Journal of Official Statistics
Vol. 2, No. 4, 1986, pp. 545–553

How Increased Automation Will Improve the 1990 Census of Population and Housing of the United States

Peter Bounpane[1]

Abstract: The U.S. Bureau of the Census will increase significantly the automation of operations for the 1990 Census of Population and Housing, thus eliminating or reducing many of the labor-intensive clerical operations of past censuses and contributing to the speedier release of data products. An automated address control file will permit the computer to monitor the enumeration status of an address. The automated address file will also make it possible to begin electronic data processing concurrently with data collection, and, thus, 5–7 months earlier than for the 1980 Census. An automated geographic support system will assure consistency between various census geographic products, and computer-generated maps will be possible. Other areas where automation will be introduced or increased are questionnaire editing, coding of written entries on questionnaires, and reporting of progress and cost by field offices.

Key words: 1990 U.S. Census of Population and Housing; increased automation; automated address control file; automated geographic support system; earlier processing.

1. Introduction

The U.S. Census Bureau began planning the 1990 Census of Population and Housing – the Bicentennial Census of the United States – several years ago. Even though April 1, 1990, is still 3 years away, an early start was necessary because of the complexity of the issues and the time needed to implement decisions. The broad range of issues addressed in census planning are described in Bounpane (1985). Our goals for 1990 are to publish more timely data products and to make the whole census process more cost-effective while at the same time maintaining a high level of accuracy. In other words, we are attempting to make the census process more productive. We hope to achieve greater productivity by automating outmoded clerical operations and by entirely rethinking the data collection and data processing stages of the census.

Over the last century, the census has played an important role in the history of automated data processing in the United States. By 1890, the U.S. census had become an encyclopedic enumeration of the American people. The 1890 Census marked a great increase over previous censuses both in the number of inquiries and the volume of data tabulated and published. Census officials, realizing that something had to be done to speed up the processing and tabulation for the 1890 Census, gave a young engineer named Herman

[1] Assistant Director, Bureau of the Census, Washington, D.C. 20233, U.S.A.

Hollerith the assignment of constructing a quicker tabulating device. The electromechanical tabulating machine Hollerith developed for the 1890 Census – which read punched cards by electrical pulses – revolutionized both census-taking and statistical tabulating. Hollerith's machine was soon used worldwide in business and census applications (Austrian (1982)).

Hollerith's invention allowed greater volumes of data to be processed, more sophisticated cross-classifications, and all in a shorter time and at less cost. His punch-card system was modified and improved by the census machine shop for each successive census over the next 60 years.

Eventually, computers replaced the tabulating machine for processing data, and the census was again at the forefront of the technological revolution. UNIVAC-1, the first major computer system for civilian use, was installed at the Census Bureau in 1951 and was used to process part of the 1950 Census. Though large, cumbersome, and slow by today's standards, UNIVAC-1 was a major advance from the Hollerith tabulating system. Computers were used to process all of the 1960 Census, and, of course, the 1970 and 1980 Censuses.

Another new device accompanied the 1960 Census: FOSDIC, a replacement for keying, was introduced for entering data into the computer. FOSDIC is an acronym for Film Optical Sensing Device for Input to Computer. Questionnaires were microfilmed by special page-turning cameras, and FOSDIC read the data from microfilm into the computer. This advance, which was developed to meet the specialized needs of the decennial census, eliminated the need for key-punchers, saved time, and improved quality. FOSDIC has been used in the last three censuses.

The point of this brief history is that the decennial census, because of its massive workload and unique character, has called forth new technology, new tabulating, computing, and automated equipment to speed up the processing of census data.

As we examined our experience from the 1980 Census, we found that while the census was generally a success there was need for improvement. We determined that much of the improvement in timeliness, accuracy, and cost-efficiency could come from taking a fresh look at automation and increasing automation in the census.

While we do not yet know whether a specific automation decision will save money, we believe that our decisions will lead to a more efficient and accurate census. We will invest in automation that could reduce costs or that is necessary for maintaining or improving the quality of the census. Automating census operations will allow us to replace labor intensive and error-prone clerical operations with automated techniques that are quicker, more accurate, and easier to control.

