PN-ABK-526
67874

JESS INTERIM REPORT
DATA MANAGEMENT AND ANALYSIS

JESS Report No 22

Prepared by

Leonard A Malczynski
Associates in Rural Development, Inc
110 Main Street, Fourth Floor
P O Box 1397
Burlington, VT 05402
U S A
Under AID contract number AFR-0134-C-00-5047-00

Date 30 November 1987

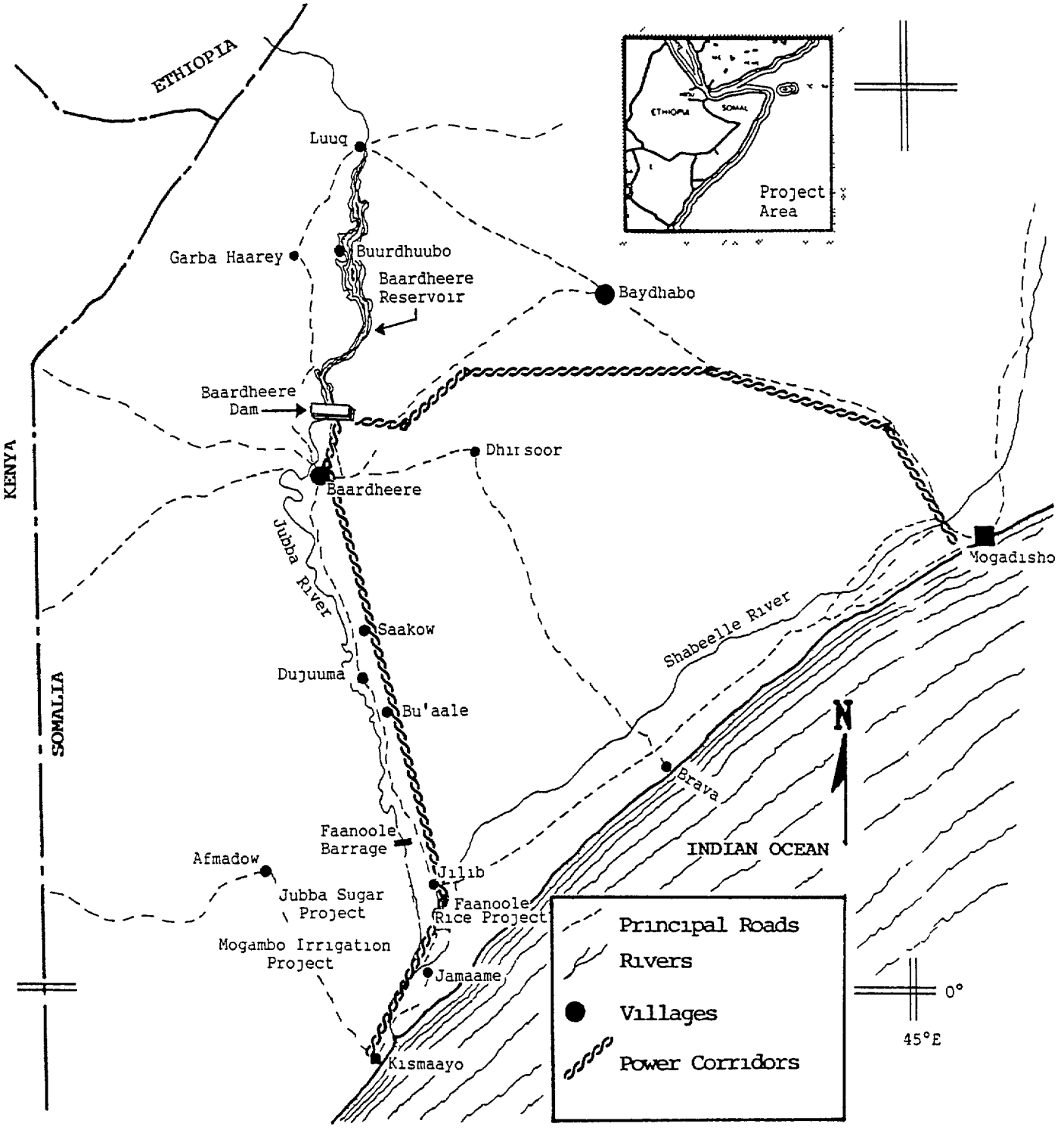
A

CONTENTS

Section	Page
Acronyms and Abbreviations	1
Preface	11
I. Executive Summary	1
II Introduction	2
III. Situation Assessment	3
A. Physical Handling of Data	3
B Data Entry	3
C. Data Analysis	5
D. Hardware	5
E. Software	6
F. Training	6
G. Supplies	7
IV. Work Completed	8
A Questionnaire Management	8
B Data Entry	8
C Data Analysis	8
D Training	9
V Recommendations	10
Appendices.	
A--List of Persons Contacted	A-1
B--Terms of Reference	B-1
C--Physical Flow of SEBS Questionnaires and Data Entry Estimation	C-1
D--JESS Data Entry and Verification Package	D-1
E--SEBS Data Entry System User's Manual	E-1
F--Preliminary SEBS Analyses	F-1
G--Analysis of Vegetation Plots	G-1
H--ASCII Conversion and File Transfer Primer	H-1

B

Area of the Baardheere Dam Project and the Jubba Valley Development



ACRONYMS AND ABBREVIATIONS

AHT	Agrar und Hydrotechnik GmbH
AID	U S Agency for International Development
ARD	Associates in Rural Development, Inc
BuRec	U S Bureau of Reclamation
DBMS	data base management system
DD	Data Dictionary
DDL	Data Description Language
DES	Data Entry System
DML	Data Manipulation Language
GSDR	Government of the Somali Democratic Republic
GTZ	Gesellschaft fur Technische Zusammenarbeit GmbH
IBM-AT	International Business Machines microcomputer
IBM-XT	International Business Machines microcomputer
JESS	Jubba Environmental and Socioeconomic Studies
JuDAS	Jubba Development Analytical Studies
MJVD	Ministry of Jubba Valley Development
MS-DOS	Microsoft Corporation's Disk Operating System
PC	personal computer
RMR	Resource Management & Research, Ltd
SEBS	Socioeconomic Baseline Surveys
USAID	U.S. Agency for International Development

PREFACE

The Jubba Environmental and Socioeconomic Studies (JESS) (number 649-0134) are jointly funded by the Government of the Somali Democratic Republic (GSDR) and the U S Agency for International Development (AID). JESS is part of a larger project funded by USAID and the GSDR, the Jubba Development Analytical Studies (JuDAS) Technical assistance and JESS project management are being provided to the Ministry of Jubba Valley Development (MJVD) by Associates in Rural Development, Inc. (ARD) of Burlington, Vermont, under AID contract number AFR-0134-C-00-5047-00. This report describes activities during the consultancy of Leonard Malczynski from 24 May to 21 June 1987.

I. EXECUTIVE SUMMARY

The data management situation at JESS upon the consultant's arrival was in less than adequate shape. While the JESS team has made considerable progress in insuring the quality and accuracy of the data collected, it has been less successful with data entry This is understandable owing to the turnover in local staff and their inexperience with data management as well as persistent power supply fluctuations and problems of hardware integrity.

This consultancy has therefore concentrated on improving the data entry system for Socioeconomic Baseline Surveys (SEBS) by writing a set of more efficient data entry programs and training the data entry staff in their use However, based on the recommendation of this consultant, all the data entered up to June 1987 (approximately 143 households) was scrapped and will require re-entry due to the number and complexity of errors introduced into the data base.

Analysis of the data was given high priority and has resulted in preliminary sets of tables for the data entered as of 14 June 1987, primarily to test the current data analysis software.

A realistic time line was reestimated for the SEBS The limits to SEBS completion are field time and coding, not data entry time. This time line assumes a Phase II report date of 1 April 1988 and an end to SEBS field data collection activities on 1 December 1987.

A preliminary data management and analysis system was also established for the vegetation plot data

II. INTRODUCTION

Data base management has been particularly challenging for the JESS program, both for technical and logistical reasons. JESS is conducting, for the first time throughout the Jubba Valley, primary-level data collection in relation to a broad range of socioeconomic and environmental variables. This includes detailed household surveys, physical sampling of vegetation, and collection of data in relation to public health, fisheries and water quality Logistically, JESS has had to procure and operate computer equipment in an oftentimes hostile environment (temperature, humidity, power outages, surges and drops) where there is limited access to trained computer specialists and electronics repair facilities

This was the second trip by the JESS data base management specialist More trips are envisioned as the processing and analysis of socioeconomic and environmental data continue during the remainder of JESS' Phase II (until April 1988) and Phase III (April through September of 1988).

Overall, the SEBS has required the most attention by this consultant. However, the consultant has also assisted the JESS ecologist in processing vegetation data and worked with other JESS fisheries and public health consultants who are using computers for data processing and analysis This report details the format of an approximate program designed within the limitations of human resources and physical infrastructure inherent in Somalia.

III. SITUATION ASSESSMENT

A Physical Handling of Data

The JESS team has established a very high quality control system for their SEBS questionnaires Each questionnaire is

- filled in the field;
- checked for errors in the field;
- coded in Mogadishu,
- checked by supervisors after coding (marked in rose to indicate coding completed, marked in yellow to indicate coding checked, and marked in brown to indicate coding problem to be examined by economist/anthropologist),
- checked again by the data center manager,
- sent to data entry (marked in green to indicate completed data entry); and
- stored safely

This system has an excellent error detection program. The questionnaires, as presented to the data entry staff, represent as accurately as possible the responses recorded in the field. Nevertheless, there exists a bit of confusion in tracing questionnaires through the data entry process. The process currently has no way of determining the number of questionnaires coded and entered. A simple method was established, but not as yet followed by the time of the consultant's departure, primarily due to the home leave absence of the team's economist.

