

Miscellanea

Under the heading Miscellanea, essays will be published dealing with topics considered to be of general interest to the readers. All contributions will be refereed for their compatibility with this criterion.

Recent Work With Microcomputers for Census Processing in Developing Countries

Vivian Toro and Kathleen Chamberlain¹

Abstract: This paper is a revision of the paper entitled "The Use of Microcomputers for Census Processing in Developing Countries" written by Vivian Toro and Thomas Melaney and presented at the American Statistical Association meetings in August 1987. The authors discuss how microcomputers offer potential solutions to many of the problems developing countries encounter when processing census data with mainframe computers. The authors also describe the Integrated Microcomputer Processing System (IMPS) developed by the International Statistical Programs Center (ISPC), U. S. Bureau of the Census, and present three case studies of the use of IMPS by statistical offices in developing countries.

The experiences of a number of developing countries confirm that microcomputers are technologically sound and cost-effective tools

for processing censuses. The continuing technological advancement and refinement of microcomputer software make their use even more advantageous.

This paper focuses on the recent experience of Burkina Faso and Senegal in using microcomputers. A summary is given of the use of IMPS by an increasing number of countries. Additionally, this paper describes the latest enhancements and future plans for IMPS. It concludes with some thoughts on the use of microcomputers in the areas of data collection and processing, as well as in the use and dissemination of census data.

Key words: Microcomputers; census; developing countries; software; data processing; data collection; integrated.

¹International Statistical Programs Center, U.S. Bureau of the Census, Washington, D.C. 20233, U.S.A.

This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are those of the authors and do not necessarily reflect those of the Census Bureau. An earlier version of this paper was presented at Symposium 88: The Impact of High Technology on Survey Taking. Ottawa, Canada, October 1988.

1. Microcomputers as Tools for Processing Census Data

Technological developments over the last two years have made the processing of large volumes of data on microcomputers even more practical. Problems normally associated with large-scale processing jobs such as population censuses are quickly disappearing. Data stor-

age devices for microcomputers are more plentiful, faster, more reliable, smaller in size, higher in capacity, and less expensive. Processing speeds of microcomputers continue to increase and memory continues to expand.

Statistical offices in developing countries have found this technology easily accessible. For the most part, finding a local microcomputer vendor is no longer a problem. Typically, three or more different vendors are likely to be represented. This more competitive market has not only expanded the number of options, but also has reduced the prices and improved the terms of maintenance contracts.

Microcomputers are now an integral part of the operation of most statistical offices in developing countries. Microcomputers are being used for processing surveys, scheduling, budgeting, word processing, and other activities. A cadre of microcomputer literates exists in most of these countries. In addition, a variety of microcomputer software packages in different languages is now available.

Instead of asking whether or not to use microcomputers, statistical offices are now wondering how best to use them. They are asking what type of microcomputer software and hardware will best meet their short- and long-term needs. They are interested in obtaining the resources to buy this hardware and software, to train their staff, to provide hardware maintenance, and to integrate microcomputers into existing mainframe environments.²

As statistical offices have become more comfortable with the microcomputer technology, many have chosen to use microcomputers for processing all or part of their census data. Burkina Faso, Senegal, and Micronesia were some of the pioneers that opted to process their census data using microcomputers. They were followed by countries such as Niger, Swaziland, Yemen, Honduras, Uruguay, and the Central

African Republic. Countries including the Philippines, Pakistan, and Côte d'Ivoire chose to use microcomputers just for the entry of the data from census questionnaires. Countries such as Tanzania and Cameroon chose to share processing between microcomputers and a mainframe computer.

Microcomputers have provided a solution to some of the problems encountered in the past by countries which used only mainframe computers to process their census data. Such countries have been able to afford the computing capability of the microcomputer within their limited budgets. The availability of several microcomputers has resolved most of the computer accessibility problems. Microcomputer maintenance is easier, and environmental standards are less restrictive. Additionally, training requirements are less demanding given the availability of software that is easier to learn and use, even for persons with little background in data processing.

The availability of microcomputer software which addresses the particular needs of census data processing has contributed greatly to the success microcomputers have had in statistical offices. Most of the countries mentioned above have used or plan to use IMPS.

2. The Integrated Microcomputer Processing System

2.1. Background

The introduction of more powerful microcomputers in the early 1980s was not accompanied by microcomputer software which adequately addressed the requirements of census processing. ISPC embarked on the development of IMPS to provide microcomputer software for the major tasks of census data processing: data entry, editing, tabulation, analysis, and operational control. Because a project of this magnitude takes considerable time and resources, ISPC decided to use existing software where

²In this paper, any reference to mainframe computers will also include minicomputers.

possible, and to provide intermediate products throughout the development of IMPS.

