Recensement Census

NATIONAL CENSUS TEST

Report No. 28

Evaluation of Processing Operations

RECENSEMENT
96
CENSUS

Statistique Canada Statistics Canada

Canada

NATIONAL CENSUS TEST

Report No. 28

Evaluation of Processing Operations

Phil Stevens
Neelam Prakash
Lorie Shinder

Special Surveys Division

July 1994

Introduction

This report discusses problems resolved in the processing (including capture) of the 1993 National Census Test (NCT) survey. It may help in the planning and development of any future Census Test, assuming the methodology of the Census to continue as a self-completion, mail-back paper-and-pencil form. On the other hand, the NCT was adapted to 1993 Labour-Force Survey (LFS) procedures; as the LFS moves to CAI (Computer-Assisted Interviewing) some of our problems may be unique to the 1993 test.

Field Operations - Background

The NCT used a November 8, 1993 reference date. Questionnaire drop-off started on Saturday October 30 and finished on Friday, November 5, 1993.

One component of the test was selection of a subsample of NCT households[1] whose responses were captured as the "Edit Failure Survey" (EFS) component of the survey as soon as received in the ROs, then sent back to interviewers for field edits and follow-up. These questionnaires were re-captured in order to test the field edits and follow-up.

Startup of data capture of EFS responses was delayed a day or two because of the November 11th holiday (Thursday) and problems encountered with the capture program written in the DC2 software. The first transmission was received from the Edmonton RO on Saturday, November 13, 1993. Data collection was completed for EFS responses by November 30, 1993. The raw EFS file consisted of 10,145 person-records.

Data capture of NCT responses started on December 10,1993 and the final transmission was received on January 26, 1994. The raw NCT file (combined LFS-based and special populations) consisted of 47,057 person-records.

Print Requirements

Labels: Labels and interviewer-assignment control-lists were generated through the LFS and had to fit in with the LFS production facilities. The household identification code on

---

[1]Sample selection was completed on the Mainframe by Mike Egan. The main sample was selected as a subset of the LFS sample from April, May and June of 1991, which "rotated out" of (i.e., finished with) the LFS in September - November 1991. It was judged that, two years after LFS participation, the households would not have retained any bias in their attitudes from the LFS experience that would significantly affect their NCT responses. A total of 17,109 dwellings was selected. The EFS sample was 1/2 of the NCT sample, i.e., 8500 dwellings. A further 3985 dwellings were picked for "special population" samples, primarily from 1991 Census files.

labels and control lists was 15 bytes long, consisting of the LFS-file fields PSU, GROUP, CLUSTER, ROTATION, LISTING and MULTIPLE. Only fourteen bytes of identifying information were generated in the label program instead of fifteen, with 'multiple' not being printed. To resolve this problem regional offices were instructed to code '0' in 'multiple' on the labels and control lists. In spite of the instructions, 'multiple' was still blank for many cases on the assignment control lists. This was problematic because they were to be linked to household-related responses from the main questionnaire to create a Household file. To recover the multiple a match to the sample file to pick up this field was performed. All multiple codes of 1 or higher were selected and assigned manually on-line.

All linkage specifications used a 20-byte id consisting of the Interviewer Assignment Number (IAN), PSU, group, cluster, rotation, listing and multiple. With hindsight, it would have been better to have used a shorter household-id code similar to the LFS RO-DOCKET. Linking on a household identifier of twenty bytes was long, cumbersome and error-prone.

During assignment planning, ROs were asked to assign "7" as the second digit of the IAN if the dwelling was picked for the LFS-based sample, and "9" for special-population samples[2]. In general, this worked and made it readily possible to separate and order the assignments for printing labels and assignment control sheets. However, there were a few errors which suggests that clean-ups must be done early in the initial stages of processing in order to verify the IAN.

Furthermore, to accomodate the IAN planning (R21 files) there needs to be adequate lead time to allocate IANs. These are allocated by PSU, GROUP and CLUSTER for the LFS rotate-out sample. We allowed three months for the completion of assignment planning in the regional offices.

A problem arose in the Special Population sample. Some EAs exceeded 75 households which was the maximum allocation for one assignment. In order for the ROs to determine how to break the EA down into meaningful geographical delineations at the household level the household id was carried on the L01 file. A hardcopy listing with address information was provided for each RO to cross reference with the household id given on the L01 file. For EAs that were too large for one assignment '01' was put in

[2]A total sample of 3985 households was selected using 1991 Census Visitation Records to locate EAs with high concentrations of the special populations of interest. The Winnipeg and Saskatoon Metis samples were obtained from provincial membership association lists. These households were in addition to the LFS-based main NCT sample.

the yield[3]. Fortunately, none of these procedures had an impact on the instruction manuals written by Survey Operations.

The Special Population samples were not weighted because they are not representative samples of the population in general, and national estimates will not be calculated from the information obtained. The major purpose of these samples is to evaluate the Census Test Questionnaire for specific groups[4] that may experience particular problems with the proposed questions.

If there is to be a Special Population component in future Census tests then careful attention is required in the preparation of the "F03" files needed by LFS (John Rowland) as input to the "S03" print files needed by ISD (Kathy Reid/ Dave Bowman). Extra processing was required to reformat Special Population records into an "F03" structured response file. This file was required by the LFS to create an "S03" print file for the production of NCT labels and control lists. For the two Métis samples selected from membership association lists we had to make up dummy Prov-FED-EA-Hhld numbers. The proper Province code was assigned, '000000' for FED-EA and then Hhld number was sequentially assigned commencing at '001'. Group and Rotation were recoded to zeros. The LFS "F03 Short" record length is 280 bytes but the Special Population mock F03 was created as 131 bytes. The LFS F03 file expects a record length of 372 so the Special Population file was zero-filled to meet this requirement.

Two problems arose with the Special Populations. The F03 interviewer assignment numbers (IANs) were not moved over correctly and consequently did not match Dave Bowman's file. This was resolved in the program that merged the F03 and R21 (assignment planning) files. The LFS system required a current IAN as well as a previous IAN to be placed on the F03 file in two consecutive places. Secondly, it was decided that Edmonton was to handle the assignment planning for Winnipeg. This complicated the print file preparation so that Edmonton could print Winnipeg's labels by requiring a matching on PSU, Listing number and last byte of IAN for RO 16 and RO 17 data.

A few days were lost during production to produce a print file. Some more fields on the

---

[3]Yield is a required field according to Labour Force procedures for updating assignments in the ROs

[4]Blacks in Halifax, Asians in Montréal, Blacks in Montréal, Latin Americans in Montréal, Asians in Toronto, Blacks in Toronto, Aboriginals in Winnipeg, Métis in Winnipeg, Métis in Saskatoon, Aboriginals in Regina, Aboriginals in Edmonton and Asians in Vancouver

372-byte F03 were identified that should have specific values[5] (see footnote below for future reference).

Data Capture

The start of EFS data capture was delayed three or four days mainly because of bugs in the DC2 software being used for the NCT. Due to time constraints Special Surveys did not have an opportunity to review keyer instructions. One instruction called for household data (Steps 1 - 7 and QQ47 onwards) to be captured only once as a part of the person-1 record. In several cases, household data were captured for somebody other than person 1[6]. An ambiguity in the instructions may have been the reason. It would have been desirable to have reviewed and commented on the instructions.

In using DC2, there was no means of control to guarantee all household members had been captured. DC2 was capable of controlling for the number of forms but there was no way of knowing whether a key operator had missed out an entire person from the form. A higher level of verification than was used would be helpful.

The design of the questionnaire was set up for vertical capture to reflect the capturing of person records within the household. No colour distinction was made on the form to assist the key operator visually to stay in the correct column on each page. A recommendation for shading should be made if the budget can handle the additional cost for printing.

Another limitation of the DC2 software was that it lacked verification flexibility. For example, 100% of a given field had to be verified by re-keying for all forms. It was not possible to verify a sample of forms. It would have been preferable to have taken a subsample of documents and verified on several or all fields. If this option had been available a better picture of the error rate incurred by the key operator would have been evident. For budget reasons we chose to verify the 20-byte Household identifier and questions 2-5 for every form. We would advise with hindsight a higher level of verification.

Processing

---

[5] survey id (pos.25,1) = 1
preprinted code (pos.52,1)=1 (=0 if special populations)
pos. 54,8 = blank (flags based on a previous month - make special populations look like births)
pos. 280,1 = 0 (there are no notes)
pos. 266 = 1 (if flagged for EFS or blank on special populations)

[6]This occured primarily in the Montréal RO.

Since a major purpose of the test was to measure errors, the capture program was written and costed to allow for multiple entries for all precoded questions, including those with instructions "mark one only". It is questionable whether subject matter people were interested enough in multiple-response errors to make this worth while. Developing and testing the data capture entry system took many person-days more than if Special Surveys' standard processing practice had been used, of capturing the first response where only one is expected. Future tests might also use RO facilities for grooming before capture.

A shortage of resources during the processing was experienced. Planners and managers of any future Census test will want to be assured of having adequate, qualified programmers to handle complex programming requirements[7]. To complete the testing and production runs for the NCT three experienced programmers were found at short notice in January 1994, each available for a few weeks only. Each programmer was assigned his own tasks. What was being tested or produced and by whom required hour-to-hour attention and direction from a coordinator[8].

Attention needs to be paid to the creation and availability of detailed test files. During the phase of program development for derived variables programmers wanted good test files. Because of the nature of the DVs these test files are best supplied by subject matter persons who are responsible for specifications for the DVs. Early advisement should be given to subject matter in order that they can prepare for this part of the processing.

Working on the same platform would have been advantageous in terms of location and management of file creation. For example, all of the processing could have been handled on our own LAN, or on the Census LAN (UNIX) or on the Mainframe. In actuality, there was a lot of uploading and downloading of production files in order to accomodate the two working environments used for the 1993 NCT[9]. From the creation of the RAW files to the PREDIT files the processing was handled on the Mainframe. Afterwards the processing was all done on the UNIX, except for one of the short-term borrowed programmers who preferred to work on the mainframe. At times the Unix was a bit slow due to maximum user capacity or space limitations. Presumably there are cost savings in using a LAN such as the Census Unix, although it might be difficult to determine just how much.

---

[7]The NCT Operational team in Special Surveys consisted of Phil Stevens -Manager, Neelam Prakash - Programmer (replacement for Mike Egan in September, 1993), and Lorie Shinder - Processing Rep.

[8]Evelyn Ryan of Census Operations controlled and monitored the assignment of tasks of the three programmers.

[9]Processing was developed and run on both the Mainframe as well as the UNIX.

Other than for income data (Q.46 in the test) there was no editing of results by Special Surveys after capture. For future tests, from a capture point of view, consideration should be given to omitting the income-question cents boxes, or to retaining the background colour in them (i.e., not dropping them out). They added to the incidence of error despite key operators being instructed not to capture the cents. Perhaps income-question instructions could be tested: "enter the amount to the nearest dollar". If future tests include editing the income data, verification on this field should be included, and there should be more edits included in the capture system. About four programmer-weeks were used in the test just to edit Q46 which became an expensive and time-consuming aspect of overall programming.

To satisfy LFS weight-correction procedures, every record in the main LFS-sample-based results file had to have an age value. Year of birth was imputed where missing, then age derived from it. Age was then copied to the EFS file. The EFS sample was a subset of the NCT LFS-based sample, with data for each person captured twice - once before field edits and follow-up in the EFS file, then after edits and follow-up in the NCT file. However, some records in the EFS file were not captured in the main NCT file, because not returned in time for NCT capture, or lost in the mail. In the step of copying age to the EFS file from the NCT file, an "unknown" value was given to EFS records not in the NCT file. Some of these EFS records nevertheless had valid year-of-birth data. It might have been desirable to allow for separate derivation of age for these records. If future tests follow the same methodology for an EFS component, this step should be taken into consideration.

