

WHAT CAN THE BUREAU OF THE CENSUS SUPPLY?

Conrad Taeuber
U. S. Bureau of the Census

It seems useful to broaden the topic question by adding the related questions, When? How? For what areas? In what form? At what cost? The answers to these questions are necessarily interrelated. Some items become available earlier than others. The form in which a user wishes to have the data and the cost involved depend to a large extent on the number and size of the areas for which data become available. Other papers will deal in some detail with the specific issues in relation to the journey-to-work data. My purpose is to describe the situation as it relates to the broad scope of census data, many of which are of direct concern to all transportation planning.

WHAT?

Before commenting on what will be available, let me underscore an important item that will not be available. Under no circumstances will the Bureau of the Census make available lists of names and addresses, with or without any individual characteristics relating to a person or to an address. Statistical tables will not be made available if the characteristics of an individual would be revealed by them. This will have some effect on those users who expect to deal with the data for the very smallest areas, for here it may be necessary at times to suppress some of the figures for an area because the population or number of housing units in the area is so small that the statistical totals would constitute a possible disclosure. We realize that at times this may seem to impose limitations on highly desirable uses of the statistics, but preserving confidentiality of the individual census data is more important than avoiding some occasional inconvenience to a serious user.

The inclusion of this topic on the conference program reflects a recognition of the importance of the volume of data that the census provides. The most widely used statistic from the census is no doubt that of the head count. Among the questions asked most frequently are the following: How many people live in each of the communities that are served? How has this number grown or declined? Is further growth or decline indicated in the age composition of the people who are living there now? Another matter of early concern is the location of the population in big cities, in suburban areas, in smaller cities or towns, on farms, or in the open country areas outside the limits of any town or city. The distribution of the population as homeowners and renters and the size and nature of the dwelling units that they occupy are matters of considerable interest. So is the growing use of trailers.

The size of families and the number of large and small families may be of considerable importance, as may also be the growth in the number and proportion of one-person households. The income levels of the residents of an area are important for the development of antipoverty and welfare programs as well as for the possibility of developing new markets. The older couples who no longer have to provide for the care and rearing of children may represent a group to which particular attention is devoted. This may also be true of the growing number of senior citizens in the population. National origins of the population and the persistence of cultural patterns brought from overseas or from some other part of the United States are often of importance. The

level of education of the people in an area and the possibility that the area continually exports its best educated youth to other areas are also of importance. The presence of persons with certain occupational skills may suggest the possibility of some industrial development or the need for new efforts at vocational training.

The age, size, and character of the housing and the facilities and equipment in the homes may indicate significant needs for the expansion of public or private utilities and services. There is much interest in the volume and the characteristics of new construction in cities and suburban areas, the rapid increase of multiunit structures in suburban areas, and the changes in value of homes for buying or renting. Businessmen are interested in the amount, size, cost, location, and equipment of vacant units. Differences in these and similar measures, such as between rural and urban areas, small and large cities, central cities and suburban areas, are all important matters for anyone concerned with providing essential services in each area.

The pattern of settlement of a community, the contrasts in density among various areas, and the proportion of units in single-family and multifamily structures have an important bearing on the generation of traffic. The presence of automobiles, the number of families with two or more workers, and the presence of children of school age are all matters that affect traffic patterns as well as the need for a variety of community services.

The list of important information can be extended considerably. The census is an inventory of the Nation at a particular point in time. This inventory can be compared with a similar inventory taken 10 years earlier, and trends can be discerned from such a comparison as well as from an analysis of the detailed data as of the census date. The number of children of preschool age is a useful indicator of the needs for schoolrooms and teachers in the near future; the number of middle-aged adults is an indicator of the future growth in the number of senior citizens. The fact is that much important information is directly available in the standard tabulations and that more can be secured if more cross tabulations are prepared. Imagination and money are really the major limitations on the amount of information that can be extracted from the census data.

