

Chapter 1

OVERVIEW OF THE DATA PROCESSING ACTIVITIES AT THE CA PROCESSING CENTER

1.1 INTRODUCTION

Two of the major activities of the First National Cambodian Census of Agriculture 2013 are the manual processing and machine processing of the accomplished census documents. Manual processing involves manual review of the entries for completeness and acceptability as well as coding some of the items. On the other hand, machine processing involves data entry, computer editing of entries and summarization of data according to predetermined tabulation formats.

The accomplished census documents, namely: Form A (Listing of Household); Form B (Basic Information on Agricultural Holding); Form C (Growing of Crops); Form D (Raising Livestock and Poultry); Form E (Aquaculture Activity); Form F (Agricultural Household and Its Socio-Economic Characteristics) and Form G (Village Questionnaire) have been folioed, edited and coded at the Provincial/Municipality level before they are transmitted to the Census of Agriculture (CA) processing center. At the CA processing center, these census documents will undergo data capture, key verification, consistency and completeness checks before the data files are processed for tabulation of results.

1.2 DEFINITION AND SCOPE OF CA DATA PROCESSING

CA data processing refers to all activities through which the census documents undergo at the CA Processing Center to prepare them for tabulation of results.

Data processing at the CA Processing Center covers the processing of the following documents:

- Form A (Listing of Household);
- Form B (Basic Information on Agricultural Holding);
- Form C (Growing of Crops);
- Form D (Raising Livestock and Poultry);
- Form E (Aquaculture Activity);
- Form F (Agricultural Household and Its Socio-Economic Characteristics) and
- Form G (Village Questionnaire)

1.2.1 Data Processing

Data Processing consists of the following activities:

- a. Receipt and control of all forms from provincial/municipal office
- b. Data entry
- c. Key verification
- d. Consistency and completeness checking
- e. Submission of data files and weekly reports to the CA Director General.

1.3 OBJECTIVES OF CA DATA PROCESSING

The objectives of the First National Cambodian Agriculture Census 2013 data processing are:

- a. To ensure that all census documents have been received at the CA processing center
- b. To perform manual verification on the census documents
- c. To key-in and key-verify the data from the census documents
- d. To check for the consistency and completeness of encoded data files
- e. To concatenate data files by commune and
- f. To prepare and transmit data files to the CA Director General for further data processing

1.4 ORGANIZATIONAL SET-UP

To effectively carry out the data processing activities at the CA processing center, the following organizational set-up is established:

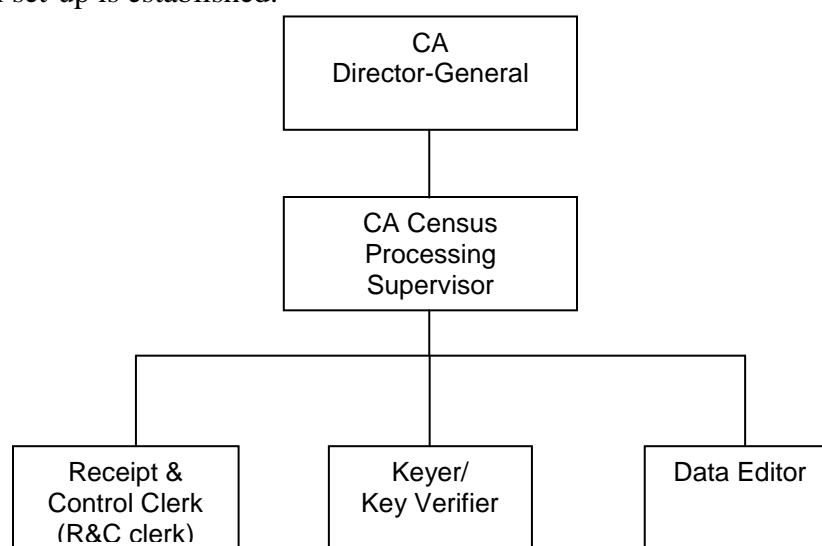


Figure 1. Personnel set-up at CA processing center

The CA census processing group is composed of the CA Census Processing Supervisor (CA CenPS), Receipt and Control Clerk (R&C clerk), Keyers, and Data Editor. The CA processing center is under the supervision of the CA Director-General.