A detailed plan of processing steps needs to be drafted as well as documented to provide an overview of the tasks at hand and in the proper sequence. For example, after the creation of many of the derived variables the NCT team became aware that the temporary and foreign residents (i.e., those checking Step 4 or Step 6 of the questionnaire) should have been dropped from the NCT and EFS files at earlier stages of processing, when duplicate and empty records were dropped. This also impacted on the coding because the write-ins had been split off and sent for coding before temporary and foreign residents were dropped. At the time of code linkage we had more codes than we could initially account for, until we recalled that the extras were from the dropped temporary and foreign residents. Up-to-date documentation is important[10].

Autocoding

An unanticipated component was the request by subject matter to allow for more than one code to be returned for a multiple response. One characteristic of the ACTR system is that it can only provide one code per write-in. Multiple responses for Ethnic origin

[10]Attached are the NCT and EFS processing flowcharts.

(Q16) and Language (Q09 and Q11) were resolved manually by subject matter. Using SAS, Special Surveys produced hardcopy to assist with the manual resolution. It contained additional information from other questions for the respondent as well as write-ins from other members of the household.
It would be helpful if Subject Matter people were to decide in advance if they wanted to include more than one code from write-ins, and if so the maximum number of entries, and how to handle situations which exceed that limit. In addition, they could specify their requirements for additional information to aid coding early in the planning for coding and whether or not a hardcopy or machine readable format is acceptable.

Resolution of multiple response might be handled better by assigning a unique code for a given multiple during initial coding and then resolving these codes after the linkage processing. The reasons would be twofold: first, Special Surveys could accurately verify code linkage back to the NCT and EFS files and second, Subject Matter would have an opportunity to look at all codes before deciding the appropriate course of action (e.g., whether to recode to more than one code or not).

The multiple response problem will continue to be problematic with the use of open-ended questions. If this style of questioning is used for the 1996 Census or the next NCT further system development will be required to handle the resolution of these cases.

APPENDIX

Processing Flow

EFS DATAFLOW

UNIX PROCESSING

EFS RAW FILE

EFS BRANCH

STEP 01:

```
SAS PROGRAM:     CAPS/COD/PROD/EFS/VLDPID.PGM
SAS LOG:         CAPS/COD/PROD/EFS/VLDPID.LOG
SAS OUTPUT:      CAPS/COD/PROD/EFS/VLDPID.OUT
SAS CARD(S):     -
INPUT FILE(S):   CAPS/COD/PROD/EFS/RAW1728.COMPLETE.DAT.00.DEC01      (10,145)
OUTPUT FILE(S): CAPS/COD/PROD/EFS/VLDPID.RAW1728.DAT.10.DEC10         (10,090)
```

STEP 02:

```
SAS PROGRAM:     CAPS/COD/PROD/EFS/RAWNDU.PGM
SAS LOG:         CAPS/COD/PROD/EFS/RAWNDU.LOG
SAS OUTPUT:      CAPS/COD/PROD/EFS/RAWNDU.OUT
SAS CARD(S):     -
INPUT FILE(S):   CAPS/COD/PROD/EFS/VLDPID.RAW1728.DAT.10.DEC10         (10,090)
OUTPUT FILE(S): CAPS/COD/PROD/EFS/RAWNDU.RAW1728.DAT.11.MAR10          (10,087)
```

STEP 03:

```
SAS PROGRAM:     CAPS/COD/PROD/EFS/NONTEMP2.PGM
SAS LOG:         CAPS/COD/PROD/EFS/NONTEMP2.LOG
SAS OUTPUT:      CAPS/COD/PROD/EFS/NONTEMP2.OUT
SAS CARD(S):     -
INPUT FILE(S):   CAPS/COD/PROD/EFS/RAWNDU.RAW1728.DAT.11,MAR10         (10,090)
OUTPUT FILE(S): CAPS/COD/PROD/EFS/NONTEMP.RAW1728.DAT.03.MAR16         (10,087)
```

EFS BRANCH

MAINFRAME PROCESSING:

FIRST PROGRAM (STEP 1): SPEC.ICT9311.SRCE(SORT)

```
INPUT FILE(S):   SPEC.ICT9311.RAWPH1.ESSDEC01                  # RECORDS:  10,145

OUTPUT FILE(S):  SPEC.ICT9311.RAWPH1.ESSDEC01.SORTED           # RECORDS:  10,145
                 lrecl: 1728,  blksize:8640
```

(STEP 2): SPEC.ICT9311.SRCE(DUPSEQ2)

INPUT FILE(S): SPEC.ICT9311.RAWPH1.ESSDEC01.SORTED — # RECORDS: 10,145

OUTPUT FILE(S): SPEC.ICT9311.RAWPH1.DUPS — # RECORDS: 0
: SPEC.ICT9311.RAWPH1.UNIQUE — # RECORDS: 10,145
lrecl: 1728, blksize:8640

(Program is checking & producing output file with duplicate records)

SECOND PROGRAM: SPEC.ICT9311.SRCE(FSPLIT)

INPUT FILE(S): SPEC.ICT9311.RAWPH1.ESSDEC01 * — # RECORDS: 10,145
* DOWNLOADED TO UNIX: /PROD/EFS/RAW1728.COMPLETE.DAT.00.DEC01

OUTPUT FILE(S): SPEC.ICT9311.RAWESS.PERSID00 — # RECORDS: 55
: SPEC.ICT9311.RAWESS.VLDPERID * — # RECORDS: 10,090
lrecl: 1728, blksize:8640
* DOWNLOADED TO UNIX: /PROD/EFS/VLDPID.RAW1728.DAT.01.DEC10

(Program is spliting records with PERSONID = '00')

PREDIT PROGRAM: FINAL FILE BEFORE GOING INTO PREDIT

THIRD PROGRAM: SPEC.ICT9311.SRCE(COMPEDIT)

INPUT FILE(S): SPEC.ICT9311.RAWESS.VLDPERID — # RECORDS: 10,090

OUTPUT FILE(S): SPEC.ICT9311.EFS.PREDIT.NONTEXT — # RECORDS: 10,090
lrecl = 92, blksize = 8832

: SPEC.ICT9311.EFS.PREDIT.TEXT — # RECORDS: 9,551
lrecl = 532, blksize = 8512

(PL1 program is going through the PREDIT step and creating 2 output files;
(1)...TEXT file, lrecl =92, blksize = 8832
(2)...NON-TEXT file, lrecl = 532, blksize = 8512

FILE BEING PREPARED FOR AUTOCODING:

SUBSEQUENT PROGRAM(S) TO CREATE TEXT FILES FOR AUTOCODING

PROGRAM 1: SPEC.ICT9311.SRCE(ALLTEXT)

INPUT FILE(S): SPEC.ICT9311.RAWPH1.ESSDEC01 — # RECORDS: 10,145
LRECL: 1728, blksize = 8640

OUTPUT FILE(S): SPEC.ICT9311.PF.RAWPH1.ALLTEXT — # RECORDS: 10,145
LRECL: 794, blsize = 8734

(Program is picking up all fields where text is and producing 1 big text file of 794 bytes which will be split in subsequent program)

PROGRAM 2 : SPEC.ICT9311.SRCE(SMALLTXT)

STEP A: SORT
INPUT FILE: SPEC.ICT9311.PF.RAWPH1.ALLTEXT — # RECORDS: 10,145
LRECL: 794

72-BYTE OUTPUT FILE(S):

SPEC.ICT9311.PF.RAWPH1.Q09TX1
SPEC.ICT9311.PF.RAWPH1.Q09TX2
SPEC.ICT9311.PF.RAWPH1.Q09TX3
SPEC.ICT9311.PF.RAWPH1.Q11TXT
SPEC.ICT9311.PF.RAWPH1.Q12TXT
SPEC.ICT9311.PF.RAWPH1.Q13TXT
SPEC.ICT9311.PF.RAWPH1.Q16TXT
SPEC.ICT9311.PF.RAWPH1.Q18TX1
SPEC.ICT9311.PF.RAWPH1.Q16TX2
SPEC.ICT9311.PF.RAWPH1.Q16TX3
SPEC.ICT9311.PF.RAWPH1.Q19TXT
SPEC.ICT9311.PF.RAWPH1.Q24TXT
SPEC.ICT9311.PF.RAWPH1.Q41TXT
SPEC.ICT9311.PF.RAWPH1.Q22TXTOC
SPEC.ICT9311.PF.RAWPH1.Q22TXTIC

97-BYTE OUTPUT FILE:
SPEC.ICT9311.PF.RAWPH1.Q22TXTIC

163-BYTE OUTPUT FILE:
SPEC.ICT9311.PF.RAWPH1.Q42TXT

RECORDS ON EACH OUTPUT FILE: 10,145

NOTE: Each of the above files created, is split into two files; non-blank & blanks; non-blanks are sent for autocoding.

STEP B1: SPLIT FOR Q09 TEXT 1 (KNOWLEDGE OF LANGUAGE)
INPUT FILE(S): SPEC.ICT9311.PF.RAWPH1.Q09TX1 — # RECORDS: 10,145
LRECL: 72, BLKSIZE: 8856

OUTPUT FILE(S): SPEC.ICT9311.PF.RAWPH1.Q09TX1.NONBLK * — # RECORDS: 902
LRECL: 72
* FILE SENT FOR AUTOCODING

: SPEC.ICT9311.RAWPH1.Q09TX1.BLANKS — # RECORDS: 9,243

STEP B2: SPLIT FOR Q09 TEXT 2 (KNOWLEDGE OF LANGUAGE)
INPUT FILE(S): SPEC.ICT9311.PF.RAWPH1.Q09TX2 — # RECORDS: 10,145
LRECL: 72

: SPEC.ICT9311.PF.RAWPH1.Q09TX2.NONBLK * — # RECORDS: 102
LRECL: 72
* FILE SENT FOR AUTOCODING

: SPEC.ICT9311.RAWPH1.Q09TX2.BLANKS — # RECORDS: 10,043

STEP B3: SPLIT FOR Q09 TEXT 3 (KNOWLEDGE OF LANGUAGE)
INPUT FILE(S): SPEC.ICT9311.PF.RAWPH1.Q09TX3 — # RECORDS: 10,145
LRECL: 72

:	SPEC.ICT9311.PF.RAWPH1.Q09TX3.NONBLK *	# RECORDS: 24
	LRECL: 72	
*	FILE SENT FOR AUTOCODING	
:	SPEC.ICT9311.RAWPH1.Q09TX3.BLANKS	# RECORDS: 10,121

STEP C:	SPLIT FOR Q11 TEXT (MOTHER TONGUE)	
INPUT FILE(S):	SPEC.ICT9311.PF.RAWPH1.Q11TXT	# RECORDS: 10,145
	LRECL: 72	
OUTPUT FILE(S):	SPEC.ICT9311.PF.RAWPH1.Q11TXT.NONBLK *	# RECORDS: 883
	LRECL: 72	
*	FILE SENT FOR AUTOCODING	
:	SPEC.ICT9311.RAWPH1.Q11TXT.BLANKS	# RECORDS: 9,262

