Chapter 5

DATA ENCODING

Data encoding is the process of converting questionnaires to computer-readable format. For the 2008 Population Census of DPRK, the data encoding will be done at the PSO processing centers.

5.1   OBJECTIVE AND PRINCIPLES OF DATA ENCODING

The objective of data encoding for this Census is to convert CP Form 2 questionnaires to digital format as quickly and accurately as possible. The keyer or data encoder should enter the data as they appear on the questionnaire. It is the duty of the keyer to check whether the geographic ID of the document given by the R&C clerk and those of the batch assigned by the PSO CenPS are exact match before proceeding with the actual encoding or key verification.

Consistency checking during data entry is generally kept to a minimum. Resolving consistency errors are resolved in a later phase using computer edit programs. The keyers do not need to be familiar with the concepts of the questionnaire. As much as possible, keyers are not given the decision to resolve data errors.
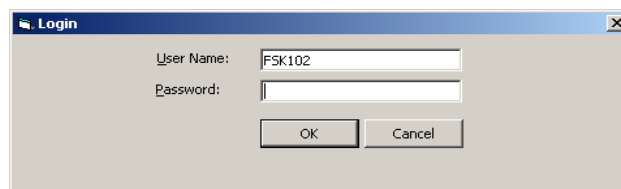
5.2   RUNNING THE DATA ENTRY PROGRAM

The PSO CenPS assigns each keyer their own user name in order to allow them access to the data entry program through the SCIPS. The user names also uniquely identify keyers in the SCIPS system.

The keyers are given the user names from **FSK102, FSK103, FSK104, …, FSKnnn.** With these user names the keyer is allowed to open a batch/folio that is assigned to him/her by the supervisor. It is not possible for the data encoder to select and open any other batch file. The mode in which the batch would be opened also depends on what mode the supervisor has assigned to him/her. The keyer is allowed only to do the **Open Batch option.**

5.2.1   Logging Into the System

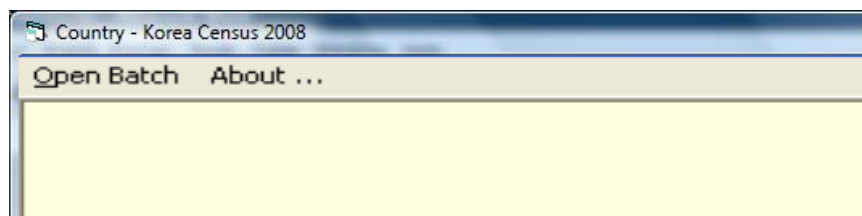- When the login prompt is displayed, the keyer enters the assigned User Name, then, click the **OK** button.

- Then the SCIPS main menu appears.

5.2.2 The SCIPS Main Menu for Keyers

A sample **SCIPS Main Menu** screen is shown below.

When the keyer selects the **Open Batch** option, the assigned batch will be opened and initially the Data Entry screen will be displayed. The keyer can now start encoding the questionnaire.

NOTE: Geographic area codes common to the questionnaires in the batch would be automatically supplied by SCIPS

5.2.3 The Data Entry Screens

Data Entry Screens

There are four (4) screens in the Data Entry program and these are: (1) *Geographic and Household Identification Screen*; (2) *Household and Dwelling Unit Information Screen*; (3) *Person Information Screen*; and (4) *Mortality Information Screen*.