The design objectives of IMPS included ease of use, availability of a common data dictionary, ability to run with a standard microcomputer configuration, and modularity. The concept of modularity was of particular importance in the early development of IMPS. The system was designed in such a way that modules could be easily swapped as software improvements were made. For example, prior to the development of the IMPS data entry component, CENTRY (Census data ENTRY), commercial data entry packages were commonly used. The substitution of CENTRY does not require modification of the other application modules such as editing and tabulation. Another advantage of modularity is flexibility. The user can choose which tasks to perform using IMPS and which to perform using other software or other hardware.

The first step in the development of IMPS was to make the software packages CONCOR and CENTS 4 available on microcomputers. CONCOR (CONSistency and CORrection) and CENTS 4 (CENSus Tabulation System version 4) are editing and tabulation packages, respectively, developed by ISPC. These packages run on most mainframe computers and were in use by over 100 developing-country statistical offices by 1985. Both packages were written in COBOL, to ease the job of adapting them to the number of different mainframe computers found in developing countries. The availability of Realia COBOL, a powerful and efficient COBOL compiler for microcomputers, facilitated the adaptation of CONCOR and CENTS 4 to the microcomputer environment.

The first microcomputer versions of CONCOR and CENTS 4 appeared in 1984. The user instruction sets for these versions were the same as for the mainframe versions. This meant that the same CONCOR or CENTS application programs could run on either the mainframe

computer or on microcomputers at statistical offices where both types of hardware were available. A number of statistical offices, including Senegal, Cameroon, and Somalia, took advantage of this flexibility.

Data capture, often a major bottleneck in census processing, was the next hurdle. ISPC evaluated commercial data entry packages to identify the best software for the capture of census and survey data. The evaluation identified two packages, ENTRYPOINT from Datalex, Inc., and RODE/PC from DPX, Inc., as meeting the performance criteria. At the time of the evaluation, RODE/PC had a slight advantage over ENTRYPOINT in terms of cost and speed. Therefore, ISPC recommended RODE/PC for census data entry on microcomputers.

The widely used Computer Programs for Demographic Analysis (CPDA), written by the Bureau of the Census in the early 1970s, was made available for microcomputers by Westinghouse Public Applied Systems under the sponsorship of the U.S. Agency for International Development (USAID). The Microcomputer Programs for Demographic Analysis (MCPDA) is one of several demographic analysis packages available on microcomputers. Currently, the Bureau of the Census is developing a more comprehensive package called Model Spreadsheets for Demographic Analysis (MSDA). Additionally, it is preparing a manual on the analysis of census data. This manual brings together the most useful methods of demographic analysis and projection, including brief descriptions, discussions of the advantages and disadvantages of different methods, and other helpful hints for interpreting the results. The manual will also discuss available software for performing each type of analysis. The MSDA will be available in about one year and the manual during the following year.

The statistical analysis package PC-CARP, developed by Iowa State University, is the statistical analysis component of IMPS. Over the

past two years it has become widely used by statistical offices in developing countries for the calculation of variances in surveys, including post enumeration surveys. Users of PC-CARP include Zambia, Peru, Egypt, Zimbabwe, Costa Rica, Haiti, and the Philippines. The user interface of PC-CARP facilitates its use by analysts with little or no data processing experience.

With microcomputer software available for data entry, editing, tabulation, demographic analysis and statistical analysis, and with access to high capacity storage devices such as Bernoulli Boxes,³ it became feasible to recommend census processing on microcomputers for countries with populations of 10 million or less. The 1985 census of Burkina Faso was the first major census to be processed entirely on microcomputers. RODE/PC, CONCOR, CENTS 4, and MCPDA were used. Burkina Faso's experience is described in detail in Section 3.

Since 1985, ISPC has added new features to IMPS. These include a common data dictionary, a more economical data entry module, improvements to CONCOR and CENTS, a census management and reporting system, and a census resource planning software.

³Bernoulli Boxes are removable cartridge-based hard-disk systems. They can serve as both mass-storage devices and backup systems. Information is stored in greater densities and higher data access speeds are allowed. Each cartridge unit has one or two cartridge readers. Depending on the model, the units can read data from either 10 or 20 megabytes cartridges. Since these cartridges are removable, they provide virtually unlimited capacity. The newer models also have up to 80 megabytes of hard-disk storage in addition to the removable cartridges. In addition, they are physically very resistant to rough treatment which would otherwise have destroyed or caused the loss of data on hard disks. They have been used for several years in developing countries and have proved to be highly reliable.

