

Prepared for:

Office of Population
Bureau of Science and Technology
Agency for International Development
Washington, D.C.
Under Contract No. DPE-3024-C-00-4063-00
Project No.: 936-3024

PN AAR-951

38401

PROCEEDINGS OF EXPERT PANEL ON MICROCOMPUTER
SOFTWARE FOR SURVEY AND CENSUS PROCESSING

by

James C. Knowles, Ph.D.

December 3-4, 1984

Edited and Produced by:

Population Technical Assistance Project
International Science and Technology Institute, Inc.
2033 M Street, N.W., Suite 300
Washington, D.C. 20036
Phone: 202/466-7290
Telex 275445/TSI UR

Report No.: 84-19-008
Published April 5, 1985

TABLE OF CONTENTS

ABBREVIATIONS IN THE REPORT	ii
BACKGROUND FOR PROCEEDINGS OF EXPERT PANEL ON MICROCOMPUTER SOFTWARE DEVELOPMENT FOR SURVEY AND CENSUS PROCESSING	iii
EXECUTIVE SUMMARY	v
PROCEEDINGS OF EXPERT PANEL ON MICROCOMPUTER SOFTWARE DEVELOPMENT FOR SURVEY AND CENSUS PROCESSING	1

APPENDICES

Appendix A: List of Participants

Appendix B: Agenda

ABBREVIATIONS IN THE REPORT

AID	U.S. Agency for International Development
BuGen	Bureau of the Census
CAI	Computer-aided Instruction
CPS	Contraceptive Prevalence Survey
DDD	Demographic Data for Development
DHS	Demographic and Health Surveys
DP	Data Processing
INPLAN	Integrated Population and Development Planning II
LDC	Less Developed Country
MSH	Management Sciences for Health
NAS	National Academy of Sciences
PRITECH	Primary Health Care Technology
RTI	Research Triangle Institute
SAS	Statistical Analysis System
SPSS	Statistical Package for the Social Sciences
UPS	Uninterruptible Power Supply
WFS	World Fertility Survey

BACKGROUND FOR PROCEEDINGS OF EXPERT PANEL ON MICROCOMPUTER
SOFTWARE DEVELOPMENT FOR SURVEY AND CENSUS PROCESSING*

Microcomputers have become an important tool for many projects supported by the Office of Population. They have been used in managing family planning programs, collecting and analyzing demographic data, projecting population growth, modeling population-development relationships, and graphically displaying these relationships.

There are many characteristics of the microcomputer that make it valuable in less developed country environments:

- sophistication and power: it can perform detailed data analysis functions
- portability: it has low shipping costs and can be hand-carried by short-term technical advisors
- user-friendliness: sophisticated operations can be performed by individuals without extensive training in computer programming
- interactive operations: the user can enter data and job commands encountered with batch processing
- environmental tolerance: it can function in wide-ranging conditions with minimal climate-control equipment
- low maintenance costs: less developed country institutions can more easily assume recurrent expenses, and
- low fixed costs: the technology can be provided at relatively low cost.

Recent technology advances have made it feasible to use microcomputers rather than mainframe systems for many data entry, editing, tabulation and analysis tasks. In the past few years, microcomputers have been used in several less developed countries to carry out entire census and survey processing activities. As the technology develops further, it is clear that microcomputers can play an even greater role in data processing work, thereby reducing fixed and recurrent computer costs and improving the speed with which data becomes available.

In response to these developments, AID and other donor organizations are exploring options for broader application of microcomputer technology to enable more efficient and cost-effective processing of census and survey data. For AID's Office of Population, this exploration is part of a larger planning exercise as the Office prepares for the 1990 round of censuses and implements a new round of demographic and family planning surveys under the Demographic and Health Surveys

* This background paper was supplied by Dr. John Crowley, Social Science Analyst, Policy Development Division, Office of Population, AID.

(DHS) Project. To assist this exploration, a panel of microcomputer experts was assembled to advise AID and other donor organizations on existing microcomputer technology and future hardware and software development needs for census and survey data processing. In particular, the panel of eight experts was asked to address five crucial issues:

1. What are the data processing needs for censuses and surveys over the next five-to-ten years?
2. What are the microcomputer hardware requirements for processing census and survey data?
3. What are the microcomputer software requirements for processing census and survey data?
4. What new developments in microcomputer technology are forthcoming which may change the way in which data collection activities are undertaken?
5. What are the training and support requirements that will accompany the adoption of microcomputers for census and survey data processing work?

This report summarizes the discussions and recommendations of the panel's deliberations.

Executive Summary

1.1 Purpose of Meeting

The Panel was convened for the purpose of reviewing both existing and projected future developments in microcomputer hardware and software suitable for processing and analyzing censuses and surveys in Less Developed Countries (LDCs) in order to identify possible gaps which would need to be filled either by new software development or by adapting existing mainframe software for use on micros. The Panel consisted of representatives from AID, the U.N., the World Bank, AID cooperating agencies and consultants. The meeting was held from December 3-4, 1985, at the National Academy of Sciences' Joseph Henry Building.

