PROGRAMS OF THE U.S. BUREAU OF THE CENSUS RELATED TO HIGHWAY PLANNING

Donald E. Church, Transportation Division, U.S. Bureau of the Census

This paper discusses the use of Census Bureau programs as input sources for socioeconomic data for automating highway planning. I am using the term automation in a broad sense that implies intensive use of computers and models for combining a wide spectrum of relevant facts or factors and distilling this mass of detail into a limited number of findings that can be used as part of the information base for making judgments and formulating policy.

Taken in this broad sense, almost all types of socioeconomic facts may be useful as input in some phase of an automation program. The most pervasive factor is population—its size, geographic distribution, and other characteristics as shown by the census of population and housing. The size, characteristics, and geographic location of economic activity also are influential. Sources of these inputs are the censuses of agriculture, manufactures, mineral industries, wholesale, retail, service industries, and construction. The census of transportation data provide a third dimension on interregional or interarea commodity and passenger flows as well as transportation.

Highway planning, by its very nature, must be dynamic and based on anticipated future situations. Benchmark data from the censuses, normally linked over time, are used widely in making projections for the future. Frequently, the projections are simple extensions of past trends; but, with the increased availability of computers and automated procedures, complex methods (including models) often are used that involve interrelationships among many basic factors. Among the census programs of value for quantifying basic relationships and factors for estimating probable trends are the well-known censuses and the current programs such as the annual survey of manufactures.

The Bureau has developed a substantial variety of computer software for its own use, some of which are applicable for automating highway planning. Among them are the "Admatch" package for assigning geographic codes to local records by matching addresses in a geographic base system. This is a method for automation of highly detailed geographic aspects within major urban areas.

The Admatch-Dime system may be characterized as "microgeographic" in contrast to the "macrogeographic" computer program, called PICADAD, which is used for automating all geographic aspects of the census of transportation. PICADAD is especially valuable for identifying origin-destination couplets in the nation as a whole, aggregating observations into selected origin-destination areas, and measuring straight-line distances between any given pair of areas that are not in the same city or locality.

Other Bureau publication programs of direct interest to automation, include the Statistical Abstract of the United States, which has been issued annually since 1878. The abstract is most commonly used as a handy source of data on a wide range of subjects. It contains an appendix entitled Guide to Sources of Statistics and source citations at the foot of roughly 1,300 tables and charts drawn from data issued by more than 200 federal, international, and private agencies. The source citation suggests the agency (or other source) that one may contact to obtain more details and further data.

A new source book is the Directory of Non-Federal Statistics for States and Local Areas, which was issued for the first time in March 1970. In addition, there are such general-purpose publications as the County and City Data Book, County Business Patterns, and Location of Manufacturing Plants.

Much of our work is related to information directly useful to the transportation field. The 1970 censuses of population and housing provide a wealth of detail on urban transportation problems. The census of population obtained the following information from a 15 percent sample of households: residence location, work location and transportation mode used to travel to work during previous week. This was supplemented by the census of housing, which obtained data on the number of automobiles owned or regularly used by members of the household.

The ability to couple the residential location with the work location provides a powerful tool for computing "travel desire lines" and a host of other urban highway planning applications. (This is the subject of the Highway Research Board Special Report 121.) The distribution of automobiles within urban areas provides a measure of the availability of private transport. The usefulness of both of these sets of data increase manyfold when combined with other small-area statistics relating to land use patterns, and income levels.

The census of transportation is designed to obtain data to fill, or at least narrow, the serious gaps that exist in the transportation data field. So far, the program (divided into three surveys) has been concerned principally with intercity and interstate transportation and aimed specifically at three "blind spots": the nation's truck fleet, transportation of commodities from point of production to market or redistribution point, and passenger travel. Each of the surveys is based on samples rather than complete enumerations. The primary objective is to obtain national data, with as much state or other area detail as feasible.

The truck inventory and use survey provides estimates of the number of trucks and truck-miles classified by such characteristics as size, body type, occupational use, area of operation, and fleet size. With respect to geography, estimates are presented for each state and are divided into three ranges of operation—local, short range (beyond local but usually not more than a 200-mile radius), and long range (more than a 200-mile radius). The 1967 final report contains nearly 700 pages—mostly tables—designed to meet general public needs. Even that many tables did not exhaust the potentially useful data obtainable from the survey. Consequently, a public-use tape is available (at cost) for special analyses.

One geographic difference between the truck survey and the other parts of the transportation program is worth emphasizing at this point: The truck survey shows the vehicle population and characteristics in specified geographic areas (states, divisions, and regions), but it does not include origin-destination or flow data. On the other hand, the main thrust of the commodity and travel surveys is flow data, showing origin and destination areas insofar as feasible.