The following are the duties and responsibilities of each member of the census processing group:

1.4.1 CA Census Processing Supervisor (CA CenPS)

The CA CenPS shall:

- Provide day-to-day supervision and coordination for the machine processing
- Serve as the LAN Administrator / Trouble-shooter
- Check the quality of the work of the R&C clerk, keyers and data editors
- Create back-up copies of data files
- Run or schedule the running of programs needed for Completeness and Consistency Checks
- Provide weekly reports to the CA Director-General on the status of processing
- Transmit Census frequency tables and data files to the CA Director General

1.4.2 Receipt and Control Clerk (R&C clerk)

The R&C clerk is responsible for:

- Reporting to the CA CenPS the progress of receipts from the provinces;
- Checking of the bundling/folioing done on the questionnaires received from provinces;
- Making sure that the geographic area code supplied in the folio cover is correct
- Issuing the documents to authorized personnel upon instruction of the CA CenPS;
- Recording of the receipt of the forms in the Receipt and Control Form
- Ensuring the smooth flow of documents within the CA Processing Center
- Keeping track of the documents' location
- Maintaining the documents in an orderly manner in the storage area
- Perform other duties as may be assigned by the CA CenPS

1.4.3 Keyer

The keyer shall do:

- Efficient conversion of the entries in the questionnaire into the computer, including
- Key verification, and when necessary;
- Correction of previously keyed data; and
- Other duties as may be assigned by the CA CenPS

1.4.4 Data Editor

The data editor is responsible for:

- Checking the completeness of codes for write-in entries in the questionnaires
- Checking the completeness of encoded questionnaires
- Cleaning or updating batch files which have unacceptable error messages in the error list
- Generating error list for ID Validation or Completeness Check upon instruction of the CA CenPS
- Generating error list for data consistency checking upon instruction of the CA CenPS
- Performing other duties as may be assigned by the CA CenPS

1.5. FLOW OF ACTIVITIES

The processing at the CA processing center is in two stages (1) Core Module (Form A and Form B) and (2) Form C, Form D, Form E, Form F and Form G. The chart on the following page shows the process flow for all forms at the CA.

The processing of questionnaires at the CA Processing Center will be by commune. It is recommended that encoding of all folios for one commune be finished before working on another commune.

The following steps provide details on the process flow for the processing of questionnaire folios of all forms:

Step 1. The R&C Clerk

- Receives questionnaire folios from province/municipality;
- Segregates folios from other submitted census forms;
- Sorts folios by geographic ID;
- Records the documents into DP Control Form 1 (R&C Control Form for CP Form);
- Arranges folios into designated racks;
- Labels the racks accordingly; and
- Together with Data Editors, check the completeness of codes for write-in entries in the questionnaires

Step 2. The R&C Clerk

- Informs the CA CenPS that folios have been submitted and are ready for encoding.

Step 3. The CA CenPS

- Assigns encoding or key verification of folio to keyer and
- Directs the R&C clerk to supply the assigned folio to the keyer.

Step 4. The R&C Clerk

- Retrieves and brings the assigned folio to keyer;
- Records the giving of folio to keyer on DP Control Form 1; and
- Requests the keyer to sign on the space provided for **Data Encoding** or **Key Verification** on the DP Control Form 1 to acknowledge receipt of the assigned folio;

Step 5. The Keyer

- Runs the Data Encoding and supply necessary parameters. Note that the keyer will be automatically shown the screen for entering the identification. The cursor automatically goes to the the ID of the questionnaire because the geographic-area codes have already been supplied.
- Encodes the data items from the questionnaires;

Step 6. The Keyer

- Affixes his signature on the space provided for **Data Encoding** or **Key-Verification** on the Folio Cover.
- Returns folio to R&C clerk upon completion of the encoding or key verification task.

Step 7. The R&C clerk

- Records the returned folio on DP Control Form 1;
- Attaches **BLUE** marker or tag on the returned folio to indicate that it has been encoded OR a **RED** marker or tag to indicate that it has been key-verified; and
- Stores folio in its rack.

Step 8. The Keyer

- Informs the CA CenPS that keying/key-verification of the assigned folio is finished.

