Future of Urban Transportation Data

EDWARD LIMOGES, *Southeastern Michigan Council of Governments*
CHARLES L. PURVIS, *Metropolitan Transportation Commission*
SHAWN TURNER, *Texas A&M University*
MARC WIGAN, *Oxford Systematics*
JEAN WOLF, *Georgia Institute of Technology*

The exciting opportunities technology is providing researchers in the field of urban transportation data come with equal challenges. These challenges include data privacy and confidentiality, fluctuating priorities on data needs, and problems for data collection budgets given the imminent need for modernizing the transportation infrastructure. Adding to these issues is the need to train a new generation of analysts to effectively analyze varying quantity and quality of data. Certainly new technologies promise to reduce the cost while improving the accuracy of collected data, but this benefit is countered to some degree by people's increasing reluctance to be monitored.

Members of the Transportation Research Board Committee on Urban Transportation Data and Information Systems lead the research community in many of the following areas:

- Using census data in transportation planning;
- Monitoring traffic and transit systems;
- Using ITS-generated data in planning;
- Using household survey data in describing traveler characteristics and metropolitan travel patterns;
- Combining data sources;
- Studying issues of data privacy and confidentiality; and
- Creating standards for meta-data (data about data) and meta-analysis (comparative data analyses).

This paper provides a perspective on these issues and a look at the opportunities and challenges that will face urban transportation planners in the 21st century. The several themes explored in this paper include

- The issues related to continuous data collection;
- The use of global positioning systems and other technologies in urban transportation data;
- The use of small area land use and socioeconomic data in metropolitan planning;
- Privacy and data ownership concerns; and
- The concept of "information from data."

CONTINUOUS DATA COLLECTION

Urban transportation planners in the United States are familiar with the "3-C" planning process— continuous, comprehensive, and cooperative. The "continuous" aspect of metropolitan planning, as envisioned by the 1962 Federal-Aid Highway Act, referred to the need to periodically update long-range transportation plans. Traditionally, traffic and transit system monitoring programs were the only continuous data collection programs in place within state and metropolitan areas. Other significant data collection programs, such as the U.S. decennial census or metropolitan area household surveys, were based on once per decade or an even more infrequent basis.

Three major trends in continuous data collection programs are of interest:

Continuous Census Data

The American Community Survey (ACS), a major initiative of the U.S. Census Bureau, is intended to eventually replace the traditional census "long form." ACS is a major change in the way transportation planners will collect census data in the 21st century. Planners will be able to get annual census data for large areas of 100,000 or more in population, and "rolling average" data for small areas such as census tracts and travel analysis zones. Understanding the statistical properties of rolling average data for small areas, and the creative and full use of such properties in transportation planning activities is an emerging research need.

Continuous Personal Travel Data

The typical large metropolitan area has experienced large-scale household travel, or home interview, surveys approximately once per decade over the past 40 years. The U.S. experience is that of shifting away from traditional in-home face-to-face interviews to more convenient mail-out or telephone surveys. These surveys were typically conducted during weekdays throughout the spring and fall of the survey years. An emerging international trend is the "continuous" household travel survey; data is collected every day of the year, every year. Examples include the Victorian Area Transport Survey in Melbourne, Australia, and a similar survey in Sydney, Australia. Other metropolitan areas are beginning to realize that "people travel throughout the year, not only during specific seasons" and are more willing to commit data collection budgets that treat travel as a year-round affair. Whether U.S. metropolitan areas can commit to a multi-year, continuous household survey collection effort, such as seen in Melbourne and Sydney, has yet to be determined.

Archiving ITS Data for Planning and other Applications

Data collected and archived by intelligent transportation systems (ITS) offer great potential in expanding the scope and types of analyses conducted in transportation planning (*1*). Archived ITS data is already being used to a great extent for transportation planning, operations, and research. These activities include congestion monitoring (*2*), origin-destination studies (*3*), planning and traffic model development and validation (*4*), and traffic flow research (*5*), to name only a few. Given the potential of ITS, there are numerous opportunities as well as challenges in this rapidly growing area.