Introductory remarks were made by Mr. John Dumm (on behalf of Mr. Steven Sinding, Director of AID's Office of Population) and by Mr. Richard Cornelius, also of AID. They explained that AID's interest in microcomputer technology stems from its potential for reducing costs of providing assistance in the area of demography and from its potential for reducing the turnaround time in processing census and survey data. They also said that the meeting was well timed in that it coincided with the start-up of the new Demographic and Health Surveys (DHS) project and with initial planning for AID's assistance to the 1990 round of censuses.

1.2 Conclusions Regarding the Practicality of Using Microcomputers

The Panel reviewed census and survey processing needs over the next decade and concluded that it would be practical to process the majority of LDC censuses, particularly in Africa and Latin America, using micros. The Panel also concluded that it would be practical to process most surveys with micros, including the 35 surveys to be carried out over the next three years under the DHS project. The Panel's conclusions regarding the practicality of using micros for these purposes was based in part on its review of several successful past attempts to use micros for such work and on its projection of future hardware and software trends.

Based on its review of past experience in using micros for data processing, the Panel felt that the use of micros would not only lower costs and improve turnaround time but would also be likely to improve the quality of the data collected. The direct involvement of analysts in a decentralized data processing system was cited as the principal reason for expecting improved quality with the use of micros.

1.3 Hardware Selection

The Panel felt that there was a need for some hardware standards to be adopted for census and survey processing. There was a broad consensus that the IBM-PC/XT/AT line of computers (as well as compatibles) is an appropriate standard. However, there was less agreement on the exact model and memory configuration appropriate for census and survey processing. In addition, there was considerable interest expressed by some Panel members in the potential usefulness of inexpensive lap-size computers, such as the Radio Shack Model 100. The Panel felt very strongly, however, that software is a much more important constraint than hardware; and accordingly, the Panel devoted most of its time to a review of the available software.

1.4 Software Selection

Adaptation of Existing Software. The Panel was in agreement that there is little need for AID to fund the development of any new microcomputer software. Instead, the Panel favored adapting commercial software for particular applications and downloading selected mainframe software, such as CENTS4 and CONCOR, for use on micros. With respect to the latter, the Panel felt that some programming might be worthwhile to develop an interactive front-end for mainframe software which is presently batch-oriented.

Stages of Census and Survey Processing. The Panel reviewed the availability of micro software for performing various stages of census and survey processing, including sample selection, data entry, editing, data displays/browsing, tabulation, management reporting, graphics, statistical analysis, demographic analysis, recodes/file manipulation, data communications and tutorials. Sample selection and editing were identified by the Panel as the areas in which there is presently no adequate micro software available. The Panel urged that CONCOR, or its equivalent, be implemented on a micro for editing and that efforts be made to identify a suitable mainframe sampling package for downloading to micros.

The Panel felt that an eclectic approach is appropriate in selecting software, depending on the requirements of each application. At the same time, however, the Panel stressed the importance of using software packages for data entry which allow the data dictionary to be encoded along with the data, so that the work of defining such a dictionary would not have to be repeated when different programs are used. The Panel also felt that certain commercial relational data base management programs, such as DBASE III and KnowledgeMan, could be potentially quite useful for purposes of data entry, editing and file manipulation.

1.5 New Directions

The Panel discussed several new frontiers in the use of micros for census and survey processing. The topic most discussed was the possibility of using lap-size micros at the interview level. Although the Panel identified several potential problems with such an approach, they nevertheless felt it was worthy of selective experimentation. The Panel also thought it would be useful to experiment with an alternative approach in which the data would be entered into a lap-size computer immediately after the actual interview, so that any apparent inconsistencies found by the computer could be resolved by the interviewer before leaving the interview site.

1.6 Additional Issues

The Panel also considered the areas of computer training, service and maintenance, and problems related to electric current. In the case of training, the Panel felt that a broad approach to microcomputer training would yield a high return over the next decade and would be more cost-effective than a narrow, project-oriented approach to training. At the same time, the Panel was in favor of experimenting with more cost-effective approaches to microcomputer training than the usual one-on-one approach used in the past. Computer-assisted tutorials, presently available with leading commercial packages, were regarded as a desirable trend; but the Panel recognized the need to have them (or similar materials) available in local languages.

In the case of service and maintenance, the Panel did not consider local service contracts to be necessarily the best approach due to high cost and undetermined reliability. As an alternative, the Panel considered providing host country institutions with a supply of spare parts (e.g., disk drives, memory chips) and training their staff to make the kind of routine repairs that are most commonly required with microcomputers. The Panel stressed the need to test all equipment thoroughly before it is shipped to developing countries.

Lastly, with respect to electric current, the Panel did not consider that the presence of either 220-volt and/or 50-cycle current in some LDCs to be an obstacle to their use of U.S.-procured computer equipment. However, the Panel felt that it was very important to protect such equipment from poor quality current by using either voltage regulators or, where necessary, uninterruptible power supplies.

PROCEEDINGS OF EXPERT PANEL ON MICROCOMPUTER SOFTWARE
DEVELOPMENT FOR SURVEY AND CENSUS PROCESSING
(December 3-4, 1984)

National Academy of Sciences
Joseph Henry Building, Room 355
2122 Pennsylvania Avenue, Washington, D.C.

1. Introduction

Introductory remarks were made by Mr. John Dumm, speaking on behalf of Mr. Steven Sinding, Director of AID's Office of Population, and Mr. Richard Cornelius, also of AID's Office of Population, who served as the Panel's Chairperson.

