Automated Conversion of Milepoint Data to Intersection/Link Network Structure: An Application of GIS in Transportation

WENDE A. O'NEILL AND BALAKRISHNA AKUNDI

Restructuring or converting network data is an essential function of the geographic information system (GIS) when adapted to transportation. Implementing effective data restructuring models in the GIS allows users to collect and maintain data in the format with which they are most familiar while allowing others to use it in a format they require. Milepoint referenced data in road inventory files provide valuable network information for transportation research. Individual records in these files represent variable-length sections of roads. A new record is created each time a highway attribute changes. Consequently, a segment of road between two intersections may be represented by one or more records in a road inventory file. Further, all attributes in these records are associated with both directions of travel along a section of road. Many transportation analysis models require networks to be represented by a node-link data structure in which nodes symbolize an intersection of two or more roads. Additionally, if a road is two directional, it is represented by two links, each of which has its own set of attributes. To utilize road inventory data in these analysis models, network information has to be converted from milepoint to an intersection/link format. This process involves aggregating and disaggregating attribute data to represent longer and shorter road segments, and also disaggregating data into bidirectional information. Data conversion efforts are needed to produce intersection/link network representations from milepoint data. A microcomputer model for data conversion is developed and application issues and model sensitivities are addressed.

Many transportation agencies currently are investigating adaptation of the geographic information system (GIS) to transportation. One issue for consideration relates to data structures, specifically representation of transportation networks in digital data bases. Data base design solutions that require distinct groups within planning agencies to conform to one specific structure are unrealistic and unsatisfactory. A model developed here demonstrates that different network structures may be maintained by individual groups. Further, various data bases may be shared among groups, thus increasing information utilization and lessening redundancy.

Road inventory files represent an extensive data base containing highway attribute information. Records in these files represent sections of roadways such that a new record is created each time an attribute changes. The data structure in road inventory files is represented by variable length segments representing several network attributes simultaneously within a segment. Consequently, there is a great deal of data redundancy in these files. The spatial component in this data base limits queries to identification of the district, county, city, or town to which a highway belongs. Milepoint information is provided for each road segment terminus that does not represent a highway intersection. Additional milepoint information in each record gives the distance, in miles, from a set reference point to the beginning of the highway segment represented by the data record.

A dynamically segmented data base structure may be addressed using the model. In this structure, each network attribute is associated with variable-length segments defined by occurrences of changes in the attribute. When several attributes are overlayed, the intersection of segments represents the milepoint data structure found in the road inventory files.

Transportation planning models typically require a network data structure in which links represent one-directional sections of roads. Nodes distinguish the beginning and end of each link and represent an intersection in which change in travel direction may occur. Attributes, such as those found in road inventory files, are associated with each link. Depending on the nature of the application, an intersection/link network may encompass every road in a region or simply a subset of roads.

To make use of road inventory or milepoint attribute information in studies requiring an intersection/link data structure, aggregation and disaggregation of network data must take place. The manual process for performing this task is labor intensive and error prone. An automated procedure has been developed to simplify this data structure conversion process. As a result, road inventory files may be utilized in a broad range of applications.

This paper describes theoretical and practical issues related to conversion from one network data structure to another. Aggregation and disaggregation impact the accuracy of information. Simple rules are defined in the model to achieve an intersection/link data base.

NETWORK DATA AGGREGATION, DISAGGREGATION ISSUES

Aggregation and disaggregation of spatial data have been addressed primarily by geographers for polygonal data structures. In transportation, sketch planning and network abstraction issues are relevant to aggregation and disaggregation of lineal data. Literature on GIS applications in transportation discusses data base design, methodologies for attaching attribute data to lines, and the need for flexibility in transportation

W. A. O'Neill, 414 Cedar Orchard Dr., Blacksburg, Va. 24060. B. Akundi, Department of Civil Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Va. 24061.

data bases to meet a variety of objectives (*1–4*). Recently, Kuykendall (*5*) has addressed the capabilities of ARC/INFO to utilize data captured using a variety of location-referencing methods. However, no consideration is given to errors induced by converting data from one structure to another. Unfortunately, few have researched the types of errors produced by aggregating and disaggregating network data and sensitivity of transportation models to these errors.

Some network attribute data are used in transportation analysis to represent available travel supply. These values determine the level of service of a network, which, in turn, influences the amount of travel predicted, as well as the distribution of travel demand through the system. Certain transportation applications, such as evaluating the impact on travel demand of increasing capacity on network links, require models to be sensitive to changes in network attributes, like increases in the number of lanes. Other applications strive for the opposite effect. For example, when investigating a subregion of a network, an abstracted representation of the infrastructure outside the study area is desired. To achieve this abstraction, several links will be replaced by one link which, ideally, represents the same amount of supply available in original data. This approach of minimizing the effects of altering a network is also required for converting data structures. A new network representation should accurately reflect transportation supply found in the existing structure.