The continuous 24 hours a day, 7 days a week fine-grained detail typical of ITS data will eventually supplement or replace traditional traffic data sampling programs in many urban areas, resulting not only in cost savings but in a dramatic increase in the amount of data available for analysis. Additionally, the large amount and type of data available through

ITS will improve existing analyses and enable new types of transportation analyses, such as those focusing on travel reliability or actual discrete choices made by (anonymous) individual travelers or shippers. Numerous data user groups will be brought together because of their interest in obtaining archived ITS data. Therefore, archived ITS data has the potential to serve as a platform for coordinating diverse transportation agency efforts and to provide a feedback loop for various transportation disciplines.

Just as there are significant opportunities for using archived ITS data, so are there significant challenges. These include

- Formulating information use and sharing policies that encourage public and private sector partnerships that return value to the sponsoring agency and its customers, yet protecting individual's privacy rights while using ITS;
- Addressing data access and management issues inherent in sharing and analyzing extremely large databases in a distributed data user environment;
- Developing analytical tools and procedures that can be used to explore and analyze trends and relationships in large, complex transportation databases; and
- Developing and incorporating automated quality control procedures into data archiving practices to ensure that end users have data of sufficient quality and accuracy.

GPS AND OTHER TECHNOLOGIES TO ENHANCE URBAN TRANSPORTATION PLANNING

The automation of urban data collection offers some significant improvements in data quality by eliminating transcription errors and by introducing on-line error checking and improved form flow. At the same time, these new technologies eliminate the expense of transcribing written documents into electronic format. Automation further provides for the capture of once burdensome details about urban travel. For example, Global Positioning System (GPS) technology automatically records trip origin and destination, start and finish times, length and duration, and all travel routes. Finally, automation provides a means to capture urban travel data previously unattainable in a survey setting. On-board engine monitors and other vehicle sensors, when installed in personal vehicles, can provide much driver behavior-related data, including constant information for vehicle speed, engine speed, throttle position, and engine load. By gaining knowledge regarding these driver-vehicle-infrastructure interactions, transportation industry modelers and practitioners can better understand the contributions of driver behavior on traffic conditions and emissions.

Within the past several decades, the traditional paper-and-pencil interview for travel diary data collection has been supplemented or replaced with computer-assisted-telephone interviews. Most recently, computer-assisted-self-interview methods are being evaluated. There are numerous references available that review the traditional diary instruments and the transition to automated data collection (*6–8*). Two recent European research projects have examined handheld PCs (*9*) and Internet-based websites (*10*) for long distance travel data collection. As the cost of handheld computers with large storage capabilities continues to fall, computer-assisted travel diaries should become increasingly commonplace.

GPSs add a new spatial dimension to trip-making data by tracking actual route choice. Research projects investigating automated diaries with GPS include the Federal Highway Administration’s Lexington study, and travel surveys in the Netherlands (*11*). These projects are among the first to combine electronic travel diaries with GPS receivers to gain exact temporal and spatial details of each trip. Several other projects have installed passive GPS receivers in automobiles and trucks (*12*) to capture travel route information. Various recent research projects have evaluated GPS receiver performance in London (*13*), Paris (*14*), and Atlanta (*15*). The cost of GPS technology has also continued to decline, making it quite feasible for a GPS component to be included in travel surveys. In fact, a number of major travel surveys scheduled for the year 2000 will contain GPS elements.

Within the last several decades, the need for accurate vehicle fuel consumption and pollutant emission models has forced both researchers and practitioners into developing automated means for capturing second-by-second vehicle and engine activity data. These projects include research conducted in France in 1983 and 1990 with 55 instrumented vehicles (*16*); a Canadian project using that country's "Autologger," a vehicle-mounted data collection device (*17*), and several instrumented vehicle studies conducted by the U. S. Environmental Protection Agency (*18*, *19*). A vehicle instrumentation package is scheduled for implementation as a sub-sample to the year 2000 Atlanta Metropolitan Regional Travel Survey (*20*). Not surprisingly, such technologically advanced data collection brings challenges. Common concerns include instrumentation costs, ease of installation, and data storage requirements.

