REPORT ON A MISSION

TO RWANDA

18 April -- 3 May 1983

Prepared by: James Otto

Under contract to U.S. AID

Contract number: 698 0662.01-C-00-3004-00

Project number: 698-0662.01, FHI

Project name: Rwanda Fertility Survey

Table of Contents.

1. Introduction and general description of the mission.

The purpose of the mission was to assess the data processing facilities in Rwanda and to develop a plan for the data processing of the National Fertility Survey to be conducted by the National Population Office (ONAPO) later this year. Shortly after my arrival, I was also asked to review the data processing requirements of the National Agricultural Survey which will be conducted from August 1983 to July 1984.

Data processing has a very short history in Rwanda. The first computer installed in the country was an NCR Century 8250 purchased by UNFPA for the processing of the 1978 population census. The machine was delivered at the end of 1978, but was not completely operational until February 1980 because of delays in preparing the computer room and lack of experienced technicians. The first year of operation was marked by a multitude of hardware, software and electricity problems which were finally resolved by the end of 1980.

The only computer application to date has been the 1978 census. The computer was used for on-line entry and validation of the census during 1980 and 1981. Tabulation began in 1982 and is scheduled for completion at the end of May 1983. System design and programming for the census were accomplished by three different U.N. experts assisted by a Rwandan counterpart. COCENTS was installed in March 1980 by a team from Delta Systems Consultants under contract to U.S. AID. At that time, five Rwandans and one of the U.N. experts were trained to use COCENTS.

It has always been the intention of UNFPA to transfer control and responsibility for the computer to the Rwandan government, but no specific arrangements were ever made for effecting the transfer. During my stay, a TELEX was received from UNFPA headquarters announcing that the computer would be transferred to the government at the end of May. Until now, the government has made no formal arrangements for the operation and ma ntenance of the computer. The data entry operators for the census have already been laid off and assignments for the remaining census data processing staff have not been determined.

A meeting was held under the direction of the Director of the National Statistics Office, Ministry of Planning on April 26 to discuss the data processing requirements of potential users of the government computer. There are four surveys which could use the computer facilities in the coming year: the National Agricultural Survey to be conducted by the Ministry of Agriculture; the National Fertility Survey to be conducted by ONAPO; the 1981 census post-enumeration survey (also called the demographic survey) carried out by the Statistics Office; and a budget survey to be done later this year, also by the Statistics Office.

From the rough estimates of data processing requirements described in the meeting, the computer should be able to satisfy the requirements of the four surveys, but additional memory and terminals may be required. The census post-enumeration survey is the unknown factor. This was a large survey (a 10% sample of the population or about 500,000 people) and it will require considerable computer resources. Financing for this survey is in doubt and there is a possibility that the questionnaires will have to be coded again.

Moreover, no one has been assigned to do the system design and programming. The other surveys have more modest requirements: the agricultural and budget surveys are multi-round surveys which will collect and process data over the course of a year; the fertility survey will require intensive data entry operations for a two or three month period followed by reduced requirements for background processing (editing, recoding and tabulation). The agricultural and fertility surveys are both in a good position to use the computer, since their data processing systems have already been designed and in the case of the agricultural survey, programming has started.

The government has an excellent opportunity to establish a national statistical computing center, but it must act quickly to staff and organize the center or chaos may result. Policies should evolve as processing requirements increase, but basic administrative and technical procedures need to be established early to allow the sharing of computer facilities and costs between multiple users. The principal issues to be resolved are:

1. scheduling of data entry operations;
2. submission and scheduling of production jobs;
3. provision of program development facilities - entry of programs, compilation and testing;
4. allocation and security of data files stored on disk;
5. allocation of maintenance and operational costs between users in a reasonable and equitable way.

The computer center will require an effective administrator, backed up by a qualified technical advisor. These people should be identified soon and if they are not available, should be requested as part of a technical assistance project to develop the computer center.

I would like to thank everyone at ONAPO, the Census Bureau, the Agricultural Survey and AID, who by their warm welcome and co-operation were of great assistance in the work of this mission. Special thanks are due Alain Mouchiroud, for his generous hospitality and willingness to help in any way possible.

Appendix 1 - List of mission contacts.

U.S. AID

Eugene Chiavaroli, AID Affairs Officer
Alan Getson, Population Program Officer
Brice Lerner, Administrative Officer

National Population Office (ONAPO)

Madame Gaudence Habima Nyirasafari, Director
Alain Mouchiroud, UNFPA expert in demography
Christophe Boneza, demographer
Monique Mukamwanzi, demographer

Census Bureau (BNR)

Silas Nyibizi, Director
Gervais Condo, Assistant Director
Carlos da Costa de Carvalhao, UNFPA expert in demography
Bernard Ntabahwana, programmer
Festus Mutarasisi, programmer

National Agricultural Survey

Serge Rwamasirabo, Director
Yvon Dejaegher, Director of field work
Jean-Marie Sehene, programmer

Other contacts

Ephrem Twagirayizu, Director of Statistics Office,
Ministry of Planning
Joseph Thiobault, SOMECA technical representative

Appendix 2 - Description of computer facilities at the Census Bur

Manufacturer: NCR Model: Century 8250

Memory size: 128K bytes

Operating system: IMOS III (uses approximately 34K bytes)

Peripheral equipment:

7 video display units
3 pairs of fixed/removable disk platters
(capacity = 4.9 MB/platter, for a total of 29.4MB)
2 tape cassette drives for programs and file backup
1 300 line/minute printer

Software:

programming language: COBOL
other software: text editor
statistical package: COCENTS

Principal use: processing of 1978 census.

Manufacturer: CROMEMCO Model: System 2 (Z-2D)

Memory size: 64K bytes

Operating system: CROMEMCO DOS

Peripheral equipment:

1 Haseltine video display unit
2 5 1/4" floppy disk drives (double sided, double density)
1 CROMEMCO character printer

Software:

programming languages: Z-80 assembler, structured BASIC,
FORTRAN IV with RATFOR, COBOL
packages: word processing, data base, text editor

Principal use: training;
interface with NCR 8250 for data transfer.

II. Proposal for the data processing of the National Fertility Survey.

A. Background.

The National Fertility Survey will be conducted by the National Population Office (ONAPO) between August and November, 1983. The survey is designed to descibe the levels, trends and determinants of human fertili y in the country and to provide other important demographic information. The questionnaires are modelled after the World Fertility Survey questionnaires used in fertility surveys in 42 developing countries between 1974 and 1984, and consist of a household schedule, a detailed individual questionnaire for all women ages 15 to 49 and questionnaire for a sub-sample of the husbands of interviewed women. The sample consists of 100 sampling units (districts) and is designed to produce an average of 60 households per sampling unit with an average of one eligible woman per household. The husband's questionnaire will be administered to a 25% sample of the husbands of interviewed women who are currently in a monogamous union. The overall sampling rate of husbands is expected to be 12.5% of interviewed women, so the sample will comprise 6,000 households, 6,000 individual respondents and 750 husbands.