The suggestion has been made at times that the Bureau of the Census should take a list of names and addresses and match it to census records, giving the sponsor a tabulation of the relevant characteristics of the cases that could be identified for comparison with census tabulations for the general population. Although such work can be arranged within the strict requirements of the confidentiality of census information, the Bureau is not prepared to undertake such projects. Experience has shown that the results of such analyses are likely to lack statistical reliability because of the difficulty of securing an adequate degree of matching. The address coding guide, which has been developed for many metropolitan areas, provides a means of determining in what geographic area any particular address falls; it gives the address ranges for every block side in the area covered. Thus, any address in the proper format can be readily allocated to the area to which it belongs. A local agency can then tabulate the characteristics of its own cases in relation to census areas such as tracts, and statistical comparisons can then be made on the basis of the data for common areas.

FOR WHAT AREAS?

The special contribution of the census is that it reaches into every part of the United States; it is the one governmental activity that involves every person throughout the country. It provides some results for every city or town or village, for every county or township or other subdivision, and for every state. Within all of the standard metropolitan statistical areas (SMSA's), census tracts have been established; each tract has an average population of about 4,000 to 5,000 persons. For cities of 50,000 and more, and their adjoining built-up areas, a limited amount of information is provided by city blocks. Arrangements have been made to have statistics provided (at cost) on the basis of city blocks for a number of smaller cities.

A number of general statements can be made about the areas for which census data will be tabulated. The questions that are asked of 100 percent of the population can be

tabulated for areas as small as city blocks. The questions that are asked on a 15 or a 20 percent basis can be tabulated for areas as small as census tracts and cities of similar size. The questions that are asked of the 5 percent sample can be tabulated for places of 10,000 or more. This refers in the main to the relatively simple tabulations. Cross tabulations, especially those involving several variables, would require somewhat larger base populations. In general, the more detailed cross tabulations will be prepared only for larger places, cities, metropolitan areas, and states.

Information will be available for rural and urban areas. In census usage, an urban area includes (a) any incorporated place of 2,500 persons or more and (b) the densely built-up areas adjoining cities of 50,000 and more, whether incorporated or not. The Bureau of the Census has also identified unincorporated settlements of 1,000 persons or more and assigned boundaries to them. If such an unincorporated place has a population of 2,500 or more, it, too, is included as urban territory. Areas that are not classified as urban are regarded as rural. Rural includes incorporated places of fewer than 2,500 persons and unincorporated places of 1,000 to 2,500 people living on farms and others living in the open country, in the developments located along highways and having no recognized boundaries, and in small clusters on the outskirts of many cities of fewer than 50,000 persons. Urban places and urban areas are recognized in many of the tabulations. Typically the rural population is divided into 2 groups, the farm and the non-farm. The farm population includes all persons living on farms, whether they are actively engaged in agriculture or not.

The fact that we have developed a system whereby each address in the urbanized area can be coded to the appropriate block side has led to an expectation that statistics might be made available for individual block sides. This is not the case. The only purpose to be served with the identification to block sides or block faces is to provide a capability of preparing tabulations for areas other than those recognized in the standard tabulations. Block sides could be combined into areas within a specified distance for shopping centers, school districts, health districts, or other areas. The standard tabulations could be prepared for such ad hoc areas at cost. Such tabulations can be prepared only at the Bureau of the Census because the summary tapes to be made publicly available will not contain the block-side identification.

The Bureau plans to prepare some tabulations by zip code areas, approximating the areas as necessary. These tabulations are not to be published but will be available on a special-order basis in the form of printouts or summary tapes. It is planned to use the 5-digit codes in the city delivery areas of major metropolitan areas and 3-digit codes elsewhere.

IN WHAT FORM?

Publication in the form of printed reports will be continued along the lines that were followed in previous censuses. However, in recognition of the fact that many users will prefer to have the data in the form of magnetic tapes, any material that appears in the printed reports will also be available in the form of summary tapes. The summary tapes will actually provide substantially more detailed data than can be included in the printed reports. The summary tapes are likely to be available somewhat earlier than the printed reports. Summary information will also be available in the form of microfilm or microfiche.