SMALL AREA LAND USE AND SOCIOECONOMIC DATA

Transportation planning requires data that is organized and tabulated by small geographic areas, usually called traffic analysis zones. Patterns of land use and transportation systems are the two key determinants of small areas. Boundaries are drawn to separate differing land uses, to identify major traffic generators, and to fit within and relate to the transportation network. Other considerations include natural features and governmental boundaries. The boundaries and patterns of these small geographic areas form a somewhat abstract mosaic that outlines the spatial structure of large regions.

Space and Activity

The physical aspect of land use and transportation is facilities and spaces, and the functional aspect is activities and occupants. The common understanding of the interrelationships between form and function is more advanced for transportation, with its roads and other publicly owned facilities, than for land use. These interrelationships represent an integration of systems and flows that include measures of capacity, levels of service, and functional classifications. However, although land is largely privately owned, it is still publicly monitored and measured for space activities. Three areas of government oversee this monitoring—the U.S. Census Bureau; state employment insurance agencies; and municipal departments that oversee building construction and property assessments. These agencies create and maintain data concerning residential and nonresidential buildings and the occupants of those structures.

There are problems with all three sources. Several key questions in the decennial census present difficulties for questionnaire respondents because concepts are unfamiliar or facts unknown. These include two key housing items: number of units in a structure (attached or multi-unit buildings can generate confusion) and year structure built (many respondents, especially renters, do not know). Deficiencies in workplace location result from workers who do not know their workplace address, but only how to get there. Aside from address problems and missing establishments, in some states insurance files of employers cannot be released to other public agencies. Also, construction permit data may not be tabulated in "geocodable" form, and assessment records may not be computerized or may lack street addresses.

Recent Developments

Several recent developments represent substantial advances for socioeconomic and land use data. These developments are

- The replacement of the Standard Industrial Classification (SIC) system by the North American Industry Classification System (NAICS)(*21*). NAICS will better reflect the current structure of the national economy and incorporate some land use differentiation. However, NAICS in Census 2000 presents multiple challenges. One issue is how to assure correct assignment of the new classes; another is how to effectively bridge data to link 1990 SIC employment and 2000 NAICS employment at the same level. Finally, a third issue is how to create a full land use breakdown of employment using post-processing (*22*).
- The initiation of ACS. During the coming decade ACS will provide continuous census sample data, eventually including small areas. It will also form the basis for a nationwide Master Address File (MAF) of residential addresses, and facilitate completion of the multi-dimensional Land-Based Classification System (LBCS) model for classifying land use based on property characteristics. LBCS will replace a system developed in the mid-1960s (*23*). Also, the Census 2000 assignment of workplaces to small areas will be improved through the use of better coding resources, including an updated TIGER file, a more complete business establishment list, and an improved geographic allocation procedure.

Future Possibilities

On the horizon is the promise of full integration of census data (both 2000 and ACS), state employer data, and local construction and cadastral parcel assessment data. The central concept will be to use data drawn from the cadastral file, which will be continuously updated to reflect construction and demolition and provide a complete count of buildings and characteristics. Choosing several large geographic areas that already have the required data files could test these possibilities. Operationally, each address in the study area's MAF would be given housing characteristics drawn from the parcel file. The Census Bureau would match these addresses to Census 2000 and ACS addresses to create small area summaries of census population and household data by housing class. The study area's housing records would be populated by the census summary data through random assignment of household characteristics. An equivalent process, using relevant files, would be used to populate nonresidential buildings and floor space. The ultimate goal of such data integration is to maximize the usefulness of public information through multiple public services.

PRIVACY AND DATA OWNERSHIP CONCERNS

Society is rapidly becoming highly dependent on extensive data exchange, and substantial actions have been taken by the United States and other major countries to privatize and secure many of these data streams. These efforts are creating pressure to withhold information in hope of economic incentives, and pitting personal privacy against the desire to be included in the data flow. Eventually, data will come to be owned solely by the individual it involves (*24*).