The data processing system proposed in this report follows the recommendations of the World Fertility Survey publication, "Data Processing Guidelines", (Basic Documentation, No. 11, Volumes 1 and 2, May 1980). It deviates from standard WFS practices for machine editing to meet the objective of producing the basic survey tabulations within four months of the completion of field work. This proposal assumes that the government computer center described above will be operational and accessible by the time processing is to begin (August, 1983). It further assumes that software developed by the World Fertility Survey will be available for use on both the WFS headquarters computer in London and the NCR computer at the government computer center.

B. Data preparation.

1. General.

Data preparation requires manual processing of the questionnaires, followed by machine processing to convert the responses in the questionnaires to a format suitable for computer processing. Manual processing will be done in the fertility survey office which should have sufficient space for the personnel required and storage of the questionnaires. The survey office should be close to the computer center to facilitate the delivery of questionnaires and resolution of administrative and technical problems.

The survey office staff should consist of a minimum of 5 editors/coders, 2 controllers/correctors and a supervisor. The office staff should attend the interviewers training course to insure they thoroughly understand the questionnaires. This training will be supplemented by specific training in procedures for office processing, data entry and error correction. As part of their training, they should transcribe 50-100 questionnaires from the pilot survey to the final version of the questionnaire. These questionnaires will subsequently be edited, coded and used for testing the computer programs developed for the survey. The procedures for office

processing will be published in a manual which should be prepared as soon as the final version of the questionnaire and the basic coding instructions are available.

Data preparation is scheduled to proceed in parallel with the field work, beginning two weeks after the field work begins and ending four to six weeks after it is completed. This is an ambitious, yet feasible, schedule which requires thorough training before the work begins and full time supervision while it is underway. The schedule is based on a processing rate of 15 households per person per day. If this rate is not achieved, extra staff must be assigned. This can be accomplished by either training more people than are required initially or by re-assigning interviewers to the office staff, when they have finished the field work.

2. Office processing.

Office processing will be done in two steps: receipt and control of the questionnaires as they arrive from the field and editing and coding of the questionnaires for data entry. As questionnaires are received from the field they will be organized by district and sequenced by household. The individual and husbands questionnaires will be sequenced by the line number of the woman and inserted inside the household schedule. At the same time, the count of complete and incomplete household, individual and husband's questionnaires is compared to the totals recorded by the field work supervisor. After all discrepancies are resolved, the totals are recorded on the district summary form and the household list containing the result of the interview for each household surveyed is completed. The household lists are used to update the district file which is stored in the computer to check data entry and prepare periodic data preparation status reports.

When the receipt and control procedures are finished, the questionnaires are passed (by district) to the editing and coding step. Editing is similar to the verification done by the field supervisors and its purpose is to verify that all questionnaires have been correctly filled in. Special attention should be paid to verifying that skip instructions have been followed and that the filter questions which control the skips are consistent with the previously recorded responses. The questionnaire is largely pre-coded, so the coding of open-ended and certain other questions will be done at the same time as the editing. Coding instructions will be prepared for each question to be coded and will show the correspondence between responses and the numeric codes to be recorded in the boxes printed on the questionnaire.

3. Data entry.

Data entry will be done on visual display units (terminals) directly attached to the computer and will be controlled by programs written specifically for the fertility survey questionnaires. There are three stages of data entry: district file creation and update, entry of the questionnaire data and verification of the questionnaire data. The district file is created from the district summary form and is updated with the information in the household lists prepared during the control of the questionnaires. The file contains sampling information and the processing status of each household in the sample. The file is created and updated directly using a video display unit.

Its contents are verified visually by displaying it on the terminal or listing it on the printer. The file is used during the entry of the questionnaire data to verify the questionnaire identification (district and household numbers) and to record the data entry status of the household. Detailed or summarized data entry status reports can be prepared by listing the district file.

The questionnaire data for a district can only be entered after the district file has been updated with the household list for the district. The district and household numbers are entered first and checked against the district file. If the household number does not appear in the district file or if it has already been entered, the household will be rejected. After the identification, the household schedule data are entered, followed by the individual and husband's questionnaires, where applicable, in sequence by the line number of the woman. There is a possibility that structure checks (see machine editing below) and checks to verify the skip pattern of the questionnaire will be incorporated in the data entry program to guarantee that the data entered is essentially consistent.

The questionnaire data entered will be stored by district. When the entire district is entered, the data will be verified by entering the questionnaire data again and comparing it to the data stored in the computer file. Discrepancies will be resolved by examining the questionnaire and modifying the originally entered data when necessary. Districts which have been completely entered and verified will be consolidated periodically (minimum of once a week) and added to the raw data file, which will eventually contain all the data from the survey.

C. Machine editing imputation and recoding.

1. General.

The World Fertility Survey data processing guidelines recommend thorough machine editing to detect inconsistencies in the data. Questionnaires with inconsistencies are corrected manually by reference to the original questionnaires and coding corrections to be applied to the raw data file. The objective of machine editing is to produce highly consistent data that is suitable for sophisticated demographic and statistical analysis. This process is time consuming (an average of 15 months per survey) and a recent evaluation of its benefits suggests that little is gained in relation to the time spent. The initial analytical needs of the Rwanda survey are less extensive than typical WFS surveys and there is a need to produce results as quickly as possible, so I propose that most of the machine editing be replaced by systematic correction of those variables considered essential for presenting the general results of the survey. Any inconsistencies detected will be resolved by using pre-defined rules which assign either a reasonable default value or a not-stated value to the variables involved. The extent of the automatic correction will be reported by printing the number of corrections made and the variables affected. The facility for manual correction will be retained by allowing the printing of questionnaires with errors as an option in the automatic correction programs.

2. Structure edit.

The structure edit verifies that all the required data, and only the required data, is present for all the questionnaires belonging to a household. The data in the questionnaire are organized into 13 record types, each of which contains different information. The structure edit program will verify that the necessary records are present and that there is consistency between the household, individual and husband's questionnaires. Errors will be corrected manually by reference to the questionnaires. If the structure checks are incorporated in the data entry programs, the structure edit program will provide independent confirmation that the data are structurally correct and acceptable for subsequent processing. When the raw data file is structurally correct, marginal distributions of all questions will be printed to prepare the editing and automatic correction rules for resolving inconsistencies.

3. Date editing, imputation and recoding.

The individual questionnaire contains the dates of all the respondent's pregnancies and unions. These dates are necessary for the construction of variables important to demographic analysis, but to be useful the dates must be consistent and stored in a standard format. WFS has developed a generalized program for the editing, imputation and recoding of the dates contained in WFS questionnaires. This program does a comprehensive consistency edit of the pregnancy and union histories. In addition, it has facilities for the imputation of missing or incomplete dates and the construction of many of the variables needed for the standard WFS tabulations. The variables constructed by the program are in a standard format which will permit the use of the standard WFS tabulation package and facilitate the preparation of tables needed for the initial survey report.

To use the date edit program, the date information must be extracted from the individual questionnaires and converted to a standard format. The program to accomplish this, called the date extract program, will also perform some basic consistency checks of the date information and make automatic corrections for any inconsistencies detected. Examination of the correction statistics printed by the date extract and date edit programs and the marginal distributions of the date recode variables should give a general idea of data quality and the need for any further editing.