Mr. Dumm explained that the funds available for population activities had been shrinking in real terms over the past few years and that this trend was expected to continue. He said that the Office of Population's interest in microcomputer technology stems largely from the technology's potential for reducing the cost of assistance provided for censuses and demographic surveys.

Mr. Dumm said that timing was another important consideration in the convening of the panel, for several reasons. First, AID was expecting to provide some assistance to the 1990 round of censuses, planning for which should begin as soon as possible. Second, the new Demographic and Health Surveys (DHS) project had begun recently, and it is expected that micro computers will be widely used on the project. Third, there is a clear need for the results of demographic surveys to become available as soon as possible, and the use of micro-computers may be able to reduce the time required to process survey data. The 1983 Philippine Demographic Survey was cited as an example of a survey which provided useful and timely information for planning the allocation of resources in the country's family planning program.

In his introductory remarks, Mr. Cornelius underlined the need for accurate and timely demographic information both for the administration of AID's population assistance program and for the Less Developed Countries (LDCs) themselves. He also stressed the importance of timing as a reason for convening the panel, citing the frequent delays experienced in the past with the processing of census and survey data (delays of 3-4 years in processing time have not been unusual). He said that limited access to mainframe computers has been an important factor contributing to such delays, as well as a general scarcity of computer support personnel. He said that he often has the impression that LDC statistical offices act as a training ground for computer personnel for the private sector. He explained that in the past AID has tried to cope with this problem by using simpler surveys, with limited questionnaires, but that this approach had frequently failed to provide the kind of information people wanted.

Mr. Cornelius explained that AID's program of assistance in the area of demography was effectively at a crossroads, marked by the start-up of the new DHS Project, in which thirty-five surveys are to be carried out during the next five years, and by the need

to plan AID's assistance to the 1990 round of censuses. He said that the purpose of the panel should be to explore the implications of utilizing decentralized computer facilities (as opposed to relying on a single central mainframe facility). He said he wanted the panel to consider the likely directions of microcomputer technology over the next five years.

A brief discussion of the agenda followed (see Appendix B). Mr. Collins (AID) felt that the agenda should be broadened to consider sectors other than population--for example, to include AID's work in agriculture and education--but there was general agreement that the agenda was appropriate, as submitted.

The various panel members introduced themselves. (A list of the panel members is provided as Appendix A)

2. Survey and Census Processing Needs

Mr. Patin and Ms. Diskin of the Bureau of the Census (BuCen) presented a background report to the panel on the feasibility of using micros to process the 1990 round of censuses. The report indicates that most African countries and about half of the Latin American countries are expected to have populations of under 5 million in 1990 and that, even with the present technology, it would be feasible to tabulate their censuses on microcomputers.

Representatives of the DHS Project (Lapham, Cushing) briefly described their project and their plans for carrying out thirty-five surveys over the next five years. They said that under the project's dissemination task they plan to prepare a 25-page policy report within a few months of completing field work (to be followed by a more substantive report within one year). They view the use of micros in data processing as an essential component in their plans to adhere to such a tight schedule. They said they expect their core questionnaire to consist of around 200 questions, with each supplementary module having fifteen or more additional questions. Sample size is expected to be in the range of 3-7 thousand households. It is their view that micros can be used to process, tabulate and analyze surveys of this size and degree of complexity.

Panel members cited examples of a number of other surveys carried out with AID funding which either have already used or would be using micros for processing and/or analysis. The surveys mentioned included Oral Re-hydration Therapy (ORT) and immunization surveys to be carried out by Primary Health Care Technology (PRITECH) in some seventeen countries; a program of education surveys to be carried out by the Latin American Bureau of AID, as a follow-on to the Kissinger Report; several AID-funded surveys on agriculture and nutrition to be carried out in Africa; and a series of demographic surveys being carried out in Sahelian countries under the sponsorship of the Institute of the Sahel (Bamako). Mr. Sadowsky (U.N.) mentioned the U.N.'s National Household Survey Capability Program, which had frequently encountered bottlenecks in the areas of computer hardware, software and trained personnel.

There was a brief discussion of past experience with micros in processing census and survey data. Mr. Patin (BuCen)

mentioned that micros have often been used to process a single survey but that once this work was completed, the micros would not be used in other work. He felt there was a need to provide broader training in connection with the use of micros for such purposes so that the benefits from the introduction of the technology would extend beyond a single survey. Mr. Muller (World Bank) said that there was a need for better information in the field about past experiences with micros, that the same mistakes tend to be repeated from one country to another. A number of panel members felt that the agricultural sector has been something of a pioneer in the use of micros for processing surveys and that, although the technology they have used is often out-of-date today, there is much useful experience upon which the population program can draw.

Ms. Cushing (Westinghouse) described recent experience with micros in Zimbabwe under the Contraceptive Prevalence Survey (CPS) Project. She said that micros were used to process data for a survey of 3,000 households and that the experience had been very positive. She said that the Survey Mate program was used for data entry and that she had written a custom BASIC program to edit the data. All of the work had been done with non-computer personnel, who learned the system very quickly. She said that one important advantage of using decentralized data-processing is that the work is not handed off from one department to another, as is typically the case when central data processing is used. With decentralized processing, those involved with the actual data collection see the processing work through to completion and are generally much more highly motivated to preserve the quality of the data than is the case when data are processed centrally. She said that they succeeded in producing a clean tape in only three months (from the time of data entry) and that this compares to a typical period of 4-6 months (and often much longer) with centrally processed CPS surveys. She emphasized the fact that it is generally the data entry and editing phases of a survey which most frequently experience delays.

