 7 Post graduate and above | 11 | 10 |
| All | 220 | 220 |

# Experience of Data quality with CAPI

| Current weekly activity status Code of Persons | CAPI data | PAPI data |
|---|---:|---:|
| 11 Worked in h.h. enterprise (self-employed): own account worker | 33 | 40 |
| 12 Worked in h.h. enterprise (self-employed): employer | 0 | 1 |
| 21 Worked as helper in h.h. enterprise (unpaid family worker) | 13 | 10 |
| 31 Worked as regular salaried/ wage employee | 53 | 52 |
| 51 Worked as casual wage labour: in other types of work | 11 | 14 |
| 61 Had work in h.h. enterprises (self-employed) but did not work | 3 | 1 |
| 71 Had regular salaried/wage employment but did not work | 1 | 1 |
| 81 Sought work/did not seek but was available for work | 1 | 0 |
| 96 Attended educational institutions/attended domestic duties/ rentiers, pensioners, remittance recipients / Others (including begging, prostitution) | 106 | 111 |
| 98 Did not work due to sickness(for casual workers only) | 4 | 0 |
| Total | 225 | 230 |

► **From the above and also few other indicators, it was concluded that data quality did not suffer due to use of CAPI, even though it was a first time experience.**

# Suggestions from Field

**(n) Suggestions for more facility on CAPI :**

- Comments of investigators & supervisors should be differently coloured.

- Provisions should be there for item-wise comment in interviewer module as in supervisor module to clarify specific information and scrutiny points.

- Incorporating  large code lists  (e.g. NIC 5 digit and NCO 3 digit codes ) in application ( as table lookup) will enhance the capability and performance of field staff.

- There should be facility of storing the information in pdf format for subsequent use.

- Concepts / Instructions relating to a question may be shown in a pop-up box, while filling up the data field.

- Display of some computed value, counts etc. helps the Interviewer.

# Conclusion

**(o) Conclusion** :

➤ As CAPI does not support generation and use of random number, second stage sampling cannot be done on CAPI. So, 2-stage survey of NSSO had to be indirectly managed. But ultimately it was successful.

➤ CAPI enables real time data analysis and much better control over the survey management at any level (supervisor/HQ). CAPI is especially good for touch questions.

➤ CAPI is being continuously upgraded including more and more features, based on feedback from users (many such from NSSO- e.g. provision for Hindi, multiple Headquarter etc. ). Its Designer platform is very user-friendly.

➤ Non-availability of grid structure is not really a limitation. Random access to chapters(blocks) and easy scrolling ensures smooth navigation.  Hence, CAPI is suitable even for large questionnaire.

# Conclusion

**(o) Conclusion – (contd) :**

➤ In the Field, very little problem has been faced relating to data sync, application downloading, connectivity, response time on device, etc.

➤ Schedule canvassing time does not significantly increase in CAPI.

➤ Comfort level of the Interviewers in using tablet and CAPI interface was on the whole much better than PAPI.

➤ CAPI may be suitable even for large-scale surveys, if it is not a multi-stage complex survey.

➤ Before switch-over to CAPI, hands-on training and mock interviews are must. During take-off of the survey, one expert helpline should be present for trouble-shooting.

# THANK YOU