4. Other editing, automatic correction and recoding.

Many of the variables required for analysis are not recorded directly, but must be constructed from one or more of the variables in the questionnaire. To simplify tabulation and subsequent analysis, these variables are constructed by a program and stored in a recode file which is then used for all further processing. To simplify inter-country comparisons and the use of standard analytical programs, WFS prescribed a standard format for the recode variables of all surveys (described in WFS Data Processing Guidelines, Volume 2, pages 169-199).

The WFS standard recode format should be used for the same reasons given above for using the WFS data edit program. The recode file will be created from the raw data file and the date recode

variables. It may not be necessary to create all the standard variables and some Rwanda specific variables will be required, but the exact variables will not be determined until the tabulation requirements are defined. Once the tabulation plan is available, specifications for recoding the required variables will be written. The specifications will include instructions for the detection and automatic correction of any inconsistencies involving the recode variables. The recode program, like the date extract program, will produce a summary of the number of corrections made and the variables affected.

D. Tabulation.

There are no tabulation specifications yet, but the initial tables produced should be a subset of the standard tables proposed for inclusion in the WFS first country reports (see WFS Guidelines for Country Report No. 1, Basic Documentation No. 8, December, 1977). This will permit the use of the WFS tabulation package, the COCENTS generator, and the standard table specifications, headings and variable labels associated with it. It will also facilitate production of additional tables after the basic survey report is completed. Tables from the household schedule can be done with COCENTS.

National Fertility Survey

Appendices

1. Questionnaire description and size estimates.
2. Data processing system flowchart.
3. Data file summary.
4. Program summary.
5. Data processing estimates.
6. Data processing calendar.
7. Data processing work plan.
8. Technical assistance proposal.

National Fertility Survey

Appendix 1 - Questionnaire description and size estimates.

The information in the questionnaires is divided for data entry purposes into 13 record types which are described below. Households are identified by sampling unit and household number. Household members are further identified by their line number on the household schedule. Both the individual and husband's questionnaires are identified by the woman's line number.

Record Type	Contents of record	Characters per section/record
10	Identification and cover page	17
11	Household member	12
20	Individual questionnaire cover page	7
	Section 1 - Respondent's background (Q101-Q120)	26
	Section 2 - Maternity summary (Q201-Q227)	29
		62
30	Section 2 - Pregnancy history (Q228-Q235)	13
40	Section 2 - end (Q236-Q242)	7
	Section 3 - Union history (Q301-Q328)	50
		57
41	Section 3 - Former unions (Q329-Q333)	10
50	Section 4 - Contraceptive knowledge/use (Q401-Q430)	48
61	Section 5 - Open interval (Q515-Q532)	31
62	Section 5 - Closed interval (Q535-Q553)	39
63	Section 5 - Open/closed interval (Q554-Q566)	30
70	Section 5 - end (Q567-Q591)	29
	Section 6 - Work history (Q601-Q627)	34
	Section 7 - Husband's background (Q701-Q719)	21
	Section 8 - Appreciation of demographic situation in Rwanda (Q801-Q810)	18
		102
80	Husband's questionnaire cover page	7
	Section 1 - Husband's background (H101-H119)	24
	Section 2 - Marriage and fertility (H201-H218)	31
		62
81	Section 3 - Contraceptive knowledge/use (H301-H317)	26
	Section 4 - Fertility regulation (H401-H427)	35
	Section 5 - Appreciation of demographic situation in Rwanda (H501-H511)	19
		80

Data entry volume and file size estimates.

Record type	Average per questionnaire	Characters to be entered	Characters in raw data file [1]
10	1	17	21
11	8	(12) 96	(22) 176 [2]
20	1	62	72
30	5	(13) 65	(25) 125
40	1	57	67
41	0.25	(10) 2.5	(22) 5.5
50	1	48	58
61	0.75	(31) 23.25	(41) 30.75
62	0.75	(39) 29.25	(49) 36.75
63	0.25	(30) 7.5	(40) 10
70	1	102	112
80	0.125	(62) 7.75	(72) 9
81	9.125	(80) 10	(90) 11.25
Total	20.25	520	735

Total data to be entered = 6000 x 520 = 3,168,000 characters

Size of raw data file = 6000 x 735 = 4,410,000 characters

[1] The difference between characters entered and characters in the raw data file consists of identification fields which are only entered at the beginning of the questionnaire, but are stored in each data record.

[2] The numbers in parentheses are the number of characters in one record.

National Fertility Survey

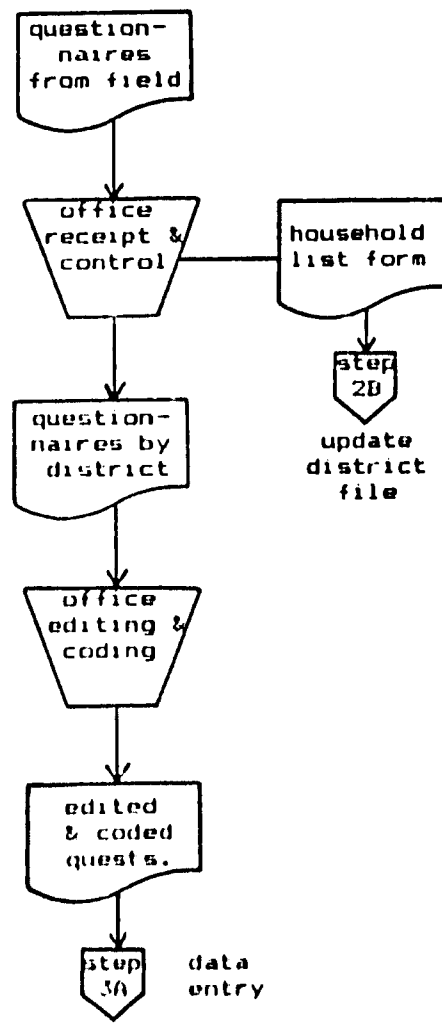
Appendix 2 - Data processing system flowchart.

1. Office processing.

A. Receipt and control.

By district

1. Sort questionnaires by household and line number.
2. Count households, individual and husband's questionnaires.
3. Compare to district control sheet and resolve differences.
4. Record totals on district summary form.
5. Complete household list form.



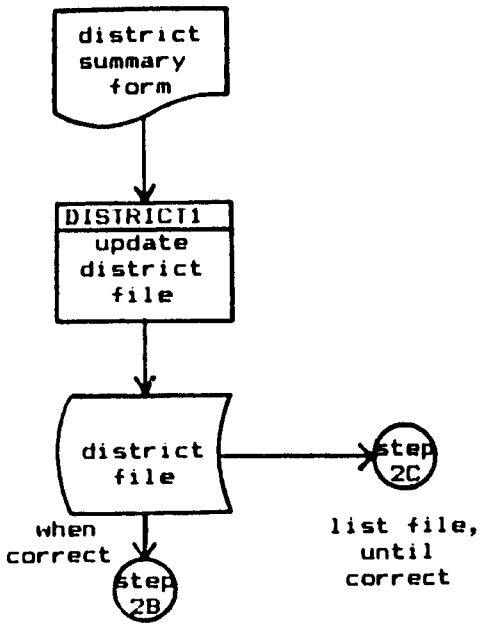
B. Office editing and coding.