2.2. *New features*

2.2.1. *Common data dictionary*

CONCOR and CENTS 4 originally were developed as stand-alone packages for mainframe computers. During their development, little attention was paid to integrating the two packages because the major concerns were portability and use of the software on computers with very limited memory capacities. As a result, the user was forced to define the data file for the CONCOR application program, then again for the CENTS 4 application program, using a different notation each time. After CONCOR and CENTS 4 were adapted to the microcomputer, the data definition became even more cumbersome because commercial data entry packages such as RODE/PC required yet a third definition of the data. Multiple data definitions were not only time-consuming to code but were also error-prone.

The task of integrating CONCOR and CENTS through a common data dictionary resulted in CONCOR version 3 and CENTS version 5 which were released in early 1988. With these versions, the definition of the data file to be edited and tabulated is done only once. The interactive module of IMPS called DATA-DICT prompts the user for a name and the characteristics of each data item. The resulting definition, the Data Dictionary, is used by both CONCOR and CENTS.

In these new versions of CONCOR and CENTS, the user interface has been greatly simplified, and the CONCOR edit reports were made more concise. Several new features were added to CONCOR such as the ability to reference repeating data items and greater flexibility in creation of the extract files.

2.2.2 *CENTRY: Software for data entry*

As the cost of commercial data entry packages rose, it became apparent that statistical offices in developing countries needed a less expensive

alternative. For example, in order to enter questionnaires within one year from a country with a population of 10 million would require approximately 30 microcomputers, meaning 30 copies of the data entry software package. At about \$600 (U.S.) per copy, data entry software becomes quite expensive. Moreover, RODE/PC and ENTRYPOINT contain features, such as extensive logic checks, that are essential for survey data entry but not needed for census data entry. For census data entry, speed is of primary importance since the volume of data is so great. Capturing what is on the questionnaire, consistent or inconsistent, is essential. Inconsistencies can be corrected later through CONCOR and should not be left up to data entry staff to correct. Most developing countries have trouble just attaining acceptable keying rates with systems that have no or very limited logic checks.

Keeping in mind the particular needs for the entry of census data, ISPC developed a package called CENTRY as the data entry module of IMPS. CENTRY is a screen-oriented menu-driven package which allows for developing data entry applications, entering and verifying data, and collecting statistics on data entry operations. CENTRY uses the same Data Dictionary as CONCOR and CENTS. Features of CENTRY which make it attractive for census data entry include: programmable data entry screens, valid value checking, automatic duplication of fields, cursor control, skip pattern control, record retrieval and modification, and operator statistics. The data entered through CENTRY are stored in an ASCII file ready for use by other packages such as CONCOR. The user-friendliness of the IMPS data dictionary and CENTRY allows a person with minimal training who is familiar with the census questionnaire to set up a data entry application in a day. CENTRY requires 256 kilobytes of memory and two floppy disk drives. It is written in the C language to promote fast execution.

CENTRY will be released in November

1988. Already several countries, including Central African Republic, Burundi, and the Philippines, are planning to use it for census data entry.

2.2.3. dCONTROL: Census management system

With a large data processing operation such as a national census, it is important to track each unit of data through the various processing phases. dCONTROL is an interactive census management and control system that helps census managers to monitor these phases. It also serves as a "check-in" facility for data by geographical unit, such as enumeration area (EA), thereby preventing duplication or omission of EAs. dCONTROL also allows for the production of preliminary count reports and other management reports.

In its current state, dCONTROL is a prototype that can be adapted to meet the particular needs of a country. It is being used in both Niger and Senegal where their statistical office data processing staff customized it to suit their specific needs. ISPC currently is developing a generalized version of dCONTROL which will be available in early 1989 in English, French, and Spanish.

dCONTROL requires a good cartographic system that accurately defines all the geographic and administrative boundaries for the census. The cartographic system serves as the basis for a geographic coding scheme that identifies each statistical area down to the EA level. dCONTROL monitors the various phases of processing at the EA level. The first phase is usually the receipt of the questionnaires from the field by the central office; the last phase is the keying of the census data.

dCONTROL is written in dBASE III PLUS programming language compiled using CLIPPER by Nantucket. It runs on an IBM PC/XT or PC/AT or compatible with 512 kilobytes of memory and at least 10 megabytes of auxiliary

storage, usually a hard disk. The amount of hard disk needed depends upon the number of EAs in the country. For a country of 8000 EAs, two megabytes of hard disk are sufficient to store the dCONTROL programs and the data.