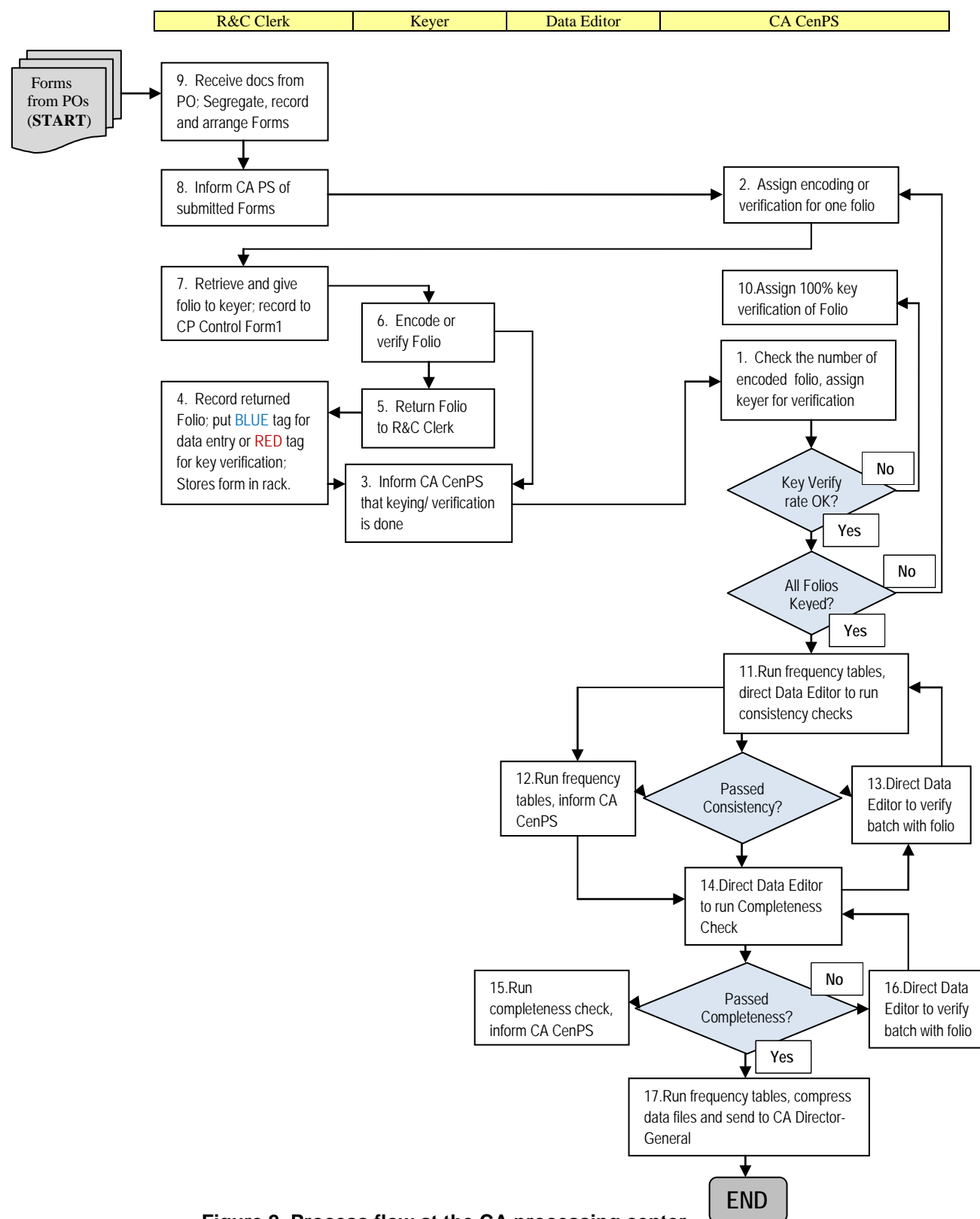


Figure 2. Process flow at the CA processing center

Step 9. The CA CenPS

- For Data Entry tasks, The CA CenPS must check the number of encoded questionnaires counted by the program and the actual number in the folio. If discrepancies are found, the folio must be returned to the keyer to insert or add missing questionnaires.
- If there are no discrepancies, the CA CenPS should assign the encoded folio to another keyer for key verification.
- For Key Verification tasks, The CA CenPS must compare the number of key-verified questionnaires counted by the program with the target number of questionnaires to be verified. If discrepancies are found, the folio must be returned to the Data Encoder to complete the verification.
- Check the verification error in CSPRO data entry statistics menu.

Step 10. The CA CenPS

- If the error rate exceeds the set tolerance limit, the batch should be re-assigned to keyer for **100% key verification**. Note that previously key verified questionnaires would be offered again for verification.

Step 11. The CA CenPS

- If the verification error rate is less than the tolerance limit, then the batch will be tagged as '*verify clean*'.
- If all folios have been verified and tagged as '*verify clean*', the CA CenPS creates a backup of the data files.
- The CA CenPS generates frequency tables using the data entry files as input. The tables are later sent to CA Director-General together with the data files.
- Once the tables are generated, the CA CenPS should subject the batch for data consistency check. The CA CenPS can run the data consistency check by himself or can direct the Data Editor to do this for him.