There is a serious bottleneck in the coding verification done by the data center manager The manager cannot keep pace with the flow of coded questionnaires produced by the enumerators/coders

B Data Entry

The following types of errors are predictable and expected. Training a data entry staff, all of whom, have never touch-typed, and in some cases, never touched a typewriter or calculator, takes time. Nevertheless, there are many errors which can be reduced.

Errors of omission have been found in the first sets of data to be entered for the SEBS. Essentially, the data entry personnel have inadvertently skipped over some of the questionnaires within the questionnaire packet, thereby not entering the data. Therefore, the integrity of the data base has not been maintained.

Related to omission error, the data entry staff has also introduced redundancy into the data base by entering several records twice

As may be expected at any data entry installation, even with highly skilled data entry personnel, typographical errors are made Many of these errors can be reduced by programming a data entry system.

Widespread dBase III file structure errors were potentially the most damaging to the data base. A data structure error is the incorrect assignment of type to a variable. Assigning a variable a numeric type when it is a character type is the most damaging

For instance, in the process of coding responses, digits are used as codes. These same codes are not intended for later use in calculation, they only represent a characteristic of a variable. If they are incorrectly assigned a numeric type, leading zeros in the code will be removed. As an example, suppose the variable "crop" has 20 possible responses, from 01 to 20 Assigning crop a numeric type would establish the codes 1 to 20 and not 01 to 20. Although the point seems minor, during analysis when the data file is selectively queried by crop, "SELECT IF crop = 01" will not select the crops coded 1 This poses a problem because, during analysis, cases will be missed that are valid.

In addition, another advantage to assigning types correctly is reduction of data entry errors. In both the data base file structure and data entry programs, types can be used to restrict the input of characters A data entry program could, for example, convert all lower case letter entries to upper case, automatically removing the need to use the shift key, saving keystrokes and ensuring that all text is capitalized (important for searches of the data base) A numeric type would prevent the assignment of any letter or symbol to that variable

Estimating the rate of data entry was difficult The data center manager, new to the job, has had to deal with problems of electrical power, absenteeism and incorrect coding. A simple rate calculation, number of questionnaires divided by number of days, is of no use unless these problems are solved

Soon after the consultant's arrival, a second shift was added to the data entry schedule. With the addition of the second shift, the data entry staff entered 40 percent of the total number of questionnaires in a period of seven working days (approximately 5 2 percent of the estimated total number of questionnaires to be administered) The additional data entry operator had a considerable effect

C. Data Analysis

This consultant has suggested to AHT/GTZ that they obtain a copy of Statpac for their Wang microcomputer They will then be able to analyze the JESS data using the same package as JESS and do additional analyses as well.

An estimation of the data entry time frame for the project was completed by this consultant and Dr Craven, the JESS economist. A probable time line was completed for the life of data collection from which data entry and analysis times were estimated (see Appendix B)

The team met with Dr Murray Watson of Resource Management and Research, Ltd. (RMR) in order to examine the software and data on Dr Watson's aerial census of the Jubba Valley The software performed satisfactorily. The JESS team has been asked by Dr Watson to review the summarized findings and indicate whether additional information should be extracted from his data base Dr. Deshmukh, the JESS ecologist, and the consultant made some suggestions for further analysis and presentation.

D Hardware

The hardware, which was originally the source of recurrent problems, is now operating well. The two IBM AT clones were tested and are working. In addition, the IBM XT, left by the BuRec team, was refurbished by removing the broken 10-megabyte hard disk and installing a 20-megabyte hardcard.

As soon as the team's other IBM AT returns from stateside repairs, they will have four working IBM ATs or clones, one XT and one PC (located at MJVD). Thus, no further need for IBM-compatible hardware is foreseen.

There still continue to be problems with electrical power supply Mogadishu line voltage at the data center is often as low as 175 volts. These low voltages are not adequate to power the JESS battery-charging system. In addition, one of the DC to AC inverters is not operating.

The consultant met with the AHT/GTZ team in order to discuss the compatibility of the computer hardware/software systems of both JESS and AHT/GTZ. AHT/GTZ is using Wang microcomputers which operate under MS-DOS, the same system used by JESS computers Any file written on the JESS equipment can be read by the AHT/GTZ equipment, assuming of course, that they possess the same software as JESS Thus, while files can be easily transferred between projects, software cannot The reason for this is that Wang was unable or unwilling to produce 100 percent IBM-compatible equipment. However, they wisely chose MS-DOS, thus permitting file transfer.

E. Software

Fortunately the AHT/GTZ team is also making use of dBase III and Lotus 1-2-3 for its analyses The word processing systems are different, but files saved in MS WORD with no format can be read by the Wang and easily converted to Wang word processing format by choosing one of the Wang menu items.

F. Training

The duration of the consultancy precluded the mastery of software by the JESS team Although considerable progress has been made by the Somali staff, the expatriate staff, with few exceptions, is still only proficient with the word processing package. The main constraint to training is the demand of fieldwork. The staff is either busy preparing for a trip or actually in the midst of one.

The only way to gain the necessary experience is to have data to manipulate Some team members now have access to a substantial amount of data. They will soon be concentrating on learning Lotus 1-2-3, dBase III and Statpac. There is however, a lack of adequate training materials for the Lotus package. The user's manual and tutorial are not written for the novice computer user Thus, additional materials should be purchased

The host-country staff has received some microcomputer training in the past year The consultant was also impressed by several individuals who have learned a considerable amount on their own. The situation at the data center is not really conducive to training and, due to the considerable data entry task ahead, little time is available for formal training Many of the data entry staff are inquisitive and capable of learning more about computer use, but JESS is not charged with formal training in computer science

G. Supplies

It was attempted to estimate the computer supplies currently available, but because of the poor early inventory system, it was a difficult task The team appears to be well supplied with computer paper and diskettes, however, printer ribbon supplies are adequate only for the Epson LQ 1500

IV. WORK COMPLETED

After initial examination of the situation, the consultant developed a work plan which addressed several points.

A. Questionnaire Management

A few small changes need to be made to the questionnaire management system. First, the physical flow of questionnaires needs to be more restricted It would be ideal to know, at any one time, the progress being made by the coding team. This can be accomplished using the straightforward checklist devised by Dr. Craven. This checklist classifies all the household numbers included in SEBS by district and village As a questionnaire passes each stage of the process (coding, checking, verification and entry), it is marked off the appropriate list. This way, progress at any stage can be measured.

B Data Entry

A data file management procedure is required Currently all the SEBS data is on one floppy diskette The data was segregated by region (Lower, Middle, Upper Jubba) to individual floppy diskettes in each of the two computer rooms (see Appendix E).

There was also a need for the JESS data processing staff to have a data entry program written to minimize the data entry errors and increase the data entry rate. The consultant spent a good part of his stay programming that package. (see Appendices C, D and E).

C Data Analysis

Using the currently available, yet unreliable keypunched data, test runs of the statistical analysis proposed by the team were performed on the SEBS and vegetative data (see Appendix G). Codebooks for the SEBS analyses (using Statpac software) and file structures for the SEBS (using dBase III software) were developed and are on record at ARD in Vermont and being used by the JESS team in Somalia. Unfortunately, the consultant was not able to complete all the data analysis on which he had planned. The team has a good basic list of descriptive statistics to generate The list needs to be finalized as soon as possible.

Due to the nature of the dBase III file structure errors, several revisions had to be made to those structures. The current structures are correct to the extent possible, barring additions or deletions of variables from the SEBS questionnaires

D. Training

A considerable amount of time was spent on the Data Entry System (DES). The data entry staff spent four days testing the programs before spending six days in actual operation The data entry staff has been able to complete the entry of half of the data scrapped Two shifts of two operators each are working, supervised by both the data center manager and the economist

During the development of the DES, the consultant was assisted in detailed error checking by data center staff members Rooda Bile and Aweys Mohamed, and by Axmed Cali, field technician for SEBS, who provided suggestions which greatly improved the programs' operation.

Several short periods were also spent on Lotus, dBase III and Statpac, primarily on a one-to-one basis Additionally, three staff members participated in a talk given by the consultant to the Mogadishu microcomputer users group on data base management.

V. RECOMMENDATIONS

Socioeconomic Baseline Study

- Limit the duration of the study to 1 December 1987 or the complete collection of the 1000-household sample, whichever occurs first.

- Establish a concrete set of descriptive analyses to be performed on the SEBS data and execute those analyses as soon as all data for a region are entered.

- Limit any inferential analyses to the period after all the data have been analyzed descriptively. One can easily get lost testing hypotheses before examining all the descriptive results.

Microcomputer Use

- All the research members of JESS must get involved in all uses of the computers--word processing, spreadsheets and statistical analysis. It is the responsibility of the research designer to carry out the analysis of his or her data collection effort. In order to accomplish this, researchers should.

  --consistently set aside time on a scheduled basis, individually or as a group,

  --consider the tasks of each team member as a group--share problems and solutions;

  --initially try not to work alone--form an individual users' group (this implies keeping several computers in the same place), and

  --the user should keep attending the microcomputer club meetings and plan to demonstrate what he/she does and does not know to that group.

- An excellent way to learn is to take a subset backup-copy of the data and attempt real analyses. Attack the data with a definite goal in mind.