2.2.4. CENPLAN: (CENSus PLANning)

Although great strides have been made in a very short time in the development of census processing software for microcomputers, there is still room for improvement. The goal is to make census results available as soon as possible after the census is taken. A number of factors contribute to the delay of census results, not the least of which is poor planning.

ISPC is developing a software package called CENPLAN which allows the census planner to determine the resources needed for census processing, given certain parameters such as population size, available computer equipment, and time constraints. Although CENPLAN cannot ensure that proper planning is done in time for a census, it provides the facilities to plan. CENPLAN uses spreadsheet software and will be available in French and Spanish, as well as English. The first release, which will address only the computer processing of a census, will be available in November 1988.

3. Case Studies

The number of countries using IMPS for the processing of census data continues to grow. Table 1 is a partial list of countries that have used, are using, or plan to use IMPS to process their census data. Statistical offices in developing countries are taking advantage of the system modularity that IMPS offers.

Some countries, including Niger, Senegal, Benin, Burkina Faso, the Central African Republic, Micronesia, Burundi, Yemen Arab Republic, Mali, Malawi, Swaziland, and the Pacific outlying areas of the U.S. have used or

plan to use microcomputers for all aspects of their census data processing. Senegal, Yemen Arab Republic, and Mali, had originally planned to use microcomputers for data entry and a mainframe computer for further processing of the data. However, when they realized they could obtain results in a more timely and cost-effective manner, they decided to process the data on microcomputers. For example, they found that CONCOR and CENTS generally execute faster on the microcomputers than on their mainframes. Even when the actual execution time was theoretically the same or slightly faster on the mainframe computer, they found they could obtain results faster on the dedicated microcomputers because they did not have to share computer resources with other users.

Other countries such as Somalia and Côte d'Ivoire have developed their CONCOR and CENTS programs using microcomputers but are running these programs against data on mainframe computers. Some countries like Ghana and Ethiopia did their data entry on microcomputers, then uploaded the data to a mainframe computer for processing. Furthermore, countries like Honduras, the Philippines, and Tanzania are still trying to decide whether to use microcomputers, a mainframe computer, or a combination of the two for processing the census data. The selection criteria are not always solely technical. Some countries are committed to using mainframe computers for processing census data because their governments or donor agencies had purchased or procured the mainframe equipment several years ago with the express objective of using it for the census. Despite a technological revolution, amortization of this equipment was necessary before decision-makers could justify the purchase of new equipment. IMPS modular design allows statistical offices in developing countries to take advantage of existing hardware and software that is appropriate for census data entry. For example, Tanzania, and Cameroon

are using minicomputers such as ICL/DRS 300 and IBM System/36 for data entry. On the other hand, Yemen Arab Republic is using microcomputers for data entry but has contracted out the development of their data entry programs because specialized Arabic data entry is needed. All these countries plan to use or are using CONCOR and CENTS for the editing and tabulation of the data.

Additionally, as other IMPS components become available, statistical offices in developing countries can plan to take full advantage of it. For example, Burundi, the Central African Republic, and the Philippines are planning to use CENTRY for the entry of their census data.

Although designed primarily for census processing, IMPS also is being used for survey processing. Gambia, Morocco, Zambia, Senegal, Portugal, Ghana, Rwanda, Egypt, and American Samoa are using one or more of IMPS modules for surveys.

The following two case studies, Burkina Faso and Senegal, were presented in the Toro-Melaney paper as representative of the benefits of IMPS for census processing activities. A follow-up of their experience follows. The Federated States of Micronesia (FSM), the third case study presented in the Toro-Melaney paper, have continued to use microcomputers successfully to process the census data for some of the states. FSM is currently discussing the establishment of a national census that would standardize the date, questionnaire, and processing activities used by all the states.

3.1. Burkina Faso

In December 1985, the National Institute of Statistics and Demography (INSD) of Burkina Faso, formerly Upper Volta, conducted their second national census. It took them only 18 months to enter and process the data of their estimated population of 9 million.

Three IBM PC/AT microcomputers with 512

kilobytes of memory, and 20 megabytes of hard disk storage were used to process the data. The data entry equipment consisted of twenty-two IBM PC microcomputers each having 512 kilobytes of memory, and two IBM PC/XT microcomputers with 10 megabytes of hard disk storage and 640 kilobytes of memory. Related peripherals included six printers, two DIGIDATA tape drive units, and three Bernoulli Boxes with 10 megabyte removable cartridges. The tape drive units were used for backup and archival purposes, while the Bernoulli Boxes were used for primary storage of the census data. RODE/PC, CONCOR, and CENTS 4 were the software packages used for data entry, editing, and tabulation, respectively.