STEP D:	SPLIT FOR Q12 TEXT (PLACE OF BIRTH)	
INPUT FILE(S):	SPEC.ICT9311.PF.RAWPH1.Q12TXT	# RECORDS: 10,145
	LRECL: 72	
OUTPUT FILE(S):	SPEC.ICT9311.PF.RAWPH1.Q12TXT.NONBLK *	# RECORDS: 950
	LRECL: 72	
*	FILE SENT FOR AUTOCODING	
:	SPEC.ICT9311.RAWPH1.Q12TXT.BLANKS	# RECORDS: 9,195

STEP E:	SPLIT FOR Q13 TEXT (CITIZENSHIP)	
INPUT FILE(S):	SPEC.ICT9311.PF.RAWPH1.Q13TXT	# RECORDS: 10,145
	LRECL: 72	
OUTPUT FILE(S):	SPEC.ICT9311.PF.RAWPH1.Q13TXT.NONBLK *	# RECORDS: 299
	LRECL: 72	
*	FILE SENT FOR AUTOCODING	
:	SPEC.ICT9311.RAWPH1.Q13TXT.BLANKS	# RECORDS: 9,846

STEP F1:	SPLIT FOR Q16 TEXT 1 (ETHNIC ORIGIN)	
INPUT FILE(S):	SPEC.ICT9311.PF.RAWPH1.Q16TX1	# RECORDS: 10,145
	LRECL: 72	
OUTPUT FILE(S):	SPEC.ICT9311.PF.RAWPH1.Q16TX1.NONBLK *	# RECORDS: 8,838
	LRECL: 72	
*	FILE SENT FOR AUTOCODING	
:	SPEC.ICT9311.RAWPH1.Q16TX1.BLANKS	# RECORDS: 1,307

STEP F2:	SPLIT FOR Q16 TEXT 2 (ETHNIC ORIGIN)	
INPUT FILE(S):	SPEC.ICT9311.PF.RAWPH1.Q16TX2	# RECORDS: 10,145
	LRECL: 72	
OUTPUT FILE(S):	SPEC.ICT9311.PF.RAWPH1.Q16TX2.NONBLK *	# RECORDS: 3,262
	LRECL: 72	
*	FILE SENT FOR AUTOCODING	
:	SPEC.ICT9311.RAWPH1.Q16TX2.BLANKS	# RECORDS: 6,883

STEP F3: SPLIT FOR Q16 TEXT 3 (ETHNIC ORIGIN)
INPUT FILE(S): SPEC.ICT9311.PF.RAWPH1.Q16TX3 # RECORDS: 10,145
LRECL: 72

OUTPUT FILE(S): SPEC.ICT9311.PF.RAWPH1.Q16TX3.NONBLK * # RECORDS: 1,334
LRECL: 72
* FILE SENT FOR AUTOCODING

: SPEC.ICT9311.RAWPH1.Q16TX3.BLANKS # RECORDS: 8,811

STEP G: SPLIT FOR Q18 TEXT (RACE)
INPUT FILE(S): SPEC.ICT9311.PF.RAWPH1.Q18TXT # RECORDS: 10,145
LRECL: 72

OUTPUT FILE(S): SPEC.ICT9311.PF.RAWPH1.Q18TXT.NONBLK * # RECORDS: 98
LRECL: 72
* FILE SENT FOR AUTOCODING

: SPEC.ICT9311.RAWPH1.Q18TXT.BLANKS # RECORDS: 10,047

STEP H: SPLIT FOR Q19 TEXT (INDIAN BAND)
INPUT FILE(S): SPEC.ICT9311.PF.RAWPH1.Q19TXT # RECORDS: 10,145
LRECL: 72

OUTPUT FILE(S): SPEC.ICT9311.PF.RAWPH1.Q19TXT.NONBLK * # RECORDS: 46
LRECL: 72
* FILE SENT FOR AUTOCODING

: SPEC.ICT9311.RAWPH1.Q19TXT.BLANKS # RECORDS: 10,099

STEP I: SPLIT FOR Q24 TEXT (LANGUAGE OF EDUCATION)
INPUT FILE(S): SPEC.ICT9311.PF.RAWPH1.Q24TXT # RECORDS: 10,145
LRECL: 72

OUTPUT FILE(S): SPEC.ICT9311.PF.RAWPH1.Q24TXT.NONBLK * # RECORDS: 59
LRECL: 72
* FILE SENT FOR AUTOCODING

: SPEC.ICT9311.RAWPH1.Q24TXT.BLANKS # RECORDS: 10,086

STEP J: SPLIT FOR Q41 TEXT (LANGUAGE OF WORK)
INPUT FILE(S): SPEC.ICT9311.PF.RAWPH1.Q41TXT # RECORDS: 10,145
LRECL: 72

OUTPUT FILE(S): SPEC.ICT9311.PF.RAWPH1.Q41TXT.NONBLK * # RECORDS: 78
LRECL: 72
* FILE SENT FOR AUTOCODING

: SPEC.ICT9311.RAWPH1.Q41TXT.BLANKS # RECORDS: 10,067

STEP K: SPLIT FOR Q22 TEXT (MOBILITY OUTSIDE CANADA)

INPUT FILE(S): SPEC.ICT9311.PF.RAWPH1.Q22TXTOC — # RECORDS: 10,145
LRECL: 72

OUTPUT FILE(S): SPEC.ICT9311.PF.RAWPH1.Q22TXTOC.NONBLK * — # RECORDS: 179
LRECL: 72
* FILE SENT FOR AUTOCODING

: SPEC.ICT9311.RAWPH1.Q22TXTOC.BLANKS — # RECORDS: 9,966

STEP L: SPLIT FOR Q22 TEXT (MOBILITY INSIDE CANADA)

INPUT FILE(S): SPEC.ICT9311.PF.RAWPH1.Q22TXTIC — # RECORDS: 10,145
LRECL: 97

OUTPUT FILE(S): SPEC.ICT9311.PF.RAWPH1.Q22TXTIC.NONBLK * — # RECORDS: 1,540
LRECL: 97
* FILE SENT FOR AUTOCODING

: SPEC.ICT9311.RAWPH1.Q22TXTIC.BLANKS — # RECORDS: 8,605

STEP M: SPLIT FOR Q42 TEXT (PLACE OF WORK)

INPUT FILE(S): SPEC.ICT9311.PF.RAWPH1.Q42TXT — # RECORDS: 10,145
LRECL: 97

OUTPUT FILE(S): SPEC.ICT9311.PF.RAWPH1.Q42TXT.NONBLK * — # RECORDS: 4,058
LRECL: 163
* FILE SENT FOR AUTOCODING BUT

DUE TO SHORTAGE OF RESOURCES PLACE OF WORK UNIT
WAS UNABLE TO PERFORM MANUAL/INTERACTIVE CODING

: SPEC.ICT9311.RAWPH1.Q42TXT.BLANKS — # RECORDS: 6,087

MAINFRAME FILES RETURNED FROM AUTOCODING:

SPEC.ICT9311.NCTAC.LANQ09.FINAL.ESS -- 1,028 records
SPEC.ICT9311.NCTAC.LANQ11.FINAL.ESS --- 883 records
SPEC.ICT9311.NCTAC.PCTQ12.FINAL.ESS --- 950 records
SPEC.ICT9311.NCTAC.PCTQ13.FINAL.ESS --- 299 records
SPEC.ICT9311.ETOQ16.FINAL.ESS ------- 13,434 records
SPEC.ICT9311.ETOQ18.FINAL.ESS --------- 98 records
SPEC.ICT9311.INDQ19.FINAL.ESS --------- 46 records
SPEC.ICT9311.M5PQ22T1.FINAL.ESS ------ 1,540 records
SPEC.ICT9311.M5PQ22T2.FINAL.ESS -------- 179 records
SPEC.ICT9311.LANQ24.FINAL.ESS ---------- 59 records
SPEC.ICT9311.LANQ41.FINAL.ESS ---------- 78 records

UNIX PROCESSING FOR EFS AFTER PREDIT

"SPEC.ICT9311.EFS.PREDIT.NONTEXT"

DOWNLOADED TO UNIX

"CAPS/COD/OLDPROD/EFS/EXCLQ46.DAT.FEB01"

UNIX PROCESSING (in SAS)

EFS FLOW OF FILES AFTER PREDIT

INPUT: ../../OLDPROD/EFS/PREDIT.EXCLQ46.DAT.FEB01 (10,087)

Pgm Function:
1 linking codes received from auto-coding including ethnic and language manual resolution.

CODLNK.PGM
CODLNK.LOG
CODLNK.OUT

OUTPUT: ../../PROD/EFS/CODLNK.DVS953.DAT.00.MAR02 (10,087)

INPUT: ../../PROD/EFS/CODLNK.DVS953.DAT.00.MAR02 (10,087)

Pgm function:
2 eliminating Personid >01 that were temporary residents and missed in the earlier stages.

NONTEMP.PGM
NONTEMP.LOG
NONTEMP.OUT

OUTPUT: ../../PROD/EFS/NONTEMP.DVS953.DAT.01.MAR16 (9,986)

INPUT: ../../PROD/EFS/NONTEMP.DVS953.DAT.01.MAR16 (9,986)

Pgm Function:
3 Fixing Q46(income question) Program written by Amir Ranjbar and also establishing Multiple Error Flags for all questions that had multiple responses.

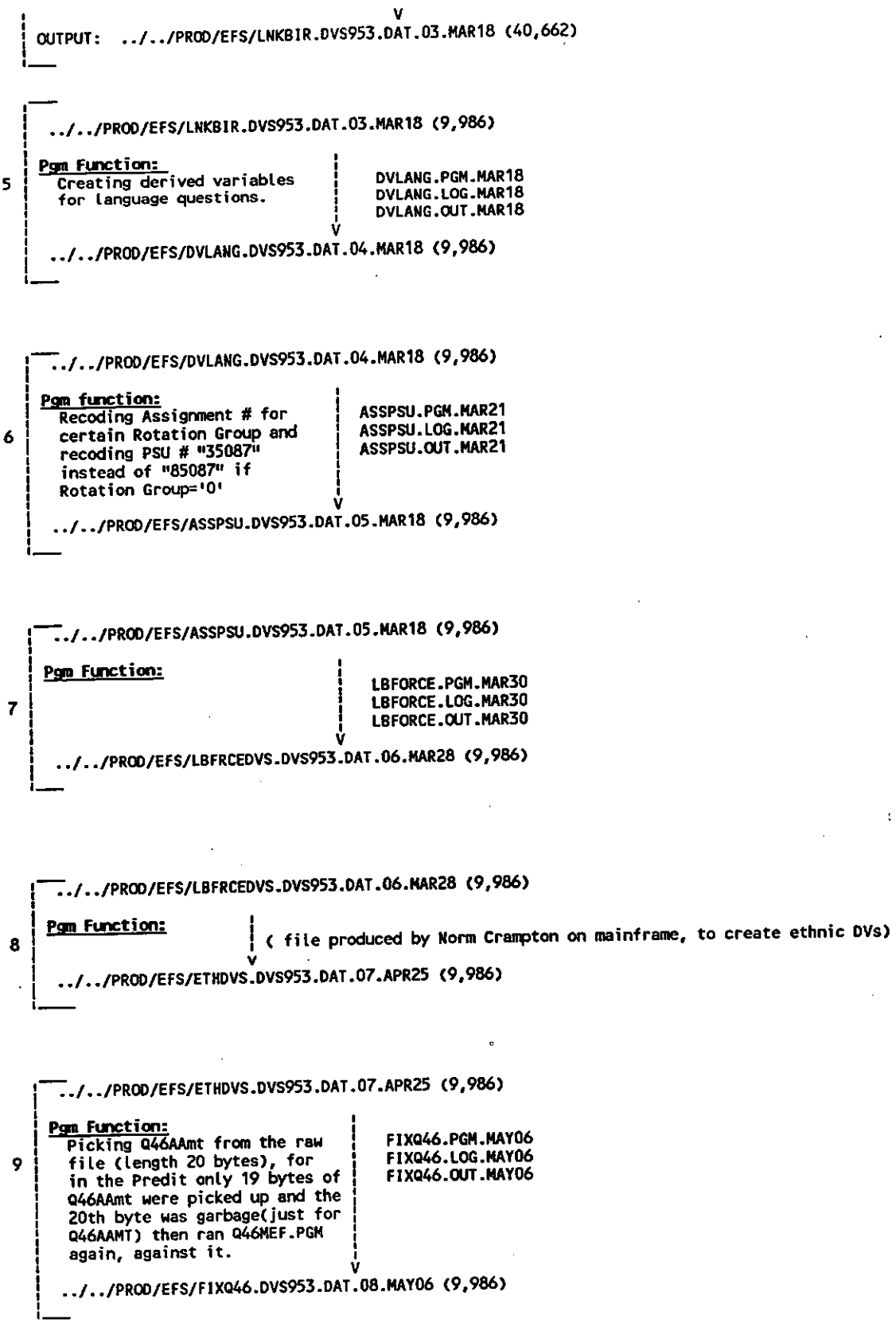
Q46MEF.PGM.MAR18
Q46MEF.LOG.MAR18
Q46MEF.OUT.MAR18