- Users should not expect others to train them, but should ask others to explain what they have done and how they accomplished it.

General Training

- Obtain a self-teaching guide to Lotus 1-2-3 (e g , Baras, Guide to Using Lotus 1-2-3, Osborne-McGraw Hill)

- Have ARD investigate the Lotus 1-2-3 training programs on videotape

- Play with the computer, e.g., games, graphics, tutorials, etc.

Data Center Operations

- Rely more on the enumerators and the data entry system for coding verification in order to eliminate the current bottleneck in coding.

- Begin to train the data entry staff in the entry of the other data sets (marketing, vegetation, etc)

- In all seriousness, change the name to Analysis Center and put up a sign to confirm it.

APPENDIX A

List of Persons Contacted

MJVD

Abdi Ali Mo'alin	permanent secretary
Aweys Haji Yusef	director of planning
Yassin Nur Osman	director of finance and administration

AHT/GTZ

Heiko Brunken	team leader
Gunter Keizer	microcomputer specialist
Wolfgang Haupt	agricultural economist

JESS Team

Gus Tillman	team leader
Rooda Mohamed Abdulaahi	data entry operator
Ahmed Ali Mohamed	field technician, SEBS
Kathryn Craven	development economist
Ian Deshmukh	ecologist
Mumina Aden Ali	data entry operator
Abdi Malik	data entry operator
James Merryman	development anthropologist
Aweys Sheik Moxamed	data entry operator
Marie Tillman	data center manager
Murray Watson	RMR
Maimuna Wehlie	data entry operator

AID

Sally Patton	project manager

12

APPENDIX B

Terms of Reference

1 The contractor will conduct an 4 week consultancy in Somalia beginning on or about mid-May 1987 for Associates in Rural Development, Inc (ARD) as a part of the Jubba Environmental and Socioeconomic Studies (JESS)

2 The general focus of this consultancy will be to continue work on a data management system for the JESS project and to develop strategies for data analysis which will be used in Phase III.

3 Prior to departure for Somalia, the consultant will receive written or telephone briefings from the ARD project manager.

4 Upon arrival in Somalia, the consultant will hold briefing meetings with the USAID project manager or designate, the JESS team leader or designate and the JESS field team to discuss the consultant's scope of work At these meetings, a means for progress reporting during this consultancy will be defined.

5 The consultant will study the physical, logistical and organizational systems being used for data entry and storage, and recommend changes that he deems necessary for a more efficient operation This will include an assessment of the physical location being used for data entry, of the methods employed to store and access data, and of number and work schedule of the data processing personnel.

6 The consultant will review the levels and capabilities of the data entry personnel and identify areas of weakness. He will then, in conjunction with the local data specialist, develop a remedial training program that will be run by the latter.

7 The consultant will evaluate the organization of all the socioeconomic data from the standpoint of statistical applications He will make recommendations for and help with the reorganization of those data for which the desired manipulations cannot be performed in the most efficient manner

8 The consultant will work with the socioeconomic team to help design the Intensive Village Studies and other special studies to facilitate future data entry and retrieval

9 The consultant will help other members of the JESS team devise systems for the best organization and retrieval of data in their respective fields

10 A word-processed draft report for this consultancy will be delivered to the JESS team leader at least one full day prior to departure from Somalia This report will be written in a format which meets guidelines which the consultant will receive from the JESS team leader This report will constitute a basis for a final briefing with the USAID project manager, the JESS team leader and project personnel from the Ministry of Jubba Valley Development (MJVD) Revisions for a final report should be completed within two weeks after receiving comments from JESS and ARD

11 At the discretion of the JESS team leader, the consultant may be required to present an one hour training seminar at the MJVD on a topic related to this consultancy.

APPENDIX C

Physical Flow of SEBS Questionnaires and Data Entry Estimation

The proposed changes to the data coding and entry system are measures to reduce data entry error and maintain organization of the questionnaires. Some of the changes are a direct result of using the data entry package, which is unforgiving in terms of coding errors.

A simple system to track the SEBS questionnaires has been devised. Since the total number of questionnaires is known with a fair degree of certainty, as each questionnaire passes three critical points (coding, validation of coding and entry), it will be checked off the master list This process will permit the research team to measure its progress and identify bottlenecks

The system uses the same color-coded scheme used to physically mark the questionnaires Three master lists are maintained by the enumerator supervisor, data center manager, and data entry staff. Each requires the color code put on the questionnaire.

The consultant calculated average keystroke savings for a hypothetical questionnaire and extrapolated the results over the remaining data entry task. For each of the questionnaires below, a total actual and a total probable number of keystrokes has been estimated The last figures indicate percent of keystroke savings "Id" indicates the identifier portion of the line, "Var" the variable portion The total number of lines is based on a 1000-household survey size and the total is the total number of keystrokes.

	Actual		Probable			#	
	Id	Var	Id	Var	savings	lines	total
Individual	26	46	16	33	13	9000	325000
Absence	27	16	17	12	4	1635	22740
Marriage	26	42	16	42	0	1596	13000
Household	26	188	16	147	41	1000	41000
Income	26	430	16	370	60	1000	60000
Livestock	26	171	16	87	111	1000	111000
Plots	26	24	16	24	0	1327	8320
Harvest 1	29	84	19	64	20	1846	60120
Harvest 2	29	84	19	64	20	1673	50860
Hired labor	27	13	17	13	0	1173	2700
					Total		694740
				Potential	% savings		31.51%

15

The estimation of data entry time proceeds from an assumption concerning the number of keystrokes per hour. This number is multiplied by the number of hours available and the number of data entry stations If all the questionnaires are properly coded, it should take no more than 10 weeks, at a rate of 100 questionnaires per week, to enter the SEBS data. Of course, this optimal scenario all depends on the flow rate of questionnaires from the field to coding, then from coding to data entry, in absence of any breakdowns in computers and power supplies.

The following tables, estimated and designed by Dr. Craven and the consultant, indicate hypothetical schedules for the SEBS coupled with data entry and coding. Two variables play major roles in the estimation--coding and checking rate, and the delay between trips. The delay allows for more coding and checking since the enumerators code only when in Mogadishu However, prolonging the delay between trips prolongs the completion date of the SEBS.

In Scenario 1, coding occurs at the rate of 10 questionnaires per day and entry at 20 per day. A backlog of questionnaires develops as the time between trips decreases The "coding backlog" indicates the number of questionnaires that will not be coded, and the "entry idle" column lists the number of days the data entry staff will be idle, with no SEBS questionnaires to enter. This leaves the staff with time available for entering vegetation, marketing or women's study data or other related work.

In Scenario 2, the coding and checking rate is doubled to 20 per day. This of course decreases the coding backlog considerably

The two scenarios both point out that coding and checking are the bottlenecks in the system. The data entry staff will be idle from 92 to 118 days over a six- to seven-month period (providing that they are not entering other data) Solutions therefore are:

- decrease the sample size of the SEBS;
- train the data entry staff to code questionnaires,
- hire additional coders;
- have the data entry staff perform a post data entry desk check of all data;
- make use of the computer idle time to do preliminary analysis;
- train the data entry staff to enter other data, and
- a combination of the above.

Hypothetical Data Coding, Checking and Entry: Scenario 1

questionnaires per day
code & check 10
enter 20

Optimistic (1 week delay)

	Begin	End	# Qs	Coded	Entered	Coding Backlog	Entry Idle Days
Baardheere	06/23	07/06	132	50		82	
Jilib & Jamaame	07/13	07/28	119	42	50	159	13 20
Garba Hareery	08/03	08/15	122	64	42	217	10.75
Luuq	08/24	09/05	127	64	64	280	11.80
Bu'aale/Sakow	09/14	09/30	76	76	64	280	14.65
Kismaayo	10/12	10/26	124	124	76	280	14 75
	12/01	12/31			124		40 90
					Total idle days		106 05

Probable (2 week delay)

	Begin	End	# Qs	Coded	Entered	Coding Backlog	Entry Idle Days
Baardheere	06/23	07/06	132	100		32	
Jilib & Jamaame	07/20	08/04	119	100	100	51	15.70
Garba Hareery	08/18	08/30	122	100	100	73	13.55
Luuq	09/13	09/25	127	100	100	100	13 55
Bu'aale/Sakow	10/09	10/25	76	76	100	100	16 40
Kismaayo	11/08	11/22	124	64	76	160	16.20
	12/01	12/31			64		24.65
					Total idle days		100 05

Pessimistic (3 week delay)

	Begin	End	# Qs	Coded	Entered	Coding Backlog	Entry Idle Days
Baardheere	06/23	07/06	132	132		0	
Jilib & Jamaame	07/27	08/11	119	119	132	0	19.10
Garba Hareery	09/01	09/13	122	122	119	0	17.60
Luuq	10/04	10/16	127	127	122	0	17.45
Bu'aale/Sakow	11/06	11/22	76	76	127	0	20 05
Kismaayo	12/13	12/27	124	35	76	89	21 20
	01/01	01/31			35		23 25
					Total idle days		118.65

Hypothetical Data Coding, Checking and Entry Scenario 2

questionnaires per day
code & check 20
enter 20

Optimistic (1 week delay)

	Begin	End	# Qs	Coded	Entered	Coding Backlog	Entry Idle Days
Baardheere	06/23	07/06	132	100		32	
Jilib & Jamaame	07/13	07/28	119	85	100	66	10.70
Garba Hareery	08/03	08/15	122	122	85	66	8 60
Luuq	08/24	09/05	127	127	122	66	8 90
Bu'aale/Sakow	09/14	09/30	76	76	127	66	11 50
Kismaayo	10/12	10/26	124	124	76	66	14 75
	12/01	12/31			124		40 90
					Total idle days		95 35