The INSD data processing staff for the census project consisted of two full-time programmers and a long-term United Nations advisor. In addition, the U.S. Bureau of the Census under an agreement with USAID, provided data processing assistance in the form of training, program development, and monitoring of the RODE/PC data entry application program, the CONCOR editing program, and the CENTS 4 tabulation programs.

The preliminary counts based on the manual counts were available by March 1986. Two shifts of 25 operators began keying the 15 % sample of the data in September 1986. Two months later, this data capture operation was complete, and by March 1987, the tabulations for the 15 % sample had been produced. The edited sample data occupy 11 Bernoulli cartridges of 10 megabytes each. (It should be noted that data entry was begun several months late due to delayed arrival of microcomputer equipment.)

The entry of the remainder of the data took place between December 1986 and May 1987. The 40 tables for the entire country were available by March 1988, 18 months after data entry began. The keyed data for each enumeration area were stored on one diskette. No more than 600 diskettes were in use at one time since

they were recycled after data verification. After the data were completely verified, all the EAs for a province (average 200 EAs per province) were transferred onto one or more Bernoulli cartridges. Approximately 10 million records (population of 9 million and 1 million house-

holds) of 52 characters in length each were saved. Sixty Bernoulli cartridges of 10 megabytes each were used to store one copy of the country's data. Five versions of the country data were kept.

Table 1. Partial list of IMPS users for population censuses

Country name	Census date	Estimated population (millions)	Operational control	Data entry	Editing	Tabulation
Benin	1989	4.3	manual	CENTRY	CONCOR	CENTS
Burkina Faso	12/85	9	part autom.	RODE/PC	CONCOR	CENTS
Burundi	8/90	5	undecided	CENTRY	CONCOR	CENTS
Cameroon	4/87	10.3	manual	IBM S/36	CONCOR	CENTS
Central African Republic	10/88	2.7	dCONTROL	CENTRY	CONCOR	CENTS
Comoros	1990	.5	manual	undecided	CONCOR	CENTS
Côte d'Ivoire	3/88	10	manual	RODE/PC	CONCOR ¹	CENTS ¹
Honduras	5/88	4.8	dCONTROL	KeyEntry III	CONCOR ²	CENTS ²
Malawi	9/87	7.4	manual	ICL mini	CONCOR	CENTS
Mali	4/87	8.4	manual	RODE/PC	CONCOR	CENTS
Mauritania	10/87	1.9	manual	UNKNOWN	CONCOR	CENTS
Micronesia (Pohnpei)	9/85	.30	manual	Entrypoint	CONCOR	CENTS
Niger	5/88	7.5	dCONTROL	RODE/PC	CONCOR	CENTS
Philippines	1990	57	undecided	CENTRY	CONCOR ²	CENTS ²
Senegal	5/88	7.5	dCONTROL	RODE/PC	CONCOR	CENTS
Somalia	11/86	7.8	manual	custom pgm	CONCOR ¹	CENTS ¹
Swaziland	8/86	.7	manual	RODE/PC	CONCOR	CENTS
Tanzania	8/88	23.5	manual	ICL mini	CONCOR ²	CENTS ²
U.S. (Pacific outlying areas)	4/90	.3	manual	CENTRY	CONCOR	CENTS
Yemen A. R.	2/86	9	part autom.	custom pgm	CONCOR	CENTS

¹ Program development done on microcomputers; production processing on mainframe computers.
² Program development done on microcomputers; have not decided what to use for production processing.

The data stored include the raw data at enumeration area level (RODE/PC file format and ASCII format), the consolidated data at province level, and two copies of the edited data.

The management of about 600 diskettes and 240 Bernoulli cartridges was accomplished effectively by using a simple external labeling system. Each diskette was clearly identified by its external label indicating an enumeration area. Equally, every Bernoulli cartridge was identified by an external label indicating a province. The task of managing and processing the census data was manageable but inconvenient, particularly because the Bernoulli Boxes were capable of holding only 10 megabytes of data at a time, and the IBM PC/AT microcomputers had only 20 megabytes of hard disk storage. Due to technical difficulties, the tape units could only be used for archival purposes and not for processing the census data.

CONCOR and CENTS do not require the entire population census file to be processed or sorted as a unit. The data can be edited batch-by-batch using CONCOR. The batches can vary in size. The edited (or cleaned) batches of data can be merged into larger batches (or geographic units) for tabulation using CENTS. The resulting cross-tabulations can be saved as much smaller files, and then consolidated to produce tables for larger geographic entities. Thus, in a system using more than one microcomputer, each machine can be processing a different geographic area simultaneously, shortening the overall processing time.