Step 12. The Data Editor

- The Data Editor or the CA CenPS initiates the data consistency check by running the '**Batch Edit**' program corresponds to the form and then selecting the data file. This will generate an error list.

Step 13. The CA CenPS

- If there are errors, the CA CenPS should assign the batch the Data Editor to verify the batch with the actual folio.

- The Data Editor will update the file by running the Modify in Data Entry module. After all errors have been corrected, the data editor should close the batch and should return the folio to the CA CenPS.

Step 14. The CA CenPS

- Once all batches in a commune have been encoded and verified with the actual folios, the CA CenPS should subject the batches to ID validation or completeness check.
- Runs or directs the Data Editor to run ID validation or completeness check.

Step 15. The Data Editor

- Initiates ID validation by running the '**Completeness Check**' program. This will generate an error list if some questionnaires have invalid geographic identification codes or some enumeration areas have no questionnaires or if the questionnaire were not sorted by its serial number. The error list generated depends on the checked items in the option menu item of the module.

Step 16. The CA CenPS

- If there are errors, the CA CenPS directs the Data Editor to verify the batches with actual folios.

Step 17. The CA CenPS

- Once all the folios for a commune have passed the consistency and completeness checks, the CA CenPS creates backup of the data files;
- Generates frequency tables;
- Concatenates data by province;
- Creates a compressed file that contains data files;
- Creates a compressed file that contains the frequency tables;
- Secured the compressed files with password; and
- Sends the compressed files to CA Director General.

1.6 THE CENSUS DOCUMENTS

Listed below are the census documents that will undergo processing at the CA Processing Center:

Core Module:

- Form A (Listing of Household)
- Form B (Basic Information on Agricultural Holding)

Other Forms:

- Form C (Growing of Crops)
- Form D (Raising Livestock and Poultry)
- Form E (Aquaculture Activity)
- Form F (Agricultural Household and Its Socio-Economic Characteristics) and
- Form G (Village Questionnaire)

These are the census questionnaires used to record information about a household. It contains information on the characteristics of household. All forms have (3) main parts, namely:

Title Panel – where heading, the date of issue, coverage, confidentiality statement.

Geographic (Area) Information Panel – which contains the Geographic Identification;

Main Questionnaire Panel – which consist of questions pertaining to agricultural activities.

The data items for this panel are summarized on **Appendix A to G**:

1.6.1 Processing at the Provincial/Municipal Office

Folioing of Census Forms

To facilitate handling, the census documents are grouped and bound into folios in the provincial/municipal office following the guidelines below:

- Form A** is folioed by EA. The forms are arranged in ascending order by serial number. One enumeration area (EA) is equivalent to one folio. One plastic envelop will contains one commune. The EAs/folios are sorted in ascending order by geographic codes.
- Form B** is folioed by EA. The forms are arranged in ascending order by household serial number having about 60 questionnaires per folio. There may be at least one folio of Form B for an EA. Folios should be grouped or tied together.
- Form C** is folioed by EA. The forms are arranged in ascending order by household serial number having about 60 questionnaires per folio. There may be at least one folio of Form C for an EA. Folios should be grouped or tied together.
- Form D** is folioed by EA. The forms are arranged in ascending order by household serial number having about 120 questionnaires per folio. One folio of Form D for an EA. Folios should be grouped or tied together.

- e. **Form E** is folioed by EA. The forms are arranged in ascending order by household serial number having about 120 questionnaires per folio. One folio of Form E for an EA. Folios should be grouped or tied together.
- f. **Form F** is folioed by EA. The forms are arranged in ascending order by household serial number having about 60 questionnaires per folio. There may be at least one folio of Form F for an EA. Folios should be grouped or tied together.
- g. **Form G** is folioed by Commune. The forms are arranged in ascending order by Geographic Information codes. One commune is equivalent to one folio. One plastic envelop will contains one commune.

Folio Cover

- a. For Form A to Form F, each folio is covered by a Folio Cover. It contains the province/municipality, district, commune, village, EA number, and number of questionnaires contained in the folio. The folio number and number of folios for the EA/village/commune/district/province are indicated in Folio ____ of ____ folios.
- b. For Form G, each folio is covered by a Folio Cover. It contains the province/municipality, district, commune, and number of questionnaires contained in the folio. The folio number and number of folios for the commune/district/province are indicated in Folio ____ of ____ folios.