Probable (2 week delay)

	Begin	End	# Qs	Coded	Entered	Coding Backlog	Entry Idle Days
Baardheere	06/23	07/06	132	132		0	
Jilib & Jamaame	07/20	08/04	119	119	132	0	14 10
Garba Hareery	08/18	08/30	122	122	119	0	12 60
Luuq	09/13	09/25	127	127	122	0	12 45
Bu'aale/Sakow	10/09	10/25	76	76	127	0	15.05
Kismaayo	11/08	11/22	124	124	76	0	16 20
	12/01	12/31			124	0	21.65
					Total idle days		92 05

Pessimistic (3 week delay)

	Begin	End	# Qs	Coded	Entered	Coding Backlog	Entry Idle Days
Baardheere	06/23	07/06	132	132		0	
Jilib & Jamaame	07/27	08/11	119	119	132	0	19 10
Garba Hareery	09/01	09/13	122	122	119	0	17.60
Luuq	10/04	10/16	127	127	122	0	17.45
Bu'aale/Sakow	11/06	11/22	76	76	127	0	20 05
Kismaayo	12/13	12/27	124	71	76	53	21 20
	01/01	01/31			71		21.45
					Total idle days		116.85

APPENDIX D

JESS Data Entry and Verification Package

Examination of the data entry process at the Jess Data Center and the typographical quality of the data files created indicated that some stricter control of data entry was necessary.

This consultant has had previous experience and success in using dBase II (Gambia River Basin Studies), KnowledgeMan and dBase III+ (Agricultural Development Support II, National Sample Survey of Agriculture, Haiti) for producing data entry programs which

- control variable type--numeric or character;
- control data accuracy with range checking,
- control data accuracy with logic checking;
- reduce data entry time; and
- reproduce the physical layout of the data collection instrument.

A brief explanation of each of the five benefits follows

Control of Variable Type

The data base management system (DBMS) permits the characterization of variables (data items) as numeric, character, date, or logical. Since all coding is done with digits (to alleviate the need for touch typists and subject the data entry staff to learning the use of a 10-key numeric keypad), it is necessary to distinguish between true numeric values (to be used in calculations) and numeric codes, such as a social security number.

Control by Range Checking

The DBMS permits the limitation of accepted item values by checking them against a range of permissible values For example, responses to a question could be

1	yes
2	no
8	don't know
9	no response

19

The DBMS would then wait until the response typed was 1, 2, 8, or 9 Of course typing 1 when the response is 2 cannot be prevented, but by narrowing the range of possible errors, a much lower frequency of error will be achieved.

Control by Logic Checking

Logic checks can be performed using IF-THEN statements which narrow the range of permissible responses. If a certain value is assigned to an item, it can signal that the following items require a specific set of responses.

Reduction of Data Entry Time

Two characteristics of questionnaires can be manipulated to advantage during data entry "answer if" questions and data record identifiers.

Most socioeconomic questionnaires contain "answer if" questions. A series of follow-up questions are asked if the response is positive, otherwise the follow-up items take on a default value, usually 0 The DBMS can automatically assign the default values to the follow-up questions thereby freeing the data entry operator from entering values that can be deduced from previous responses

A data record identifier is usually composed of several variables (often combined into a composite variable) This identifier, which could be, for instance, the number (in sequence) of the survey or the name of the person interviewed, serves to identify each record uniquely within the scope of all data collected The identifier is ubiquitous, and not unique to computer-processed surveys.

The nature of the identifier differs from questionnaire to questionnaire Some questionnaires record data on one "unit of analysis," a household for example Others permit the recording of data on one or several "units of analysis," a demographic questionnaire for example Often a portion of the identifier is repeated for each unit of analysis. The DBMS can accept and verify the repeated portion of the identifier and attach it to each data record thereby reducing the number of keystrokes required to record the information. The savings in number of keystrokes can approach 50 percent[1]. Below, is an example from the SEBS questionnaire on individuals

---

[1] The Gambia River Basin Studies' daily activities questionnaire had an identifier of 13 digits and a variable line of 11

70

Identifier is composed of

item	# of characters
questionnaire type	2
date of interview	8
interviewer number	2
region	1
district	1
beel	2
village	1
household	2
Total	19

These 19 characters need only be entered once regardless of the number of individuals in the household--there is one item which holds the individuals' assigned number (the data record identifier)

Reproduction of Questionnaire Format

The format of the JESS SEBS questionnaires does not lend itself well to on-screen reproduction. This is not at all a hindrance Some of the individual pages are reproduced to some extent The DBMS has a screen design facility which, coupled with the data entry programs, provides a good facsimile of the questionnaire format In most cases the data entry staff becomes so proficient that they begin to ignore the screen formats, having memorized the keystroke sequences.

Programming the Data Entry Package

The actual programming of the data entry package is more tedious than complicated. Essentially the DBMS[2] has a Data Manipulation Language (DML) which permits applications programs to be written. The data entry program consists of a series of tests for data type and code range. The tests narrow the possible response to each question by limiting the range of response according to the SEBS codebook

---

[2] DBMS will be used to refer to dBase III + even though, in the technical sense, dBase III + is not a DBMS.

A sample portion of a program is reproduced here The comments following the (*) are the consultant's and not a part of the DBMS DML

```
err = .T.                    * set the variable err equal to True
DO WHILE err       * perform the following until err not True
   @ 3,6 GET e4 PICTURE '9'     * accept e4 as a single digit
   READ                                  * read e4 into memory
   IF e4 $ '12389'              * if e4 is 1, or 2, or 3 etc
      err =  F.                        * set err equal to False
   ENDIF                             * terminate the IF clause
ENDDO                       * if err is False go on, else redo
```

Thus an operation loops infinitely until the keyed response to "e4" meets specific criteria. Data entry systems of this sort are not forgiving on data coding errors Even though each questionnaire is checked three times, given the total number of variables (approximately 515), some coding errors reach the data entry stage. The data entry program can discover them and the questionnaire is marked for further verification.

Program Listings

WARNING: dBase III + programs are stored in ASCII format
If they are written on or in any way altered with a word
processor, they must be saved in ASCII format This is
possible with MS Word (formatted (No)) and in Wordstar
(option N, non-document mode) among others.

Program listings are printed in formatted MS Word Therefore the page breaks in these listings are not found in the ASCII versions of the programs Each new program will have a header beginning with * name PRG

The program listings are not provide here as the listings exceed 100 pages in length Both computer disk and hard copy versions are kept at JESS offices in Mogadishu.

APPENDIX E

SEBS Data Entry System User's Manual

A. Program flow
B. Diskette management
C. Program maintenance

A. Program flow

The data entry staff is presented with a series of data entry screens which permit the operator to transfer data from the questionnaires to the computer floppy diskette. Each screen is reproduced below in the sequence in which they occur. Some explanatory notes are included.

The data entry menu allows the operator to select the questionnaire. It controls for strap responses and calls the appropriate data entry program. Once a questionnaire is completed the data entry screen notes the composite identifier next to the questionnaire name (e.g. Individuals [illegible]).

```
U S A I D                    JESS PROJECT                    M J ' D
-------------------------------------------------------------------------
                    D A T A   E N T R Y
-------------------------------------------------------------------------

                    (A)  Individuals          [illegible]
                    (B)  Absence
                    (C)  Marriages
                    (D)  Household information
                    (E)  Household income
                    (F)  Livestock
                    (G)  Plots
                    (H)  Most Recent Harvest
                    (I)  Previous Harvest
                    (J)  Labor
                    (R)  Return to main menu
                                  Your Choice
```

Individual questionnaire data entry template

```
QUESTIONNAIRE CODE    1        DATE  00/00/00        INTERVIEWER        KT
REGION                DISTRICT                 BELL              VILLAGE
HOUSEHOLD             COMPOSITE IDENTIFIER
```

```
ID NUMBER            MARITAL STATUS              OCCUPATION
RELATION             DISTRICT OF BIRTH           SECOND OCCUPATION
AGE                  YEAR OF DEATH               FORAN LEVEL
SE                   SAME DOUYS                  SCHOOL LEVEL
                     SAME VILLAGE

READ                 ITALIAN              LAST SERIOUS ILLNESS
WRITE                OROM GALA            YEAR        SEASON      BEDFIDDEN
MAYMAY               AMHARIC              WHO SEEN (1)        WHO SEEN (2)
MAXAD TIRY           OTHER                TREATMENT 1         TREATMENT 2
SAWAHILI
MUSHONGOLI
ARABIC                                    NUMBER IMMUNIZATIONS
ENGLISH                                   TYPE OF IMMUNIZATION
```


24

Absence questionnaire data entry template

QUESTIONNAIRE CODE	1	DATE 00/00/00	INTERVIEWER	KF
REGION	DISTRICT	BEEL	VILLAGE	
HOUSEHOLD	COMPOSITE IDENTIFIER		ANY ABSENCE	

ID	SEND MONEY OR GOODS
AGE	TYPE OF AID
SEX	HOW OFTEN
LENGTH OF ABSENCE	
REASON FOR ABSENCE	RECEIVE MONEY OR GOODS
DISTRICT NOW RESIDING	TYPE OF AID
LIVING WITH	HOW OFTEN
WORK WITH HOUSEHOLD	