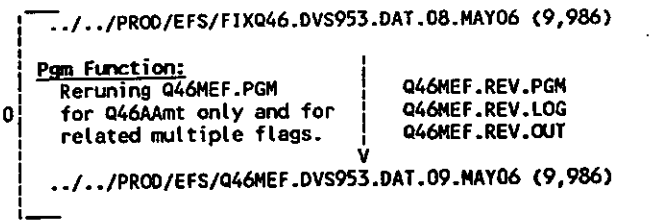
OUTPUT: ../../PROD/EFS/Q46MEF.DVS953.DAT.02.MAR18 (9,986)

INPUT: ../../PROD/EFS/Q46MEF.DVS953.DAT.02.MAR18 (9,986)
+
../../PROD/NCT/DVLANG.DVS953.DAT.02.FEB28 (40,915)

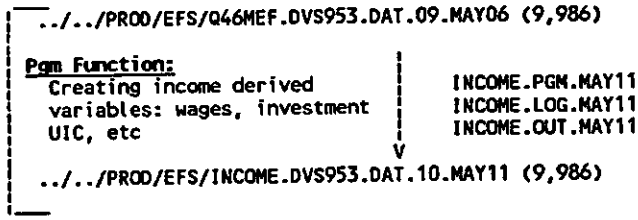
Pgm Function:
4 Loading date of birth, sex DVAGE2 from NCT to EFS when they are blank, for they were imputed on the NCT file.

LNKBIR.PGM.MAR18
LNKBIR.LOG.MAR18
LNKBIR.OUT.MAR18

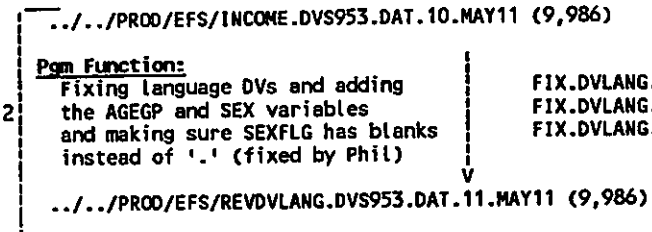

OUTPUT: ../../PROD/EFS/LNKBIR.DVS953.DAT.03.MAR18 (40,662)
../../PROD/EFS/LNKBIR.DVS953.DAT.03.MAR18 (9,986)
Pgm Function:
Creating derived variables for language questions.
DVLANG.PGM.MAR18
DVLANG.LOG.MAR18
DVLANG.OUT.MAR18
../../PROD/EFS/DVLANG.DVS953.DAT.04.MAR18 (9,986)
5
../../PROD/EFS/DVLANG.DVS953.DAT.04.MAR18 (9,986)
Pgm function:
Recoding Assignment # for certain Rotation Group and recoding PSU # "35087" instead of "85087" if Rotation Group='0'
ASSPSU.PGM.MAR21
ASSPSU.LOG.MAR21
ASSPSU.OUT.MAR21
../../PROD/EFS/ASSPSU.DVS953.DAT.05.MAR18 (9,986)
6
../../PROD/EFS/ASSPSU.DVS953.DAT.05.MAR18 (9,986)
Pgm Function:
LBFORCE.PGM.MAR30
LBFORCE.LOG.MAR30
LBFORCE.OUT.MAR30
../../PROD/EFS/LBFRCEDVS.DVS953.DAT.06.MAR28 (9,986)
7
../../PROD/EFS/LBFRCEDVS.DVS953.DAT.06.MAR28 (9,986)
Pgm Function:
(file produced by Norm Crampton on mainframe, to create ethnic DVs)
../../PROD/EFS/ETHDVS.DVS953.DAT.07.APR25 (9,986)
8
../../PROD/EFS/ETHDVS.DVS953.DAT.07.APR25 (9,986)
Pgm Function:
Picking Q46AAmt from the raw file (length 20 bytes), for in the Predit only 19 bytes of Q46AAmt were picked up and the 20th byte was garbage(just for Q46AAMT) then ran Q46MEF.PGM again, against it.
FIXQ46.PGM.MAY06
FIXQ46.LOG.MAY06
FIXQ46.OUT.MAY06
../../PROD/EFS/FIXQ46.DVS953.DAT.08.MAY06 (9,986)
9


../../PROD/EFS/FIXQ46.DVS953.DAT.08.MAY06 (9,986)
Pgm Function:
Reruning Q46MEF.PGM
for Q46AAmt only and for
related multiple flags.
Q46MEF.REV.PGM
Q46MEF.REV.LOG
Q46MEF.REV.OUT
../../PROD/EFS/Q46MEF.DVS953.DAT.09.MAY06 (9,986)

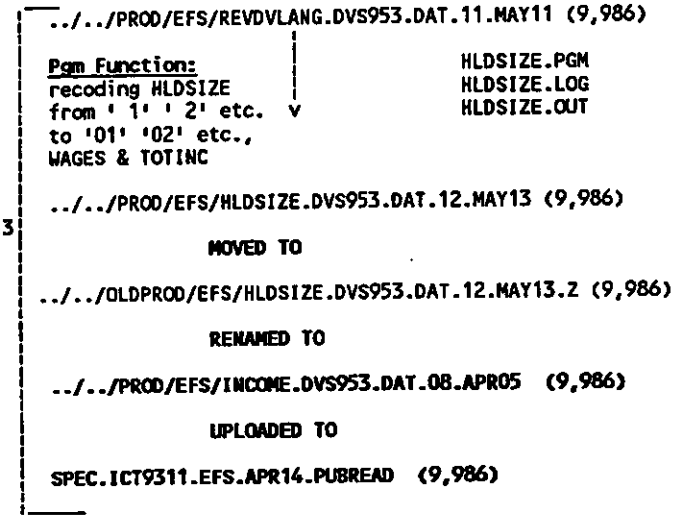
10


../../PROD/EFS/Q46MEF.DVS953.DAT.09.MAY06 (9,986)
Pgm Function:
Creating income derived
variables: wages, investment
UIC, etc
INCOME.PGM.MAY11
INCOME.LOG.MAY11
INCOME.OUT.MAY11
../../PROD/EFS/INCOME.DVS953.DAT.10.MAY11 (9,986)

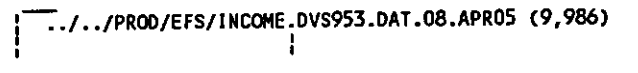
11

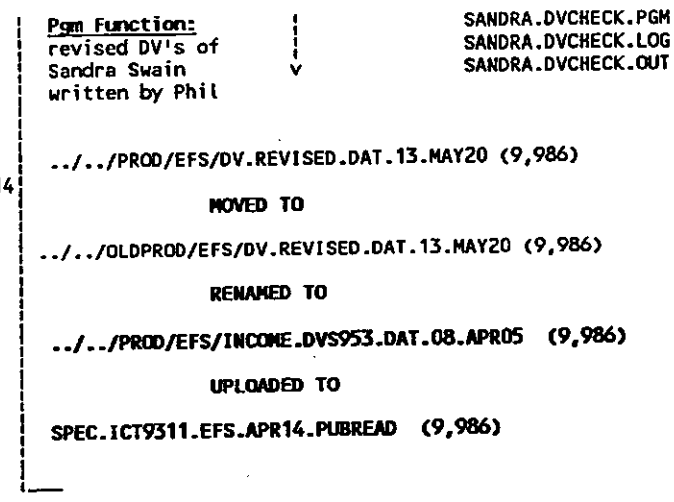

../../PROD/EFS/INCOME.DVS953.DAT.10.MAY11 (9,986)
Pgm Function:
Fixing language DVs and adding
the AGEGP and SEX variables
and making sure SEXFLG has blanks
instead of '.' (fixed by Phil)
FIX.DVLANG.PGM
FIX.DVLANG.LOG
FIX.DVLANG.OUT
../../PROD/EFS/REVDVLANG.DVS953.DAT.11.MAY11 (9,986)

12


../../PROD/EFS/REVDVLANG.DVS953.DAT.11.MAY11 (9,986)
Pgm Function:
recoding HLDSIZE
from ' 1' ' 2' etc.
to '01' '02' etc.,
WAGES & TOTINC
HLDSIZE.PGM
HLDSIZE.LOG
HLDSIZE.OUT
../../PROD/EFS/HLDSIZE.DVS953.DAT.12.MAY13 (9,986)
MOVED TO
../../OLDPROD/EFS/HLDSIZE.DVS953.DAT.12.MAY13.Z (9,986)
RENAMED TO
../../PROD/EFS/INCOME.DVS953.DAT.08.APR05 (9,986)
UPLOADED TO
SPEC.ICT9311.EFS.APR14.PUBREAD (9,986)

13


../../PROD/EFS/INCOME.DVS953.DAT.08.APR05 (9,986)


Pgm Function:
revised DV's of
Sandra Swain
written by Phil
SANDRA.DVCHECK.PGM
SANDRA.DVCHECK.LOG
SANDRA.DVCHECK.OUT
../../PROD/EFS/DV.REVISED.DAT.13.MAY20 (9,986)
14
MOVED TO
../../OLDPROD/EFS/DV.REVISED.DAT.13.MAY20 (9,986)
RENAMED TO
../../PROD/EFS/INCOME.DVS953.DAT.08.APR05 (9,986)
UPLOADED TO
SPEC.ICT9311.EFS.APR14.PUBREAD (9,986)

NCT DATAFLOW

UNIX PROCESSING

NCT RAW FILE

NCT BRANCH

STEP 01:

SAS PROGRAM: CAPS/COD/PROD/EFS/VLDPID.PGM
SAS LOG: CAPS/COD/PROD/EFS/VLDPID.LOG
SAS OUTPUT: CAPS/COD/PROD/EFS/VLDPID.OUT
SAS CARD(S): -
INPUT FILE(S): CAPS/COD/PROD/EFS/RAW1728.COMPLETE.DAT.00.JAN26 (47,057)
OUTPUT FILE(S): CAPS/COD/PROD/EFS/RAW1728.VLDPID.DAT.00.FEB10 (40,917)

STEP 02:

SAS PROGRAM: CAPS/COD/PROD/EFS/RAWNTEMP.PGM
SAS LOG: CAPS/COD/PROD/EFS/RAWNTEMP.LOG
SAS OUTPUT: CAPS/COD/PROD/EFS/RAWNTEMP.OUT
SAS CARD(S): -
INPUT FILE(S): CAPS/COD/PROD/EFS/RAW1728.VLDPID.DAT.00.FEB10 (40,917)
+
: CAPS/COD/PROD/EFS/RAW1728.COMPLETE.DAT.00.JAN26 (246)

OUTPUT FILE(S): CAPS/COD/PROD/EFS/RAWN1728.NONTEMP.DAT.02.MAR17 (40,)

UPLOADED TO MAINFRAME

'SPEC.ICT9311.RAW1728.NONTEMP.MAR17'

(Program is removing all duplicate records and all non_temporary residence)

NCT BRANCH

MAINFRAME PROCESSING:

FIRST PROGRAM : SPEC.ICT9311.SRCE(DUPSEQ2)

INPUT FILE(S): SPEC.ICT9311.RAWNCT.BKUP0126.PUBREAD # RECORDS: 47,057

OUTPUT FILE(S): SPEC.ICT9311.RAWNCT.DUPS # RECORDS: 86

SPEC.ICT9311.RAWNCT.UNIQUE # RECORDS: 46,971
(LRECL: 1728, BLKSIZE: 8640)

(Used SORT & PL1 to create two files: DUPLICATE RECORD FILE & UNIQUE RECORD FILE)
(Program is checking duplicate records)

SECOND PROGRAM: SPEC.ICT9311.SRCE(FSPLIT3)

INPUT FILE(S): SPEC.ICT9311.RAWNCT.DUPS # RECORDS: 86

OUTPUT FILE(S): SPEC.ICT9311.RAWNCT.DUPS.DROP # RECORDS: 30
: SPEC.ICT9311.RAWNCT.DUPS.KEEP # RECORDS: 56
(LRECL: 1728, BLKSIZE: 8640) (28 DUPLICATES)

(Program is spliting households that could not be manually resolved vs. those that can be resolved. 30 vs 50 respectively. set of 28 duplicates will be merged to the UNIQUE File).

THIRD PROGRAM: SPEC.ICT9311.SRCE(RECODE)

INPUT FILE(S): SPEC.ICT9311.RAWNCT.UNIQUE # RECORDS: 46,971

OUTPUT FILE(S): SPEC.ICT9311.RAWNCT.RECODE.HHLD6.FEB3 # RECORDS: 46,971
. LRECL: 1728, BLKSIZE=8640

(14 records were recoded using PGM=RECODED
6 from Phil's frequency dump, where PERSONID = '00' but should have been '01'
8 from Lori's memo dated January 21, 1994 "DOCUMENTATION FOR NCT PHASE2); records with PERSONID 12,13,14)

FOURTH PROGRAM: (STEP1) SPEC.ICT9311.SRCE(VLDMERGE)

INPUT FILE(S): SPEC.ICT9311.RAWNCT.DUPS.KEEP # RECORDS: 56

OUTPUT FILE(S): SPEC.ICT9311.RAWNCT.DUPUNIQ # RECORDS: 28

(STEP2) SPEC.ICT9311.SRCE(VLDMERGE)

INPUT FILE(S): SPEC.ICT9311.RAWNCT.DUPUNIQ # RECORDS: 28
: SPEC.ICT9311.RAWNCT.RECODE.HHLD6.FEB3 # RECORDS: 46,971

OUTPUT FILE(S): SPEC.ICT9311.RAWNCT.MERGE.DUPREC.FEB3 # RECORDS: 46,999
(LRECL: 1728, BLKSIZE: 8640)

(SAS Program is picking up from 56 records(duplicate) a unique record by using IF>LAST>ID function of SAS and then concatenating the .DUPUNIQ. file with the recoded .HHLD6.FEB3. using DBSCOPY.

FIFTH PROGRAM (STEP 1): SPEC.ICT9311.SRCE(FSPLIT2)

INPUT FILE(S): SPEC.ICT9311.RAWNCT.MERGE.DUPREC.FEB3 # RECORDS: 46,999

OUTPUT FILE(S): SPEC.ICT9311.RAWNCT.NOHHLD.DATA # RECORDS: 39
: SPEC.ICT9311.RAWNCT.HHLD.DATA # RECORDS: 46,960
(LRECL: 1728, BLKSIZE: 8640)

(STEP 2): SPEC.ICT9311.SRCE(FSPLIT2)

INPUT FILE(S): SPEC.ICT9311.RAWNCT.NOHHLD.DATA # RECORDS: 39

OUTPUT FILE(S): SPEC.ICT9311.RAWNCT.NOHHLD.KEEP # RECORDS: 2
: SPEC.ICT9311.RAWNCT.NOHHLD.DROP # RECORDS: 37
(LRECL: 1728, BLKSIZE: 8640)

(Program is spliting PERSONID > '00' & no data beyond pos. 45;
39 records were found & dumped; 2 out of 39 were selected to keep and merge with good file).

(STEP 3): SPEC.ICT9311.SRCE(FSPLIT2)

INPUT FILE(S): SPEC.ICT9311.RAWNCT.HHLD.DATA # RECORDS:
46,962
: SPEC.ICT9311.RAWNCT.NOHHLD.KEEP # RECORDS: 2

OUTPUT FILE(S): SPEC.ICT9311.RAWNCT.MERGE.DUPREC.FEB4 # RECORDS:
46,962
(LRECL: 1728, BLKSIZE: 8640)

(Program is using DBSCOPY to concatenate 2 above files together and produce 1 output file)

SIXTH PROGRAM: SPEC.ICT9311.SRCE(ST4ST6)

INPUT FILE(S): SPEC.ICT9311.RAWNCT.MERGE.DUPREC.FEB4 # RECORDS:
46,960

OUTPUT FILE(S): SPEC.ICT9311.RAWNCT.WITH.ST4ST6.DROP # RECORDS: 263
: SPEC.ICT9311.RAWNCT.WITHOUT.ST4ST6 # RECORDS:
46,716
(LRECL: 1728, BLKSIZE: 8640)

(Program is spliting the good file into 2 files
-- one with step4 = 3 or step6 = 5 (246 records)
-- one without step4 = 3 or step6 = 5;
(temporary residents and usual place elsewhere in Canada)

NOTE: 17 RECORDS HAD BOTH STEP4 = 3 & STEP6 = 5; THAT ACCOUNTS FOR THE DOUBLE COUNT IN 263 RECORDS.

SEVENTH PROGRAM: SPEC.ICT9311.SRCE(PER00)

INPUT FILE(S): SPEC.ICT9311.RAWNCT.WITHOUT.ST4ST6 # RECORDS:
46,716

OUTPUT FILE(S): SPEC.ICT9311.RAWNCT.VLDPERID # RECORDS:
40,934
: SPEC.ICT9311.RAWNCT.PERID00 # RECORDS:
5,782
(LRECL: 1728, BLKSIZE: 8640)

(Program is spliting PERSONID ='00' into a separate file & keeping the rest of the records into a VLDPERID file & dumping a handful of records to see the actual records with PERSONID = '00'.)

EIGHTH ROGRAM: SPEC.ICT9311.SRCE(POSTALBL)

INPUT FILE(S): SPEC.ICT9311.RAWNCT.VLDPERID # RECORDS:40,934

OUTPUT FILE(S): SPEC.ICT9311.RAWNCT.VLDPERID.BLANK # RECORDS: 17
: SPEC.ICT9311.RAWNCT.VLDPERID.NONBLK # RECORDS:

40,917
(LRECL: 1728, BLKSIZE: 8640)

(Program is spliting input file into 2 output files)
-- one which has no information beyond Postal Code eg. nothing in Pos.51 to next 154 bytes, these are being deleted from good file.
-- one with information.

NINTH ROGRAM: SPEC.ICT9311.SRCE(DUPSEQ4)

INPUT FILE(S): SPEC.ICT9311.RAWNCT.VLDPERID.NONBLK # RECORDS: 40,917

OUTPUT FILE(S): SPEC.ICT9311.RAWNCT.VLDPERID.UNIQ # RECORDS: 40,915
(LRECL: 1728, BLKSIZE: 8640)

(Program is removing 2 duplicate records from the file)
NOTE: This step was done after the file had gone through the PREDIT. These 2 duplicatae records were picked up a week later when QQQTXT was sent for Auto-Coding.

PREDIT PROGRAM: FINAL FILE BEFORE GOING INTO PREDIT

TENTH PROGRAM: SPEC.ICT9311.SRCE(COMPEDIT)

INPUT FILE(S): SPEC.ICT9311.RAWNCT.VLDPERID.NONBLK # RECORDS: 40,917
: SPEC.ICT9311.SRCE(PREEDIT6)

OUTPUT FILE(S): SPEC.ICT9311.NCT.PREDIT.NONTEXT.FEB7 # RECORDS: 40,917
lrecl: 532, blksize: 8512

: SPEC.ICT9311.NCT.PREDIT.TEXT.FEB7 # RECORDS: 247,197
lrecl: 92, blksize: 8832

(PL1 program is going through the PREDIT step and creating 2 output files;
(1).. TEXT FILE , Lrecl = 92, Blksize = 8832
(2).. NON-TEXT FILE, Lrecl = 532, Blksize = 8512)

FILES BEING PREPARED FOR AUTOCODING:

PROGRAM 1: SPEC.ICT9311.SRCE(SORT4)

INPUT FILE(S): SPEC.ICT9311.NCT.PREDIT.TEXT.FEB7 # RECORDS: 247,197
LRECL = 92, BLKSIZE = 8832

OUTPUT FILE(S): SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED1 (Q09T1, Q09T2, Q09T3) # RECORDS: 7,022
: SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED2 (Q16T1, Q16T2, Q16T3) # RECORDS: 57,560
: SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED3 (Q22T1, Q22T2, Q22T3) # RECORDS: 12,225
: SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED4 (Q42T1-Q42T3, Q42PC) # RECORDS: 58,968
: SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED5 (Q11TX) # RECORDS:

6,041 : SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED6 (Q12TX) # RECORDS:

5,959 : SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED7 (Q13TX) # RECORDS:

2,105 : SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED8 (Q18TX) # RECORDS: 670
: SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED9 (Q19TX) # RECORDS: 588
: SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED10 (Q24TX) # RECORDS: 138
: SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED11 (Q41TX) # RECORDS: 447
: SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED12 (Q02TX) # RECORDS: 613

: SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED13 (Q34TX) # RECORDS:

32,.085 : SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED14 (Q35TX) # RECORDS:

20,890 : SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED15 (Q37TX) # RECORDS:

21,228 : SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED16 (Q38TX) # RECORDS:

20,262 : SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED17 (Q43TX) # RECORDS: 396
: SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED18 (Q46TA-Q46TK) # RECORDS:

(Using SORT to create 18 output files according to the required fields from the input file)

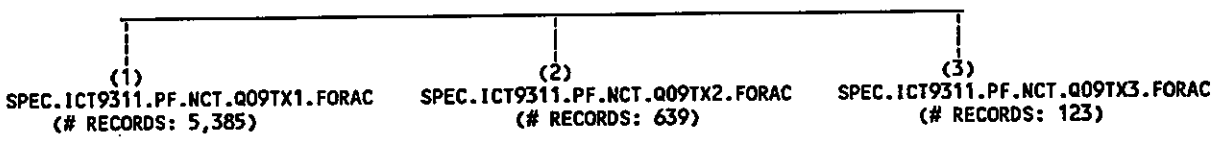
SUBSEQUENT PROGRAM(S) TO CREATE TEXT FILES FOR AUTOCODING