1. Verify contents of questionnaire.
2. Code open-ended and and other questions.

2. District file system.

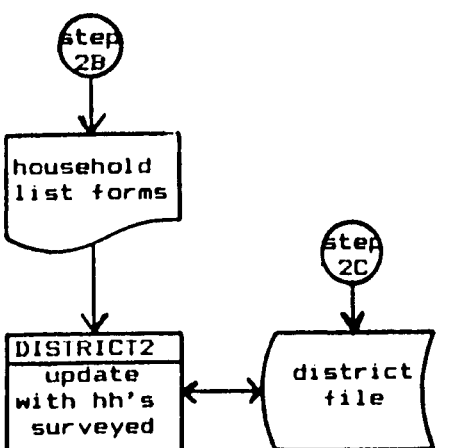
A. Create/update district file.

1. Create district file from district summary form.
2. List file and verify manually.
3. Correct errors.



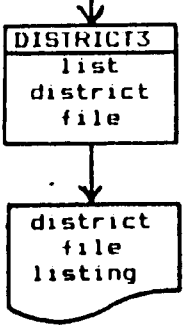
B. Update district file with households surveyed.

1. Accumulate household lists until sufficient number for entry.
2. List file and verify manually.
3. Correct errors.
4. Must be done before questionnaire data entry for district.



C. List district file.

1. For manual verification.
2. To show data entry status.



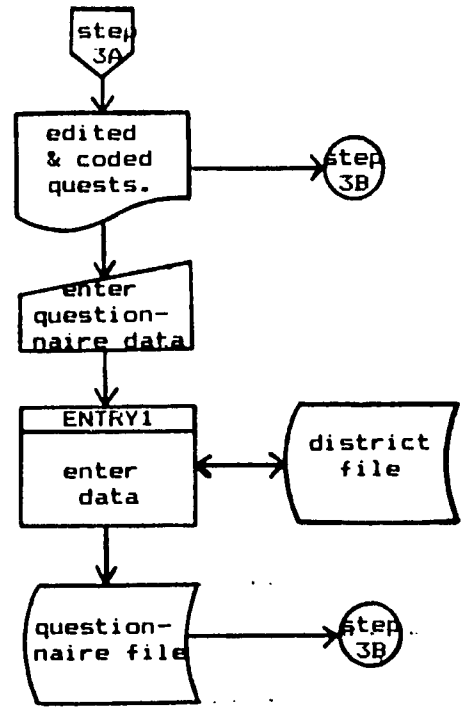
3. Data entry and consolidation.

A. Data entry by district.

1. Verify district.

2. Verify/update district file for each household entered.

3. Create district questionnaire file.

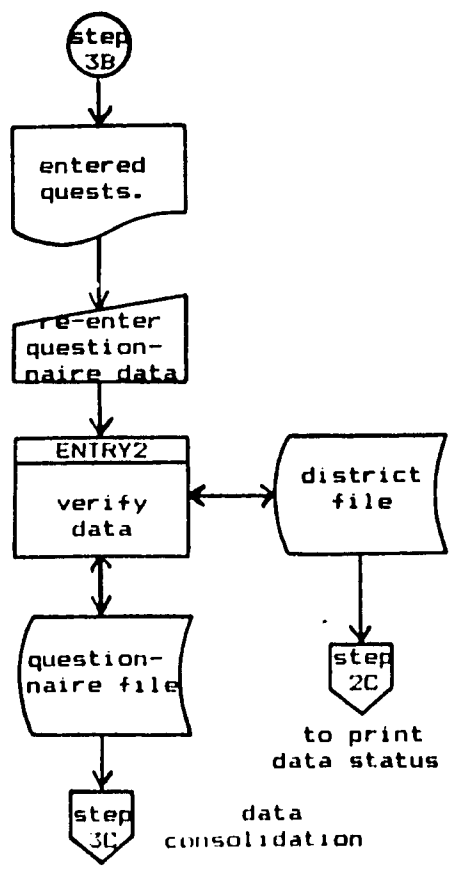


B. Verify data by district.

1. Re-enter data.

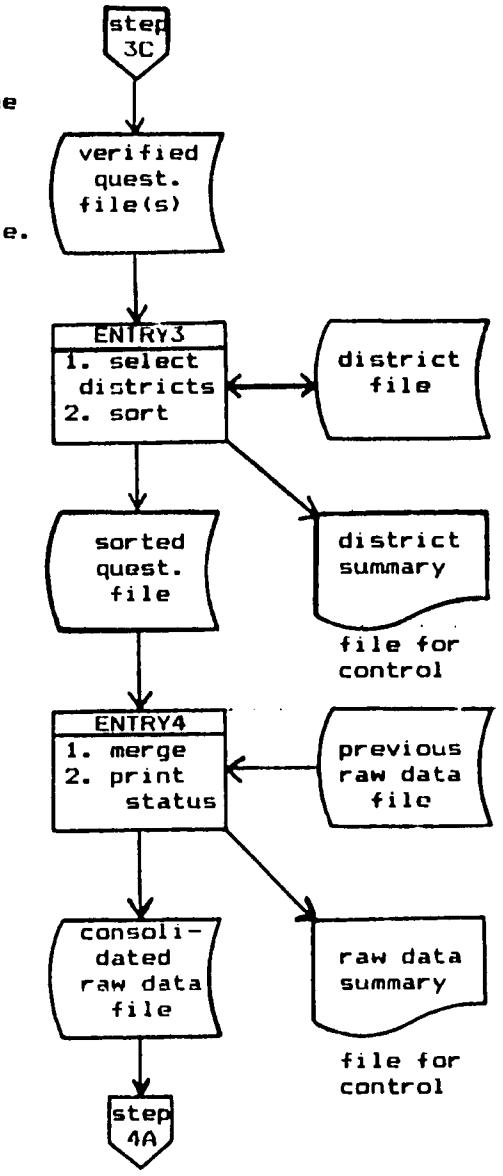
2. Compare to original and resolve errors.

3. Accumulate statistics.



C. Data consolidation.

1. Select verified districts.
2. Sort by district, household, line record type, sequence.
3. Print district summary.
4. Merge with previous raw data file.
5. Print raw data summary.