A major challenge for the transport and data communities is to participate and guide this process of information management and ownership to the best community outcome. Having people freely providing their own data and opinions must be balanced with the public accessibility to aggregated results that are in the community's interest. The rapid rise in GIS data has the potential to upset this delicate balance, and ways of involving the community in the ownership and use of its data must be found.

Privacy is a separate issue, but it is becoming closely linked through the identification, surveillance, and charging systems now moving towards integration through GIS databases, GPS use, smartcards, and the growth of ITS data interchange standards. Only if the two major issues of data ownership and privacy can be navigated successfully will transport and traffic data become more targeted, appropriate, reliable, and efficient.

INFORMATION FROM DATA

Even before the wave of massive future information flows became apparent, the value of digested data had become obvious through books focusing on the implications of transport data (*25*). Twenty years later the flood of inaccessible and disorganized transport data has only just been met and matched by the Bureau of Transportation Statistics in the United States, and has flowed on to the user community with some isolated examples of similar forms of interpretation. The mindset that links metropolitan planning organization data flows to transport planning is shifting and will continue to do so.

The big growth areas will be in traffic, land use, and economic activity data as transport data balance more directly with community needs. Methods have yet to be found for digesting such data and then communicating it to the general and professional communities in a way that will allow for informed response and understanding. This objective has barely been met in transport, although GIS should permit such communication within several years.

The new types of data will place great pressure on the transport and planning communities to widely increase research efforts through the use of interactive tools. As this occurs, transportation and planning leaders will come to respect the opinions and expertise available in the general community.

CONCLUSIONS

In *Data Smog: Surviving the Information Glut*, author David Shenk offers his "first law of data smog"—"Information, once rare and cherished like caviar, is now plentiful and taken for granted like potatoes" (*26*). Urban transportation planners and data analysts in the 21st century will be faced with data that is more continuous in nature as well as more conflicting. Reconciliation of conflicting data sets will be more important in the decades to come. Equally important will be the need to effectively reduce and synthesize data from various sources, including continuous census data, continuous data from ITS efforts, and other emerging technologies that will collect data on traveler behavior and transportation systems.

REFERENCES

1. Margiotta, R. *ITS as a Data Resource: Preliminary Requirements for a User Service*. Report FHWA-PL-98-031. U.S. Department of Transportation, 1998.

2. Ishimaru, J.M., and M.E. Hallenbeck. Central Puget Sound Freeway Network Usage and Performance. Draft Research Report. Washington State Department of Transportation, Washington State Transportation Center, Seattle, 1999.
3. Dixon, M., and L.R. Rilett. *Evaluation of ITS Automatic Vehicle Identification for Origin-Destination Estimation*. TRB Preprint 991459. Presented at the 78th Annual Meeting of the Transportation Research Board, Washington, D.C., January 1999.
4. Bunch, J., K. Wunderlich, and G. Nelson. Development of Representative Day Scenarios to Capture the Impacts of Variation in Conditions for Transportation and ITS Planning. Presented at the National American Travel Monitoring Exhibition and Conference. Charlotte, North Carolina, May 1998.
5. Zhou, M., and F.L. Hall. Investigation of the Speed-Flow Relationship Under Congested Conditions on a Freeway. Presented at the 78th Annual Meeting of the Transportation Research Board. Washington, DC, 1999.
6. Stecher, C., S. Bricka, and L. Goldenberg. Travel Behavior Survey Data Collection Instruments. In *Conference Proceedings 10: Conference on Household Travel Surveys: New Concepts and Research Needs*, TRB, National Research Council, Washington, D.C., 1996, pp. 154-169.
7. Kalfs, N., and W.E. Saris. New Data Collection Methods in Travel Surveys. In *Activity-based Approaches to Travel Analysis*, (D.F. Ettema and H.J.P. Timmermans, eds.). Pergamon Press, 1997.
8. Sarasua, W., and M. Meyer. New Technologies for Household Travel Surveys. In *Conference Proceedings 10: Conference on Household Travel Surveys: New Concepts and Research Needs*, TRB, National Research Council, Washington, D.C., 1996, pp. 170-182.
9. Haubold, I., and K. Axhausen. Handheld PC-Based Travel Data Collection by Self-Administered Interviewing. Presented at the Association of Survey Computing Conference on New Methods for Survey Research, Southampton, England, August 1998.
10. Plaxton, J., P. Jackson, J. Polak, and D.Wofinden. Using the Internet to Collect Data on Travel Behaviour. Presented at the 78th Annual Meeting of the Transportation Research Board, Washington, D.C., January 1999.
11. Draijer, G., N. Kalfs, and J. Perdok. *Possible Application of Global Positioning System (GPS) for Collecting Travel Data.* Project Report. Transport Research Centre of the Ministry of Transport, Public Works, and Water Management, The Netherlands, 1998.
12. Wagner, D., J. Seymour, N. Malcosky, J. Kinateder, and J. Nguyen. Determining Heavy Duty Truck Activity Using Global Positioning System (GPS) Technology. Presented at the 8th CRC On-Road Vehicle Emissions Workshop, San Diego, April 1998.
13. Ochieng, W. *ATT Positioning Requirements: GPS Performance in the City of London*. Project Report. Centre for Transport Studies, Imperial College, London, December 1998.
14. Flavigny, P-O., J-P. Hubert, and J-L. Madre. Analysis of Road Traffic Data Observed by GPS and Comparison with Other Sources. Presented at ADEME-INRETS Conference, Number 9815, September 1998.