Marriage data entry template

QUESTIONNAIRE CODE	2	DATE 00/00/00	INTERVIEWER	KF
REGION	DISTRICT	BEEL	VILLAGE	
HOUSEHOLD	COMPOSITE IDENTIFIER			

SPOUSE ID	BRIDEWEALTH
LAST RESIDENCE	
IN SAME DISTRICT	CATTLE
MALE = LAST RESIDENCE	CAMELS
FEMALE = LAST RESIDENCE	SHEEP/GOATS
MALE = AGE	CASH
FEMALE = AGE	
Nth MARRIAGE FOR FEMALE	
Nth MARRIAGE FOR MALE	COST OF SETTING UP HOUSE
	YEAR OF MARRIAGE
	YEAR OF DIVORCE
	YEAR OF DEATH

26

Household data entry templates

```
QUESTIONNAIRE CODE    1        DATE  00/00/00        INTERVIEWER        KK
REGION                DISTRICT                LEEL              VILLAGE
HOUSEHOLD             COMPOSITE IDENTIFIER
```

```
# CONSUMPTION UNIT                  0.0
# PRODUCTION UNITS
WIPE CONTRIBUTES TO FOOD

MAIN SOURCE OF DRINKING WATER
                    OTHER SOURCE
                            WHAT
                             WHY
PAY FOR WATER
          WHEN
DISTANCE TO NEAREST MAIN WATER
```

```
        FACILITIES
    BATHE
    URINATE
    DEFECATE
      CHANGE
       WHERE
         WHY
```

```
                  OWN              LET              RENT
MUNDOULE            0
CARIICH
 # OF ROOMS
HOURIS
BARAAKU
 # OF ROOMS
SAR
 # OF ROOMS
OTHER

      # OF THE ABOVE WITH TIN ROOFS
```

GRAIN STORAGE FACILITIES

	#	CAPACITY
BAMBOO		
WOODS		
CLAY POTS		
TREES		
BARK		
OTHER		

HOUSEHOLD INVENTORY

Radios	Fish traps
Watches	Canoes
Bicycles	Spears
Tractors	Other fishing equipment
Vehicles	Other hunting equipment
Animal carts	Boats
Hand carts	Sheep
Bamboos	Chickens
Axes	Other fowl
Ranges	Draft donkeys
Fish lines/hooks	Burden donkeys
Gill nets	

27

CATTLE

LACTATING FEMALES
NON-LACTATING FEMALES
MATURE MALES (BULLS)
CASTRATED, MATURE MALES
IMMATURE FEMALES
IMMATURE MALES

NO. OF THE ABOVE DRAFT ANIMALS
NO. OF THE ABOVE BURDEN ANIMALS

CAMELS

LACTATING FEMALES
NON-LACTATING FEMALES
MATURE MALES (BULLS)
CASTRATED, MATURE MALES
IMMATURE FEMALES
IMMATURE MALES

NO. OF THE ABOVE DRAFT ANIMALS
NO. OF THE ABOVE BURDEN ANIMALS

OWNS FRUIT TREES

#

MANGOES
PAPAYA
BANANA
LIMES
AVOCADO
TAMARIND
GUAVA
OTHER

Household income data entry templates

QUESTIONNAIRE CODE	5	DATE	[illegible]	INTERVIEWER	[illegible]
REGION	DISTRICT		FEEL	VILLAGE	
HOUSEHOLD	COMPOSITE IDENTIFIER				

INCOME

AGRICULTURAL	PROPERTY SALES
LIVESTOCK	RENTAL INCOME
HUNTING GATHERING	REMITTANCES
CRAFTS	PRIDEWEALTH
BUSINESS	INHERITANCE
WAGE INCOME	GIFTS

HEARD OF LAND REGISTRATION	BAKTER	DO YOU FARM
REGISTERED YOUR LAND	WHAT FOR WHAT	OWN A FARM
IF, WHY NOT		

	Gu	Agaa	Dc r	Vilaal	Total
Ag inputs					
Live-lock					
Food					
Miscellaneous					
Housing					
Clothing					
Education					
Gifts Loan					
Fees/Taxes					
Other					

29

QUESTIONNAIRE CODE = DATE 00/00/00 INTERVIEWER HF
REGION DISTRICT BEEL VILLAGE
HOUSEHOLD COMPOSITE IDENTIFIER

ANY CAMELS ? # YEARS SINCE BEEFING CAMELS
EVER OWNED CAMELS WHY STOPPED
MAXIMUM NUMBER

IN THE LAST YEAR

CAMELS BORN # CAMELS SLAUGHTERED
CAMELS PURCHASED TYPE OF CAMELS
TYPE OF CAMELS # CAMELS DIED
TOTAL PRICE PAID TYPE DIED
WHAT CAUSED DEATH

CAMELS SOLD
TYPE OF CAMEL CAMELS GIVEN TYPE
TOTAL PRICE PAID CAMELS RECEIVED TYPE

ANY CATTLE ? # YEARS SINCE BEEFING CATTLE
EVER OWNED CATTLE WHY STOPPED
MAXIMUM NUMBER

IN THE LAST YEAR

CATTLE BORN # CATTLE SLAUGHTERED
CATTLE PURCHASED TYPE OF CATTLE
TYPE OF CATTLE # CATTLE DIED
TOTAL PRICE PAID TYPE DIED
WHAT CAUSED DEATH

CATTLE SOLD
TYPE OF CATTLE CATTLE GIVEN TYPE
TOTAL PRICE PAID CATTLE RECEIVED TYPE

```
ANY GOATS  0                    # YEARS SINCE KEEPING GOATS
    EVER OWNED GOATS                              WHY STOPPED
    MAXIMUM NUMBER
```

```
                    IN THE LAST YEAR

# GOATS BORN                          # GOATS SLAUGHTERED
# GOATS PURCHASED                     TYPE OF GOATS
TYPE OF GOATS                         # GOATS DIED
TOTAL PRICE PAID                      TYPE DIED
                                      WHAT CAUSED DEATH
# GOATS SOLD
TYPE OF GOATS                         GOATS GIVEN          TYPE
TOTAL PRICE PAID                      GOATS RECEIVED       TYPE
```

```
ANY SHEEP  0                    # YEARS SINCE KEEPING SHEEP
    EVER OWNED SHEEP                              WHY STOPPED
    MAXIMUM NUMBER
```

```
                    IN THE LAST YEAR

# SHEEP BORN                          # SHEEP SLAUGHTERED
# SHEEP PURCHASED                     TYPE OF SHEEP
TYPE OF SHEEP                         # SHEEP DIED
TOTAL PRICE PAID                      TYPE DIED
                                      WHAT CAUSED DEATH
# SHEEP SOLD
TYPE OF SHEEP                         SHEEP GIVEN          TYPE
TOTAL PRICE PAID                      SHEEP RECEIVED       TYPE
```

```
EXCHANGE ANY LIVESTOCK  0
TYPE OF EXCHANGE
AMOUNT OF EXCHANGE
```

```
QUESTIONNAIRE CODE    7        DATE   00 00 00        INTERVIEWER        KF
REGION                DISTRICT                 BEEL              VILLAGE
HOUSEHOLD             COMPOSITE IDENTIFIER
```

```
PLOT #
LAND TYPE
AREA
TENURE
HOW OBTAINED
YEAR
                         CROPS GROWN ON PLOT
 GRAIN  FRUIT  VEGETABLE  BEANS  OIL SEEDS  LISIM  COTTON  TABACCO

   IRRIGABLE PLOTS                          FALLOW
   WHICH ONE                                LAND TYPE
   OWNER                                    WHY
```

```
                                                  Num
QUESTIONNAIRE CODE    _        DATE   00 00 00        INTERVIEWER        KF
REGION                DISTRICT                 BEEL              VILLAGE
HOUSEHOLD          SEASON        YEAR          COMPOSITE IDENTIFIER
```

```
CROP/TREE                  BARTERED                  PURCHASED
AREA                       EXCHANGED                 LOCATION
INTERCROP                                            TYPE
PROPORTION                 GIFTS                     PRICE
LOSSES                                               VOLUME OF LARGEST
 CAUSE                     ZAKAT                      PURCHASE
PRODUCTION
RENT                       STORED
CONSUMED                   STORAGE LOSS
SOLD                       REASON
REVENUE                    RECEIVED
SALES
SALES     5
LOCATION            TYPE
VOLUME
PRICE
SEASON
```


32

QUESTIONNAIRE CODE	[illegible]	DATE	[illegible]	INTERVIEWER	[illegible]
REGION	DISTRICT		REFL	VILLAGE	
HOUSEHOLD	COMPOSITE IDENTIFIER			USED LABOR	

INFI	TYPE	SALARY PER DAY
SECTION	# OF PEOPLE	FOOD VALUE
	# OF DAYS	# OF MEALS

1. Diskette management

In order to reduce the loss of data the SEES data is divided among 5 diskettes for storage. There are two computer rooms and each has its set of diskettes. The data is further divided by region. Therefore each computer has a set of 21 diskettes, one per region and its backup copy. The diskettes are kept in their respective rooms.

The IFS checks to ensure that the correct diskette is placed in the correct disk drive before permitting data entry. In addition there is a data backup procedure for the data diskettes instituted in the SEI.