17

4. Structure edit and corrections.

A. Structure edit.

1. Test for presence of correct records.

2. Print error questionnaires.

B. Error corrections.

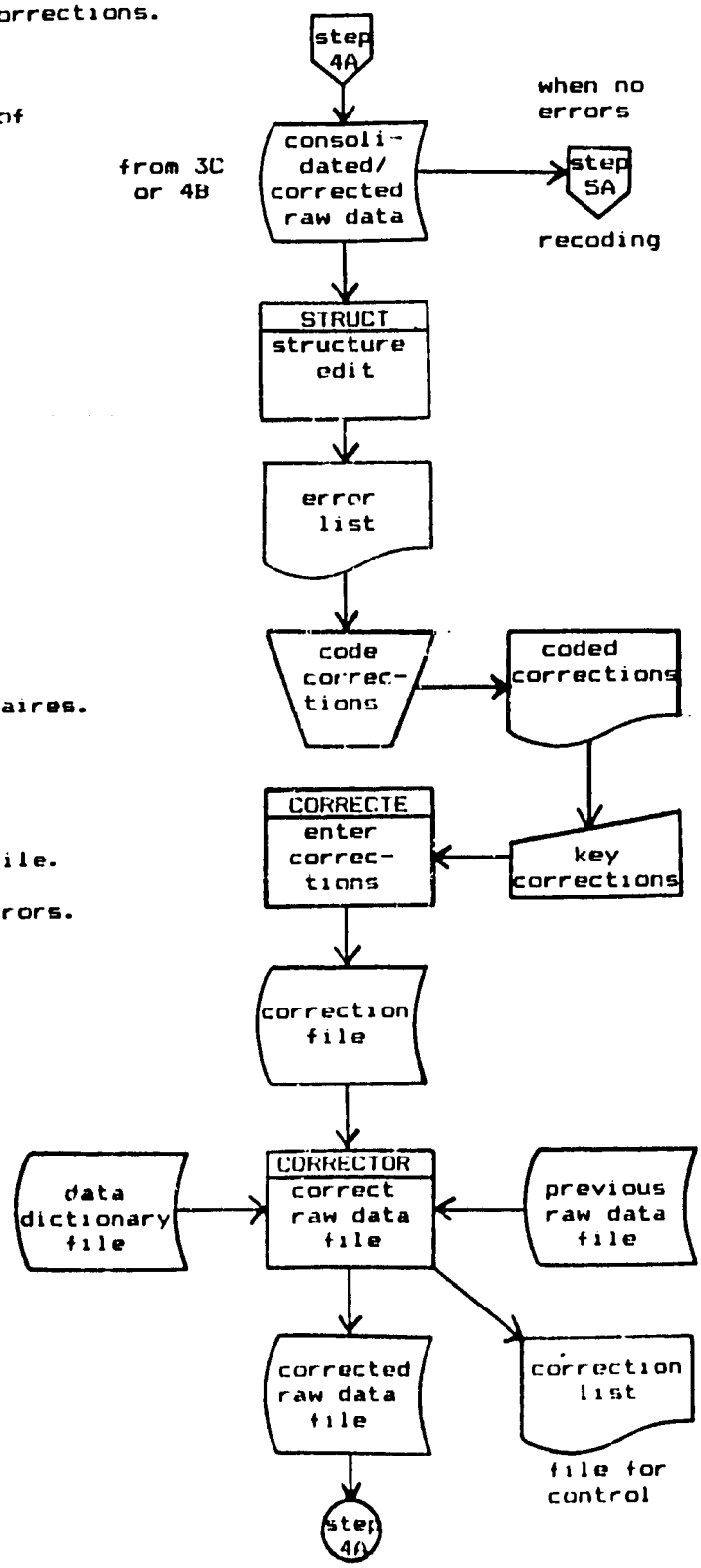
1. Retrieve questionnaires.

2. Code corrections.

3. Enter corrections.

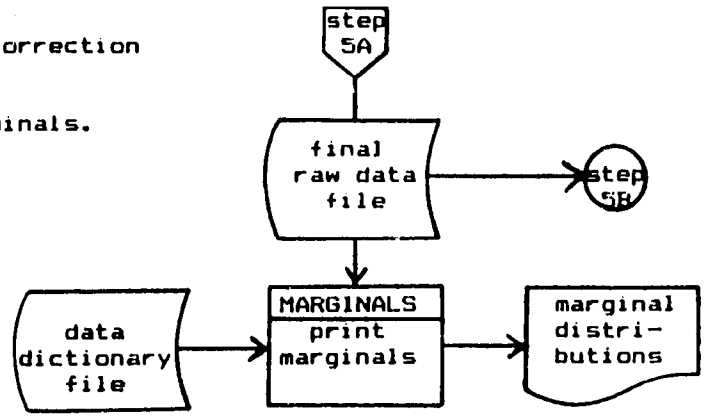
4. Correct raw data file.

5. Repeat until no errors.



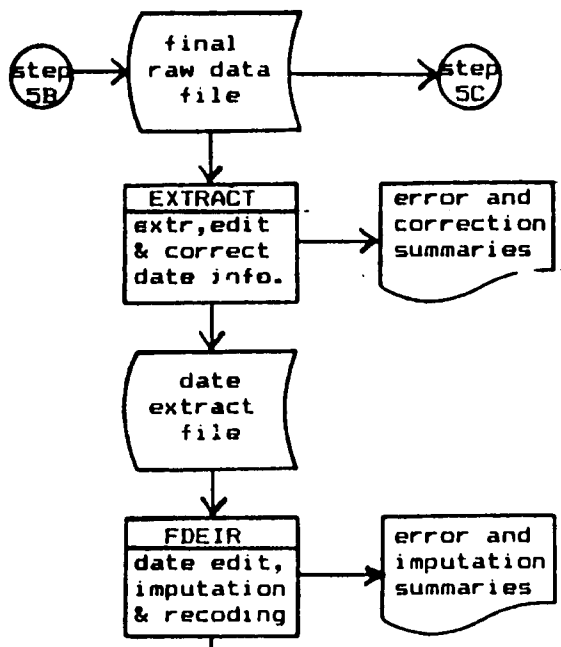
5. Editing, automatic correction and recoding.

A. Print raw data marginals.



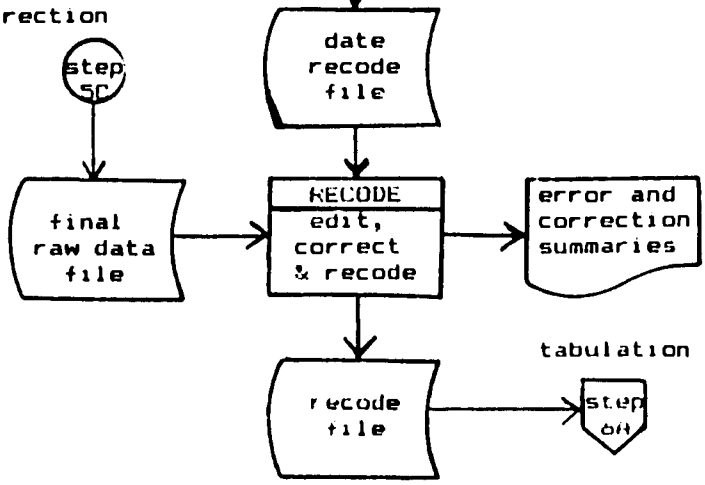
B. Date editing, imputation and recoding.

1. Extract date information.
2. Edit and correct date format errors.
3. Edit and impute dates.
4. Recode dates.
5. Print error and correction/imputation summaries.