Mr. Del Pinal (Westinghouse) described his favorable experiences in Honduras using Survey Mate to enter data from a survey on a region by region basis. Mr. Kelly described his experiences as a public health consultant working in Senegal, where he uses a Radio Shack TRS-80 Model 100 lap-size portable computer to enter data on a village level.

In discussing these experiences, the panel was in general agreement that micros had been successfully used already in a number of countries for processing surveys and that there was no reason to believe that they should not be used more broadly and systematically for this purpose in the future. There was also general agreement that where constraints exist they are primarily of a software, rather than of a hardware, nature. The panel felt that a lag of one to three years had been generally observed between the introduction of a fundamentally new microcomputer (e.g., the IBM-PC, the Apple Macintosh) and the development of software which could take full advantage of its features. For that reason, it is unlikely that projected future developments in the hardware area would have much impact on the type of micro systems to be used in LDCs over the next five years.

In summary, the main conclusions of the discussion of survey and census processing needs were as follows:

- 1) there should be considerable scope for using micros to process and analyze surveys and censuses over the next 5-10 years;
- 2) a good deal of success has already been experienced in LDCs using micros for work of this type;
- 3) using micros to process survey data should result in both a shorter turnaround time and better quality data; and
- 4) the availability of suitable software is likely to be a much greater problem than the availability of suitable hardware.

3. Hardware Selection

The panel was in general agreement over the need for some degree of standardization in hardware systems, and IBM-PC compatibility was the preferred standard. (A few panel members discounted the need for a standard, arguing that incompatibility was becoming less and less of a problem.) It was suggested by some that an IBM-PC with 256k RAM and two floppy disk drives be regarded as a working standard for purposes of evaluating software and for developing new software. Other panel members felt that, over a five-year period, a more powerful standard should be adopted. Mr. Knowles (The Knowles Corporation) pointed out that commercial software developers were already developing products requiring more than 256k RAM and that some existing products, such as SPSS, were not practical to use on systems with less than a 20-megabyte hard disk.

Several panel members felt that the real potential for micros in LDC survey processing lay with the use of large numbers of relatively inexpensive, simple systems, such as the Model 100 system used by Mr. Kelly in Senegal. With such systems it would be possible to enter data at (or at least close to) the interview site, possibly even performing simple consistency checks and then resolving any inconsistencies on the spot. There was general agreement that a system such as the Model 100 was well-suited for such operations. (Later in the day, Mr. Kelly demonstrated his system to the panel, in particular, the techniques he had developed for entering data interactively and for subsequently transferring these data from the Model 100 to his Apple II desk-top computer).

The panel concluded that it was unlikely that a single hardware configuration would meet all LDC needs in this area. Instead, a two- or three-tiered system was envisaged in which fairly simple, Model 100-type systems, might be used for field-level operations, floppy-disk based desk-top systems might be used for data entry and editing, and hard-disk based systems, such as the IBM-AT, might be used for tabulations and analysis. Although it was generally acknowledged by the panel that data entry and editing bottlenecks had accounted for most of the delays in processing past surveys, it was also recognized that not much meaningful analysis, particularly of a multivariate nature, had been carried out at all with most past surveys and

that the lack of access to computers and/or easy-to-use software on the part of analysts had been a major contributing factor to this lack of analysis. It was recognized that the hardware requirements for decentralized analysis were likely to be much greater than that required for decentralized data entry and editing.

During the course of the discussion, a number of other points were made by individual panel members. Ms. Diskin (BuCen) stressed the value of having redundant hardware with micro systems. If one component fails, the whole system does not become inoperable as is typically the case with mainframes and minis. She also mentioned that micro systems typically require less service and maintenance than larger systems and that this is particularly important in an LDC context. She said that there is still a role for traditional programming languages in micro systems, if only to patch things together from time to time. However, she saw possible difficulties in maintaining standards and quality with decentralized systems of data processing.

Mr. del Pinal suggested that the new three and one half-inch floppy disks were very appealing for use in LDCs for decentralized data entry because of the fact that they could hold more data than typical floppy diskettes and were enclosed in a protective case. Mr. del Pinal also pointed out that it is possible to over-load a micro system by trying to do too much with it; he suggested that the uses for each system should be explicitly prioritized before it is installed and that the training provided should be consistent with these priorities.

Mr. Knowles suggested that the panel give some attention to the possible role of multi-user systems and local area networks, particularly in relation to their potential use for data entry and editing. Mr. Muller urged the panel to consider the growing importance of the Unix operating system on micros and the large body of existing software that would become accessible with the use of Unix.

In summary, the panel concluded that no single hardware configuration is likely to meet LDC needs for decentralized data entry, processing and analysis; instead, a two- or three-tiered approach to hardware selection should be adopted.

4. Software Demonstrations

4.1 Software Demonstration

Following the lunch break, several presently available systems for the processing of survey data on micros were demonstrated to the panel. They included the Survey Mate program, which was demonstrated by Mr. Henry Elkins, Management Sciences for Health (MSH); the Entry Point and CENTS4 programs, which were demonstrated by Mr. Jeff Rosen (BuCen); and the package of programs developed by Mr. Patrick Kelly for the Model 100 during the course of his work in Senegal.