SPLIT FOR Q09 TEXT (KNOWLEDGE OF LANGUAGE)

INPUT: SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED1 # RECORDS: 7,022
SPEC.ICT9311.NCTAC.LANQ09.FINAL.ESS # RECORDS: 1,028

OUTPUT(1): SPEC.ICT9311.PF.NCT.Q09TXT.WITHAC # RECORDS: 875
lrecl = 99, blksize = 8811

(2): SPEC.ICT9311.PF.NCT.Q09TXT.FORAC # RECORDS: 6,147
Lrecl = 92, Blksize = 8832



SPLIT FOR Q11 TEXT (MOTHER TONGUE)

INPUT: SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED5 # RECORDS: 6,041
SPEC.ICT9311.NCTAC.LANQ11.FINAL.ESS # RECORDS: 883

OUTPUT(1): SPEC.ICT9311.PF.NCT.Q11TXT.WITHAC # RECORDS: 756
lrecl = 99, blksize = 8811

(2): SPEC.ICT9311.PF.NCT.Q11TXT.FORAC # RECORDS: 5,285
lrecl = 92, blksize = 8832

SPLIT FOR Q12 TEXT (PLACE OF BIRTH)

INPUT: SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED6 # RECORDS: 5,959
SPEC.ICT9311.NCTAC.PCTQ12.FINAL.ESS # RECORDS: 950

OUTPUT(1): SPEC.ICT9311.PF.NCT.Q12TXT.WITHAC # RECORDS: 808
lrecl = 99, blksize = 8811

(2): SPEC.ICT9311.PF.NCT.Q12TXT.FORAC # RECORDS: 5,151
lrecl = 92, blksize = 8832

SPLIT FOR Q13 TEXT (CITIZENSHIP)

INPUT: SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED7 # RECORDS: 2,105
SPEC.ICT9311.NCTAC.PCTQ13.FINAL.ESS # RECORDS: 299

OUTPUT(1): SPEC.ICT9311.PF.NCT.Q13TXT.WITHAC # RECORDS: 237
lrecl = 99, blksize = 8811

(2): SPEC.ICT9311.PF.NCT.Q13TXT.FORAC # RECORDS: 1,868
lrecl = 92, blksize = 8832

SPLIT FOR Q16 TEXT (ETHNIC ORIGIN)

INPUT: SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED2 # RECORDS: 57,560
SPEC.ICT9311.NCTAC.ETOQ16.FINAL.ESS # RECORDS: 13,434

OUTPUT(1): SPEC.ICT9311.PF.NCT.Q16TXT.WITHAC # RECORDS: 11,869
lrecl = 99, blksize = 8811

(2): SPEC.ICT9311.PF.NCT.Q16TXT.FORAC # RECORDS: 45,691
lrecl = 92, blksize = 8832

(1) SPEC.ICT9311.PF.NCT.Q16TX1.FORAC (# RECORDS: 31,338)

(2) SPEC.ICT9311.PF.NCT.Q16TX2.FORAC (# RECORDS: 10,608)

(3) SPEC.ICT9311.PF.NCT.Q16TX3.FORAC (# RECORDS: 3,745)

SPLIT FOR Q18 TEXT (RACE)

INPUT: SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED8 # RECORDS: 670
SPEC.ICT9311.NCTAC.ETOQ18.FINAL.ESS # RECORDS: 94

OUTPUT(1): SPEC.ICT9311.PF.NCT.Q18TXT.WITHAC # RECORDS: 75
lrecl = 99, blksize = 8811

(2): SPEC.ICT9311.PF.NCT.Q18TXT.FORAC # RECORDS: 595
lrecl = 92, blksize = 8832

SPLIT FOR Q19 TEXT (INDIAN BAND)

INPUT: SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED9 # RECORDS: 588
SPEC.ICT9311.NCTAC.INDQ19.FINAL.ESS # RECORDS: 46

OUTPUT(1): SPEC.ICT9311.PF.NCT.Q19TXT.WITHAC # RECORDS: 37
lrecl = 99, blksize = 8811

(2): SPEC.ICT9311.PF.NCT.Q19TXT.FORAC # RECORDS: 551
lrecl = 92, blksize = 8832

SPLIT FOR Q22 TEXT (MOBILITY INSIDE/OUTSIDE CANADA

INPUT: SPEC.ICT9311.RAWNCT.VLDPEID.NONBLK (40,917)
OUTPUT: SPEC.ICT9311.NCT.Q22TXT (40,917)

SPEC.ICT9311.NCT.Q22TXTIC (40,917)

SPEC.ICT9311.NCT.Q22TXTOC (40,917)

SPEC.ICT9311.NCT.Q22TXTIC.NONBLK (5,763)

SPEC.ICT9311.NCT.Q22TXTIC.BLANKS (35,154)

SPEC.ICT9311.NCT.Q22TXTOC.NONBLK (1,210)

SPEC.ICT9311.NCT.Q22TXTOC.BLANKS (39,707)

INPUT: SPEC.ICT9311.Q22TXTIC.NONBLK (5,763)
OUTPUT: SPEC.ICT9311.Q22TXTIC.UNIQUE (5,761)

INPUT: SPEC.ICT9311.Q22TXTIC.UNIQUE (5,761)
SPEC.ICT9311.NCTAC.M5PQ22T1.FINAL.ESS (1,540)

OUTPUT: SPEC.ICT9311.PF.NCT.Q22TXTIC.WITHAC (892)
lrecl = 104, blksize = 8840

SPEC.ICT9311.PF.NCT.Q22TXTIC.FORAC (4,869)
lrecl = 97, blksize = 8827

INPUT: SPEC.ICT9311.Q22TXTOC.NONBLK (1,210)

OUTPUT: SPEC.ICT9311.Q22TXTOC.UNIQUE (1,210)

INPUT: SPEC.ICT9311.Q22TXTOC.UNIQUE (1,210)
SPEC.ICT9311.NCTAC.PCTQ22T3.FINAL.ESS (179)

OUTPUT: SPEC.ICT9311.PF.NCT.Q22TXTOC.WITHAC (138)
lrecl = 99, blksize = 8811

SPEC.ICT9311.PF.NCT.Q22TXTOC.FORAC (1,072)
lrecl = 92, blksize = 8832

SPLIT FOR Q24 TEXT (LANGUAGE OF EDUCATION)

INPUT: SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED10 # RECORDS: 138
SPEC.ICT9311.NCTAC.LANQ24.FINAL.ESS # RECORDS: 59

OUTPUT(1): SPEC.ICT9311.PF.NCT.Q24TXT.WITHAC # RECORDS: 25
lrecl = 99, blksize = 8811

(2): SPEC.ICT9311.PF.NCT.Q24TXT.FORAC # RECORDS: 113
lrecl = 92, blksize = 8832

SPLIT FOR Q41 TEXT (LANGUAGE OF WORK)

```
 INPUT: SPEC.ICT9311.NCT.PREDIT.TEXT.SORTED11          # RECORDS: 447
        SPEC.ICT9311.NCTAC.LANQ41.FINAL.ESS            # RECORDS: 78

OUTPUT(1): SPEC.ICT9311.PF.NCT.Q41TXT.WITHAC           # RECORDS: 52
           lrecl = 99,  blksize = 8811

      (2): SPEC.ICT9311.PF.NCT.Q41TXT.FORAC            # RECORDS: 395
           lrecl = 92,  blksize = 8832
```

SPLIT FOR Q42 TEXT (PLACE OF WORK)

```
 INPUT: SPEC.ICT9311.NCT.VLDPERID.UNIQUE

OUTPUT: SPEC.ICT9311.NCT.Q42TXT

 INPUT: SPEC.ICT9311.NCT.Q42TXT
OUTPUT: SPEC.ICT9311.NCT.Q42TXT.NONBLK
      : SPEC.ICT9311.NCT.Q42TXT.BLANKS

 INPUT: SPEC.ICT9311.NCT.Q42TXT.NONBLK                 # RECORDS:  17,177
      : SPEC.ICT9311.PF.RAWPH1.Q42TXT.NONBLK           # RECORDS:   4,058
OUTPUT: SPEC.ICT9311.PF.NCT.Q12TXT.WITHOAC             # RECORDS:   1,425
        lrecl = 163,  blksize = 8802

        SPEC.ICT9311.PF.NCT.Q12TXT.FORAC               # RECORDS:  15,752
        lrecl = 163,  blksize = 8802
```

UNIX PROCESSING FOR NCT AFTER PREDIT

"SPEC.ICT9311.NCT.PREDIT.NONTEXT.FEB7" (40,917)

DOWNLOADED TO UNIX

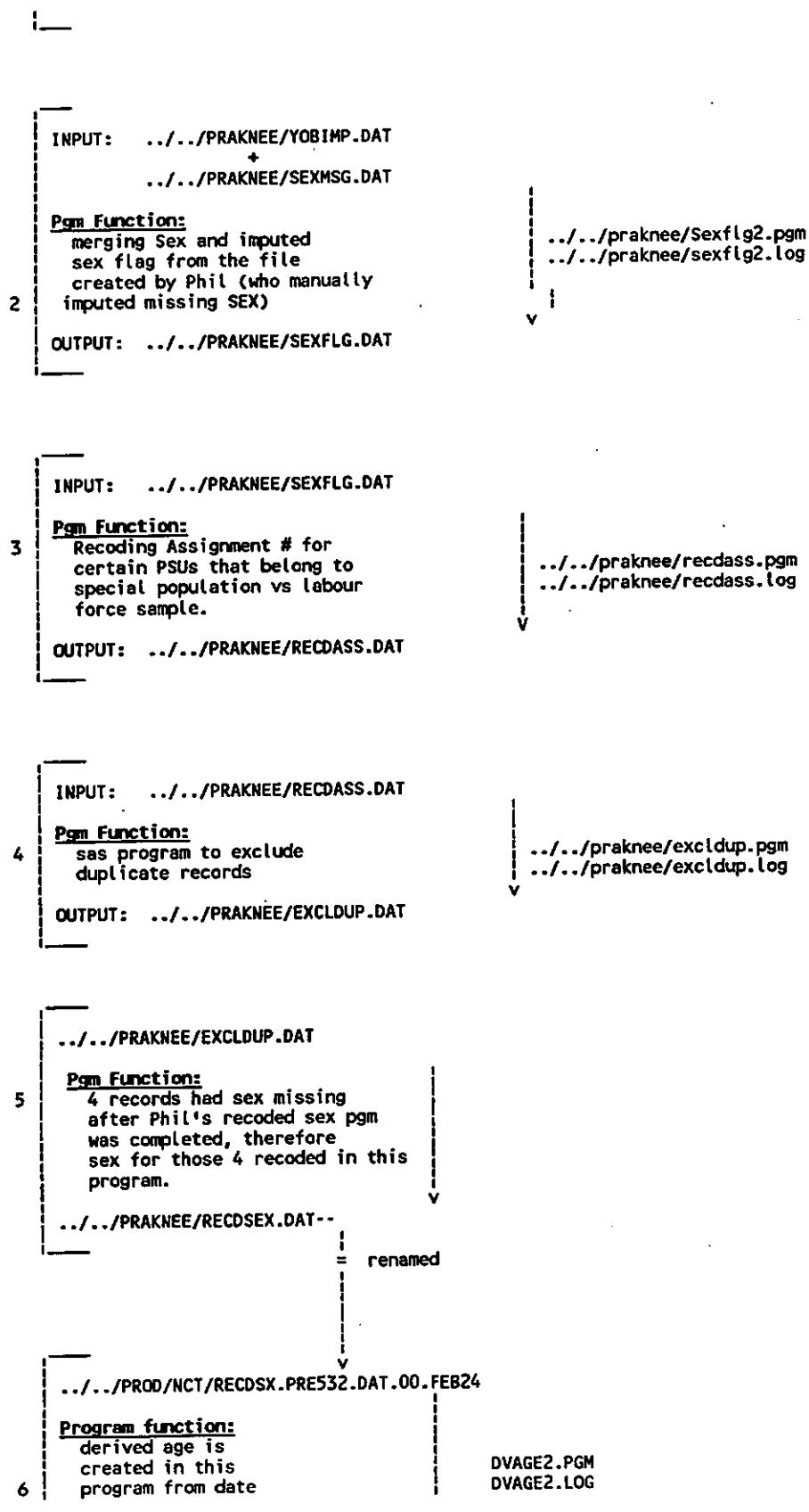
"CAPS/COD/STEVENS/NCT/NCT.PREDIT.EXCLQ46" (40,917)