A pilot survey of state departments of transportation (DOTs) has been conducted to determine methods used for data structure conversion. The Virginia DOT is working with consultants to convert milepoint data to an intersection/link structure. All data are being stored as offset information and pointers are used to indicate direction of travel to which some attributes apply. It is left to the user to aggregate and disaggregate information to fit the particular application (unpublished data). The Maryland DOT has developed a data base structure that accommodates both milepoint and intersection/link queries. The Florida DOT recently completed conversion of milepoint data to an intersection/link structure (unpublished data). More information on these systems is forthcoming. The Wisconsin DOT has written Fortran routines to convert data structures. Smoothing algorithms used within these routines are currently being analyzed (unpublished data).

Although data base conversion is taking place in transportation agencies, the underlying issue of how to attach attributes to a new structure without compromising accuracy of existing information has not been sufficiently researched. A typical example of the situation faced in data structure conversion is shown in Figure 1. These roads are in James City County, Virginia. Note that a single line has been indicated for the link structure when in fact each line represents two links, one for each travel direction. Only three geographical locations of intersections and milepoint terminus correspond. Consequently, network attribute information found in Segments 1, 2, and 3 must be mapped to Links 1 and 2. Data relevant to Segment 5 must be reflected in attributes of Links 3, 4, and 5. A few milepoint segment attributes found in road inventory files are presented in Table 1. For this network segment, a decision must be made on how to determine the number of lanes and surface width for Link 1. This decision impacts link capacity, which, in turn, affects travel demand assigned on the link. The model developed here for data conversion allows sensitivity testing of aggregation/disaggregation

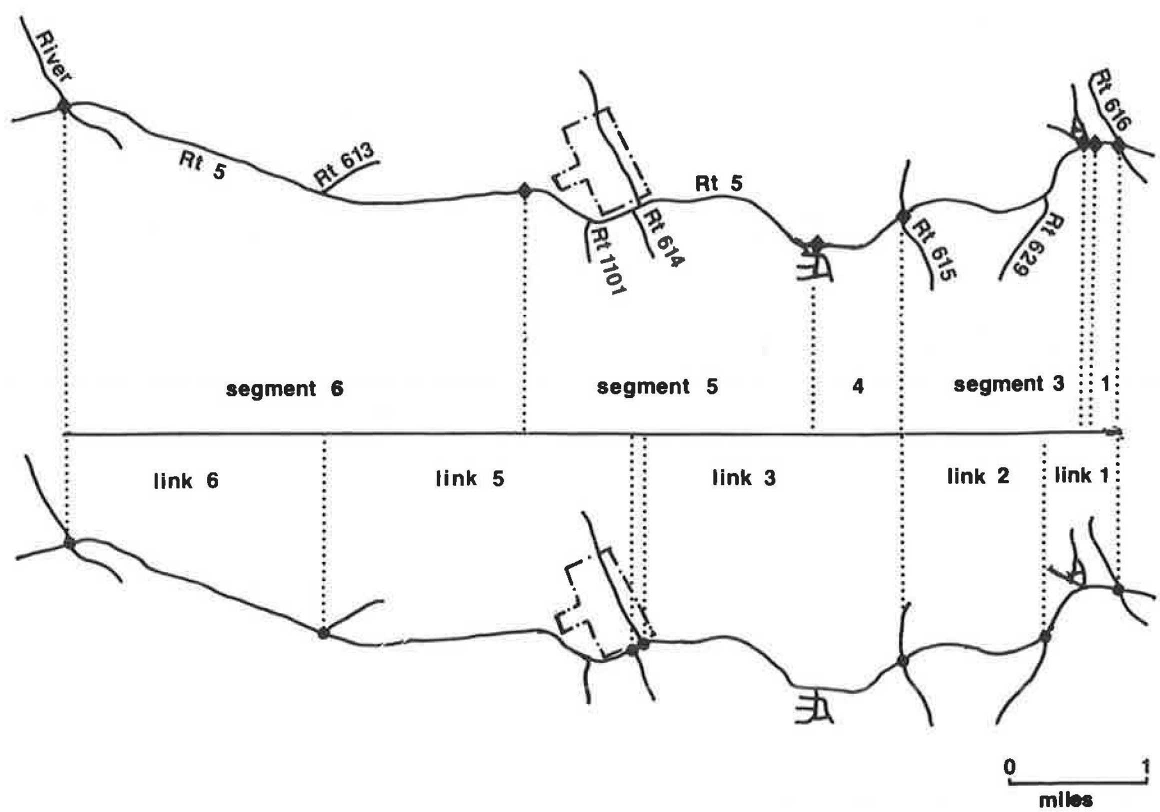

FIGURE 1 Network link and milepoint road segments in James City County, Virginia.

