Commercial Data Sources for Transportation Planning

Robert C. Stuart, Division of Environmental and Urban Systems, Virginia Polytechnic Institute and State University, Blacksburg

There is no universal best source of data for transportation planning. In many instances, commercial data are less expensive, more current, and more quickly available than government data. Furthermore, commercial data sources are unaffected by government paperwork-reduction programs that are affecting many data-collection projects. Although there are still problems in using commercial data sources, they are of the same relative magnitude as those of public sources. Commercial data have been used with some success in compiling household, employment, and trip information. In particular, R. L. Polk and Dun and Bradstreet data have already seen considerable use in urban transportation planning. As commercial data sources become more competitive, transportation planners in small and medium-sized communities would be well advised to consider this information in deciding on the means of creating and maintaining data files.

Commercial sources of data have taken on new importance for urban transportation planning. Transportation planners are more frequently selecting these sources for particular situations because they may be the least expensive, the most current, or the most quickly available. Some of these data also have other advantages: They may present a time series providing trends and early indications of change or be comparable with other subareas and urban areas. Some of them are complete with computer graphics for instant analytical use and presentation. These data are often useful for functions other than transportation, thus spreading the cost. They are also less affected by the confidentiality limitations on the census and other publicly gathered data. In addition, commercial data sources are unaffected by governmental paperwork-reduction programs that under current Office of Management and Budget guidelines limit federal statistical data-collection programs.

Despite these advantages, commercial data are not the most appropriate for all transportation planning purposes. The purpose of this paper is to introduce the most widely used commercial data sources and compare them with other sources. This will afford some guidance to the transportation planner in selecting the best source for his or her situation.

The paper is organized around three principal files used in transportation planning—the household file, the employment file, and the trip file. For each file, data sources are identified and commercial sources are described in detail. Finally, the strengths and weaknesses of the data are compared, using criteria that often guide selection of a source.

HOUSEHOLD FILE

The household file provides the basis for generation of trip productions in traditional transportation models. Today, the file may also be expected to serve additional functions, which places new requirements on the character of the data. For example, it may be used to monitor existing neighborhoods for indications of changes in travel behavior (e.g., an increase in female heads of household), to monitor new developments that may require improvements in transportation services (e.g., apartment projects, subdivisions, planned unit developments, and such that may be incorporated into a short-range two-to-five year forecast), or to identify market segments for which customized transit or paratransit services can be designed.

Data Sources

Five alternative data sources could be considered for creation of a household file:

1. The decennial and mid-decade U.S. census,
2. A primary survey,
3. Local permits and approvals,
4. International Data and Development, Inc. (IDD), and
5. R. L. Polk and Company.

International Data and Development, Inc.

IDD produces a tract-level or zonal level set of current estimates of home-related data that may meet many of the data needs of transportation planners. For each tract or zone, it includes the following items:

1. Households by structure type—single family and multifamily;
2. Households by income—five income categories;
3. Households by automobile ownership—zero, one, two, three and more; and
4. Total population.

The estimates are derived by using data from the files of the Reuben H. Donnelley Corporation. Two different data sets for the years 1972, 1974, and 1976 contain tract-level aggregates of geocoded automobile registrations and telephone listings for each standard metropolitan statistical area (SMSA) in the United States. There are some states in which the automobile registration data have not been made available; in these states, the automobile-ownership item is not available. The geocoding operation is the key to the availability of annual tract-level data. It uses proprietary, computerized coding guides in each SMSA that have been developed and are maintained for that purpose. The dual independent map encoding and geographic basic file programs (DIME/GBF) of the U.S. Census Bureau are not used.

Neither of these data bases—automobile households and telephone households—are true universes of all households. IDD's current estimates, however, are based on the theory that the growth rates in these two very large samples of total households can be used to estimate the growth rate in the universe of total households. The growth rate over time for a tract is established by using the growth rates for automobile households and telephone households for the years 1972, 1974, and 1976. This rate is then backcast to 1970 and applied to the 1970 census figures for the tract to produce 1976 estimates. The next edition of the file will use 1978 data.

Households by structure type and by automobile ownership are inferred from the file of automobile registrations and telephone listings. The method uses proprietary algorithms that compare last name and address of adjacent listings in a presorted file. The distribution of households by income is based on a regres-

sion model that uses automobile-ownership characteristics and 1970 demographics to produce an initial income estimate for a household. These estimates are aggregated into a tract-level distribution. IDD cautions that the estimates may vary from reality in tracts that have exclusively high or very low incomes.