The commodity transportation survey provides data on the intercity shipment of commodities originated at manufacturing plants. Information includes means of transport, size of plant, size of shipment, and distance and other spatial relationships between point of production and destination. The 1967 survey was based on a probability sample of about 1.4 million shipping records, and the final report contained roughly 2,700 pages. As judged by the number of requests for additional details or different breakdowns, that voluminous report did not nearly exhaust the potential of the survey. Public-use tapes have been created that provide maximum detail without disclosing activities of individual establishments or companies.

The creation of public-use tapes as well as special-purpose tapes or special tabulations, especially in the commodity transportation and travel areas, leads to a problem we all share. One side of the problem is the fact that the demand for origin-destination detail is almost infinite. The other side of the problem is the equally realistic and stubborn fact that the supply of origin-destination and related detail is limited, in relationship to the demand.

The main factors that limit data supply are the following:

1. Budget;
2. Response effort or reporting burden;
3. Avoidance of disclosure of individual, plant, or company activities; and
4. Ability of respondents to understand the questions asked and supply reliable answers.

Budgetary issues and the need for minimizing response burdens are so well known that we do not need to discuss them further at this time. Legal and other reasons for avoiding disclosure are also well known but worth some brief comments.

The need expressed for origin-destination data rarely is for "all-commodity" aggregates from point A to point B. Generally the need is for specific commodities between those two points and often for further breakdowns by size of shipment and means of transport. The main source of disclosure in this type of situation is the identification of a commodity at a specific origin because that information alone often discloses the producer. The most frequent solution is to combine commodities until a sufficient mixture of different shippers are involved to avoid disclosure or to cluster geographic points into broader areas, or to do both. Although this solution is less than ideal, it is often unavoidable even is the sample includes most (or all) of the plants in the area.

Another factor that limits data supply is the ability of respondents to interpret the questions asked and to supply reliable answers. This concept is so obvious that you may wonder why I mention it, but I assure you that it is an extremely serious limitation and often is not recognized by the survey statistician or data user. For example, several years ago in a travel survey, we asked people to answer questions regarding trips they had taken "since the beginning of last month"—a recall period of only about 6 weeks because the interviewing was done at about midmonth. The recall period was short, and we defined the term trip so clearly that there seemed to be no chance of confusion. The public cooperation was excellent, and the responses appeared to be good, complete, and reasonable. However, we later found that most people apparently did not know when "the first of last month" actually occurred. We discovered this by running two independent samples with partially overlapping time periods.

In a subsequent travel survey, we found that the clear definition of a trip—defined as being out of town overnight or going to a place 100 miles away—was logically precise and unambiguous but actually was not fully believable by respondents. For example, one person reported that he had not taken a trip during the last three reporting months. However, we subsequently interviewed him on a quality check and found that he had spent almost every weekend with his mother, who lived some 70 miles away. He said that he did not think we would consider those visits a trip, and he did not want to give us "bad" information. We found that the idea of merely being out of town overnight is not considered to be a trip in so many instances that we now define travel only in terms of a significant distance—100 miles away from home. Travel for distances nearer home must be measured in some other manner. Another example is the general absence of records that could be sampled to obtain data on such things as elapsed time in transit, transportation costs, and travel expenditures.

In brief, I have indicated two types of situations in which the necessary data cannot be obtained and published—one because of disclosure limitations, the other because reliable replies are not obtainable, because of either response errors or lack of records. In both instances, it may be feasible to construct estimates with acceptable accuracy (not necessarily close precision) by using data that can be collected and by applying factors to estimate data that cannot be collected or at least published. For example, a strong need has been expressed for origin-destination commodity flow data in terms of value of products. Bills of lading and most other shipping documents show weight but not value. The reporting burden on shippers, in my opinion, would be unreasonably high if they were asked to report the value of each shipment in a sample needed for measuring traffic flows. The most promising alternative is to obtain a set of factors, such as approximate value per pound, that can be applied to the tons of commodities shipped, provided the need by data users for value of shipment statistics is sufficiently great to offset the cost of obtaining and processing adequate factors.

A similar situation exists with respect to travel, in which there is a demonstrated need for travel expenditure data, especially with respect to increases or decreases in expenditures to be anticipated from future variations in the volume and characteristics of travel. The essential raw data are not a matter of record and, in my opinion, cannot be obtained in a mass survey with adequate precision. The most feasible alternative is the development of expenditure factors to be applied to data in the national travel survey, such as the number of trips by traveler income level and purpose of trip.

Some progress has been made in applying this type of solution, largely under the name of "model building" or "simulation" or "automation." This approach doubtless will continue to gain momentum, especially if survey statisticians and model builders cooperate more closely. The survey statistician should not only be the supplier of input data needed to obtain realistic answers but also the proposer of modifications to the models so that existing or obtainable data are more effectively used. Unfortunately, too many simulations are only exercises in logic and mathematical procedures, because they require inputs that cannot be measured in the real world. I also hasten to add that too many survey statisticians have not made a real effort to obtain the specific data needed for input to existing models. Progress in this type of solution necessarily will be slow, but it promises great dividends.