Information for some 34,600 census tracts will be published in printed reports, one for each SMSA. Because the tabulations based on the 100 percent questions will be available earlier than those based on the sample questions, tapes or printouts of the 100 percent materials will be made available to the local census tract committees as soon as the tabulations for the area have been completed. They will be free to publish these data on their own, or to make copies for distribution as they see fit. Copies of these printouts may also be secured at cost from the Bureau of the Census. The information contained in these printouts will later be included in the census tract reports.

One other form in which some census information will be made available is through public-use samples. As in 1960, the public-use samples will consist of basic records, without any names and addresses, and with limited geographic information. Information

in these records is coded in such a way that no identification of any individual is possible.

All information to be made available for public use, whether in the form of printed reports or of magnetic tapes, will be carefully screened in order to avoid disclosure of information that the Bureau of the Census is required to hold in confidence. Should there be any possibility of a danger of direct or indirect disclosure, the data will be suppressed or two or more areas will be combined.

Maps are available for all areas for which tabulations will be made available. As a part of the preparatory work for the 1970 census, the Bureau undertook the preparation of metropolitan area maps, compiling these from the variety of maps that were available for each of the areas. For the urbanized areas, these maps are developed to a common scale and with a uniform treatment of the culture shown on the maps. Copies of these maps will be available at cost.

WHEN?

The first preliminary results, in the form of simple head counts, became available on a flow basis in May, 1970, and the last of these was issued in September. Beginning in August 1970 and continuing through December, the final official head count figures in a series of advance reports were made available. These, too, were on a flow basis. Before the end of that series, the first state reports were made available; these give data on age, sex, race, and relationship to head of household. The data on housing characteristics, issued at the same time, simply give the count of housing units for each place. The reports giving somewhat more detail will be available between October 1970 and May 1971.

At this point it may be useful to emphasize an important element in the entire census process that has a major effect on when data become available. There were 3 census questionnaires. One of them was to be filled out for 80 percent of the population, one for 15 percent, and the other for 5 percent. The questions that were asked of 80 percent of the population were also included on the 15 and 5 percent questionnaires. In other words, there was a core of questions that everyone was expected to answer. In census jargon we refer to these as the 100 percent questions. When you filled out your questionnaire you may or may not have noticed that these 100 percent questions required no manual coding. All of the responses were to be given in the form of filled-in circles. This was not true of all of the questions on the sample questionnaires. There, some of the answers were to be given in written form, such as state or country of birth, occupation and industry, and place of residence 5 years ago. These require manual coding into a form that the electronic equipment can read. The consequence for timing is that the tabulation of the sample information cannot be started as rapidly as that for the information collected on a 100 percent basis.

As soon as the questionnaires are received at the Jeffersonville, Indiana, processing office, the 100 percent portions are microfilmed. By means of an optical scanner, called FOSDIC (Film Optical Sensing Device for Input to Computer), the entries on the microfilm are converted to entries on magnetic tape, and then the tabulation process starts. In the meantime, manual coding is done on those questionnaires that require it. They are later microfilmed, and then the tabulation of the sample data is started.

The priority given to the 100 percent items is not only a mechanical one. There is an important legal element as well. By law, the official population totals for states must be delivered to the President by December 1. This is necessary in order to ensure the proper flow of actions that will lead to the reapportionment of the seats in the House of Representatives for the elections of November 1972.

Although priority is given to the 100 percent items, the tabulation of the sample items receives early attention. The issuing of reports based on the sample data begins February 1971 and proceeds rapidly until these reports are completed.

There is one exception to the priority of tabulations of the 100 percent data. That relates to the items to be published by city blocks. The block statistics cannot be produced as part of the first run of the 100 percent materials. They require a separate run through the electronic computers. The timing for these reports is January to June

1971. The fact that this will require the printing of a line of data for each of about 1.5 million blocks will give some concept of the size of this particular job.

Timetables at best reflect targets; only after the fact is it possible to judge how closely we have met them. The December 1 date is one that we must meet; the others we expect to meet, but we are not required to do so. Until the field enumeration is completed we are dealing with a series of events that we can control only partially. It is possible to say that that stage of programming and debugging of programs is sufficiently far advanced that we have no misgivings on that score, and the availability of the necessary computer capacity is also assured. We feel confident that we will be able to hire, train, and retain the necessary complement of clerical workers to complete the manual coding on time. Manual coding is expected to require some 200,000 clerk-days in order to handle the coding of 20 items on questionnaires relating to more than 40 million persons. Not all items apply to each of them, but the sheets relating to each of these persons must be scanned in order to identify and code the relevant items.