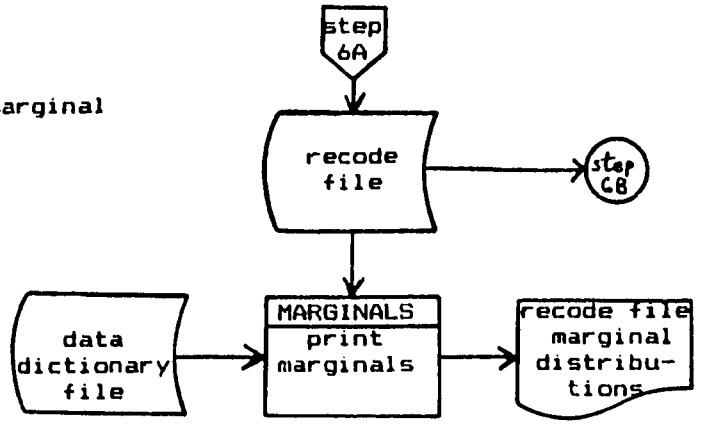
C. Edit, automatic correction and recoding.

1. Edit and correct data.
2. Recode data.
3. Print error and correction summaries.



6. Tabulation.

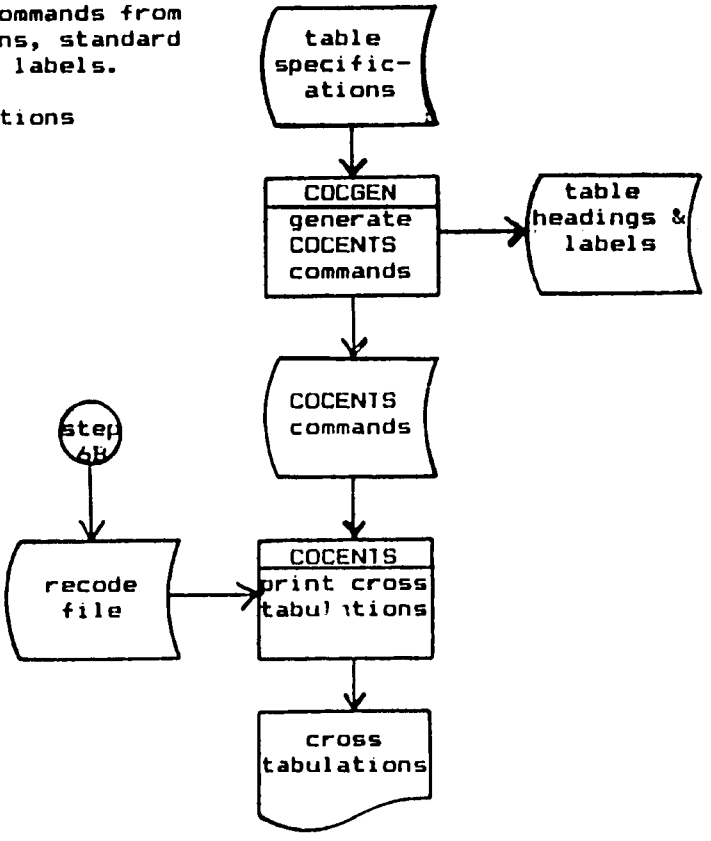
A. Print recode file marginal distributions.



B. Print cross-tabulations.

1. Generate COCENTS commands from table specifications, standard table headings and labels.

2. Print cross-tabulations with COCENTS.



National Fertility Survey

Appendix 3 - Data file summary.

1. District file.

This is an indexed sequential file containing one record for each sampling unit (district). The file contains master sampling information, interview and data entry status for each household in the district. The file is used to verify the questionnaire identification and to control data entry. Record size is 200-250 characters and there are 100 records.

2. Questionnaire file.

This is a data entry work file. A file is created by the data entry program for each district entered. It is read (and modified where required) by the data verification program. After verification, the data is copied to the raw data file and the questionnaire file for the district is deleted. This is a sequential file with variable length records. The average size of a district file is 1200 records (42,000 characters).

3. Raw data file.

This is a sequential file with variable length records used to accumulate the questionnaire data after it has been entered and verified. Data are added to the file by the data entry consolidation program and updated by the correction program. The final file will contain approximately 120,000 variable length records (21-112 characters) and approximately 4.4 million characters. Since this is a sequential file, three generations of the file will be maintained on separate disks for security.

4. Correction file.

This is a sequential file with variable length records containing corrections for errors detected in the structure edit. This is a temporary file and will contain a maximum of 4,000 variable length records (25-400 characters) and 40,000 characters.

5. Data dictionary files.

There are two sequential dictionary files to define the raw data and recode data files. These files are used to document the data files and to define their contents for generalized software. The files contain fixed length records (80 characters) and will have approximately 2000 records each (160,000 characters).

6. Date extract file.

This is a sequential file with variable length records used as a work file by the date edit, imputation and recode program. The file will contain approximately 15,000 records (24-432 characters) and approximately 1.4 million characters.

7. Date recode file.

This is a sequential file with fixed length records containing date dependent recode variables created by the date edit, imputation and recode program. The file contains 6000 records of 190 characters each or 1.2 million characters.

8. Recode file.

This is a sequential file with fixed length records created from the date recode and raw data files by the recode program. This file is used for tabulation and analysis. It contains 6000 records of 512 characters each or 3.1 million characters.

9. Table specification and library files.

These are sequential files containing the table specification and headings/labels for use by the COCENIS generator tabulation package. Both files have fixed length records (80 characters) and will contain approximately 6,000 records (480,000 characters) each.

National Fertility Survey

Appendix 4 - Program summary.

1. Programs to process the district file.

   A. DISTRICT1 - create district file from district summary.

   B. DISTRICT2 - update district file with households surveyed.

   C. DISTRICT3 - list district file.

2. Data entry and consolidation.

   A. ENTRY1 - enter questionnaire data.

   B. ENTRY2 - verify questionnaire data, resolve errors.

   C. ENTRY3 - extract data from verified questionnaire files.

   D. ENTRY4 - consolidate entered data, print data entry status.

3. Structure edit.

   A. STRUCT - check questionnaire structure, print errors.

   B. CORRECTE - enter corrections for structure errors.

   C. CORRECTOR - correct raw data file.(1)

4. Editing, automatic correction and recoding.

   A. EXTRACT - extract, edit and correct date information.

   B. FDEIR - date edit, imputation and recoding.(1)

   C. RECODE - edit, correct data, create recode file.

5. Tabulation.

   A. MARGINALS - print marginal frequency distributions.(1)

   B. COCGEN - COCENTS generator: create COCENTS commands from simplified table specifications.(1)

   C. COCENTS - generalized cross-tabulation package.(2)

(1) Generalized program or package available from World Fertility Survey.

(2) Already installed at Census Bureau.

National Fertility Survey

Appendix 5 - Data processing estimates.

1. Office processing.

6000 questionnaires @ 15 quest./person/day = 400 person days

400 person days / 5 editors/coders = 80 days (4 months)

Note: If the rate of processing is less than 15 questionnaires/day additional editors/coders will have to be added.

2. Data entry and verification.

A. District file creation and update.

24,000 characters @ 4,000 characters/hour = 6 hours (1 day)

B. Questionnaire data entry and verification.

3,200,000 characters @ 4,000 characters/hour = 800 person hours

800 person hours x 2 (entry and verification) = 1,600 person hours

1,600 person hours @ 30 hours/person/week = 53 person weeks

53 person weeks / 4 operators = 13 weeks (3 months)

3. System design and programming.

A. Data entry and structure edit.

Activity	Elapsed time
Prepare codebook and data dictionary.	1 week
Design district file programs.	1 week
Design data entry programs.	2 weeks
Design structure edit program.	1 week
Write, test and document district file programs.	1 week
Write, test and document data entry programs.	4 weeks
Write, test and document structure edit program.	2 weeks
Modify, install and test correction program.	2 weeks
Train correctors.	1 week
Modify, install and test marginals program.	2 weeks
Total	12 weeks

Note: Actual time estimates are not given because there is considerable overlap between activities. As a result, the total is not the sum of the individual estimates.