IDD computes an uncontrolled population estimate for a tract by multiplying the estimated total households in it by its 1970 occupancy factor (persons per dwelling unit), adjusted to account for changes in the factor since 1970. The adjustment is based on the statewide change in the occupancy factor obtained from the 1976 survey of income and education.

The population and income distribution by tracts is then adjusted to agree with the latest census bureau revenue-sharing figures. The size of this adjustment, reported by IDD to be usually less than 3 percent, is dependent primarily on the portion of the county that has rural route or U.S. Post Office Box addresses that cannot be geocoded.

The availability of IDD data for transportation planning is very new (1978) and has not yet been subject to the kind of critical evaluation that transportation planners have accorded the more widely used Polk and Dun and Bradstreet (D&B) data sets.

R. L. Polk and Company

R. L. Polk and Company prepares city directories in more than 2000 U.S. cities and towns ranging from very large cities such as Birmingham, Alabama; Atlanta, Georgia; San Francisco, California; Houston, Texas; and Pittsburgh, Pennsylvania, to towns of fewer than 3000 persons. Polk also compiles annually on a nationwide basis Reports of Cars and Trucks on the Road and New Motor Vehicle Registrations.

To produce its city directories, Polk conducts annual door-to-door interviews of households and nonresidential activities. These use a field listing method similar to that often employed in land-use surveys conducted by transportation planners. One important difference is that the interviewer is equipped with a computer-prepared printout of the previous year's listing and, essentially, records only changes. This not only improves the efficiency and accuracy of the survey but also provides indicators of changes during the previous year.

As a by-product of the directories, the computerized survey results can be, and for many areas have been, summarized at the census-tract, traffic-zone, or other desired subarea level. The result is not only the usual cross-section-in-time data set produced by land-use or housing surveys, but also a set of socioeconomic indicators of change that can be used for monitoring. Polk calls its commercial data package Profiles of Change. The package includes, for census tracts or other subareas, the statistical summaries, computer-made maps, and magnetic tapes. Profile reports for 1974 and 1975 were distributed to 318 U.S. cities under a grant from the U.S. Department of Housing and Urban Development.

The data items reported in Profiles include the tract population and the number of households by

1. Sex of head,
2. Number of children,
3. Owner versus renter,
4. Occupation of head,
5. Retired head,
6. Jobless head,
7. Income index,
8. Number of adults,
9. Number of children less than 18 years old,
10. Vacancies,
11. Two-canvass vacancies,
12. New construction,
13. Number of units in structure,
14. Mobility rate—housing units that have had a change of occupants, and
15. Number of demolitions.

Because the name of the employer is asked, a journey-to-work trip table can be prepared. This, as well as the data gathered on business establishments, is discussed below.

The data on automobiles and trucks, based on registrations, are available by tract for all tracted SMSAs. On new vehicle registrations, an indicator of change is available by tract for about 100 SMSAs. Data are supplied either on tape or as printout.

Evaluation

It is possible to point out some relative advantages and disadvantages of each of the data for use in a household file.

1. Availability: The Polk files are not available for some metropolitan areas. In others, they may be available for some jurisdictions but not for others and therefore are not suitable for areawide transportation planning. Other sources are a possibility in all areas.
2. Coverage of urban fringe: The developing fringe of the metropolitan area is a place where travel volumes and patterns can change rapidly and drastically. The census covers the fringe well but is quickly outdated. The same would be true of a primary field survey. Tabulations of building permits and subdivision and rezoning approvals can be an important source of data in this area. This source has the further advantage of lending itself to use in a short-range-development forecasting process of considerable accuracy and value for anticipating travel demand. IDD covers the fringe by using current estimates and is thus intermediate in currency between the census and local sources. These estimates may, in a few instances, be seriously inaccurate because of geocoding problems. Polk data generally do not extend much beyond the built-up urban area. In some regions, a hybrid system that combines the Polk file for the built-up area and local tabulations and selective field checks of new development would appear to provide a multipurpose file of great cost-effectiveness.
3. Processing time: The time between the survey and the availability of data is relatively long for the census and for complete primary surveys and relatively short for commercial sources.
4. Currency: The census is now conducted every five years. Add the processing time to this, and its currency can be rated as fair or variable. Comprehensive primary surveys for transportation data have generally been infrequent, and the trend seems to be toward even fewer. The other three sources can provide annual reports.
5. Cost: Generally, the costs of primary surveys are highest, those of the census are lowest, and those of commercial sources are toward the low side.
6. Multiple use: When the cost of a data set can be divided among several users, the cost per use can be substantially reduced. The census is an outstanding example of this. The Polk file may have the greatest potential in this feature and is already being used a variety of ways.