UNIX PROCESSING (in SAS)

NCT FLOW OF FILES AFTER PREDIT

```
   INPUT:   ../../STEVENS/NCT/NCT.PREDIT.EXCLQ46
                          +
1          ../../PRAKNEE/YOBIMPUT.DAT

   Pgm Function:
     merging Year of Birth and                    ../../praknee/Yobimput.pgm
     imputation flag for year of birth            ../../praknee/Yobimput.log
     from the 274 records file
     created by Matthew Briggs, SSMD
   (who manually imputed missing y.o.b.)
                                                  V
   OUTPUT:  ../../PRAKNEE/YOBIMP.DAT
```


INPUT: ../../PRAKNEE/YOBIMP.DAT
+
../../PRAKNEE/SEXMSG.DAT
Pgm Function:
merging Sex and imputed sex flag from the file created by Phil (who manually imputed missing SEX)
../../praknee/Sexflg2.pgm
../../praknee/sexflg2.log
2
OUTPUT: ../../PRAKNEE/SEXFLG.DAT
INPUT: ../../PRAKNEE/SEXFLG.DAT
Pgm Function:
Recoding Assignment # for certain PSUs that belong to special population vs labour force sample.
../../praknee/recdass.pgm
../../praknee/recdass.log
3
OUTPUT: ../../PRAKNEE/RECDASS.DAT
INPUT: ../../PRAKNEE/RECDASS.DAT
Pgm Function:
sas program to exclude duplicate records
../../praknee/excldup.pgm
../../praknee/excldup.log
4
OUTPUT: ../../PRAKNEE/EXCLDUP.DAT
../../PRAKNEE/EXCLDUP.DAT
Pgm Function:
4 records had sex missing after Phil's recoded sex pgm was completed, therefore sex for those 4 recoded in this program.
5
../../PRAKNEE/RECDSEX.DAT--
= renamed
../../PROD/NCT/RECDSX.PRE532.DAT.00.FEB24
Program function:
derived age is created in this program from date
DVAGE2.PGM
DVAGE2.LOG
6

of birth by
Christian Branconnier, COD.
Derived variable is
DVAGE2 (pos.670,2)

DVAGE2.OUT

../../PROD/NCT/DVAGE2.DVS953.DAT.01.FEB24

../../PROD/NCT/DVAGE2.DVS953.DAT.01.FEB24 (40,915)

Program function:
derived language
variable are created
here from Q9
KOL (672,2)- knowledge of language
HLN (674,1)- home language
MTN (675,2)- mother tongue
7 LGEDUC(677,1) - language of education
LGWORK (678,1)- language of work
OL (679,1) - knowledge of official language
NOL (680,1)- knowledge of non-official language
OLNOL(681,1)-knowledge of official/ non-official language

DVLANG.PGM
DVLANG.LOG
DVLANG.OUT

../../PROD/NCT/DVLANG.DVS953.DAT.02.FEB28 (40,915

../../PROD/NCT/DVLANG.DVS953.DAT.02.FEB28 (40,915)
+
../../PROD/NCT/ST4ST6.DAT (263)

Program function:
8 created data file with 263 rec. & 3 variables-(hhldid,step4,step6) This program is therefore removing households from the DVLANG file that have temporary residences through merge process by eliminating those records that were in the st4st5 data file.

NONTEMP.PGM
NONTEMP.LOG
NONTEMP.OUT

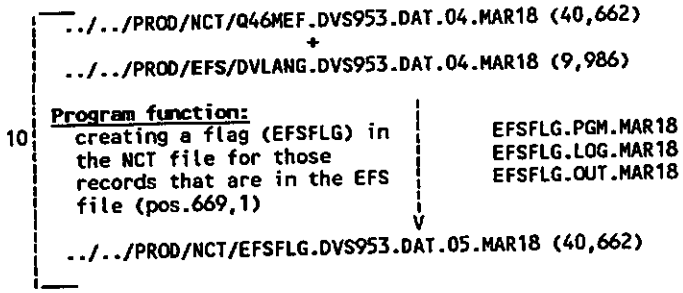
../../PROD/NCT/NONTEMP.DVS953.DAT.03.MAR16 (40,662)

../../PROD/NCT/NONTEMP.DVS953.DAT.03.MAR16 (40,662)

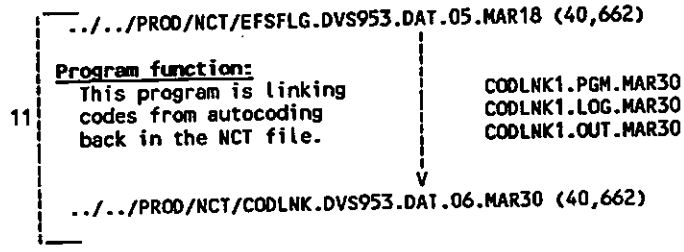
Program function:
9 Q46 (income question) and Multiple error flags for all related questions have been created here in this program

Q46MEF.PGM.MAR18
Q46MEF.LOG.MAR18
Q46MEF.OUT.MAR18

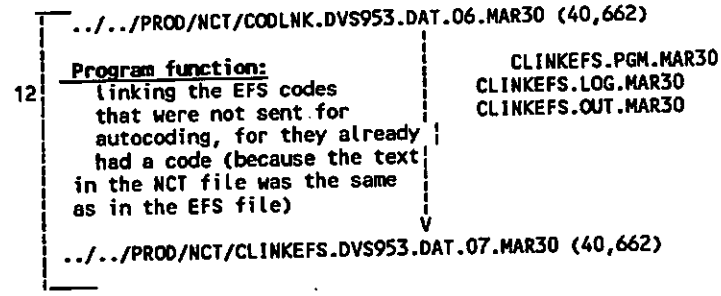
../../PROD/NCT/Q46MEF.DVS953.DAT.04.MAR18 (40,662)


../../PROD/NCT/Q46MEF.DVS953.DAT.04.MAR18 (40,662)
+
../../PROD/EFS/DVLANG.DVS953.DAT.04.MAR18 (9,986)
Program function:
creating a flag (EFSFLG) in the NCT file for those records that are in the EFS file (pos.669,1)
EFSFLG.PGM.MAR18
EFSFLG.LOG.MAR18
EFSFLG.OUT.MAR18
../../PROD/NCT/EFSFLG.DVS953.DAT.05.MAR18 (40,662)

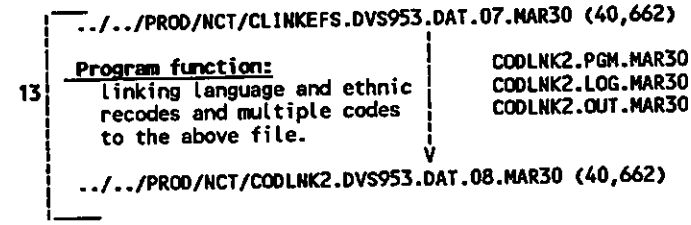
10


../../PROD/NCT/EFSFLG.DVS953.DAT.05.MAR18 (40,662)
Program function:
This program is linking codes from autocoding back in the NCT file.
CODLNK1.PGM.MAR30
CODLNK1.LOG.MAR30
CODLNK1.OUT.MAR30
../../PROD/NCT/CODLNK.DVS953.DAT.06.MAR30 (40,662)

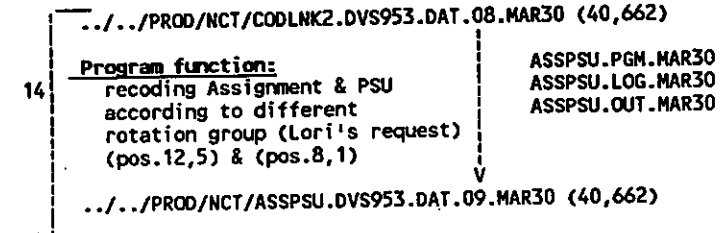
11


../../PROD/NCT/CODLNK.DVS953.DAT.06.MAR30 (40,662)
Program function:
linking the EFS codes that were not sent for autocoding, for they already had a code (because the text in the NCT file was the same as in the EFS file)
CLINKEFS.PGM.MAR30
CLINKEFS.LOG.MAR30
CLINKEFS.OUT.MAR30
../../PROD/NCT/CLINKEFS.DVS953.DAT.07.MAR30 (40,662)

12


../../PROD/NCT/CLINKEFS.DVS953.DAT.07.MAR30 (40,662)
Program function:
linking language and ethnic recodes and multiple codes to the above file.
CODLNK2.PGM.MAR30
CODLNK2.LOG.MAR30
CODLNK2.OUT.MAR30
../../PROD/NCT/CODLNK2.DVS953.DAT.08.MAR30 (40,662)

13


../../PROD/NCT/CODLNK2.DVS953.DAT.08.MAR30 (40,662)
Program function:
recoding Assignment & PSU according to different rotation group (Lori's request) (pos.12,5) & (pos.8,1)
ASSPSU.PGM.MAR30
ASSPSU.LOG.MAR30
ASSPSU.OUT.MAR30
../../PROD/NCT/ASSPSU.DVS953.DAT.09.MAR30 (40,662)

14

../../PROD/NCT/ASSPSU.DVS953.DAT.09.MAR30 (40,662)

Program function:
15 creating labour market derived variables from pos (790 to 804)

LBFORCE.PGM.MAR30
LBFORCE.LOG.MAR30
LBFORCE.OUT.MAR30

../../PROD/NCT/LBFORCE.DVS953.DAT.10.MAR30 (40,662)

../../PROD/NCT/LBFORCE.DVS953.DAT.10.MAR30 (40,662)

Program function:
16 weights are added to the Labour force file in this file (pos.925,8).

LBFORCE.WGTED.PGM.MAR30
LBFORCE.WGTED.LOG.MAR30
LBFORCE.WGTED.OUT.MAR30

../../PROD/NCT/LBFORCE.DVS953.WEIGHTED.DAT.11.MAR30 (40,662)

../../PROD/NCT/LBFORCE.DVS953.WEIGHTED.DAT.11.MAR30 (40,662)

Pgm Function:
17 (file produced by Norm Crampton on mainframe, to create ethnic DVs)

../../PROD/NCT/ETHDVS.DVS953.DAT.12.APR22 (40,662)

../../PROD/NCT/ETHDVS.DVS953.DAT.12.APR22 (40,662)

Pgm Function:
18 Picking Q46AAmt from the raw file (length 20 bytes) for in the Predit only 19 bytes of Q46AAmt were picked up and the 20th byte was garbage, in order to run Q46MEF.PGM again just for Q46AAmt in the next step.