B. Editing, automatic correction, recoding and tabulation.

Activity	Elapsed time
Prepare editing, automatic correction and recoding specifications.	4 weeks
Prepare tabulation specifications.	1 week
Modify tabulation libraries.	2 weeks
Write, test and document date extract program.	2 weeks
Modify, install and test data edit, imputation and recode program.	2 weeks
Write, test and document edit, automatic correction and recode program.	3 weeks
Modify, install and test CODENIS generator program.	2 weeks
Test tabulation specifications.	3 weeks
Total	12 weeks

4. Computer time requirements.

A. Program compilation and testing.

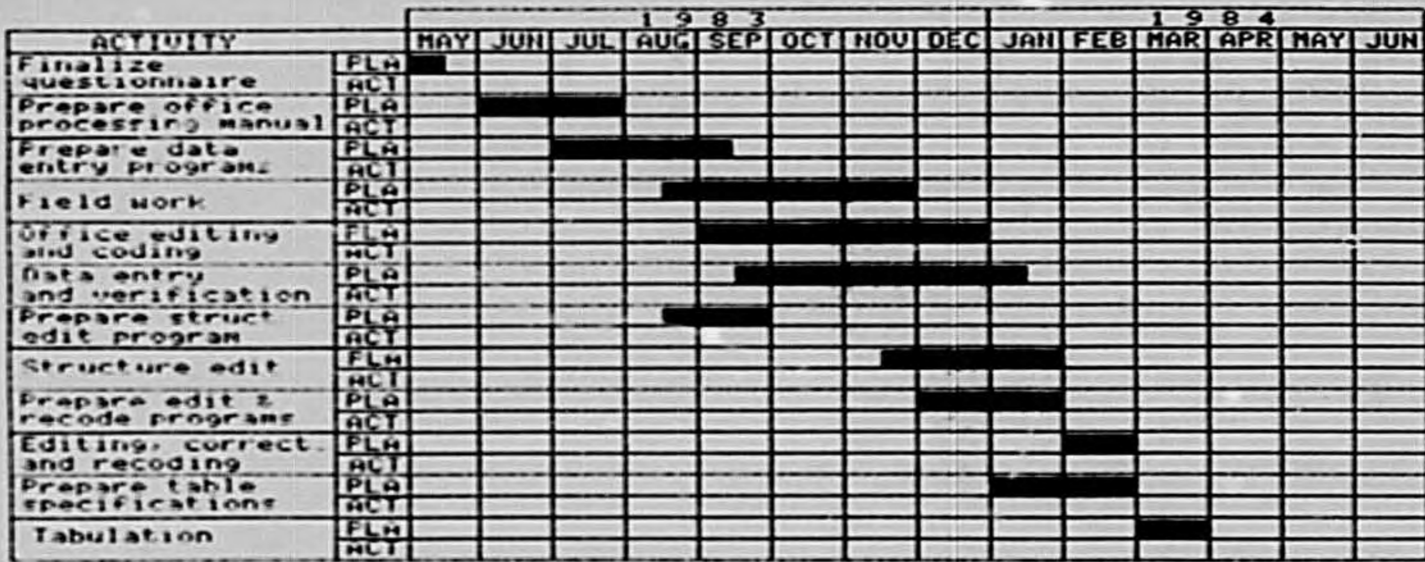
Activity	Elapsed time
Data entry and structure edit.	60 hours
Editing, automatic correction, recoding and tabulation	60 hours
Total program development (over 4 months)	120 hours

B. Data processing.

Activity	Elapsed time
Create/update district file (6 hours data entry)	10 hours
Questionnaire data entry (using 4 terminals simultaneously)	800 hours
Structure editing (5 runs x 1.5 hours/run)	7.5 hours
Corrections (5 runs x 1.5 hours/run)	7.5 hours
Raw data marginals	6 hours
Data extract and edit (4 runs x 2 hours/run)	8 hours
Edit, automatic correction and recode (4 runs x 2 hours/run)	8 hours
Recode file marginals	8 hours
Tabulation (100 tables: 15 runs x 1 hour/run)	15 hours
Total data processing (less data entry)	60 hours

National Fertility Survey

Appendix 6 - Data processing calendar.



PLA = planned ACT = actual

National Fertility Survey

Appendix 7 - Data processing work plan.

Activity	Resp.	Dates
1. Design final version of questionnaire.	ONAPO	May 83
2. Prepare office processing manual.		
A. Editing and coding manual.	ONAPO	May-June 83
B. Codebook and data dictionary.	DP	July 83
3. Design district file, data entry and structure edit programs.	DP	July-Aug. 83
4. Train office editors and coders. (transcribe 50-100 questionnaires from pilot survey to final questionnaire)	ONAPO/DP	July-Aug. 83
5. Complete district summary form.	ONAPO	Aug. 83
6. Write, test and document district file programs.	DP	Aug. 83
7. Create district file.	DP	Aug. 83
8. Write, test and document data entry programs.	DP	Aug.-Sept. 83
9. Train data entry operators.	DP	Sept. 83
10. Write, test and document structure edit program.	DP	Sept. 83
11. Install and test correction program.	DP	Sept. 83
12. Train correctors.	ONAPO/DP	Sept. 83
13. Install and test marginals program.	DP	Sept. 83
14. Prepare draft tabulation plan.	ONAPO/DP	Aug.-Oct. 83
15. Prepare editing, automatic correction and and recoding specifications.	DP	Dec. 83
16. Prepare tabulation specifications and modify libraries.	DP	Dec. 83
17. Write, test and document date extract program.	DP	Jan. 84
18. Install and test date edit and imputation program.	DP	Jan. 84
19. Write, test and document edit, automatic correction and recode program.	DP	Jan.-Feb. 84
20. Install and test COCENIS generator program.	DP	Feb. 84
21. Test tabulation specifications.	DP	Feb. 84
22. Train local staff to use tabulation system.	DP	Feb. 84

National Fertility Survey

Appendix B - Technical assistance proposal.

Two technical assistance visits of eight weeks duration are proposed: the first to write, test, install and train the ONAPO staff to use the data entry and structure editing portion of the system; the second has the same objective for the editing, automatic correction and tabulation portion of the system. Each visit is to be preceded by four weeks preparation in Europe, including two weeks at the World Fertility Survey headquarters in London.

Visit 1: proposed dates August - September 1983.

Preparation: July, 1983.