Unanswerable at this time is the question, "When will the data for the state, city, SMSA, or group of blocks in which a person is particularly interested become available?" The work flow is necessarily such that it is not possible to start with a rigid order of states. Instead, work units will be processed as they become available, which means that the order initially depends on the time when the field collection is completed. It will be some time after that before a definite timetable for each area can be established. Experience in previous censuses suggests that the first states to be completed will be some of the less heavily populated ones and that the states with the larger metropolitan areas may well be last.

One element that affects the timing of access to the census results is that, with the equipment being used, the customer who wishes to have his data on computer tape can have them somewhat earlier than the customer who wishes to have printed reports. This is not a matter of deliberate discrimination against those persons who are dependent on the old-fashioned methods of reading and assembling data. Even though much of the printing will be done by the Government Printing Office with the most advanced electronic printing devices, it will still be possible to supply tape copies earlier than the printed reports, by a matter of a few weeks.

WHAT WILL IT COST?

Predictably the answer to the question of cost is bound to be "That all depends." The user who sends in an order for every summary tape that will be made available from the 1970 census had better prepare himself for a total of 2,054 reels of tape. At the standard price of \$60 per reel, the total cost would be \$123,240.

However, it is not necessary to secure all the tapes for every area in the United States. One can be selective. If a user wishes to have only the tapes available for Kentucky, he would order only 31 tapes at a cost of \$1,860. However, he might wish to have only certain tapes for Kentucky. If his needs are limited to one or another count, he might order only 1, 2, 3, or 4 reels. The tapes will be so arranged that data for only one state appear on any one reel or a given set of reels.

All information relates to a specific area. Insofar as these areas and their boundaries are readily known, there is no problem in relating the data to their appropriate areas. However, if small areas are to be used, it may be necessary to have maps that will aid in locating them. Such maps are available from the Bureau of the Census. The need for maps applies especially to the situations in which tapes relate to enumeration districts or block groups (in those cities in which they serve in place of the standard enumeration districts). Data for enumeration districts are not to be published, but summary data for these small areas will be made available on computer tape or as printouts from such tape. An enumeration district is the basic small area that is used to control the work in the field, and this serves as the work unit in the processing stages as well. Enumeration districts are small areas, generally with a population of about 800 persons. The person who wishes to use enumeration districts as the building blocks for his own geographic entities will need to secure also an appropriate set of maps showing what the boundaries of these districts are. He will also wish to have the

Master Enumeration District List (MEDLIST) that furnishes area and place names corresponding to the numeric identification codes that are used on the tapes. The MEDLIST and the enumeration district maps for a state the size of Kentucky are expected to cost about \$400.

It is not expected that the needs of all users will be satisfied through the materials that will be publicly available. The Bureau of the Census expects to be able to prepare special tabulations to meet special needs. The charge for such tabulations will be the actual out-of-pocket cost incurred by the Bureau. These will need to be estimated separately for each job.

It will be possible to secure data and also special compilations of the summary data through access to summary tape processing centers that can be found in all parts of the United States. Many of these are local nonprofit organizations, including some chambers of commerce, university research bureaus, and metropolitan planning organizations. A number of profit-oriented groups and service bureaus also plan to acquire tapes and to perform special compilations from these summary tapes. The Bureau of the Census will be glad to supply a list of organizations that have informed it of their intention to render such services to the public. The summary tapes that will be available to such organizations include only information that can be made publicly available. Under no circumstances will they have records from which any individual or any address can be identified.

CONCLUSION

To give a complete description of the materials that will be available on the basis of the 1970 census requires far more space than is available here. Those who wish to pursue any of the possibilities mentioned or have other inquiries should get in touch with the Bureau of the Census, Washington, D. C., 20233, for suggestions on ways that specific needs can be met.