In the case of Burkina Faso, both editing and tabulation were done at the province level. The CONCOR program took an average of 20 minutes to edit each of the 30 provinces. The execution of the CENTS tabulation programs for each province took about 15 minutes. Through CENTS, the intermediate province-level tabulations were merged to produce the country-level tabulations. Three IBM PC/AT microcomputers with 20 megabytes of storage were used for editing and tabulation.

The keying of the country data in nine months instead of the estimated one year was due primarily to the high keying rates attained. The keystrokes ranged from 9 000 to 14 000 per hour with an error rate of 0.3 %. Since INSD was not able to give monetary incentives to the keyers, they provided them with other types of motivation. For example, the best keyers could choose to work during the shift (morning or evening) that was more convenient to them and schedule their leave more freely. They also were promised a permanent job at the INSD after the census was over. Given the high unemployment rate in Burkina Faso, most keyers were motivated just to keep a relatively well paying job. Out of 500 applicants only 50 were selected after taking a written test and being interviewed. Above all, the keyers were hard working and took pride in their work.

Maintenance of the microcomputers was accomplished by the data processing team. A supply of spare parts allowed them to keep downtime to only 4 %. The diskette units and the keyboards caused most of the problems. In spite of these and some initial electrical problems, the equipment maintenance was done in a timely and nondisruptive fashion. As a pioneer, INSD demonstrated that processing a population census of 9 million persons using microcomputers is not only feasible and cost-effective, but an effective approach to the production of timely results.

3.2. *Senegal*

The Direction de la Statistique (DS) took the second national population census of the Republic of Senegal in May 1988. The estimated population was 7.5 million.

The DS chose to purchase microcomputers for the data entry operation instead of renting IBM 3742 equipment because of the reasonable cost of microcomputer hardware, the availability of adequate data entry software, and the potential use of microcomputers for other pro-

jects after the keying operations are finished. The editing and tabulation of the data were scheduled to be done using an IBM 370/145 at the Computer Center of the Ministry of Finance, a data production facility used by other government agencies.

The original microcomputer configuration included 18 IBM PC microcomputers with 256 kilobytes of memory for data entry, two IBM PC/XT microcomputers with 10 megabytes of hard disk storage for program development, five printers, and a tape unit used for data backup. Uninterrupted Power Supply (UPS) units are used for power supply protection.

The RODE/PC software package was selected for data entry. The data would be edited using CONCOR and tabulated using CENTS 4. The CONCOR and CENTS 4 programs could be developed using the microcomputers, then transferred to the mainframe computer for production runs.

During a three-week technical assistance visit to Dakar in January 1986, data processing advisors from ISPC, under an agreement with USAID, conducted a CENTS 4 workshop, installed CONCOR and CENTS 4 on the IBM mainframe computer, and discussed tabulation plans and specifications with DS personnel.

Originally, the data processing staff responsible for the census operation consisted of the head of the data processing section and the head of the data entry section of the DS. They furthered their training during a work-study visit in early 1986 at ISPC in Washington, D.C. During a six-week stay, the data entry supervisor, who had no programming experience, learned to use menu-driven RODE/PC and assisted in developing the data entry application for the census questionnaire. The data processing chief, an experienced systems analyst with prior CONCOR and CENTS 4 training, remained at ISPC for three months. In addition to learning RODE/PC, she wrote the CONCOR edit and imputation programs. She also completed much of the tabulation program-

ming using CENTS 4. All the work was done on an IBM PC/XT.

Upon their return to Senegal, they were able to implement modifications to the programs easily when significant changes were made to portions of the questionnaire. The data for the pilot census, conducted in March 1987, were keyed using microcomputers. The CONCOR and CENTS 4 programs were transferred to the mainframe computer. No changes to the code were required. The keyed pilot census data also were transferred from 5 1/4-inch diskettes to 8-inch diskettes to facilitate transfer to the mainframe computer.

Although they encountered no major problems while processing the pilot census data on the mainframe computer, the data processing staff became frustrated due to the inconvenience. They had to physically transport themselves and the data to the mainframe computer site several miles away. After conducting a test, they found that because of the need to share the computer facilities, it took longer to run CONCOR and CENTS 4 programs on the mainframe computer than it did on the dedicated microcomputers. The data processing staff also found the interaction with the IBM PC much friendlier than with the IBM mainframe computer.