While the automation advances we plan for the 1990 Census will not involve the development of new technologies, they will be based on innovative applications and refinements of existing technologies. The Census Bureau has embarked on a vigorous program to examine automation alternatives in test censuses before making choices for the 1990 Census. Since we are contemplating significant changes in automation for 1990, I will first describe how the 1980 Census was taken so the departures will be more easily understood.

2. 1980 Census

The 1980 Census was taken using the mail-out/mail-back procedure in areas of the country that contained 95 percent of the population. We purchased address lists for some of these areas and listed addresses ourselves elsewhere. In all cases, the address lists were then checked

and updated by the U.S. Postal Service and our own field personnel. The USPS delivered questionnaires to each housing unit a few days before census day and householders were asked to fill them out and mail them back to a temporary census district office on April 1st. The aim of this approach was to complete as much of the census as possible by the less costly mail method and then to do the costly and time-consuming follow-up of those housing units that did not return a questionnaire. We had received questionnaires for about 83 percent of the households within 2 weeks of our initial mailing. A large work force (270 000 at peak) visited nonresponding housing units and vacant units. In sparsely populated areas where mail-census procedures were not suitable, census enumerators went door-to-door to take the census (Cho and Hearn (1984, pp. 241–263)).

We set up 409 temporary district offices to carry out data collection. Most of the operations were done manually. For each office, a large number of clerks were hired to make changes (additions, deletions, corrections) to the address lists, check in mail-returned questionnaires and edit the questionnaires for completeness and consistency, assign housing units for follow-up, monitor the enumeration of the nonresponding units, and tally preliminary counts. Many of these operations can be considered "processing," but processing did not begin in earnest until the collection offices completed their work, closed, and shipped their questionnaires to one of three processing centers. The offices generally closed 5–7 months after census day.

At the processing centers, the questionnaires were microfilmed and the data read into the computer by FOSDIC. Though processing center operations were largely automated, written entries for many questionnaire items (e.g., ancestry and occupation) were manually given numeric codes prior to computer processing.

This system worked very well considering the amount of manual work involved and the sharp division between data collection and data processing. First, the Census Bureau met the deadline dictated by law for the release of apportionment and redistricting counts. Apportionment is the process whereby a state is awarded a share of the 435 seats in the House of Representatives based on its population; redistricting refers to the process of redrawing the boundaries of legislative districts within states based on the principle of "one person/one vote." Second, many of the small-area data were issued earlier than for the previous census. For example, the 1980 Census data for 2.5 million blocks used in redistricting were produced in less than 12 months. For the previous 1970 Census, similar data for 1.7 million blocks took 18 months to produce. Third, many more data, especially for race and Spanish-origin groups, were published. Still, we did not release some of the data products, particularly those based on the sample questions, as quickly as planned. (This delay was due in part to budget problems that forced us to cut staff and temporarily suspend sample coding operations.)

For the 1990 Census, we want again to meet our deadlines and we want to release other data products more quickly than before, as well as keep costs reasonable and make the counts as accurate as possible.

3. Automation Plans for 1990

We have identified a number of areas that are candidates for automation, and have already begun to test some of them.

3.1. Geography

Geographic materials are essential to a successful census for two reasons: First, having correct and legible maps helps our enumerators find every housing unit so that we have a complete

count; and second, having correct boundaries and geographic information helps us assign each housing unit and the people who live there to the appropriate land area. One of our problems in the 1980 Census was that our geographic materials, including the maps, were produced in separate operations involving a great deal of clerical work. This process was slow and error-prone, leading to delays in production and inconsistencies in some of the products.

For 1990 we are automating our geographic support system, which we are calling TIGER (Topologically Integrated Geographic Encoding and Referencing system). TIGER will integrate all the geographic information that was produced in separate operations in 1980. This will allow us to produce the geographic products and services for 1990 from one consistent data base, and will help us avoid some of the 1980 Census delays and inaccuracies. Having computer generated maps that match the geographic areas in our tabulations will be an improvement over the clerical operations of the 1980 and earlier censuses. For a full discussion of the automated geographic support system, see Marx (1986).

3.2. *Address control file*

Another improvement planned for the 1990 Census is the development of an automated address control file. Since we will again use the mail-out/mail-back methodology, an accurate and up-to-date address control file is essential. In 1980, although the initial control list of addresses was computerized, changes in the address file during the census were made manually. For 1990, we will have continuous access to the automated address control file so that we can keep the list current.