TABLE 1 SAMPLE MILEPOINT SEGMENT ATTRIBUTES

Rt.	From	To	Length	SW	ShW	Lanes	AADT	ST	MP
5	Rt199	18MWRt199	0.18	50	6	4	6260	6	5.01
5	18MWRt199	24MWRt199	0.06	24	6	2	6260	6	5.19
5	24MWRt199	Rt615	1.55	24	6	2	6260	6	5.25
5	Rt615	Rt1438	0.67	24	6	2	4590	6	6.80
5	Rt1438	53MWRt1101	2.28	24	6	2	4590	6	7.47
5	53MWRt1101	Chickahominy	3.43	24	6	2	4590	6	9.75

SW = Surface Width, ShW = Shoulder Width, ST = Surface Type, MP = Milepoint

rules so that accuracy in characterizing network supply is maximized.

DATA STRUCTURE CONVERSION MODEL

Input Requirements

Graphic display of a network is becoming commonplace in many transportation analysis models. Inherent in the definition of GIS is the ability of a system to graphically display and manipulate data. To enhance user-friendliness and simplify computational complexity of the conversion procedure described here, graphic images of highway networks are used by the microcomputer model. However, road inventory files contain highway attribute or descriptive data from which it is difficult to construct a map. Locational information found in road inventory files is summarized as follows. Each record contains (a) district, county, city or town identifiers, (b) the route number of the road segment, and (c) 12-character labels identifying termini-from and termini-to locations of the segment. No digital coordinates are provided in these files. Consequently, visual representation of the underlying network is not easily obtainable from this information.

A digital data base representing a map of the study area network is used to simplify data conversion. This digital representation conforms to an intersection/link network structure. Information on route number or name of each link accompanies the x,y coordinates of each node. Route numbers of links intersecting at the beginning and end node of each link are necessary. Depending on the nature of the application, other locational information may include county and analysis zone identifiers.

Several options are available for creating or obtaining a digital map to be used in the conversion process. Two procedures are described. If the study area is small and a digitizing table is available, software is provided in the data conversion system to create the required data base. Several base maps, such as county highway maps and U.S. Geological Survey 7.5-min quadrangle maps, may be used to identify an intersection/link structure that is matched with the location referencing method applied in road inventory files. Users may capture x,y coordinates for network nodes alone or digitize several shape points to represent curves in each link. A label, entered during digitizing, is associated with each link in the file. This label is similar to identifiers found in road inventory files. It indicates route name and the names of intersecting routes. Providing names of intersecting routes during digitizing is not necessary but may easily be done when the study area is small. Otherwise, this information can be constructed from the data base, assuming each link has been labeled. Finally, a county code may be entered that matches county code numbers found in road inventory files. County codes are used in the search process to eliminate sections of a road outside the area of interest.

An alternative method for obtaining a digital representation of the study area is to use an existing data base such as highway network (HNET) files [generated in Urban Transportation Planning System (UTPS)] containing link and node information, or TIGER files created by the Census Bureau, or possibly Digital Line Graphs (DLGs) produced by the U.S. Geological Survey. Any existing data base will require editing to incorporate link label information and county codes if this information is not currently present. In urbanized study areas, HNET files represent an ideal source as long as network information has been maintained. However, in many instances HNET files are incompatible with road inventory files that contain statewide data on primary and secondary roads, Interstates, and any road with a federal project number. In this event, TIGER or DIME files or DLGs may be used. These files require some restructuring to achieve an intersection/link highway network data structure. Further, users should be aware that a common criticism of purchased or externally acquired data bases is that information often is outdated, so, for example, newly constructed roads may not be in the file.

Data Structure Conversion

A two-step process is undertaken for converting milepoint information to an intersection/link structure. The first step involves variable or attribute definition. The second step

converts data based on rules defined in the previous step. Details on each of these procedures follow.

Variable Definition

Two types of variables found in road inventory files may be classified as measured and descriptive. A measured variable is one that may be aggregated or disaggregated using mathematical equations without significant loss in accuracy. This assumption is not valid for descriptive variables. With descriptive variables, aggregation of information will result in unacceptable loss in accuracy.

