

USAID - AID REFERENCE
CENTER

6150170⑨
PO-AAF-283-

14p AI

RURAL ROADS SYSTEMS PROJECT

AID/AFR c1495

SECOND SIX MONTH REPORT FOR USAID

AND FAO

Kenya Rural Roads Project

615-0168
nd

615-0170

Prepared by: Harvey Herr

14 February 1960

INTRODUCTION

The second 6-month report to FAO and USAID contains 4 sections and 3 appendices. The report is organized as given below:-

- I. Accomplishments during the second 6 months
- II. Major problem areas detected.
- III. Proposed activities for the next 6 months
- IV. Chronological narrative of activities.

Appendix I. Current administrative organization staff

- II. Schedule of support for DEVRES and Department of Transport development economist.
 - III. The WANG MINI-COMPUTER
-

I. ACCOMPLISHMENTS DURING THE SECOND SIX-MONTHS

A. FAO NUTRITION MONITORING SURVEY:

1. Social Amenities Records collected together with nutrition data were extracted from the IRS-IV nutrition data file, placed on diskette and converted to WANG=FORMAT for use on that machine.
2. Mr. David Alnwick has completed the second nutritional survey report utilizing the WANG MINI-COMPUTER. He utilized the original IRS-IV nutritional data which has been further modified and field checked.
3. Rural Roads Nutritional data round 2 will be collected in the field. The schedules have been prepared.
4. Labour Force Survey Data (the Urban portion) is being validated. These data will later be linked to the Urban Households in the Nutrition Survey.

B. RURAL ROADS IMPACT STUDY

1. The Rural Roads Master Data File has been amended 19 times with new data and corrections. The file contains 49,109 records.
2. All cyclical data through December has been received. Data through November is on the file.
3. Validation rules for cyclical data determined.
4. Validation program for cyclical records 40-47 written; a separate program for cyclical records 31,50,60,61 was also written.
5. A program was written to summarize the number of records of each type that had been collected and passed to the Master-File for each case.
6. Three Validations on the cyclical data were completed.
7. A rigid data control procedure has been established for all new data collected for rural roads.

8. Weighted opening areas tables sent to DEVRES
9. Initial discussions held with Development Economist from Ministry of Transport.
10. WANG Evaluation for possible system expansion completed.
11. Revalidation of IRS-IV data completed, weights for this survey reviewed and subsequently replaced on the final data file.
12. Impact of the quarterly collection of data reviewed in conjunction with DEVRES table requests.
13. Mr. AOBIE MACHARIA has been assigned by CBS to learn the Rural Roads Impact Study data processing system
14. Significant effort will be required by the data processing section to design a system of data control, data validation and data analysis for the forthcoming National Sample. The National Sample will be carried out with enough geographic detail so that small area comparative results with Rural Road Impact areas can be achieved.

II MAJOR PROBLEM AREAS DETECTED

1. Limited CBS Expertise in Statistical data processing and job control under IBM DOS/VS has caused my time to be fragmented by others asking for assistance.
2. Validations of cyclical data has been time consuming although rigorous.
3. A significant amount of time was spent on the maintenance of generalized software (FPS, COCENTS, SPSS, CENTS-III, CENTS-AID II, CONCOR) that is used for accessing Rural Roads data and other files.
4. Work requests from users are accepted without regard for existing priorities, causing staff to become fragmented.

5. The archiving, documentation and use of already produced data is not systematically controlled.
6. There is no systematic inclusion of data processing staff in the development of survey instruments.
7. There are indications that training received by the staff at ISPC was felt to be inadequate. Three of the staff recently returned have indicated a desire to resign, or transfer.
8. The necessity for strong D.P. management within CBS remains.

III. PROPOSED ACTIVITIES FOR THE NEXT SIX MONTHS

A. FAO NUTRITION MONITORING SURVEY:

1. Merge nutrition data with Labour Force Survey and IRS-IV Household data.
2. Transfer the above data to the WANG MINI-COMPUTER for analysis by David Alnwick.
3. No new data collection for nutrition (other than Rural Roads) is envisaged during the next 6 months.