EMPLOYMENT FILE

Employment files, which are required for generating

trip attractions for transportation models, are increasing in use for other purposes such as economic development planning, econometric modeling, transportation market segmentation, and civil defense planning.

Data Sources

There is no universal best source of data for an employment file. All sources require considerable processing, considerable supplementing, or both. The sources considered here include

1. State employment-security files—the most widely used base,
2. A primary field, telephone, or mail survey, and
3. Commercial data—Dun and Bradstreet and R. L. Polk.

In addition, a wide range of other public and commercial sources are used, usually to supplement one of the above. These include city and manufacturing directories, commuter computer files, and census place-of-work data.

Dun and Bradstreet

D&B is the most widely used commercial source of data for creating employment files for transportation planning. These data are by-products of the credit-rating and the business information compiled on approximately 3.8 million establishments in the United States. Information of use in transportation-planning employment files, for each location of a company, includes the following items:

1. Legal business name;
2. Secondary business name (i.e., trade styles, divisional name, and such) where appropriate;
3. Street address;
4. Mailing address;
5. Standard industrial classification (SIC) code (primary and up to five secondary codes);
6. Line of business (narrative of primary function at the establishment location);
7. Annual sales volume (if available);
8. Number of employees at the location;
9. Total number of employees;
10. Identification of the facility as a branch, subsidiary, or headquarters unit;
11. Identification of the facility as manufacturing or nonmanufacturing;
12. Area code and telephone number;
13. Name and title of chief executive;
14. Geographic codes for state, county, city or town, zip, and SMSA;
15. Year that the business was started;
16. D-U-N-S identification number of the establishment;
17. D-U-N-S identification number of the headquarters (if the establishment is a branch);
18. D-U-N-S identification number of the parent (if the establishment is a subsidiary); and
19. D-U-N-S identification number of the ultimate firm (the topmost firm in a multiestablishment company).

The data are updated regularly, more frequently for the larger firms. D&B can code the information to census tracts inexpensively by using the proprietary Donnelley Corporation address-coding guide (the Donnelley Corporation is a subsidiary of D&B).

R. L. Polk

The Polk city directories list businesses as well as households. Data on each commercial establishment, which can be aggregated to the census tract, include the following items:

1. Address,
2. Five-category SIC classification,
3. Turnover,
4. Change of occupants since previous canvasses,
5. Number of vacancies,
6. Number of two-canvass vacancies,
7. Number of demolitions, and
8. New construction.

Evaluation

Compared with other sources, the most serious weakness of the D&B data is underreporting, especially of small firms and particularly in the services classifications. This can be resolved by factoring up tract totals by SIC category to aggregate to known county totals. Public and semipublic (e.g., church) employment is completely omitted and requires a supplementary survey. The advantages of the D&B data compared with the employment-security files are in its (a) greater provision of employment by location in multiple-location businesses, (b) precoding by census tract, and (c) provision of exact employment rather than by a system of categories.

The principal, and very serious, weakness of Polk data for creating an employment file is the lack of an employment figure. Some data on this can be derived from tallying the persons reporting each employer. However, it is impossible to categorize separate locations of the same employer. For various reasons, underenumeration of employment occurs. The advantage of the Polk data is that they produce the most complete list of establishments by address and showing separate locations.

TRIP FILE

The Polk data are unique in their ability to provide transportation planners with current journey-to-work data. Because approximately half of all transit passengers are making work trips, Polk provides a useful, unique data source for bus-route planning and the identification of market segments for which customized service can be planned. The other two sources of trip tables—the census place-of-work, journey-to-work questions and the origin-and-destination travel survey—cannot provide current data for two reasons: They are conducted too infrequently, and they require, typically, a year or two or more to process the data for use.

CONCLUSIONS

The R. L. Polk and the D&B data have already seen considerable use in urban transportation planning, enough that some evaluation has been made by users. The verdict is mixed at this time; each source has both enthusiastic supporters and skeptics.