With an automated address file, it will be much easier to determine whether or not we included a specific address in the file. It also will be possible to update the file where we missed an address in earlier operations. We can imprint bar codes (like those on supermarket items) on the questionnaires and use electronic equipment to read the information in the bar codes. Thus, we can use the computer for checking in and keeping track of census questionnaires instead of doing check-in manually as we did in 1980. As a result, it will be easier for our enumeration staff to identify nonrespondents' addresses.

Finally, with an automated address list, we can update the list and use it in future Census Bureau operations. In our 1985 and 1986 test censuses, we successfully implemented an automated address control file and automated check-in.

3.3. *Earlier data conversion*

One of the most promising ways to take advantage of automation in the census, and our biggest challenge, is to convert the data on the questionnaires into a computer-readable format earlier in the census process than in past censuses. This approach is essential if we are going to take full advantage of automation and release data products quicker.

In 1980, the conversion of data to machine-readable form did not begin until after the district offices completed all enumeration, edits, and follow-ups and shipped all questionnaires to one of the three automated processing centers. This was a sequential process. This meant that many completed questionnaires that could have been automatically processed early in the census, lay around for several months until the district offices closed. Also, because we did not have an automated address control file, we had to process all the questionnaires for an enumeration district in one batch.

The automated address control file for the 1990 Census will allow us to conduct flow processing, and to do it concurrently with data collection. An earlier start in 1990 (5–7 months ahead of the 1980 schedule) will allow more time for review and correction and will

enable the computer to assist in certain census operations. It will contribute to the early identification of enumeration problems. Also, by converting questionnaire data to machine-readable form sooner, we can minimize the loss of data when original questionnaires are accidently damaged or destroyed. Finally, and perhaps most importantly, it will help us meet our goal of disseminating data products more quickly.

Planning for concurrent processing in the 1990 Census has centered on two major questions: Where and how would it be done? The "where" issue involves the number of processing offices and the degree of centralization or decentralization. In 1980 we processed the census questionnaires sequentially and had three processing centers. With concurrent processing, having so few centers probably would not be feasible because of the need to move materials quickly between processing and collection offices. Greater centralization of processing activities also places greater staffing burdens on the center, i.e., the need to hire more employees in one area.

We weighed these concerns against problems related to decentralization – the need for more hardware and the difficulties of controlling and supporting many processing offices.

The "how" issue involves the technology we will use to convert questionnaire data into a computer-readable format. In the 1980 Census, we employed the FACT-80 system (with FOSDIC technology as the base) to convert microfilm directly to computer tape. FACT is an acronym for FOSDIC and Automated Camera Technology. The complete data-conversion system consists of high-speed cameras that film the questionnaires, film developers to process the rolls of microfilm, and the FOSDIC machines that read the data from microfilm to computer tape.

We also looked at key-entry as a primary data conversion methodology. Both FOSDIC and keying are tested methodologies that have proved workable over the years. Because there are technical limitations to how many FOSDIC systems we can build and maintain for 1990, we had considered data keying to give us maximum flexibility in decentralization. Keying was not considered as a viable option as the sole data conversion technology for the entire census because of the large numbers of keyers and key stations that would be required.

Earlier in our planning we had also considered a third technology – optical mark recognition (OMR). OMR provides direct input of data into the computer, whereas with FOSDIC the questionnaires must be filmed first. As with keying, we considered OMR to allow us more flexibility in decentralizing our processing. We tested OMR in our 1985 Census in Tampa, Florida. Based on some of the problems experienced with OMR in this test, and on other concerns about cost, timing, environmental controls, and so on, we decided not to pursue further testing of OMR technology for use in 1990. We will, however, consider testing OMR and other technologies in 1990 for possible use in the 2000 Census.

In April 1986, after reviewing these two main issues at planning conferences and in internal working groups, we were able to reach some decisions. We have decided to set up eleven processing centers for the 1990 Census where we will use FACT 90 (an update of the 1980 system, still with FOSDIC as the base) to convert the data to machine-readable format.

We determined that having two primary data conversion technologies (FOSDIC and keying) would have excessively complicated our processing system for 1990. We will use keying only as a supplement to FOSDIC for entering some of the handwritten data on the questionnaires into computer-readable form.