B. RURAL ROADS IMPACT STUDY

1. Incorporate second round of Nutrition Survey Data onto the Rural Roads Master File and validate it.
2. For the original and the forthcoming Rural Roads Nutrition data carryout the same analyses completed for the FAO Nutrition Monitoring Survey Data on the WANG
3. Transfer the regular updating and amendments to the Rural Roads Master File to the Kenyan staff.
4. Review the tabulation and analyses requests submitted by the development economist and DEVRES
5. Begin a Management Plan and Systems Design for utilizing the WANG on the Rural Roads Data and the new national sample (contingent on an enhanced WANG with additional storage space and memory).

6. Work more closely with local staff to prepare a CBS analysis of Rural Roads Impact (Mr. Akach, Mr. Kekovole)
7. Utilize WHO package program to determine new standard weight and height for age percentages. This program was recently used on nutrition data for FAO study.
8. Continue to monitor validations on the IRS IV data for possible future creation of extract files for Rural Roads Analysis.

IV CHRONOLOGICAL NARRATIVE OF ACTIVITIES

AUGUST 1979

A data control form was created using the WANG mini-computer. The form provides for monitoring the number of cases (households) that each month for each road and strata have submitted completed schedules. The data are batched by road and strata. When a batch of schedules are received the date of receipt is noted, the number of schedules in each strata are counted and the date the batch was sent to data entry noted. When data are returned from data entry the diskette number is noted for each batch and the date is recorded. A data control clerk Mr. Philip Wairunge carries out this function expertly.

The second and third validation runs on opening areas data completed. The master file was updated twice with corrections from the validation runs on opening areas. Additional amendment for nutrition data were also run. Preliminary opening areas tables were prepared. These early tables showed that the data file contained an inaccurate count of households. This occurred because of inaccurate coding or entry of household identification numbers. To overcome this we created a master list of correct household numbers which were entered onto the data file as household header records. There should be no records on the file that do not have a household header record preceding it. A validation program RARLIST was designed to detect household with no "member" records or records without a household header.

SEPTEMBER

The RARLIST was written and tested although not successfully run against the data file until the following month.

Validation rules for cyclical data were received from Mr. Akach and incorporated into two separate validation programs. The stock and crop change records were grouped as were the non-farm activity records with the road use and expenditure records. This was done to assist in the management of the data validation. In addition to tests for reasonable values and ranges, we test for proper accounting in the crop and stock change records. For example we test that crops remaining in the store at the end of the last cycle equals that reported in the store last time on the next cycle. Road use records are examined for rate of travel versus method of travel.

OCTOBER

Design of the program to produce DEVRES table requirements began. It was decided to create an extract file from the Rural Roads master file aggregating the cyclical data into quarterly data and creating a file accessible with SPSS. SPSS is a generalized statistical package with which we can respond to data requests in an ad hoc way. The actual writing of this program was begun.

During this month it was discovered that certain of the rural roads had been re-routed so that they no longer passed through the listed impact areas. A field review confirmed that for one road (Kakamega) the listed impact area was only adjacent to the proposed road. It was proposed that an additional area be added and a baseline survey be carried out.

The first of the cyclical data were actually received from the data entry section early in October. When these data were entered a program flaw was detected. This flaw in the program caused some of the cyclical data records to be perceived as duplicates when they were not. These were the crop production records and the road use record; for these record types household identification, record code and cycle number were

insufficient to uniquely identify each record. For example on road use during any one cycle the same household member may take several different type trips and for crop production the same crop may be in a mixed or pure stand. This problem was resolved by increasing the number of sort key fields. This change then required that all data previously entered had to be reformatted. This was done during October.

NOVEMBER

Examination of cyclical data returns pointed out that the collection of these data in a staggered fashion had implications for the tables and data analyses requested by DEVRES. The assumption by DEVRES (and myself) was that one-half the households were interviewed monthly while the remainder were done quarterly. In fact one-quarter of the households are monthly interviewed with the remaining three-quarters having staggered quarterly interview periods. The implications were that full quarterly data from all households would not be available from the field until after November. The DEVRES request was reviewed and they were notified of possible implications.

The program for validating completeness of response for cyclical and opening areas data (RARLIST) was run seven times. The output from these programs were used to eliminate incorrectly coded households and to indicate which households have incomplete sets of records. In addition the validation program which perform range and inter-variable, inter-record validation was run three times. The rural roads master file was updated nine times with new cyclical data and record correcting amendments.