1. Prepare questionnaire codebook.

2. Design district file, data entry and structure edit programs.

3. At WFS headquarters, London.

   A. Prepare raw data dictionary and COBOL file description.

   B. Review program design and prepare preliminary versions of programs.

   C. Modify WFS software packages, MARGINALS and CORRECTOR, for NCR 8250.

   D. Transfer software and data dictionary to NCR disk.

4. At Kigali, Rwanda.

   A. System orientation for ONAPO staff.

   B. Complete, test and document district file, data entry and structure edit programs.

   C. Create district file.

   D. Train data entry operators.

   E. Install and test CORRECTOR and MARGINALS programs.

   F. Train correctors.

   G. Assist with preparation of draft tabulation plan.

Visit 2: proposed dates January - February 1984.

Preparation: December, 1983.

1. Prepare editing, automatic correction and recoding specifications.

2. Design editing, automatic correction and recoding programs.

3. At WFS headquarters, London.

   A. Prepare recode file dictionary and COBOL file description.

   B. Review edit, automatic correction and recoding specifications.

   C. Modify WFS software, FDEIR and the COCENTS generator, for NCR 8250.

   D. Transfer software and data dictionary to NCR disk.

4. At Kigali, Rwanda.

   A. System orientation for ONAPO staff.

   B. Complete, test and document date extract and automatic correction/recode programs.

   C. Install and test WFS software packages, FDEIR and the COCENTS generator.

   D. Test tabulation specifications and prepare sample tables.

   E. Train ONAPO and other local staff to use tabulation system.

Technical assistance budget.

Item		Local cost	Exchange rate	Cost US $	Remarks
Salary (12 weeks)				9,600	
Per diem					
London (12 days @ 50 pounds)		600	1.60	960	
Kigali (56 days @ $72/day)				4,032	see note 1
(56 days @ $84/day)				4,704	see note 1
Travel					
Paris-London-Paris	UK pounds	60	1.60	96	
Paris-Kigali-Paris	FFR	8630	7.30	1,180	see note 2
Computer time and supplies					
WFS computer time and software				?	see note 3
NCR computer time				?	see note 4
NCR disk				?	see note 4
Total per visit				15,868 or 16,540	
x 2 visits				31,736 or 33,080	

Notes:

1. Per diem rate is $72/day if U.S. embassy apartment is used and $84/day otherwise.
2. This is the excursion fare.
3. No rate is established, but should be able to negotiate a preferential rate.
4. NCR computer expenses will have to be determined from NCR offices in London. Costs for computer time and a disk, should not exceed 150 pounds sterling.

III. Comments on data processing of the National Agricultural Survey.

The National Agricultural Survey will be conducted by the Ministry of Agriculture from July 1983 to July 1984. The survey will cover 150 sampling units with 15 respondents per sampling unit for a total of 2250 respondents. Data will be collected by weekly visits to all of the farms. Information is recorded in seven different questionnaires at different periods during the course of the survey and much of the information will be summarized manually. Only four of the questionnaires will be processed by computer. A summary of the data to be processed with estimates of the volume to be entered and the time when it will be available appears in Appendix 1.

A series of proposals for data processing involving various computer hardware, chiefly micro-computers, were made by different agencies between 1980 and 1982. The current proposal, and the first for which any programming has been done, was made by the International Statistical Programs Center of the U.S. Census Bureau. The system is designed to be implemented on the NCR Century 8250 computer currently being used for processing the 1978 population census. The system was designed by ISPC and the programming is being shared between ISPC and a Rwandan programmer who completed the ISPC data processing course in 1982. Programs have been written for processing the pilot survey, in Washington and Kigali. The work in Kigali has been done on the computer located at the office of the NCR agents, SOMECA. This work will be transferred to the government computer as soon as the census processing is completed.

The data processing system is complete, but appears to have been designed without consulting the survey organizers. Basic tabulations and the necessary data validation checks and recoding instructions to produce tabulations were prepared by the data processing advisors. The system could also be improved by integrating the data files and some of the processing operations, thus reducing the number of files and programs which make up the system. The Rwandan programmer assigned to the survey is well qualified, but lacks experience. He is also required to perform every data processing activity himself, from entry of the programs he has written to supervision of the data entry operators. As a result, program development is falling behind schedule. The system designers should try to get the survey organizers more involved in data processing, perhaps by giving a general presentation of the system capabilities and the responsibilities of everyone involved. They should also consider more integration of system components to streamline the system and reduce the number of programs required.

The survey staff have expressed the desire to have a micro-computer to do some of the survey processing and the budget provides funds for the purchase of computer hardware. Such a computer would be useful as backup for data entry, statistical analysis which cannot be done on the NCR computer, word processing for the preparation of reports and manuals, and for budgeting and financial applications. There are three important considerations in selecting a micro-computer for the survey. First, software must already exist for the applications required, because the limited programming resources available are already totally committed to the development of programs for the primary data processing system. Second, the problem of maintenance and repair needs to be resolved, either by purchasing several independent systems which can share components and provide

mutual backup; or by purchasing a supply of spare parts and training someone to isolate and correct machine faults; or by making arrangements with a micro-computer dealer in East Africa, most likely Nairobi, for maintenance and repairs. Third, provisions should be made for the transfer of data between the micro-computer and the NCR computer. This is straightforward in principle and the CROMEMCO system was installed two years ago at the Census Bureau for this purpose. Software is required to provide a communications interface between the micro-computer and the NCR computer and this software has not been written. Another aspect of the data transfer problem is the compatability of diskettes, which are the primary storage medium for micro-computers. There is a wide variety of diskette formats available and these are usually incompatible, often even on micro-computers made by the same manufacturer. There are two options to insure compatability of diskettes which will permit the transfer of information. The micro-computer purchased must use (or be capable of producing) the same diskette format as the CROMEMCO system already installed at the Census Bureau or a second micro-computer of the same model as that selected for the survey must be installed at the government computer center. The second option will also provide the redundancy and backup mentioned under maintenance above.

National Agricultural Survey

Appendix 1 - Data summary and estimates.

Questionnaire	Contents
Household data (collected at start of survey)	Characteristics of the head of household Summary of household members Agricultural labor: 5 activities by 10 crops Inventory of agricultural equipment and livestock Agricultural marketing
Seasonal data (collected every 3 months)	Labor inputs by number of days and persons involved Animal production Agricultural and livestock marketing
Field measurements (collected each planting season)	Total size of the holding For a sample of 4 cultivated fields: size and characteristics of field, crops planted and previous production For a sample of 1 non-cultivated field: size, characteristics and use of field
Production data (collected each harvest season)	Total production for 14 crops

Data entry estimates

Type of data	Available	Estimated volume Chars./ guest.	Estimated volume Total characters	Estimated time Person hours	Estimated time Person weeks
Household schedule	Sept-Oct 83	273	600,000	100	3 1/3
Seasonal data 1	Nov-Dec 83	58	120,000	20	2/3
Field measurements 1	Jan-Feb 84	315	700,000	120	4
Seasonal data 2	Mar-Apr 84	58	120,000	20	2/3
Production data 1	Mar-Apr 84	58	120,000	20	2/3
Seasonal data 3	May-June 84	58	120,000	20	2/3
Field measurements 2	July-Aug 84	315	700,000	120	4
Seasonal data 4	Sept-Oct 84	58	120,000	20	2/3
Production data 2	Sept-Oct 84	58	120,000	20	2/3
Total			2,720,000	460	15 2/3