FIXQ46.PGM.MAY02
FIXQ46.LOG.MAY02
FIXQ46.OUT.MAY02

../../PROD/NCT/FIXQ46.DVS953.DAT.13.MAY02 (40,662)

../../PROD/NCT/FIXQ46.DVS953.DAT.13.MAY02 (40,662)

Pgm Function:
19 Rerunning Q46MEF.PGM for Q46AAmt only and for related multiple flags.

Q46MEF.REV.PGM
Q46MEF.REV.LOG
Q46MEF.REV.OUT

../../PROD/NCT/Q46MEF.DVS953.DAT.14.MAY02 (40,662)

../../PROD/NCT/Q46MEF.DVS953.DAT.14.MARY02 (40,662)

Pgm Function:
Creating income derived

INCOME.PGM.MAY02

20 variables: wages, investment income
UIC, total income, etc.

INCOME.LOG.MAY02
INCOME.OUT.MAY02

../../PROD/NCT/INCOME.DVS953.DAT.15.MAY02 (40,662)

../../PROD/NCT/INCOME.DVS953.DAT.15.MAY02 (40,662)

Pgm Function:
Fixing language DVs and adding
21 the AGEGP (pos.666,2) and SEX
(pos.668,1) variables.
(fixed by Phil)

FIX.DVLANG.PGM
FIX.DVLANG.LOG
FIX.DVLANG.OUT

../../PROD/NCT/REVDVLANG.DVS953.DAT.15A.MAY04 (40,662)

../../PROD/NCT/REVDVLANG.DVS953.DAT.15A.MAY04 (40,662)

Pgm Function:
Cleaning income derived variables
like wages, self emplyment, UIC,
with ' ' instead of '.' and
Spiliting the file into 2 parts
the Labour force file and the
Special Population file.

INCOME.FIX1.PGM
INCOME.FIX1.LOG
INCOME.FIX1.OUT

../../PROD/NCT/DVS953.LFSAMP.DAT.16.MAY04 (32,696)

22 ../../PROD/NCT/PART1.MAY04 (10,000)
../../PROD/NCT/PART2.MAY04 (10,000)
../../PROD/NCT/PART3.MAY04 (10,000)
../../PROD/NCT/PART4.MAY04 (2,696)

../../PROD/NCT/SPECPOP.DVS953.DAT.17.MAY04 (7,966)

../../PROD/NCT/PART5.MAY04 (7,966)

LABOUR FORCE FILE

../../PROD/NCT/DVS953.LFSAMP.DAT.16.MAY04 (32,696)

RENAMED TO

../../PROD/NCT/DVS953.LFSAMP.DAT.14.APR08 (32,696)
(in order to stick with the file name SM people had already been told to use)

UPLOADED TO MAINFRAME AS

INPUT: "SPEC.ICT9311.NCT.LFSAMP.APR14.PUBREAD"
(recoding the HLDSIZE from blank to '0')

OUTPUT: "SPEC.ICT9311.NCT.LFSAMP.HHSIZE.FIX"

23 INPUT: "SPEC.ICT9311.NCT.LFSAMP.HHSIZE.FIX"
(recoding Q46, WAGES, SELF_EMP, & INVEST_INC)

OUTPUT: "SPEC.ICT9311.NCT.LFSAMP.RECODE.WAGES"

RENAMED ON MAINFRAME TO

"SPEC.ICT9311.NCT.LFSAMP.APR14.PUBREAD"

DOWNLOADED TO UNIX AS

../../PROD/NCT/DVS953.LFSAMP.DAT.16.MAY16 (32,696)

RENAMED TO

../../PROD/NCT/DVS953.LFSAMP.DAT.14.APR08 (32,696)
(in order to stick with the file name SM people had already been told to use)

NOTE: PSU RECODE WAS DONE ON LFSAMP FILE BUT NOTHING WAS RECODED, FOR THERE WAS NO PSU# ='85087' IN THE FILE TO BE RECODED. HOWEVER, THE STEP WAS DONE IN ORDER TO CHECK ITS EXISTENCE AND THERFORE DOCUMENTED IN REVPSU.PGM/REVPSU.LOG/REVPSU.OUT BUT WAS NOT CARRIED OUT (PROCESSED).

../../PROD/NCT/DVS953.LFSAMP.DAT.14.APR08 (32,696)

Pgm Function:

LABFRCE.DVCHECK.PGM
LABFRCE.LOG.MAY19
LABFRCE.OUT.MAY19

../../PROD/NCT/LABFRCE.DV.REVISED.DAT.17.MAY19 (32,696)

24

MOVED TO

../../OLDPROD/NCT/LABFRCE.DV.REVISED.DAT.17.MAY19.Z (32,696)

RENAMED TO

../../PROD/NCT/DVS953.LFSAMP.DAT.14.APR08 (32,696) <---------
(still sticking with the file name SM people had already been told to use)

SPECIAL POPULATION FILE

../../PROD/NCT/SPECPOP.DVS953.DAT.17.MAY04 (7,966)

RENAMED TO

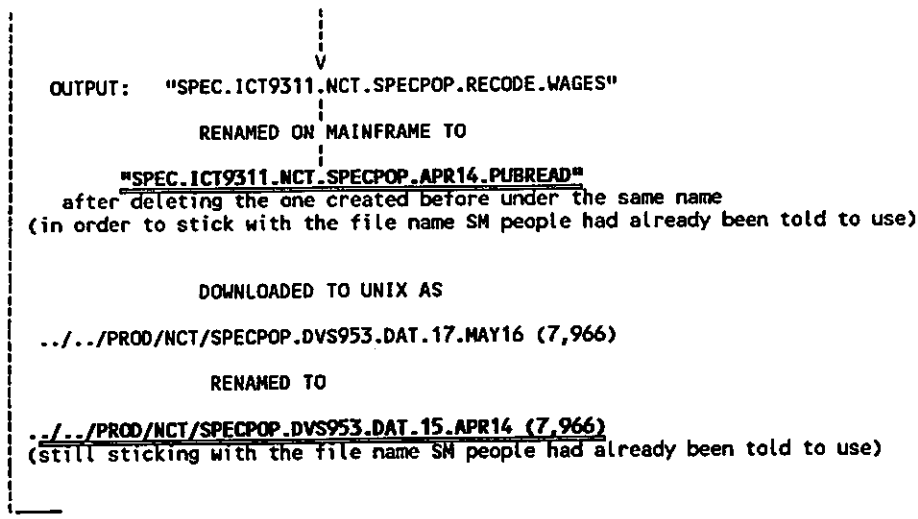
../../PROD/NCT/SPECPOP.DVS953.DAT.15.APR14 (7,966)
(in order to stick with the file name SM people had already been told to use)

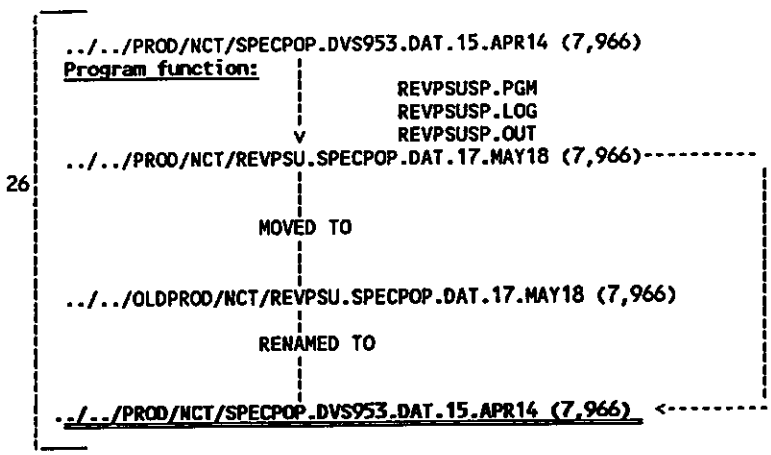
UPLOADED TO MAINFRAME AS

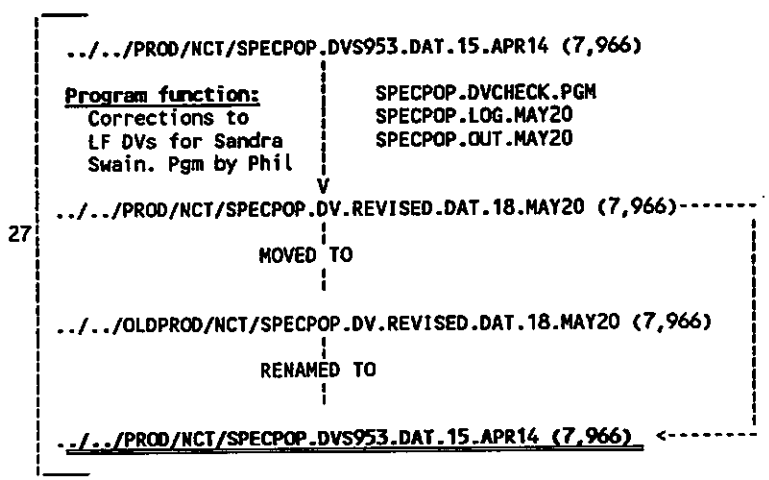
INPUT: "SPEC.ICT9311.NCT.SPECPOP.APR14.PUBREAD"
(recoding the HLDSIZE from blank to '0')

25 OUTPUT: "SPEC.ICT9311.NCT.SPECPOP.HHSIZE.FIX"

INPUT: "SPEC.ICT9311.NCT.SPECPOP.HHSIZE.FIX"
(recoding Q46, WAGES, SELF_EMP, & INVEST_INC)


OUTPUT: "SPEC.ICT9311.NCT.SPECPOP.RECODE.WAGES"
RENAMED ON MAINFRAME TO
"SPEC.ICT9311.NCT.SPECPOP.APR14.PUBREAD"
after deleting the one created before under the same name
(in order to stick with the file name SM people had already been told to use)
DOWNLOADED TO UNIX AS
../../PROD/NCT/SPECPOP.DVS953.DAT.17.MAY16 (7,966)
RENAMED TO
../../PROD/NCT/SPECPOP.DVS953.DAT.15.APR14 (7,966)
(still sticking with the file name SM people had already been told to use)


../../PROD/NCT/SPECPOP.DVS953.DAT.15.APR14 (7,966)
Program function:
REVPSUSP.PGM
REVPSUSP.LOG
REVPSUSP.OUT
../../PROD/NCT/REVPSU.SPECPOP.DAT.17.MAY18 (7,966)
26
MOVED TO
../../OLDPROD/NCT/REVPSU.SPECPOP.DAT.17.MAY18 (7,966)
RENAMED TO
../../PROD/NCT/SPECPOP.DVS953.DAT.15.APR14 (7,966) <---------


../../PROD/NCT/SPECPOP.DVS953.DAT.15.APR14 (7,966)
Program function:
Corrections to
LF DVs for Sandra
Swain. Pgm by Phil
SPECPOP.DVCHECK.PGM
SPECPOP.LOG.MAY20
SPECPOP.OUT.MAY20
../../PROD/NCT/SPECPOP.DV.REVISED.DAT.18.MAY20 (7,966)
27
MOVED TO
../../OLDPROD/NCT/SPECPOP.DV.REVISED.DAT.18.MAY20 (7,966)
RENAMED TO
../../PROD/NCT/SPECPOP.DVS953.DAT.15.APR14 (7,966) <---------