(1) Geographic and Household Identification Screen

The Geographic and Household Identification Screen allows the encoding of geographic codes and dwelling unit and household serial numbers that uniquely identify a household. The data items to be supplied in this screen are:

- Province   (2 digits)
- City(District) / County   (2 digits)
- Ri (Up /Gu /Dong)   (3 digits)
- Enumeration Area No.   (2 digits)
- Dwelling Unit Serial No.   (3 digits)
- Household Serial No.   (3 digits)

The following figure shows the Geographic and Household Identification Screen:

(2) Household and Dwelling Unit Information Screen

The Household and Dwelling Unit Information Screen captures four characteristics of the household and ten physical features of the dwelling unit. These are

- H1. Number of members in the household
  - Male   (2 digits)
  - Female   (2 digits)
- H2. Type of household (Household or Institution)   (1 digit)
- H3. Class of labor of household head   (1 digit)
- H4. Previous class of labor of household head   (1 digit)
- H5. Type of dwelling unit of the household   (1 digit)
- H6. Whether the household is the main occupant of the   (1 digit)
  dwelling unit
- H7. Floor area of dwelling unit   (3 digits)
- H8. Number of rooms in the dwelling unit   (1 digit)
- H9. Presence of water tap   (1 digit)
- H10. Source of water supply   (1 digit)
- H11. Kind of toilet facility   (1 digit)
- H12. Heating system established in the household   (1 digit)
- H13. Heating system used in the household   (1 digit)
- H14. Fuel used for cooking   (1 digit)

The following figure shows the Household and Dwelling Unit Information Screen:

제1편 세대및 살림집등록

1.이세대에 현재 살고있는 인구수가 모두 몇명입니까? 남자 [ ] 녀자 [ ]	4.원거세대 입니까? 1.예 2.아니 □	9.현재 어떤 위생실을 사용합니까? 1.수세식개인위생실 2.수세식공동위생실 3.재래식개인위생실 4.재래식공동위생실 5.없음 □
2.세대류형 1.가정세대 2.집체생활단위 □	5.살림집의 총면적은 얼마입니까? [ ]	
3.세대주의 직업은 무엇입니까? 1.로동자 6.부양 2.사무원 7.학생 3.농장원 4.사회보장 5.년로보장 □	6.살림방은 몇칸입니까? 1.1칸 2.2칸 3.3칸 4.4칸 혹은 그이상 □	10.이집이 어떤 난방체계입니까? 1.중앙 및 지역난방 2.전기난방 3.전기난방과 함께 리용하는 기타난방 4.가정용 석탄난방 5.가정용 나무난방 6.기 타 □
4.세대주(또는 부모)의 전직업은 무엇입니까? 1.로동자 2.사무원 3.농장원 4.협동농장원 5.협동농장원 □	7.집에 수도가 있습니까? 1.예 2.아니 □	11.주되는 난방형태는 무엇입니까? 1.중앙 및 지역난방 2.전기난방 3.전기난방과 함께 리용하는 기타난방 4.가정용 석탄난방 5.가정용 나무난방 6.기 타 □
	8.현재 주되는 음료수원천은 무엇입니까? 1.집안수도 2.공동수도 3.풀장 4.보호막이있는우물 5.보호막이있는샘물 6.못,강 7.기타 □	12.현재 주되는 취사용연료는 무엇입니까? 1.전 기 2.가 스 3.석 유 4.석 탄 5.나 무 6.기 타 □
5.살림집이 어떤 형식입니까? 1.1동1세대 2.1동다세대 3.아빠트 4.4칸 혹은 그이상 □		

(3) Person Information Screen

The Person Information Screen allows keying of characteristics of household members. The screen is laid out like a spreadsheet (or a roster) and lists the description of data items in the first column while the next columns are allotted for the keying of characteristics of household members. One column represents one member.

There are 29 data items to be captured in the Person Information Screen and these are:

- P1. Individual serial number   (2 digits)
- P2. Whether the individual is registered with the household   (1 digit)
- P3. Place where currently registered
  - Province   (2 digits)
  - County   (2 digits)
  - RI   (3 digits)
- P4. Relationship to the household head   (1 digit)
- P5. Sex   (1 digit)
- P6. Year and month of birth and age
  - Year of birth   (4 digits)
  - Month of birth   (2 digits)
  - Age   (3 digits)
- P7. Nationality   (1 digit)
- P8. Indicator on seeing   (1 digit)
- P9. Indicator on hearing   (1 digit)
- P10. Indicator on walking/climbing stairs   (1 digit)
- P11. Indicator on concentration/remembering   (1 digit)