Mr. Macharia was selected to become the Kenyan counter-part for Rural Roads data processing. In addition Mr. Macharia is responsible for Hotel Statistics, Migration Statistics, Air Statistics plus ad hoc requests for data analysis as required. Mr. Macharia has attended the ISPC data processing courses in the United States. Mr. Macharia is familiar with the kind of programming thus far developed for the Rural Roads impact study.

While he will not be able to devote his full time to this work he will be able to respond to user's queries by creating tables using one of the generalized programs available (COCENTS, CENTS, FPS, CENTS-AID),

DECEMBER

DEVRES was supplied with copies of Nutrition tables, the earlier published nutrition report and weighted baseline tables. Work on validation listings continued although reduced manpower during the holidays showed output. We discovered that Road 1 and Road 4 had incorrect household numbers for nutrition and social amenities records. These records were deleted and re-entered once the correct household numbers were determined. Most of the November cyclical data had been entered by the end of December, however new errors were introduced through incorrect coding of household identification and by the end of December our master-file showed 884 households and 41,380 records. The correct number of households should be 830.

The program RARLIST was revised to assist in the process of eliminating the incorrectly coded households. A second new program was begun directed toward displaying records by type received to date for each household during each cycle.

A number of problems emerged with the generalized softwares. The Cocents package was earlier revised to increase the table cell size, to accept variable blocked files and to provide for larger dimensioned tables. It had been earlier determined that two versions of Cocents be maintained - one with expanded cell size features and another with expanded table size features. Neither of these changes were well documented and when a recompilation of both version was required neither would execute without error. Nearly two weeks were consumed in correcting these complex programs. The file processing system (FPS) had faults reported in the print reformat program for the latest release version - we were unable to correct this error, but reverted to an earlier program version which satisfies our requirements.

Discussions began on the possible expansion of the WANG mini-computer system for work with the forthcoming National Samples and other suitable CBS data files including the Rural Roads Master-file. A consultant was hired to assist us in evaluating the feasibility of this undertaking both financially and technically. Data were collected by me from the various users about file sizes, run times, and types of runs. These data were analysed by the consultant who prepared estimates of hardware requirements.

JANUARY

The basic program used to place the rural roads data on the master file was modified to accept the second round of opening areas data to be collected during February and March. The data dictionary was modified to reflect minor changes to the schedules. These changes included coding the period of survey as monthly, quarter 1, quarter 2, quarter 3 rather than as simply monthly or quarterly. This assists us in finding the correct household schedule during validation.

The RARLIST program was run three times during January reducing the number of households on the masterfile from 874 to 834. Three up-dates were completed in January increasing the file size to 49709 records. Three validations followed for the cyclical data, file amendments resulting from these programs are now being processed

Data from the integrated rural survey were discovered to be faulty and I was asked to review it. I asked that several validation runs be prepared by Mr.Nguli. This was done and I reviewed the results passing along my comments to Mr.Kirimi for action. A small program was written (similar to one prepared for Rural Roads) to examine crop cyclical data records and the sample weighting. The listing showed clearly that incorrect weights had been applied. Mr.Nguli thereafter corrected these weights and some reasonable results from the IRS-IV were achieved.

The forthcoming national sampling framework is being listed in the field. This frame will become the basis for all future rural sampling work as well as a census follow-up study. Results from this sampling frame will most certainly be useful in the future for rural roads analysis as have the integrated rural surveys of the past. Systems design and data control for the sample listing (post census survey) were begun.

The amount of time required for data validation has been underestimated for cyclical data. The DEVRES tables which ought to have been prepared during January-February will be delayed at least one month, but certainly by the middle of March.

Appendix I

CURRENT ADMINISTRATIVE ORGANIZATION

No changes have taken place in the organization since the scheme presented in the last report. However lack of full time management has meant that implementation of this scheme is lacking. Productivity, training and morale of staff have not improved. We are still faced with many ad hoc requests and can only meet these demands at the expense of other ongoing activity. All in all we are a very fragmented activity responding first to this request then another.

The workload is quite substantial while the current capabilities of the staff are low. The staff who are capable tend to receive all the work leaving little time for upgrading less capable staff through training. In addition training received by the staff in the U.S. at ISPC has been criticized by some of them. Three of the staff have submitted requests for transfer. One member of the staff has quit working altogether in protest for lack of promotion. With few exceptions there is a general lack of initiative among the staff.