Hierarchy of Editing Marks

The Census forms have undergone manual processing at the provincial/municipal offices prior to receipt by the CA Processing Center. Several editors and reviewers have analyzed and reviewed the questionnaire entries, thus, it is possible that an entry may have been corrected more than once. In such cases, the color of the ink used identifies the reviewer who made the correction. Below is the color hierarchy from the lowest to highest:

Color	Used by
Black	Enumerator
Blue	Team Supervisor
	Manual Processor
Red	CA Manual Processing Supervisor

Following the above hierarchy, if an entry (originally recorded in black ink) is corrected more than once, the correction entry written in the ink with a higher color hierarchy prevails.

1.7 THE DATA PROCESSING CONTROL FORMS

The flow of documents in the CA Processing Center is to be monitored through the use of control forms. These control forms are to be prepared before the start of data processing in the CA Processing Center.

The main purpose of these control forms is to provide the processing center with a permanent record on the status of the phase(s) of the data processing. There are two types of data processing control forms: (1) **Weekly Report Forms**; and (2) **Receipt and Control Forms**.

The **Weekly Report Forms** are used to monitor the status of data processing on a weekly basis. There are two weekly report forms and these are: (1) DP Control Form 1 (Manual Processing Weekly Report Form) and (2) DP Control Form 2 (Machine Processing Weekly Report Form). These forms are used to show information such as:

- Number of expected EAs to be processed by EA/Village/Commune/District/province
- Number of actual EAs manually processed by EA/Village/Commune/District/province
- Number of actual EAs encoded by EA/Village/Commune/District/province
- Number of actual EAs that have undergone machine edit

The **Receipt and Control Forms** are used to record those who have had access to a particular folio at any particular time. Receipt and Control Form is DP Control Form 3 (R&C Control Form). This form is used to register information such as:

- The date when the documents for an EA are received for a certain phase,
- The persons responsible for accomplishing a particular activity (such as data encoding, key verification and completeness checking) or
- The date when the documents for an EA are returned to the R & C clerk after completion of the phase.

The specific control forms for the corresponding phases of data processing are as follows:

PHASE	FORM NUMBER	FORM DESCRIPTION
-------	-------------	------------------

Manual Processing	DP Control Form 1	Manual Processing Weekly Report Form
Machine Processing	DP Control Form 2	Machine Processing Weekly Report Form

The CA CenPS submits the updated Machine Processing Weekly Report Form (DP Control Form 2) to the CA Director General every Monday.

The CA CenPS is also in-charge of generating the receipt and control forms. Each type of form should be fastened in a folder arranged in ascending geographic code (Commune), that is, the lowest code should be on top and highest code should be at the bottom.

The CA CenPS should label the folder by printing on the front page the type of form.

1.8 TIMETABLE OF ACTIVITIES

The timetables of activities for Census of Agriculture processing are shown below. The CA CenPS should oversee the activities during the entire data processing so that the timetable for the different phases of work is followed accordingly.

1.8.1 First National Cambodian Agricultural Census 2013 Timetable

The following timetable for data processing shall be observed to ensure that the census results are available at the soonest possible time:

Activities	Beginning Date	Ending Date	Outputs/Remarks
1. Recruitment of personnel for data processing	July 01, 2013	July05, 2013	Manual processing – regular NIS/MAFF staffs; data entry – hired data encoders
2. Training of data processing personnel a. manual editing b. data encoding c. verification	July 15, 2013 July08 2013 July 22, 2013 July 29, 2013	July 24, 2013 July 12, 2013 July 26, 2013 July31, 2013	Trained data processing personnel
3. Manual completeness check, coding and editing	July 15, 2013	October31, 2013	Checked and edited census data
4. Data entry of checked census results	August 01, 2013	February 15, 2014	Encoded census data
5. Verification of encoded data	August 26, 2013	February21, 2014	Verified census data
6. Training of NIS-MAFF staff	January 02, 2014	February 07, 2014	1training module as planned in the TF project document
7. Hiring of TCDC Expert on IT	February 01, 2014	April 12, 2014	2-month mission of the TCDC Expert in IT under TF project
8. Consistency checking and machine editing	February03, 2014	February 28, 2014	Machine edited and consistency checked data

9. Generating and checking marginal census results	February 03, 2014	February 28, 2014	<i>Generated marginal census data</i>
10. Generation of statistical tables from the clean census data	March 01, 2014	March 21, 2014	<i>Generated statistical tables of the census results</i>
11. Consolidation of all census results into public use file and other back-up media	March 24, 2014	April 12, 2014	<i>CA databases and back-up census data</i>
12. Preparation of the different types of analysis to be applied in the tabulated census results including the 8 thematic papers	April 01, 2014	April 21, 2014	<i>Plans and prepared outlines and topics for NCAC 2013 using the tabulated census results</i>