4.2 Software Adaptation/Development

The panel was of the opinion that mainframe software could be fairly easily downloaded to micros, even with the presently available technology. BuCen reported that they had encountered little difficulty in downloading CENTS4 to the IBM-PC, once they had selected the appropriate COBOL compiler. Similarly, Mr. del Pinal reported that the Demographic Data for Development (DDD) Project had experienced little difficulty in downloading the package of FORTRAN demographic programs developed by BuCen. The panel also recognized that there is a large quantity of mainframe software presently available for downloading to micros and that it is much less expensive to download mainframe software than to develop micro software from scratch.

However, the panel was also of the opinion that the downloading of mainframe programs was at best an interim solution. Such programs tend to be relatively difficult to use (i.e., the user interface tends to be batch-oriented), and they therefore fail to provide effective computing tools to a wider group of users (e.g., analysts as well as data processing (DP) professionals). Mr. Muller was concerned about the use of downloaded mainframe software on micros because the micro does not provide an equivalent operating environment for error-checking. Mr. del Pinal said that one problem with downloading existing software is that it does not lead to any advance in the state of the art. Mr. Knowles said that it had been his experience that LDC users are very happy to have downloaded software because it is so much better than having nothing at all. Mr. Muller suggested that it is pretty easy to develop a front-end, menu-driven environment for most batch-oriented programs.

Mr. Sadowsky, in contrast to the rest of the Panel, felt that there is definitely a need for new software development and that an appropriate balance should be sought between the two alternative approaches. At the same time, he recognized that there is a considerable amount of existing software available for downloading and that one should not automatically think of new system design as the primary development route.

In summary, there was a broad consensus among the Panel (with the exception of Mr. Sadowsky's dissent) that there is little need for developing new micro software. Some software development might be required, however, to customize existing software or to develop a menu-driven, interactive environment for batch-oriented programs.

Working from the background paper prepared by Ms. Diskin, the panel proceeded to review the various stages of survey and census processing to see what types of software, both commercial and public domain, were presently available in an effort to identify any major gaps. The various stages of survey processing were identified as follows:

SAMPLE SELECTION

DATA ENTRY (including some initial editing and corrections)

EDITING (including the generation of initial frequency distributions and possible imputation of missing values)

DATA DISPLAYS/BROWSING

TABULATION

MANAGEMENT/REPORTING (Control, PERT charts, etc.)

GRAPHICS

STATISTICAL ANALYSIS (including calculation of error variances)

DEMOGRAPHIC ANALYSIS

RECODES/FILE MANIPULATION (including construction of subfiles)

COMMUNICATIONS (Data)

TUTORIALS

The deliberations of the panel with respect to each of the above topics were as follows:

Sample Selection. The panel felt that there had not been much progress in developing generalized packages for sample selection, even on mainframes. Clusters, a program developed by the World Fertility Survey (WFS), was available for this purpose; but unfortunately, no one is presently supporting this program. Mr. Muller reported that he has been working on menu-driven sample selection software in BASIC for the IBM-PC. The Panel felt that efforts should be made to identify a suitable mainframe sampling package for downloading to micros.

Data Entry. Mr. Rosen reported to the panel that he found Entry Point to have most of the necessary features for data entry but that it tends to be slow. Mr. Muller suggested that people consider using a commercial package such as KnowledgeMan, for data entry since its data definition table would automatically perform certain error checks while others could be programmed using KnowledgeMan's built-in command language. DBASE III was another package mentioned as presenting similar capabilities. Mr. Muller also suggested that certain statistical programs, such as Minitab, BMDP and SAS could be used to perform higher-level consistency checks, some of which could even be built into their data dictionaries.

Mr. Elkins said that Survey Mate allows range checks and a limited number of consistency checks across fields. He said that a new version of the program is being developed in the C language which will have many more features. Ms. Cushing said that she preferred Survey Mate to Entry Point because of the speed difference. Mr. Rosen explained that Entry Point tends to be slow because modules are constantly being swapped in and out of memory from a floppy disk. He said that it would probably perform much better if the modules were loaded onto a RAM disk. Mr. del Pinal said that he found Survey Mate easy to learn but that the consistency checks were insufficient. He said that he did not find the limitation of a maximum of four digits per field to be a serious constraint.

Ms. Diskin pointed out that census processing requirements were typically quite different from survey processing requirements. Census data sets tend to be very large, with relatively simple questionnaires. Analysts do not have enough time to perform edits. Instead, editing has to be done by programs, such as CONCOR. She noted that optical scanning may be practical to use in census processing in large countries.

The panel concluded that no single data entry software package was likely to be most appropriate for all surveys. Instead, the panel favored an eclectic approach in which one of several acceptable products might be selected in a given

application. At the same time, the panel saw certain problems with the use of several different software packages for a single survey. The panel felt very strongly, for example, that data entry packages should permit the encoding of the data dictionary along with the data in such a way that it would not be necessary to re-define the dictionary when passing the data from one processing program to another.