15. Wolf, J., S. Hallmark, M. Oliveira, R. Guensler, and W. Sarasua. Accuracy Issues with Route Choice Data Collection Using GPS. Presented at the 78th Annual Meeting of the Transportation Research Board, Washington, D.C., January 1999.
16. André, M. Vehicle Uses and Operating Conditions: On-Board Measurements. In *Understanding Travel Behaviour in an Era of Change*, Pergamon-Elsevier, Oxford, 1995, pp. 469-481.
17. Taylor, G. AUTOLOGGER: A Long Duration Vehicle Use Data Collection System. In *Conference Proceedings for the 6th International Conference on Travel Behaviour*, May 1991, Vol 2, pp. 432-443.
18. DeFries, T., and S. Kishan. Light-Duty Vehicle Driving Behavior: Private Vehicle Instrumentation. Final Report. Radian Corporation. Presented to the U.S. Environmental Protection Agency, Office of Mobile Sources, August 1992.
19. Ross, C., M. Meyer, S. Barker, and Y. Zemere. Analysis of Travel Behavior Using Three-Parameter Dataloggers. In *Journal of Transportation Engineering*, Vol 121, No. 4, 1994, pp. 338-344.
20. Wolf, J., R. Guensler, S. Washington, and L. Frank. The Use of Electronic Travel Diaries and Vehicle Instrumentation Packages in the Year 2000 Atlanta Regional Household Travel Survey. Presented at the “Personal Travel: The Long and Short of It” Conference, Washington D.C., June 1999.
21. North American Industry Classification System: United States, 1997. Executive Office of the President, Office of Management and Budget, Washington, D.C., 1998.
22. Limoges, E. Employment Data for Transportation Planning. Objectives for Census 2000. In FHWA-CTPP 2000 Status Report. U.S. Department of Transportation, 1998, pp. 3-5.
23. Land-Based Classification Standards, Second Draft, April 18, 1999. American Planning Association, Research Department, Chicago, 1999.
24. Wigan, M.R. Data Ownership. In *Managing Information Technologies, Organisational Impact II*. (R. Clarke and J. Cameron, eds.). North Holland, Amsterdam, pp. 157-169, 1992.
25. Wigan, M.R. *Special Report 38: Australian Personal Travel Characteristics*. Australian Road Research Board, Vermont, Victoria, Australia, 1987.
26. Shenk, D. *Data Smog: Surviving the Information Glut*. HarperCollins, New York, 1997.