- P12. Residence 5 years ago
  - Whether same place   (1 digit)
  - Other place (Province)   (2 digits)
  - Other place (County)   (2 digits)
  - Whether Up/Gu/Dong or RI   (1 digit)
- P13. Indicator on whether can read and write simple   (1 digit)
  message
- P14. Indicator on whether currently attending school   (1 digit)
- P15. Level of schooling currently attending   (1 digit)
- P16. Indicator on whether attending full-time regular   (1 digit)
  course
- P17. Highest level of education   (1 digit)
- P18. Kind of educational certificate   (1 digit)
- P19. School and field of study of highest educational   (3 digits)
  attainment
- P20. Indicator on whether attained the qualification with (1 digit)
  full-time regular course
- P21. Usual activity during the last 6 months   (1 digit)
- P22. Indicator on usual activity of the following mainly for
  own consumption in the last 6 months:
  a. Fruit/vegetable gardening   (1 digit)
  b. Fishing/raising livestock/poultry   (1 digit)
  c. Gathering firewood   (1 digit)
  d. Fetching water   (1 digit)
  e. Making handicraft   (1 digit)
  f. Others   (1 digit)
- P23. Average number of hours a day spent for activities in (1 digit)
  P22
- P24. Kind of industry   (2 digits)
- P25. Occupation   (3 digits)
- P26. Class of labor   (1 digit)
- P27. Marital status   (1 digit)
- P28. Age of first marriage   (2 digits)
- P29. Indicator on whether has a live birth during the period
  Oct 1, 2007 to Sep 30, 2008
  - Whether had a live birth   (1 digit)
  - Number of males   (1 digit)
  - Number of females   (1 digit)

The following figure shows the Person Information Screen:

(4) Mortality Information Screen

The Mortality Information Screen allows keying of entries for an indicator on incidence of death during the specified period and for ten (10) questions related to records of mortality in the household during the period Oct 1, 2007 to Sep 30, 2008. The data items to be captured in this screen are:

- Mortality Indicator (1 digit)
- M1. Serial number of the deceased (2 digits)
- M2. Sex of the deceased (1 digit)
- M3. Date of birth of the deceased
  - Year (4 digits)
  - Month (2 digits)
  - Day (2 digits)
- M4. Date of death
  - Year (4 digits)
  - Month (2 digits)
  - Day (2 digits)
- M5. Age (2 digits)
- M6. Indicator on whether the deceased was pregnant (1 digit)
- M7. Indicator on whether member died while having an abortion or miscarriage within 42 days (1 digit)
- M8. Indicator on whether member died while giving birth or within 42 days (1 digit)
- M9. Place of death (1 digit)
- M10. Indicator whether member had a live birth during the period Oct 1, 2007 to Sep 30, 2008

        o   Whether had a live birth   (1 digit)
        o   Number of males   (1 digit)
        o   Number of females   (1 digit)

The following figure shows the Mortality Information Screen:

5.2.4 Instructions for Data Entry

General Instructions in Data Entry

The keyer fills out the screen with the corresponding data in the current questionnaire by typing the entries in their respective data cells. If the cell is fully filled out, there is no need to press the enter key since the cursor automatically goes to the next data cell. This means that if the allotted digits for the cell is 2 and when the entry '02' is typed, the cursor moves to the next cell. But if the data encoder keys in '2' only, the **'Enter'** key must be pressed to make the cursor go to the next cell. In this last example, the encoded number '2' will be stored as '02' since the allotted space is for two digits.

Entering the Geographic and Household Identification

In the *Geographic and Household Identification Screen*, there is no need to enter the codes for geographic areas (province, county, RI, enumeration area) that are common in a batch/folio since these are automatically supplied by SCIPS.