Editing. BuCen representatives (Diskin, Patin) suggested that there was a real need for a micro version of a program like CONCOR since the type of tasks it performs can not be done very practically with available commercial software. At the same time, Bucen recognized that CONCOR is not very easy to use since, in its current form, it can only be used by a trained programmer. Mr. Muller suggested that a front end be developed for CONCOR. He also suggested that the conceptual basis of other editing software, such as the CANEDIT program developed by Statistics Canada, be examined (i.e, that the conceptual approaches to editing found in CONCOR not be regarded as definitive).

Mr. Knowles suggested that CONCOR-type software was most useful for editing censuses and very large surveys and that software for editing smaller surveys should provide for interactive, screen-oriented editing. He said that such editing was presently available with certain mainframe statistical packages, such as SAS, but that it had not been widely accepted because of the high cost of interactive editing on mainframes. He said that a micro would be ideal for such work.

The Panel also felt that a number of commercial programs, such as DBASE III and KnowledgeMan, could be adapted for use in editing survey data.

Data Displays/Browsing. The panel felt that survey processing software should provide for case-by-case, selective access to the data for display purposes. Such capability is already present with Survey Mate, Entry Point and with a wide variety of commercial micro software, such as DBASE III. It is also available with SAS. Mr. Muller suggested that S, a software product developed by Bell Labs for the Unix operating system, be examined, along with other Unix software developed by the Livermore group in California.

Tabulation. The panel noted that there are already a wide range of programs available for performing tabulations on micros, including Survey Mate, CENTS4, SPSS, P-STAT, and Minitab. It was observed by the panel that CENTS4 and SPSS are the packages most widely used in LDCs and that this is relevant when one considers the opportunity they present for tapping into existing experience. The advantages and disadvantages of CENTS4 were noted by the Panel. On the plus side, CENTS4 makes camera-ready tables to user specifications, while on the minus side, it requires considerable user effort and skill.

Management Reporting. The panel was aware of a program, Control, developed with AID funding, which can be used to track the various stages of a survey's field work and processing. In addition, Mr. del Pinal suggested that commercial project management packages, such as PertMaster and Harvard Project Manager, could be used for this purpose. However, Mr. Rosen pointed out that requirements for data management are somewhat

different from those of project management.

Graphics. The panel was of the opinion that there is already a great deal of graphics software available for micros (e.g., Lotus 1-2-3) and that it is continually improving in quality.

Statistical Analysis. The panel felt that there are already a large number of statistical packages available for use on micros, including such standard mainframe packages as SPSS, Minitab, P-STAT, and BMDP. One of the most extensive statistical packages for micros, STATPRO, was originally very expensive but its price has recently decreased significantly (a trend which everyone thought would continue with such software). Mr. Muller also mentioned the program, Daisy, which was developed by Bell Labs and is available for use on micros at a cost of only \$150.

For purposes of calculating error variances, Ms. Diskin mentioned Supercarp, a program developed by Iowa State, which she considered to be very good for this purpose. Although it is presently available for use only on mainframes, she felt that Iowa State would develop a micro version with a little encouragement. The WFS program, Clusters, can also be used for this purpose; but again, its problem is that it is presently unsupported.

Demographic Analysis. The DDD Project has downloaded the BuCen package of demographic programs for use on IBM-PC/XT's. Mr. Knowles pointed out that electronic spreadsheets, such as Lotus and Supercalc, can be used to facilitate a wide variety of demographic calculations. Mr. del Pinal reported that the U.N. was considering downloading at least some of the National Academy of Sciences (NAS) demographic programs for use on micros. Mr. Cressman, Research Triangle Institute (RTI), mentioned that the new Integrated Population and Development Planning II (INPLAN) project was developing a multi-regional population projection model for use on the IBM-PC/XT.

Recodes/File Manipulation. The panel felt that many commercially available relational data base management programs, such as DBASE III, can be used to recode variables and to create subfiles. Several statistical packages also provide similar capabilities.

Data Communications. The panel noted that there are several possibilities available for transferring data between micros and mainframes, including hard-wiring the units to one another through asynchronous ports, communicating over dedicated telephone lines with modems, and using compatible magnetic storage media (e.g., standard IBM formatted half-inch tapes and/or 8-inch floppies). Mr. Cressman reported that the error rate tended to be quite high with some of the more popular communications protocols, such as XMODEM and Kermit. He mentioned Blast as a better protocol, but said that he still favored the use of compatible magnetic media, such as half-inch tapes, for the transfer of large amounts of data. He mentioned some of RTI's favorable experience in using such media in LDCs. The general consensus of the panel was that data communications are not much of an obstacle at present and that they would become even easier in the future.

Tutorials. The panel noted that tutorials, utilizing computer-aided instruction (CAI) techniques, were presently

available for most commercial software, either from the software vendors themselves or from third parties. However, there was a consensus that to be truly useful such tutorials would have to be available in the language of the host country. The panel felt that tutorials would be particularly useful in the area of editing. Mr. Knowles suggested that wherever possible tutorials should cover common mistakes as well as correct procedures to use. Ms. Diskin mentioned that there was even software on the market for authoring tutorials.