When the two disks per region, one from room 1 and one from room 2, need to be concatenated dBase III provides a simple procedure. The steps are as follows:

1. Start dBase III.
2. Put a blank formatted double density disk in drive B.
   Put the disk of the region (1, 2, or 3) from ROOM 1 in drive A.
4. Using the dBase Tools menu, copy, the appropriate file (e.g. LVSTCH01.dbf) from drive A to drive B, giving it the same name.
5. Choose the dBase Tools Menu and Set Drive to B.
6. Push the escape key to exit dBase (ESC).
7. Take the diskette out of drive A.
8. Put in the diskette from ROOM 2 and the same region in drive A.
9. At the dot prompt type USE LV_01_01 (for example).
10. Then type APPEND FROM A LVSTCH01 (for example).
11. When the drive lights are off type USE.
12. Type ASLIST.

You must follow this procedure for every one of the ten SLES files. However you may be able to fill all of one region's data on one disk. In that case you will not be putting a new formatted disk in drive B. You start with a new disk when you change regions.

2. Program maintenance

Maintenance of computer programs often amounts to 50% of the total life cycle cost of the program. In most cases these costs are the result of:

1) poorly documented code
2) inadequate analysis and design

A substantial amount of time was put into the analysis and design of the SEES data entry system. There are however many improvements that could be made. In the absence of a

34

ull-time computer programmer improvements will not be made However the programms may require maintenance by one of the JESS staff The rest of this report will explain, in detail, how to maintain the SEBS data entry system

The data entry system is composed of three parts 1) the data base file structures, 2) the data entry and verification programs, and 3) the data entry screen templates Each one will be examined and explained using the example of the individual unit of analysis data entry

1 Database file structure

The following structure has been retrieved from dBase III Plus, exactly as it would appear on the dBase III Plus screen The procedure for modifying the file structure is explained in the dBase III manual on page ****** The Dbase commands for structure modification are executed from the ASSIST mode by choosing Modify Database file N B That any changes made to the spelling of any filed will cause the data in that field in all the records to be irretrievably lost In addition, since dBase III Plus is not a true DBMS changes made to the file structure will require corresponding changes to the data entry program appropriate to that file

Structure for database 6 IND01 dbf
Number of data records 99
Date of last update 06/16/87

Field	Field Name	Type	Width	Dec
1	QUESCODE	Numeric	2	
2	DATE	Date	8	changing the width
3	INTRVIEWER	Numeric	2	is permissible and
4	RE FUNCHER	Numeric	1	requires programming
5	REGION	Character	1	changing the type
6	DISTRICT	Character	1	is permissible but
7	BELL	Character	2	could cause loss of
8	VILLAGE	Character	1	some data and
9	HOUSEHOLD	Character	2	requires programming
10	COMP ID	Character	7	changing the Field
11	ID NO	Character	2	Name causes loss
12	RELATION	Character		of all data for
13	AGE	Numeric	2	that field or it
14	SEX	Character	1	requires programming
15	MARITALSTA	Character	1	
16	DIST_BIRTH	Character	2	
17	YEAR_DEATH	Numeric	2	
18	SAMEQOYS	Character	1	
19	SAMEVILL	Character	1	
20	OCCUPATION	Character	7	
21	SECOCUPTON	Character	2	
22	KOFANLEVEL	Character	1	

35

23	SCHOOLEVEL	Character	2
24	READ	Character	1
25	WRITE	Character	1
26	MAYMAY	Character	1
27	MAXAD TIPT	Character	1
28	SAWAMILI	Character	1
29	DU MONCOLT	Character	1
30	ARABIC	Character	1
31	ENGLISH	Character	1
32	ITALIAN	Character	1
33	OPON GALA	Character	1
34	AMHARIC	Character	1
35	OTHER	Character	1
36	LAST2EFILL	Character	
37	YEAR	Numeric	2
38	SEASON	Character	1
39	BEDRIDDEN	Character	1
40	WHOSEEN	Character	
41	SPOSEEN	Character	2
42	TREATMENT	Character	1
43	SECTREAT	Character	1
44	IMMUNIZ	Numeric	1
45	TYPEIMMUNIZ	Character	1
** Total **			75

The following block of text contains an annotated section of one of the data entry programs written in dBase III Plus.

Due to space restrictions the annotations will be identified by ** and will refer to the line number if not placed on the same line. N.B. dBase III + programs do not have line numbers, they are added here for clarity.

In order to make changes you should read this entire program first. Then to reference specific changes use this line number directory.

change	line
change acceptable codes	74-78, [illegible], 105, [illegible]
change acceptable region numbers	70
change variable names	51-61
change variable sizes or types	1-8

```
1     * EINDIV.PRG
      SET SAFETY OFF
2     SET DELIMITERS TO DEFAULT
4     SET DELIMITERS ON
5     PRIVATE err, choice, indi , flag1, flag2, filen
6     STORE              TO mindi
      flag1 =  N
7     DO WHILE flag1 =  N
```

```
        CLEAR
10      filen =
11      @ 21,5 SAY  *   to leave the program
12      err = 1
13      DO WHILE err
*/ check for correct region number/*
14      @ 10,27 SAY  Number of Region      GET filen PICTURE
          *
15      @ 13,5 SAY  Regions are
16      @ 1 ,5 SAY  1
17      @ 16 5 SAY  2
18      @ 19,5 SAY  3
19      READ
20         IF filen = 1   OR  filen = 2   OR  filen = 3
              OR  filen = *
21            err    *
22         ENDIF
23      ENDDO filen
24      IF filen = *
25         CLEAR
26         RETURN
        ELSE
*  create the predetermined file name using the region  *
*  number entered check to see if the correct disk is  n /*
   drive n /*
28            filen = 2 INDO  filen *   OUT
29               IF FILE(filen)
30                  USE filen
31                  flag1 = 1
32                  CLEAR
33               ELSE
34                  @ 21,5 CLEAR
35                  7 CHR(7)+CHR(7)
36                  @ 21,5 SAY  The disk in drive B is not
37                  for Region
38                  @ 21,47 SAY SUBSTR(filen,3,1)
39                  WAIT  Press any key to continue
40               ENDIF
41      ENDIF
*    ENDDO

*  The following line initiate all values to zero in the /*
*/ file  Note that if the temporary variables are numeric /*
*/ there are no " " around the value  If the variables    /*
*/ character the exact number of  zeroes must be used      /*
*/ between the " "                                         /*
*/ If you change the width or type of a variable in the   /*
*  dBase III structure you must make the appropriate      /*
*  changes here! The e3 corresponds to the true variables /*
*/ in the dBase structure  You have to figure out which of/*
*/ the e3 represents which true variable                  /*
*  These temporary variables are for the IDENTIFIER only  /*

43   STORE 0 TO a1 a2 kp
```

37

```
44    STORE 1 to a1
45    STORE  00,00,00  TO a2
46    STORE  0  TO a4,a5,a7
47    STORE SUBSTR(filen,7,1) TO a4
48    STORE  00  TO a6,a8
49    STORE  0000000  to a9
50    CLEAR

*  this is the start of the data entry loop                    *

51    flag1 = 0
52    DO WHILE flag1 =  Q
53       @ 0 1 CLEAR
54       flag2 =  C
55       DO EINDIV FNT
56          indiv =  1
57       DO WHILE flag2 =  C

*  line 55 calls a sub program READID which permits the  *
*  entry of the file's identifier                          *

58          DO READID
59          err =  T
60          DO WHILE err
61             @ 24,20 Sn  (C) Correct, (Q) Quit, (S) Save
62    GET flag3 PICTURE  1
63             READ
64             IF flag2 =  Q  OR  flag2 =  C  OR  flag2 =
65               S
66                err =  F
67             ENDIF
68          ENDDO err1
69          @ 24,0 CLEAR
70       ENDDO flag2
71       DO WHILE flag3 =  S
72          flag2 =
73          DO WHILE indiv =  1
*/ The following lines initiate all values to zero in the  *
*  file  Note that if the temporary variables are numeric /*
*  there are no " " around the value  if the variables    /*
*  character the exact number of zeroes must be used      /*
*/ between the " "                                          *
*  If you change the width or type of a variable in the   /*
*/ dBase III structure you must make the appropriate      /*
*  changes here! The e3 corresponds to the true variables /*
*/ in the dBase structure  You have to figure out which of/*
*/ the e3 represents which true variable                  /*

74             STORE 0 TO e3,e7,e27,e34
75             STORE  0  TO e4,e5,e9,e9,e13,e14,e15,e16,e17
76             STORE  0  TO e18,e19,e20,e21,e22,e23,e24
77             STORE  0  TO e25,e28,e29,e32,e33,e35
78             STORE  00  TO e1,e2,e6,e10,e11,e13,e26,e30,e31
79             choice=  C
```