Nevertheless, some safe conclusions can be drawn. One is that there is no universal best source of data for all situations. Another is that, for certain situations, a commercial data source will be the best choice. Third, notable progress is being made in overcoming some of the evident problems and shortcomings. Fourth, although there are serious, difficult problems involved in using commercial data sources, they are of the same

magnitude as the problems of using the alternative public sources of data. Finally, commercial data sources promise to be increasingly competitive with other sources. As such, they justify the intensive, innovative efforts currently being made by the providers, the users, and the funding agencies to further improve them. Thus, transportation planners in small and medium-sized communities should investigate commercial data sources in deciding on the means of creating and maintaining their files.

Iowa Urban Transportation Area: Traffic-Counting Program

Robert D. Studer, Iowa Department of Transportation, Ames

An overview of the traffic-counting program for urban areas in Iowa is presented. The Iowa Department of Transportation has an extensive program that covers 15 urban areas in the state that have populations of 25 000 or more. The types of traffic data collected, the problems and difficulties encountered in collection and processing, and the varied uses that are made of the data are reviewed. Preliminary results of the Sioux City counting program are included.

The Iowa Department of Transportation (IDOT) conducts traffic-counting programs in the 15 Iowa urban areas that have a population of 25 000 or more. The content of the counting program varies with the needs of the particular urban area and those of IDOT. The urban-area counting program is part of a statewide traffic-counting program that includes traffic surveillance at specified time and location intervals on approximately 179 200 km (112 000 miles) of Iowa streets and highways.

The 15 urban areas have a total of approximately 8874 km (5546 miles) of streets and highways, of which 1013 km (633 miles) are under state jurisdiction. Because IDOT is responsible for the highways under state jurisdiction, more emphasis has been placed on maintaining a current record of traffic data on these highways. These data and additional supportive inventory data on items such as roadway conditions and serviceability index are used to determine the need for and magnitude of improvements for highways.

The development of the federal-aid realignment study in 1976 has placed more emphasis on maintaining traffic data on the federal-aid road system in urban areas. The IDOT traffic-counting program now includes the monitoring of traffic data on the federal-aid urban system. In the 15 urban areas, there are 2451 km (1532 miles) of federal-aid urban highways.

IDOT also supports a cooperative counting program between urban areas and the state. Under the cooperative program, the urban area and IDOT coordinate the locations to be counted and the methodology and type of count to be taken. The costs involved in collecting the data are shared by the urban area and IDOT, and IDOT assumes the responsibility of processing the data. The urban area and the state also cooperate on the cost of installation and maintenance of permanent traffic recorders.

Because many of the urban areas in Iowa involve more than one jurisdiction, problems often arise as to the sharing of responsibility among the different jurisdictions. Typical of this is the Des Moines urbanized area, which includes eight different cities. However, continued efforts by IDOT on the exchange of traffic data between the state and the jurisdictions have kept these problems to a minimum.

TYPES OF TRAFFIC COUNTS

Six types of traffic counts are made:

1. Permanent-traffic-recorder counts,
2. Cordon and screen-line counts,
3. Turning-movement counts,
4. Portable-recorder counts,
5. Vehicle-classification counts, and
6. 24-h control-station counts.

Permanent Traffic Recorders

IDOT maintains 85 permanent traffic recorders at 64 locations on the highway system and 21 at 13 locations in urban areas. These permanent recorders provide continuous traffic data on highways that have different traffic characteristics, functional classifications, areas served, and types of system. Hourly, daily, weekly, monthly, and annual traffic volumes are obtained for each recorder and stored on magnetic tape.

The data obtained are compiled from each recorder and a 3-year average is computed. Factors are developed from these 3-year averages that expand the short-term counts to an average annual daily traffic volume.

For example, for a 7-11 a.m. and 2-6 p.m. manual count taken on a July weekday, these factors ranged from 1.665 for a location on I-235 in Des Moines to 1.960 for a local city street in a residential area of Ames.

Data from these recorders are also used in computing the 30th-highest-hour volume at urban-area recorders. These data are expressed as percentages of the average annual daily traffic and used in highway design.

Cordon and Screen-Line Counts

These counts are conducted on a 2-year cycle in urban areas. The cordon-count locations are established at all highways entering the urban-area boundary, and the screen-line-count locations are established along the boundary of a natural barrier such as a river or a railroad within the city. The counts are generally mechanical recorder counts taken at the specified location for a period of 24-48 h.

In urban areas such as Iowa City or Ames where a state university is a major traffic generator, the counts are conducted during the school year.