We will have two types of district offices for which the questionnaire flows will be different. For district offices in certain high population

density areas the processing centers will receive the questionnaires, perform automated check-in using laser sorters, immediately convert the questionnaires to computer-readable form, and thereby perform an automated review (edit) of the questionnaires. The district offices covered by these processing offices will likely correspond to some of our "centralized" offices in 1980 – the more hard-to-enumerate urban cores where recruiting enough temporary census workers can be difficult. These district offices will not need to hire many office clerical workers and can concentrate on field follow-up activities for households that did not mail back their questionnaires or that mailed back incomplete questionnaires.

District offices in the rest of the country will receive the returned questionnaires; use pencil-shaped, electronic "wands" attached to micro computers to read the bar codes on the questionnaries and, thus, perform automated check-in; and conduct clerical edits for completeness. Once questionnaires pass the edit, they will be sent on a flow basis to a processing center for data conversion (using FACT-90).

This decision represents a careful balance of staffing, equipment, and workload considerations as they relate to the processing and collection offices. We will have an automated address control file and automated check-in for the entire area covered by the mail-out/mail-back census, and we will achieve our goal of concurrent processing by converting questionnaire data to computer-readable format on a flow basis, several months earlier than for the 1980 Census.

So far I have discussed our plans with regard to automating geographic materials and the address control file and beginning data conversion earlier. We will increase or improve automation in other areas to help speed up the census and make it more accurate, and I will discuss briefly a few of these areas.

3.4. *Computer edits*

One area is questionnaire edit. Edit is a repetitive and monotonous job better suited for computers than people. Entering data from the questionnaires to the computer earlier in the census process will allow computer editing of the questionnaire data earlier than ever before. These edits will check the completeness and consistency of the data. In 1980, the questionnaires were manually edited in the district offices, basically to check that they had been answered completely; then, once the questionnaires went through the FOSDIC machines, the computer edited them for completeness and consistency. For 1990, manual editing would be eliminated in some district offices and replaced by computer edits.

3.5. *Automated coding*

Another promising automation technique relates to the coding of handwritten entries on the questionnaire. In 1980, manually coding the handwritten entries on questionnaires involved a large, time-consuming, and costly clerical operation. For 1990, we might be able to key handwritten responses into the computer and develop software that would assign the appropriate computer-readable codes. We cannot eliminate all clerical involvement in coding, because some handwritten responses will be incomplete or uncodable and will have to be handled by our referral units. We will, however, be able to significantly reduce the amount of manual work and, thus, save time and improve the quality of the data. Instead of a clerk having to look up the occupation "statistician" in a reference manual, find the numerical code, and fill the appropriate coding box on the questionnaire, the clerk can type in the word "statistician" and the computer will automatically assign a code and enter that onto a computer record. Thus, the

time-consuming looking-up and circle-filling are eliminated. At this time, we do not know precisely the extent that the computer will be able to assign codes without clerical intervention.

3.6. Management and administration

We will also use automation to help us plan and monitor the census. The Census Bureau is developing an elaborate automated management information data base to see that we meet important dates in making decisions for the 1990 Census. The management information system was used to help us keep track of operations for our 1985 and 1986 test censuses. In addition to serving as an aid in planning the 1990 Census, the management information system will give us up-to-the-minute cost and progress data so that we can monitor actual 1990 Census operations. In 1980, cost and progress reports were not integrated with other management reports, and some of the cost and progress information was several days old by the time managers received it.

Automation will help us control and monitor many other administrative functions. We will have an automated payroll system, as in 1980. And for 1990, we will also have, on a microcomputer, a new automated employee file that will help us organize needed information about our large temporary work force. (We did this in our 1985 test census.) For instance, we will know whether we are meeting our hiring goals in each enumeration area and we can use the file to help us make enumerator assignments. We will also have a new automated inventory control system to manage the procurement and distribution of the large volume of specialized supplies needed to take the census.

3.7. Data products

Finally, we are looking at further automation of our tabulation and publication operations for the 1990 Census. The actual tabulation of data was fully computerized for the 1980 Census, but the design and review of specifications and the review of test data was largely manual. We want to use the computer in our development of specifications and the analytical review of the tabulated data for 1990. This review, which looks for errors and anomalies in the data, is essential to maintaining the quality of our data products. Using the computer will improve this analysis.