We will attempt to respond to these problems in the following ways:

- 1) Redefine the objectives of the CBS data processing section.
- 2) Set priorities for the major work we undertake (1979 Census, National Sample, Census of Education, Rural Roads Impact Study, Labour Enumeration, Kenya Fertility Survey, etc.)
- 3) Provide each officer with a detailed work plan for the next 6-months.
- 4) Implement staff training
- 5) Investigate the use of statistical clerks to supplement the work of statistical officers.

The list of priorities will be presented to the Director of Central Bureau of Statistics for his approval. Objectives and work plans will be discussed with staff individually and jointly in our working group meetings. We anticipate completing this work in March.

Appendix II

Schedule of Support for DEVRES and the Department of transport Development Economist

During February 1980 the Development Economist Ayse Kodat reported for work at the Ministry of Transport. She thereafter came to CBS to discuss the data and her requirements with me. Initially she received the first tabulations sent also to DEVRES. Further requests from the Rural Roads master file were also tabulated. These were tables showing the differences in male and female headed households. Additional requests for her from the Rural Roads Master file are pending.

In our conversation she expressed the need to analyse the Rural Road Evaluation Data that is prepared by the Ministry of Transport. Her objective was to simplify road evaluation criteria by selecting the most critical variables through regression analysis. The data for some 470 roads were entered and a series of regression analyses performed critical variables being identified through the stepwise linear regression procedure. The results of this analysis were indicative however examination of residuals and other tests might be required to show the results conclusive.

I

The development economist and/have worked closely together. The results she sought from the data were obtained through the development of the analysis by a series of changes in the program. This is the general pattern of data analysis, progress is made though step by step changes in the program in response to going data analyses culminating in acceptable results. The danger of course is that without a pricing mechanism the step by step change become endless.

As indicated in the last 6-month report, Appendix III work for DEVRES is likely to take on the same step by step changes. It has been suggested that DEVRES send someone here for several months to participate first-hand in the process. The initial tables as requested will be completed and the step by step process can begin therefrom.

Where are we now? The DEVRES tables have not yet been prepared. This delay stems from longer and more elaborate validations for cyclical data. First tables (not DEVRES) will help us analyse the quality of responses - these are the tables to be prepared for Mr. Akach and Ayse Kodat. The tables for Mr. Akach will examine quarterly versus monthly modes of data collection for non-farm activities, area under crops from the cyclical crop production records, type and number of trips from the road use record, and characteristics of expenditures. Ayse's tables are extensive and are oriented toward household production, expenditure and consumption. We believe that at this point the data for the first quarter are clean; however experience with the similarly complex Integrated Rural Surveys shows that first tables often bring to light data inconsistencies. DEVRES tables are to be prepared this month (March), followed by tables for Ayse and Mr. Akach.

Appendix III

The WANG MINI-COMPUTER

In February a consultant completed an evaluation of options available for expanding the WANG 2200 MVP mini-computer. The Central Bureau of Statistics desired to transfer some of its statistical data processing from the IBM 370 DOS/VS onto the interactive WANG mini-computer. Among the items for potential transfer are the new expanded national samples, rural roads impact data, nutrition data and the market price information system. The consultant was required to evaluate hardware requirements, systems software, applications software and file and data management capability. Money for this consultancy was provided through FAO.

Three options were developed by the consultant. The first was simply to expand disk space, memory, terminals and software on the 2200 MVP. As discussions were held with the vendor a second option was provided - To replace the MVP model with a VS model. The VS model is a virtual storage machine capable of providing large amounts of memory to users through the technique of "paging". The VS also has much improved systems software and it supports high level languages such as FORTRAN and CUBOL plus is capable of handling much larger files. The capability of supporting many more terminals than the MVP is also a feature.

The third option was to expand the interactive capability of the IBM 370 DOS/VS by providing two dedicated disk drives, four terminals, a printer and system software to dynamically allocate tape driver to the CICS (interactive) partition. Of the three options the conversion to the WANG VS was cheapest and provided most enhancements. Staff have reviewed the consultants' report and are preparing questions for the WANG software representative. Funds are being sought.