38

```
            DO EINDIV FIN
            @ 1,21 SAY a1 PICTURE  79
            @ 1,27 SAY a2 PICTURE  xxxxxxxx
            @ 1,33 SAY a3 PICTURE  79
            @ 1,34 SAY kp PICTURE  F
            @ 2,14 SAY a4 PICTURE  x
            @ 2,22 SAY a5 PICTURE  C
            @ 2,47 SAY a6 PICTURE  79
            @ 2,67 SAY a7 PICTURE  F
            @ 2,11 SAY a8 PICTURE  79
            @ 2,44 SAY a9 PICTURE  ==  ==
            DO WHILE choice = 0
               @ 3,15 GET e1 PICTURE  99
            IF LEFT(e1,1) =
                  e1 = 0 + e1
               ENDIF
*  This is the start of a typical range or value check    *
*  one of the variables. Essentially a loop is executed /*
*  until the correct value is entered. In this case the   *
*  e2 is a character variable with a width of two thus   /*
*  the PICTURE  99  which accepts two numeric characters/*
*  only. The value of the characters must be = 0 or  44/*
*  otherwise the loop continues until the response is     *
*  within range. In order to add a new valid response     *
*  eg the response 44 you would use NE WORD and change   /*
*  the 44 to 45. Please see endnotes                      *

               err = T
               DO WHILE err
                  @ 7,15 GET e2 PICTURE  99
                  READ
                  IF VAL(e2) = 0 AND VAL(e2) 44
                     err = F
                  ENDIF
               ENDDO err2
               @ 8,15 GET e3 PICTURE  ==
               READ
               err = T
               DO WHILE err
                  @ 9,15 GET e4 PICTURE  9
                  READ

*/ This is a special case, e4 is a character variable of    *
*  width one. The 1 permits you to include all one         /*
*  character responses -or e4. dBase III checks to make    *
*/ sure that e4 is one of them                             /*

                  IF e4 $ 12
                     err = F
                  ENDIF
               ENDDO err4
               err = T
               DO WHILE err
                  @ 3,42 GET e5 PICTURE  9
```

39

```
            READ
            IF VAL(e5)   0  AND  VAL(e5)   5
               err =  F
            ENDIF
         ENDDO err5
         @ 7,42 GET e6 PICTURE  ==
         READ
         @ 3,42 GET e7 PICTURE  ==
         READ
         err =  T
         DO WHILE err
            @ 3,42 GET e7 PICTURE  =
            READ
            IF  e1(e8)   4
               err =  F
            ENDIF
         ENDDO err8
         err =  T
         DO WHILE err
            @ 10,42 GET e9 PICTURE  =
            READ
            IF val(e9)   1
               err =  F
            ENDIF
         ENDDO err9
         err =  T
         DO WHILE err
            @ 7,68 GET e10 PICTURE
            READ
            IF  VAL(e1 )   = 0  AND  VAL(e10)   1
                OR  VAL(e10)   97
               err =  F
            ENDIF
         ENDDO err10
         err =  T
         DO WHILE err
            @ 7,12 GET e11 PICTURE  =
            READ
            IF VAL(e11)   = 0  AND  VAL(e11)   1
                OR  VAL(e11)   97
               err =  F
            ENDIF
         ENDDO err11
            @ 8,82 GET e12 PICTURE  -
            READ
         err =  T
         DO WHILE err
            @ 9,68 GET e13 PICTURE  -
            READ
            IF VAL(e13)   0  OR  VAL(e13)   95
               err =  F
            ENDIF
         ENDDO err13
         err =  T
```


40

```
DO WHILE err
   @ 12,15 GET e14 PICTURE
   READ
   IF VAL(e14)  4  OR  VAL(e14)  7
      err = F
   ENDIF
ENDDO err14
err = T
DO WHILE err
   @ 13,15 GET e15 PICTURE
   READ
   IF VAL(e15)  4  OR  VAL(e15)  7
      err = F
   ENDIF
ENDDO err15
err = T
DO WHILE err
   @ 14,15 GET e16 PICTURE
   READ
   IF VAL(e16)  4  OR  VAL(e16)  7
      err = F
   ENDIF
ENDDO err16
err = T
DO WHILE err
   @ 15,15 GET e17 PICTURE
   READ
   IF VAL(e17)  1  OR  VAL(e17)
      err = F
   ENDIF
ENDDO err17
err = T
DO WHILE err
   @ 16,15 GET e18 PICTURE
   READ
   IF VAL(e18)  4  OR  VAL(e18)  7
      err = F
   ENDIF
ENDDO err18
err = T
DO WHILE err
   @ 17,15 GET e19 PICTURE
   READ
   IF VAL(e19)  4  OR  VAL(e19)
      err = F
   ENDIF
ENDDO err19
err = T
DO WHILE err
   @ 18,15 GET e20 PICTURE
   READ
   IF VAL(e20)  4  OR  VAL(e20)
      err = F
   ENDIF
```


41

```
                ENDDO err20
                err = T
                DO WHILE err
                   @ 1-,15 GET e21 PICTURE  9
                   READ
                   IF VAL(e21)   4  OR  VAL(e21)   -
                      err = F
                   ENDIF
                ENDDO err21
                err = T
                DO WHILE err
                   @ 12,24 GET e22 PICTURE  -
                   READ
                   IF VAL(e22)   4  OR  VAL(e22)   -
                      err = F
                   ENDIF
                ENDDO err22
                err = T
                DO WHILE err
                   @ 12,24 GET e23 PICTURE  -
                   READ
                   IF VAL(e23)   4  OR  VAL(e23)   7
                      err = F
                   ENDIF
                ENDDO err23
                err = T
                DO WHILE err
                   @ 14,24 GET e24 PICTURE  9
                   READ
                   IF VAL(e24)   4  OR  VAL(e24)
                      err = F
                   ENDIF
                ENDDO err24
                err = T
                DO WHILE err
                   @ 15,24 GET e25 PICTURE  9
                   READ
                   IF VAL(e25)   4  OR  VAL(e25)
                      err = F
                   ENDIF
                ENDDO err25
                err = T
                DO WHILE err
                   @ 12,64 GET e26 PICTURE  9
                   READ

*  In this case e26 is numeric and therefore a simple    /*
*  numeric comparison is used  There is no need for     /*
*/ VAL(  ) in this case                                  /*

                   IF VAL(e26)   64  OR  VAL(e26)   97
                      err = F
                   ENDIF
                ENDDO err26
```

```
IF e[illegible] # [illegible] OR e[illegible] # 0
      err = T
      DO WHILE err
         @ 1[illegible],4 GET e[illegible] PICTURE [illegible]
         READ
         IF e[illegible] [illegible] OR e[illegible] =
            err = F
         ENDIF
      ENDDO err[illegible]
      err = T
      DO WHILE err
      @ 1[illegible],35 GET e[illegible] PICTURE [illegible]
      READ
         IF VAL(e[illegible]) [illegible]
            err = F
         ENDIF
      ENDDO err[illegible]
      err = T
      DO WHILE err
      @ 1[illegible],[illegible] GET e[illegible] PICTURE [illegible]
      READ
         IF VAL(e[illegible]) [illegible]
            err = F
         ENDIF
      ENDDO err[illegible]
      err = T
      DO WHILE err
      @ 1[illegible],[illegible] GET e[illegible] PICTURE [illegible]
      READ
         IF VAL(e[illegible]) [illegible]
            err = F
         ENDIF
      ENDDO err[illegible]
      err = T
      DO WHILE err
      @ 1[illegible],[illegible] GET e[illegible] PICTURE [illegible]
      READ
         IF VAL(e[illegible]) [illegible]
            err = F
         ENDIF
      ENDDO err[illegible]
      err = T
      DO WHILE err
      @ 15,[illegible] GET e[illegible] PICTURE [illegible]
      READ
         IF VAL(e[illegible]) [illegible]
```

```
                        err = F
                    ENDIF
                ENDDO err[illegible]
            ELSE
                [illegible] CHR(7)+CHR(7)
                [illegible] = [illegible]
            ENDIF [illegible] check
                @ 18,[illegible] GET [illegible] PICTURE [illegible]
                READ
            IF [illegible] 4 [illegible] 0
                @ 1[illegible],[illegible]4 GET [illegible] PICTURE [illegible]
                READ
            ELSE
                [illegible] CHR(7)
                [illegible] = [illegible]
                @ 1[illegible],[illegible]4 [illegible] PICTURE
            ENDIF
        err = [illegible]
        DO WHILE err
        @ [illegible]4,[illegible]0 SAY [illegible] (C)orrect [illegible] (Q)uit [illegible] (S)a[illegible]
          GET choice PICTURE
        READ
          IF choice [illegible] [illegible]
            err = F
                IF choice [illegible] [illegible]
                APPEND BLANK
*  Here we replace the true variables in the dbase III  *
*  structure with the values that we have [illegible] or  /*
*   [illegible] the names in the structure have been    /*
*/ changed here then they need to be changed here      /*

                REPLACE questcode WITH a1, date WITH
                 [illegible](a[illegible]), intrviewer WITH a[illegible],
                 keypuncher WITH [illegible],
                 region WITH a4, district WITH a5,
                 [illegible] WITH a[illegible]
                REPLACE village WITH a7, household
                 WITH a8
                REPLACE comp id WITH a4+a5+a[illegible]+a[illegible]
                REPLACE id no WITH e1
                REPLACE relation WITH [illegible],
                 age WITH e[illegible] [illegible] WITH e4, maritalsta
                 WITH e5
                REPLACE dist_birth WITH e6, year death
                 WITH e7, samego[illegible] WITH e8,
                 same ill WITH e9, occupation WITH
                 e10, secocupton WITH e11
                REPLACE koranlevel WITH e12,
                 schoolevel WITH e13, read WITH e14,
                 write WITH e15, maymay WITH e1[illegible],
                 ma[illegible]ad tiry WITH e17
                REPLACE sawahili WITH e18, mushungali
                 WITH e19, arabic WITH e[illegible],
```

```
                    english WITH e21, italian  WITH e22,
                    orom gala WITH e23
                  REPLACE amharic  WITH e24, other WITH
                    e25, lastserill WITH e26,,
                    ear WITH e27, season WITH e28,
                    bedriggen WITH e29
                  REPLACE whoseen WITH e30, secseen WITH
                    e31, treatment WITH e32,
                    sectreat WITH e33, immunic WITH e34,
                    i peimuni  WITH e35
                  ENDIF
               ENDIF
            ENDDO err replace
            @ 24,0 CLEAR
         ENDDO chp
         err = T
         DO WHILE err
         @ 24,1 SAY  Continue with another individual
          (Y/N)   GET indiv PICTURE  Y
         READ
            IF indiv = Y  OR  indiv = N
               err = F
            ENDIF
            IF indiv = N
               STORE a  TO mindiv
               SAVE ALL LIKE m* TO mindiv
            ELSE
               CLEAR
            ENDIF
         ENDDO err
         @ 24,0 CLEAR
      ENDDO indiv
   ENDDO flag2
   @ 24,0 CLEAR
   IF indiv = N
      flag1 =
   ENDIF
ENDDO flag1
CLEAR
CLOSE DATABASES
RETURN
```