The keyer has to enter only the dwelling unit serial number and the household serial number.

Entering data items for Household and Dwelling Unit Information Screen

After encoding the data items for the Geographic and Household Identification, the program displays the Household and Dwelling Unit Information Screen. Usually the cursor moves to the next data item once the data for current field has been entered. But for item H2, if the entry is '**2**' referring to Institution Unit type of household, the data entry program moves immediately to the Person Information Screen without prompting for other data items in the Household and Dwelling Unit Information Screen. This is called *automatic skip*.

The other data item with automatic skip is H3 where the cursor skips to H5 when the entry for H3 is either 1, 2 or 3.

NOTE: If in case there is inconsistency between the questionnaire entries and the automatic skips, the keyer should consult the PSO CenPS for proper action.

Entering data items for Person Information Screen

In the *Person Information Screen*, data in the questionnaire would be keyed-in for one member at a time. One column of entries refers to the characteristics of one household member.

All data about a member should be encoded before one could key-in the data about the next member. Since the windows displayed on the screen is limited to a number of rows, the user must use the vertical scroll bar to see all the rows in the table (CSPro calls it a *roster*).

When all information for all members of the household have been encoded, *the user must press 'Ctrl + /'* to proceed to the next screen, that is, the Mortality Information Screen.

To go back to previously entered pages, the usual page-up key could be used. Page-down key is used to go to the next pages that have been filled out. Sometimes, page-up key doesn't seem to work. If this case happened position the cursor to the first cell in the data screen by clicking on it and then press the Up-arrow key.

There are several items in the Person Information Screen that use automatic skips and these are:

Data Item with Automatic Skip	Skip to	Condition to Trigger Skip
P2	P4	P4 = 1
P12	P13	P12 (same place) = 1
P14	P17	P14 = 1, 2, 3 or 4
P17	P21	P17 = 1, 2, 3 or 4
P23	P27	P21 not = 1
P29	Next Member	P29 (live birth indicator) = 2

Some sections of the Person Information Screen also use automatic skips depending on age and sex of the member. The data items from P8 to P17 are filled up only for members 5 years old and over. Data items from P18 to P28 should have entries only for members 16 years old and over. The data item P29 is filled up only for female members 15 to 49 years old.

Entering data items for Mortality Screen

In the Mortality Information Screen, the first data item (Mortality Incidence Indicator) acts like a screening question. If the entry for this field is 2, the program automatically prompts the keyer whether to save the batch.
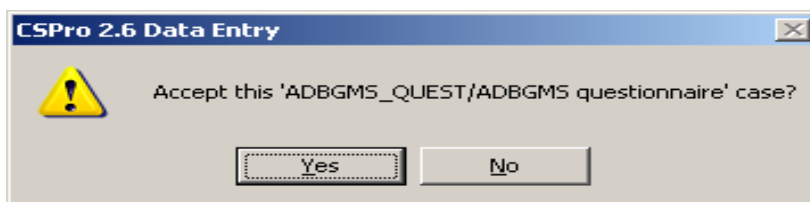
The data items with automatic skips in this screen are:

Data Item with Automatic Skip	Skip to	Condition to Trigger Skip
M6	M9	M6 = 1
M7	M9	M7 = 1
M8	M10	M8 = 2
M10	Next Person	M10 (live birth indicator) = 2

The data items from M6 to M10 should be entered only for female members who died between the ages 15 and 49.

5.2.5 Saving A Batch

After encoding entries in the last section the data entry program will confirm if the questionnaire would be accepted (saved) or not by displaying the prompt shown below.

Select **"Yes"** if the questionnaire is to be saved and go on to the next questionnaire. If **"No"** is selected, the cursor will go to the first data cell in the last data entry screen to enable the data encoder to review the correctness and completeness of encoded data.

Note: To cancel (not saving the questionnaire) the current encoded questionnaire, press ESC key and select "Discard".