As a result of their increasing success with microcomputers, the data processing staff decided to process the data for the full census using microcomputers instead of using mainframe computers. To upgrade the current microcomputer configuration to allow in-house processing for the full census, the data processing staff acquired extended mass storage consisting of three 20-megabyte Bernoulli Boxes and a tape drive unit. Two of these Bernoulli Boxes contain a hard disk with 80 megabytes of memory.

Although the file management of diskettes, Bernoulli cartridges, and tapes will be a critical component of the processing, the data processing staff have judged the file management

problems less formidable than the inconveniences of working in the mainframe computer environment. Additionally, the head of the data entry section feels confident about the handling of census data files because he developed and implemented the magnetic media management system used for the previous population census.

The DS acquired six additional IBM PC/AT microcomputers. Four are being used by the data processing staff for program development and for production processing. The other two microcomputers will be used by the demographers and statisticians for the analysis of the census data. DS plans to use the MCPDA for the demographic analysis of the census data.

The head of the data processing section has transferred her microcomputer expertise to three system analysts who also are working on the processing of the census data. Additionally, ISPC has been providing technical assistance to the DS in all aspects of the census processing activities. One of these areas is the monitoring of census processing activities and the production of management reports, including the preliminary counts report. The DS staff adapted the dCONTROL prototype to meet Senegal's specific needs. The preliminary counts reports which are based on manual counts of the May-June 1988 enumeration were available by October 1988. The data base and dCONTROL programs occupied less than two megabytes.

The DS plans to process a 10 % sample of the data first. However, the availability of a permanent site for storing the questionnaires and for coding and keying the data has delayed the start of the processing of the census data. The RODE/PC data entry program and the CONCOR editing programs have been tested. The CENTS tabulation programs are being finalized. A system to monitor operator performance based on operator statistics such as key-strokes per hour and keying error rates has been developed. This dBASE III PLUS system produces reports with operator statistics based

on the information found in the ASCII file produced by the RODE/PC data entry program.

In the meantime, the DS staff have been using RODE/PC and CENTS for other surveys. The DS staff's enthusiasm for microcomputer use extends even further. As part of the analysis and publication stage, the DS plans to use microcomputers for thematic mapping. This is the computerized creation of maps that reflect regional variations in characteristics of the population. It requires mapping software and peripheral equipment such as a digitizer and a plotter. The U.S. Bureau of the Census will provide training.

Over the last three years, the DS has become more confident and enthusiastic about microcomputers. Not only are they using microcomputers to process the census data, but they also are using them for many other projects and surveys. Most systems analysts have microcomputers on their desks and are providing microcomputer training to junior programmers. Demographers and statisticians also are using microcomputers for their work.

With the subject-matter expertise gained during the last census, the growing microcomputer literacy, and staff enthusiasm, the prognosis for the census processing is good.

4. General Prognosis

The experiences of Burkina Faso and Senegal are representative of the positive effect microcomputers and IMPS are having on the census processing activities of many countries. Three years ago these countries were pioneers in the use of microcomputers for censuses. Now over 18 countries are using or plan to use microcomputers for some aspect of their census processing. This number is increasing rapidly as statistical offices and donor agencies realize the advantages of microcomputers.

Operational control and file management are still a challenge. As recording densities of mass

storage devices continue to increase, operational control and file management will be easier, perhaps even easier than on mainframe computers. The dCONTROL census management and reporting system is helping to resolve this challenge.

In spite of improvements, data capture is still the most time-consuming operation in census data processing. There is a limit to the rate at which data can be keyed, regardless of the speed of the software. Although one could add work stations to diminish the total time for keying, there is a point at which the logistics of such an operation become highly cumbersome. As optical mark reading (OMR) technology advances, it should become a more cost-effective data entry alternative for developing countries within the next five years.

The availability of more cost-effective tools for processing census data should make it possible for statistical offices in developing countries to focus their attention and resources on the use and dissemination of census data. Many tools are becoming available to accomplish this objective, including software packages for the production of thematic maps and desktop publishing. Additionally, hardware improvements in optical disk technology allow for easier access and transfer of large volumes of data.

Census data users themselves have more tools available which should, in principle, help relieve statistical offices of some of the responsibility they have as the principal disseminators of census data. Since many census data users have never been able to obtain census results or have obtained them five or more years after enumeration, rendering their usefulness to almost null, the availability of any data is a big improvement. Census data users will welcome the ability to access a subset of the census data on diskettes that they can use on their own microcomputer systems. It is hoped that statistical offices and census data users will focus on the analysis and dissemination of reliable data and not be distracted by the many presentation

alternatives provided by the new technology.