4.3 Summary

The panel's discussion on software produced the following conclusions:

1) it was more cost-effective to download existing mainframe software to micros than to develop completely new software;

2) in cases where downloaded software is batch-oriented, it would generally be worthwhile to provide the software with an interactive, menu-driven front-end;

3) an eclectic approach should be used in selecting particular software products for various functions; it was considered unlikely that any single program would have enough desirable features in all applications to justify its adoption as a standard;

4) the functions of "sample selection" and "editing" were seen as the two areas where existing micro software is most deficient; the panel urged that CONCOR, or its equivalent, be implemented on a micro;

5) data entry software should encode the data dictionary in such a way that it would not have to be re-entered manually for use with editing, tabulation and analysis software; and

6) there is considerable scope for adapting commercially available relational data base management programs, such as KnowledgeMan and DBASE III, to perform the functions of data entry, editing and file management (including recodes).

5. New Directions

The panel felt that one of the most promising new directions to be explored involved the use of lap-size portables at the interview level. Nevertheless, the panel foresaw several possible problems with the use of computers during the actual interview. First, it was questioned whether the computer would be too distracting, focusing attention on the computer rather than on the substance of the interview. A related issue was whether having a computer present at an interview would be culturally acceptable. However, Mr. Hawkins (The Futures Group) observed that these same issues were considered when tape recorders were first used in interviews, but few if any problems actually arose with their use.

The panel also foresaw possible problems involved in trying to administer an interview with a computer. For example, although it was likely that the computer would be able to identify certain inconsistencies in responses provided at the

point of interview, it was also clear that no computer would provide the same possibilities for browsing through a questionnaire, locating and simultaneously comparing responses to two or more related questions, that a physical questionnaire provides. The need to program the questionnaire in the various local languages used in a survey was another possible problem identified by the panel.

The panel concluded that the use of computers at the interview level was certainly worth selective experimentation, but that it was not an obviously superior method of administering surveys. As another possible approach, the panel felt that it might be desirable to experiment with the use of lap-size computers to enter data immediately following the interview while still at the interview site, so that whatever inconsistencies were found by the computer could be resolved before leaving the interview site.

6. Additional Issues

The panel briefly considered each of the following additional issues: training requirements, service and maintenance requirements, and electric current-related problems.

6.1 Training

A major question considered by the panel was how much general training in the use of microcomputers, apart from the training necessary to carry out specific projects, should be supported by AID. Most panel members were of the opinion that taking a broader approach to microcomputer training would turn out to be cost-effective over a period of ten years. For example, it was suggested by Mr. Patin that effective training in the use of microcomputers over the next five years might substantially lower the amount of resources subsequently required in support of the 1990 round of censuses. Mr. Cornelius pointed out, however, that AID was constrained organizationally by sector and project divisions and that it was often impractical to provide umbrella support for institutional development efforts.

Mr. Knowles expressed the view that more cost-effective means should be found for providing training. He said that in his experience one-on-one training was very expensive and of limited effectiveness. He suggested the use of improved tutorials and the possible use of video training courses as more cost-effective approaches to training. He also suggested that some of the content of traditional training courses be built into the software--so-called expert systems--so that an LDC user would be guided intelligently in the use of the computer by the program itself. He argued that the use of such training technologies would circumvent at least some of the problems created by the high turnover of computer-trained personnel.

6.2 Service and Maintenance

The panel felt strongly that local service agreements were

not necessarily the best approach to the service and maintenance needs of micro computers. Such service agreements tend to be very expensive and there is no assurance that local personnel have either the necessary expertise or spare parts on hand to meet repair needs. As a possible alternative, the panel felt that non-specialists (e.g., ministry staff, researchers) could be trained to make routine repairs--such as adjusting the speed or alignment of a disk drive or replacing a disk drive, disk drive controller, video display controller, RAM memory chips, or power supply--and that such simple repair procedures would take care of ninety percent of the service problems likely to be encountered. Keeping on hand a supply of the necessary spare parts to make such repairs would cost much less than a local service contract.

The panel also felt that it was absolutely necessary for both hardware and software to be tested very carefully before being shipped to LDCs.

6.3 Electric Current-related Problems

The panel noted that some problems with electric current had been encountered in the past but that fewer were occurring at present. They saw no reason why computer equipment could not be procured in the U.S., even if it was designed for use with 110-volt, 60-Hz current. In cases where LDC current is 220-volt, step-down transformers could be installed at little cost to transform the current from 220 to 110. It was noted that almost all computer equipment, including IBM-PC/XT and related peripherals, either operates on DC current (e.g., all logic boards and disk drives)--so that the issue of frequency (i.e., 50 or 60 cycle) does not arise--or is able to function with either 50/60 cycle current (most monitors).

The panel felt that while the type of current (i.e., voltage, frequency) found in LDCs is not really an obstacle, the quality of the current is a very important factor to take into consideration. The panel felt that expensive computer equipment should not be run directly from wall current, even if a transformer is used. At a minimum, the equipment should be furnished with a voltage regulator, which would at the same time stabilize the voltage and protect the computer from surges and spikes. The use of a voltage regulator should also reduce the need for repairs. In areas with frequent power outages (e.g., more than one or two per week), the computer (but not necessarily peripherals such as a printer) should be connected to an uninterruptible power supply (UPS) which would provide battery back-up.