New automation techniques will also play a part in the dissemination of our data products for the 1990 Census. While the Census Bureau will continue to produce paper reports and large summary computer tape files, we must also address the needs of small computer users who will want products on floppy disks. Another new development we will consider for 1990 will be an online data base in which users can access summary data from their office computers using the telephone. The Census Bureau has already implemented such a system, called CENDATA, on a limited basis. There may be other developments in the next few years – such as improvements in laser disks – that we will be able to take advantage of for the 1990 Census. Fortunately, our final decisions on tabulations and data products can be made later in the decade, so we can take advantage of new technologies.

4. Closing

There is a sense of excitement at the Census Bureau about these automation possibilities, but some words of caution should be added. The systems developed must be simple, because they will be operated by a temporary work force with minimal training. The systems must be fully tested, proven reliable, and essentially "fail safe" to avoid crippling breakdowns. The equipment must be reasonably

priced and should either continue to have value to the Census Bureau or be marketable to someone else upon completion of the census.

Most of all, as we look to increasing automation in the census, we must take care to ensure that the confidentiality of the data we collect is maintained both in fact and in appearance. Only by maintaining the confidentiality of the census process can we ensure a high level of public trust and cooperation. The Census Bureau is proud of its record of protecting confidentiality and is constantly looking for ways to maintain and improve that protection.

The Census Bureau does not release data about individuals to anyone, including other Federal government agencies. But the sometimes menacing implications of technology require that we increase our efforts to convince individuals that they cannot be harmed by answering the census and that the information they provide is strictly confidential by law.

Automation is one of the key areas we are examining as we plan the 1990 U.S. Census of Population and Housing. There are many other issues, of course, that go into making a successful census: the basic procedures we will use to collect the data, the content of the questionnaires, hiring good temporary staff, and promotion of the census, including contacts and consultation with various groups and individuals interested in the census. However, automating many of the census tasks performed clerically in 1980 and previous censuses can help us to take the census more quickly, allowing us to meet our legal mandates for releasing apportionment and redistricting counts and to release other data products quickly. Automation could also help us introduce cost-efficiencies into many areas, improve accuracy, and also allow for better control of the census process.

Traditionally, U.S. census data collection and much of the census data processing (e.g., questionnaire check-in against the address control list, edit of questionnaires for completeness, and coding of handwritten responses) have been paper- and people-intensive tasks. The use of automated equipment can help to deal with the mountains of paper and the thousands of clerical tasks in a much more efficient and controlled way. Hiring, training, and finding space for all the people who have been needed to perform the numerous operations in past censuses have required a lot of time and money. While the 1990 Census will also require a large number of temporary workers, we are looking at ways to cut down on the number of labor-intensive activities and to use automated systems to control the census process.

We have been working on our automation plans for some time now. We tested some new approaches in our test censuses in 1985 in Tampa, Florida, and in Jersey City, New Jersey, and conducted further tests of automation this year in part of Los Angeles County, California, and in several counties in east central Mississippi. These tests are very important as laboratories where we can try out optional approaches. There will be further testing in 1987 and a dress rehearsal in 1988.

While there are many decisions yet to be made and problems to be worked out, we have progressed far enough in our automation planning to say this: there will be significantly more automation in the 1990 Census than in any previous census. We will make innovative use of automation techniques to perform data-entry earlier than ever before. We will have an automated geographic support system. We will edit questionnaires by computer. And we have already implemented an automated address control file, automated questionnaire check-in, and an automated management information system in our test censuses, and plan to have these features in 1990. Thus, we are optimistic that we are on the verge of important advances in applying automation to

census-taking. That is fitting since 1990 will mark the 200th anniversary of the first U.S. census in 1790.

5. References

Austrian, G.D. (1982): Herman Hollerith: Forgotten Giant of Information Processing. Columbia University Press, New York.

Bounpane, P.A. (1985): Plans and Issues Facing the 1990 Decennial Census. Government Information Quarterly, Vol. 2, No. 4, pp. 369–387.

Cho, L.J. and Hearn, R.L., (eds.) (1984): Censuses of Asia and the Pacific: 1980 Round, pp. 241–263. East-West Population Institute, Honolulu.

Marx, R.W. (1986): The TIGER System: Automating the Geographic Structure of the United States Census. Government Publications Review, Vol. 13, pp. 181–201.

Received July 1986
Revised December 1986