As computer technology advances, data processing needs and software are changing. IMPS is the result of years of experience by ISPC staff, but is also the outcome of requests by census data processing managers in developing countries. As resources allow, ISPC will continue to support and modify the various modules of IMPS by critically observing its use by statistical offices in developing countries, listening to IMPS users regarding suggestions for improvements, and by taking full advantage of the microcomputer environment available in statistical offices in developing countries.

ISPC's primary objective is to make these packages usable by persons who are not highly skilled or experienced computer programmers. ISPC will make modifications to CONCOR and CENTS in the area of user-friendliness over the next few years. One goal is to make at least the definitional aspects of these packages interactive so that users need not learn a procedural language in order to use them. Some aspects of the packages, such as the definition of complex consistency edits are best stated through a procedural language. Others, such as the definition of table formats, are best done interactively. ISPC also will develop tutorials and extensive census examples to facilitate the learning process.

Enthusiasm for new technology should not diminish the effect that planning, politics, and communication have on the processing of a census. After all, microcomputers are only tools for people to use; they are not a substitute for careful planning and close coordination.

5. Conclusions

The experiences of developing countries like Burkina Faso and Senegal have made the use of microcomputers to process censuses of 10 million persons a reality. Technological developments over the last three years are making

some of the problems normally associated with large-scale processing quickly disappear. As more cost-effective tools become available for processing census data, statistical offices in developing countries should be able to focus their attention and resources on the use and dissemination of census data. We eagerly await the effect of future technological development on large-scale data processing.

6. Bibliography

- Anderson, P. and Ondra, T. (1987): Trip and Assessment Report to Pohnpei, Federated States of Micronesia. Unpublished report, U.S. Bureau of the Census, Washington D.C.
- Chamberlain, K. (1988): Trip Reports to Mogadishu, Somalia. Unpublished report, U.S. Bureau of the Census, Washington D.C.
- Chamberlain, K. (1988): Trip Reports to Dar Es Salaam, Tanzania. Unpublished report, U.S. Bureau of the Census, Washington D.C.
- Gomez, E. (1988): Trip Reports to Tegucigalpa, Honduras. Unpublished report, U.S. Bureau of the Census, Washington D.C.
- Gomez, E. Cuevas, M., and Peterson, L. (1987): Trip Reports to Tegucigalpa, Honduras. Unpublished report, U.S. Bureau of the Census, Washington D.C.
- Hie, J.-M. (1988): Trip Report to Ouagadougou, Burkina Faso. Economic Commission for Africa. Unpublished report, United Nations, New York.
- Le, T. and Toro, V. (1988): Trip Report to Dakar, Senegal. Unpublished report, U.S. Bureau of the Census, Washington D.C.
- Le, T. (1988): Trip Reports to Bujumbura, Burundi. Unpublished report, U.S. Bureau of the Census, Washington D.C.
- Megill, D. (1988): Trip Report to Niamey, Niger. Unpublished report, U.S. Bureau of the Census, Washington D.C.
- Megill, D. and Toro, V. (1988): Trip Report to Niamey, Niger. Unpublished report, U.S. Bureau of the Census, Washington D.C.
- Melaney, T. and Toro, V. (1988): Trip Report to Dakar, Senegal. Unpublished report, U.S. Bureau of the Census, Washington D.C.
- Rowland, S. and Melaney, T. (1988): Trip Report to Bujumbura, Burundi. Unpublished report, U.S. Bureau of the Census, Washington D.C.
- Sawaya, S. and Banerjee, K. (1987): Trip Report to Abidjan, Côte d'Ivoire. Unpublished report, U.S. Bureau of the Census, Washington D.C.
- Stroott, M. (1988): Trip Report to Sana'a, Yemen Arab Republic. Unpublished report, U.S. Bureau of the Census, Washington D.C.
- Toro, V., Le, T., and Megill, D. (1987): Trip Report to Niamey, Niger. Unpublished report, U.S. Bureau of the Census, Washington D.C.
- Toro, V., Stanecki, K., and Le, T. (1987): Trip Report to Dakar, Senegal. Unpublished report, U.S. Bureau of the Census, Washington D.C.
- Toro, V. and Melaney, T. (1987): The Use of Microcomputers for Census Processing in Developing Countries. Paper presented at the American Statistical Association Meetings, San Francisco, California. Unpublished.
- U.S. Bureau of the Census (1988): Population Census Support Status of Sub-Saharan Africa. Unpublished report, U.S. Bureau of the Census, Washington D.C.

Received November 1987
Revised March 1989