It was noted that, unlike other computer equipment, voltage regulators and UPSs are often sensitive to current frequency. It was recommended that 220-volt, 50-cycle units be used in areas where the wall current is of this type and that a step-down transformer then be used to transform the stabilized current from 220- to 100-volt. The panel also cautioned against using cheap units from relatively unknown vendors because there may be important technical differences between various units purporting to do the same job (e.g., this is particularly true of UPS units which utilize wall current directly under normal operation but

which switch over to a battery back-up when power goes out, as opposed to UPS units which isolate the computer from the wall current at all times). Mr. Knowles and Mr. Sadowsky both reported having had favorable experience with Sola UPS models, and Mr. Knowles reported having had favorable experiences as well with Sola voltage regulators.

6.4 Summary

The discussion of additional issues produced the following conclusions:

1) Micro-computer training should be as broad as AID project and sectoral constraints permitted, since this approach would be most cost-effective over the next ten years;

2) Attempts should be made to use more cost-effective approaches to training than the one-on-one approach so widely used in the past;

3) Local service contracts for servicing and maintaining micros may not be the best solution; as an alternative, it might be better if non-specialists were trained to make simple, routine repairs, utilizing a local stock of spare parts;

4) All equipment and software should be thoroughly tested in the U.S. before being shipped to LDCs;

5) The type of current (i.e., voltage, cycle) used in LDCs should not prevent microcomputer equipment from being procured in the U.S.;

6) Given the poor quality of LDC current, computer equipment should be protected from the wall current, at a minimum, with a voltage regulator and, if necessary, with the use of a UPS; and

7) Only high-quality voltage regulators and UPS units should be used in LDCs, and only those equipped to operate with the type of wall current (e.g., 220-volt, 50-cycle) encountered in a particular area.

APPENDIX A
LIST OF PARTICIPANTS

Appendix A

LIST OF PARTICIPANTS

Mr. John Dumm, Associate Director, AID/S&T/POP
Mr. Richard Cornelius (Chairperson), AID/S&T/POP
Mr. Jack Lawson, AID/S&T/POP
Mr. John Crowley, AID/S&T/POP
Mr. Scott Radloff, AID/S&T/POP
Mr. Richard Collins, AID/PPC/DIU
Mr. John Daly, AID/SCI
Mr. Larry Patin, U.S. Bureau of the Census (International Division)
Ms. Barbara Diskin, U.S. Bureau of the Census (International Division)
Mr. Jeff Rose, U.S. Bureau of the Census (International Division)
Mr. James Knowles, The Knowles Corporation (Chapel Hill, NC)
Mr. Mervin Muller, The World Bank (Washington, D.C.)
Mr. Peter Gutterman, The World Bank (Washington, D.C.)
Mr. George Sadowsky, United Nations Statistical Office (New York)
Ms. Jean Cushing, Westinghouse Health Systems (Columbia, Maryland)
Mr. Robert Lapham, Committee on Population, N.A.S. (Wash., D.C.)
and Westinghouse Health Systems (Columbia, MD)
Mr. Jorge del Pinal, Westinghouse Health Systems (Columbia, MD)
Mr. Henry Elkins, Management Sciences for Health (Boston, Mass.)
and Henry Elkins Associates (Bronxville, New York)
Mr. Patrick Kelly, Independent Public Health Consultant (Senegal)
Mr. Gordon Cressman, Research Triangle Institute (Durham, NC)
Mr. Steve Hawkins, The Futures Group (Washington, D.C.)

AGENDA

ISTI ADVISORY PANEL
ON MICROCOMPUTER SOFTWARE DEVELOPMENT
FOR SURVEY AND CENSUS DATA PROCESSING

December 3-4, 1984

National Academy of Sciences
Joseph Henry Building, Room 355
2122 Pennsylvania Avenue, Washington D.C.

MONDAY

9:30-10:00 Introduction

1. Opening Remarks by Steven W. Sinding
2. Introduction by Richard Cornelius

10:00-11:00 Survey and Census Processing Needs

1. How many developing countries are taking surveys and censuses under the FHDS and 1990 round of censuses?
2. What are data processing requirements in terms of observations and items for survey and census activities?
3. What portion of the data collection activities can be done with microcomputers?

11:00-12:00 Hardware Selection

1. What are the computational requirements for various processing tasks and how can they be met?
2. What are the storage requirements for various processing tasks and how can they be met?
3. What are the support requirements for various processing tasks and how can they be met?
4. Which products are most appropriate for these data processing tasks?

12:00-1:00 Lunch

1/2

1:00-2:30 Demonstrations

1. CENTS4 tabulation software.
2. Survey-mate survey processing software.
3. Lap computer software for field entry/analysis.

2:45-5:00 Software Adaptation/Development

1. What are the advantages/disadvantages for downloading mainframe software programs?
2. What commercial microcomputer software exists and what are the advantages and disadvantages for data processing tasks?
3. What are the advantages/disadvantages for developing new software for various data processing tasks?

TUESDAY

9:00-12:00 Software Adaptation/Development (continued)

12:00-1:00 Lunch

1:00-2:00 New Directions

1. What are the applications and feasibility of using lap computers for various processing tasks?
2. What other technologies are on the horizon and what are their applications and feasibility for use in various processing tasks?

2:00-3:00 Additional Issues

1. What are the training requirements?
2. What are the maintenance requirements?

3:00-4:00 Wrap-Up

1. Summary of discussion and major recommendations by Richard Cornelius and James Knowles.

11