ENDNOTES

1 Before making any changes be sure to have a backup copy of the programs

2 **Use MS WORD to make the changes BE SURE to save the file UNFORMATTED, otherwise it will not work!!!!!**

3 Test the program with sample data

4 In order to add variables to the questionnaire you will have to

a add the new variables to the dBase structure

b add the appropriate new lines of code to each of the programs that require it

45

c redesign the data entry screens for the appropriate questionnaire

d This appendix cannot adequately teach you all you need to know to do these things

46

APPENDIX F

Preliminary SEBS Analyses

Although an assessment of the data currently keypunched indicates that data should be re-entered, the data were used for test runs of the statistical analysis package, Statpac. Statpac operates via an easy-to-use menu-driven system. The researcher prepares a codebook describing the data, the data themselves (transferred from dBase III in this case), and an analysis file

The consultant suggested that the SEBS staff prepare a list of 10 "analyses" to perform on each of the 10 questionnaires in the SEBS questionnaire packet The following lists are preliminary analyses suggested by the Team

Q1--Individual	Table 1 - Number of people in the HH by age/sex
	Table 2 - Marital status by age/sex
	Table 3 - Major work performed by age/sex
	Table 4 - Koranic level by age/sex
	Table 5 - School level by age/sex
	Table 6 - Literate read or write by age/sex
	Table 7 - Frequency of illnesses by age/sex and district
	Table 8 - Frequency of medical help sought (both first and second)
	Table 9 - Frequency of treatments given
	Table 10 - Average number of immunizations per person
Q2--Absence	Table 1 - Frequency of absences by age grp (age grps: 0-9, 10-15, 16-30; 30+)
	Table 2 - Frequency of absence by sex
	Table 3 - Average length of absence by age
	Table 4 - Average length of absence by sex

47

	Table 5 - Frequency of reasons for absence
	Table 6 - Frequency of those absent work HHH enterprise
	Table 7 - Frequency of aid to HHH
	Table 8 - Frequency of aid type
	Table 9 - Frequency of aid sent by HHH
	Table 10 - Frequency of aid type sent by HHH
Q3--Marriage	Table 1 - Frequency of same village
	Table 2 - Did male move districts since?
	Table 3 - Did female move districts since?
	Table 4 - Average age of marriage for 1st, 2nd, 3rd, and 4th marriages by sex
	Table 5 - Average number of marriages by sex
	Table 6 - Average number of divorces by sex
	Table 7 - Average number of widowhoods by sex
	Table 8 - Total bridewealth by year wed
Q4--Household	Table 1 - Frequency of water source
	Table 2 - Average distance to water
	Table 3 - Average number of latrines
	Table 4 - Frequency of house types
	Table 5 - Frequency of storage type
	Table 6 - Summary table of average numbers of household items #26-70

Q5--Income	Table 1 - Average total income by source
	Table 2 - Average expend by group
	Table 3 - Average net income
Q6--Livestock	Table 1 - *

*To be continued by JESS staff

APPENDIX G

Analysis of Vegetation Plots

Dr. Ian Deshmukh provided the consultant with a list of information he would like extracted from the vegetation plot data His list is reproduced here along with.

1) dBase III + structure of the file
2) Statpac codebook of the file
3) Statpac analysis control file required to complete the listed analysis (to the extent possible)

For each plot.

1. For each species (within plot)

total stem area
stem area of trees
stem area of shrubs

total canopy area
canopy area of trees
canopy area of shrubs
canopy area by height classes overall
canopy area by height classes of trees
canopy area by height classes of shrubs

descriptive statistics (mean, s, median, etc , etc.) for

height overall
height of trees
height of shrubs
height overall >2m
height trees >2m
height shrubs >2m

mean canopy diameter overall
mean canopy diameter trees
mean canopy diameter shrubs
mean canopy dia overall >2m
mean canopy dia trees >2m
mean canopy dia. shrubs >2m

2. For each plot (all species)

all same things as per species
proportion of total canopy area as trees
proportion of total canopy area as shrubs
proportion of total canopy area for each species

overall, as trees, as shrubs
proportion of individuals, trees
proportion of individuals, shrubs
proportion of individuals by species

3 Another plot sorting variable for vegetation type (e.g , forest, woodland, bushland, bushed grassland, etc).

Structure for database· A:FOREST.dbf
Number of data records 0
Date of last update : 06/19/87

Field	Field Name	Type	Width	Dec
1	PLOT_NBR	Character	2	
2	QUADRANT	Character	1	
3	PLOT_ID	Character	3	
4	SIZE	Numeric	3	
5	FLOODPLAIN	Character	1	
6	PLOT_TYPE	Character	1	
7	COMBINE	Character	2	
8	SPECIES	Character	8	
9	PLOT_SPEC	Character	10	
10	TREE_SHRUB	Character	1	
11	STEM_DIA	Numeric	5	1
12	CRWN_DIA_A	Numeric	4	1
13	CRWN_DIA_B	Numeric	4	1
14	CRWN_START	Numeric	4	1
15	HEIGHT	Numeric	4	1
16	STEMS	Numeric	2	
17	MEAN_C_DIA	Numeric	4	1
18	STEM_AREA	Numeric	7	3
19	CROWN_AREA	Numeric	5	1
20	CROWN_DPTH	Numeric	4	1
21	CROWN_VOL	Numeric	7	1
22	HGHT_CLASS	Numeric	1	
23	COVERCLASS	Numeric	1	
24	VEG_TYPE	Character	2	
** Total **			87	

Statpac codebook for vegetation study

```
FIXED                    * key
PLOT NUMBER              * variable name
1                        * starting column
2                        * width in characters
A                        * variable type Alphanumeric/Numeric
@                        * end of variable marker
```

```
QUADRANT
3
1
A
@
PLOT IDENTIFIER (PLOT NUMBER + QUADRANT)
4
3
A
@
PLOT SIZE (SQUARE METERS)
7
3
N
@
FLOODPLAIN
10
1
A
Y=YES
N=NO
@
PLOT TYPE
11
1
A
W=
L=
D=
d=
H=
A=
O=
@
COMBINE (FLOODPLAIN + PLOT TYPE)
12
2
A
YW=
YL=
YD=
Yd=
YH=
YA=
YO=
NW=
NL=
ND=
Nd=
NH=
NA=
NO=
@
```

```
SPECIES
14
8
A
@
PLOT SPECIFICATION (PLOT NBR + SPECIES)
22
10
A
@
TREE OR SHRUB
32
1
A
T=TREE
S=SHRUB
@
STEM DIAMETER (CENTIMETERS)
33
5
N
@
CROWN DIAMETER A (METERS)
38
4
N
@
CROWN DIAMETER B (METERS)
42
4
N
@
CROWN STARTING HEIGHT (METERS)
46
4
N
@
HEIGHT (METERS)
50
4
N
@
NUMBER OF STEMS
54
2
N
@
MEAN CROWN DIAMETER (METERS)
56
4
N
@
```

Examples

Both dBase III and Lotus 1-2-3 version 2 provide easy means for converting to each other's formats. Here are examples of the other methods listed above.

Lotus 1-2-3 to ASCII

1) Load the Lotus file to be transferred to ASCII
2) Using the Lotus command menu, choose Print File.
3) Assign a name to the file (it's best to use the same name under which the file is currently stored, Lotus will attach an extension of .PRN).
4) Delimit the Range to be converted.
5) Choose Options Other Unformatted.
6) Set the Left, Top, and Bottom margins to 0 and the Right margin to the total width of the cells to be converted
7) Execute the Align and Go commands
8) After leaving Lotus rename the file so that the extension is .TXT (to permit transfer to dBase III, etc)

dBase III to ASCII

1) Load the dBase III file, using ASSIST or the USE command.
2) Leave dBase ASSIST
3) Type COPY TO filename SDF (choose a filename, including a drive specification if desired) N.B.· dBase will assign an extension of TXT to the file.

ASCII to Lotus 1-2-3

1) Make sure the text file contains numbers only or letters only and that its extension is PRN.
2) Start Lotus 1-2-3.
3) Place the Lotus cursor in the upper left corner of the range in which the ASCII file will be placed
4) Choose File Import from the command menu
5) Choose Text or Numbers as appropriate.
6) Choose the file name if it appears in the list or enter it including the drive specification.
7) Do not forget to save the file once the Import is completed

ASCII to dBase III

1) Make sure the ASCII file has the extension TXT
2) Prepare a dBase III structure that exactly matches that of the ASCII file to be converted This preparation is done using the Create command. Often the only way to do this is to print the ASCII file and count the number of spaces occupied by each variable.
3) Answer N to Enter data now (Y/N)?
4) If necessary, exit Assist.

5) Type APPEND FROM filename txt SDF
   (filename txt is the name of the ASCII file)
6) Exit dBase III.