An example will clarify the variable classification methodology required to distinguish data types. Figure 2 shows a link that comprises three milepoint road segments. Each segment has a variable or attribute associated with it that takes on values *A*, *B*, and *C*. Suppose the variable is system domain and a value of *A* represents private land, *B* represents state agencies, and *C* represents national park service. If this information is aggregated on the basis of weighting formula, the value of the variable stored in the intersection/link structure is *C* because it is the value associated with the longest segment. When the new data base is queried to identify all links with value *A*, this link would not be selected because this information is lost from the system. However, if for each of these segments, the values for *A*, *B*, and *C* represent the number of structures within a certain distance from the shoulder, these values may be summed into a single link value without loss of accuracy.

Examples of measured data in road inventory files include length, number of at-grade railroad crossings, and number of structures. Descriptive data include surface and base type, route signing, and functional class. A group of variables, like surface and shoulder width, number of lanes, and average annual daily traffic (AADT) cannot be classified strictly as measured or descriptive. Mathematically aggregating values of these variables over a link is appropriate for some applications and inappropriate for others. A complete list of variables in road inventory files may be found in Table 2. Each variable has been labeled as descriptive or measured. Some records in road inventory files represent sample sections for the FHWA's Highway Performance Monitoring System (HPMS). For these sample sections, much more comprehensive data are maintained. A list of additional variables found in HPMS sections is provided in Table 3 as well as default classifications as measured or descriptive variables.

In the data conversion computer model, users are allowed to select which variables in road inventory files to convert to an intersection/link structure. The specific application for which data are required will dictate this selection process. For instance, applications using data to calculate link capacity from the *Highway Capacity Manual* formulation may use lane and shoulder width, number of lanes, and percent trucks attributes in road inventory files. However, applications that derive link capacity from the UTPS look-up table use facility and area type data.

Users also may change default classifications of selected variables. Classification of a variable dictates how and what type of information is kept in the new data base. If a variable is classified as descriptive, each link contains offset information along with the value of the variable associated with each segment. For example, referring to the link in Figure 2, attached to this link in the new data base is information indicating that this variable has value *A* for 0.4 mi, value *B* for 0.3 mi, and value *C* for 0.6 mi. If, however, the value of a descriptive variable does not change between segments, redundant information is not stored. For example, if the variable in Figure 2 has value *A* instead of *B* in Segment 2, the new data base saves this information as follows: value *A* for 0.7 mi, value *C* for 0.6 mi. Inclusion of a large number of descriptive variables significantly increases data base size. To simplify manipulation of information, separate data files are created for each descriptive variable that contain pointers to network links.

Variables classified as measured have default aggregation/disaggregation rules associated with them. For instance, length, number of structures, and number of RR crossings for each segment are summed. If AADT is aggregated, a weighted averaging method is provided. As this system is enhanced, users will be allowed to enter conversion equations as well as define new variables as mathematical and logical combinations of road inventory variables.

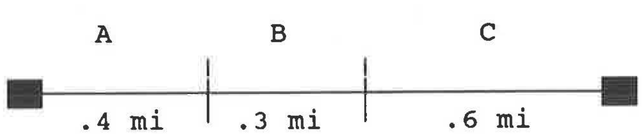

FIGURE 2 Schematic link representation in milepoint segments.

Data Conversion

Actual conversion of a milepoint data base into an intersection/link structure occurs in an interactive manner. Ideally, if all road segments fit perfectly into a link, conversion would be rather simple. Complications arise from the two input sources, namely, the road inventory files and the intersection/link digital data base. In road inventory files, segments do not necessarily terminate at all intersections. Further, users may not have selected all roads in the network as part of their digital data base. Consequently, users must be queried on how to handle special cases that occur during link building.

Before conversion can take place, road inventory files must be downloaded from the mainframe to a PC. Due to the size of these files, it is recommended that data are sorted first by county and stored in separate files. Users may then download any county of interest when needed.

The digital map of the study area is displayed on the computer screen throughout the process. Two methods are provided for identifying links for conversion. Either the program will convert links sequentially in the file or users can select individual links for conversion. Before converting data structures, the digital network file is sorted by route number to facilitate a sequential conversion approach. The second technique, individual link selection, provides users the flexibility of displaying a regional or statewide network though actually building a data base for a subarea. A zooming function is provided in the model for more precise display of the area of concern.

TABLE 2 ROAD INVENTORY FILE FORMAT

Column(s)	Variable	Classification
46 - 50	Length	Measured
51 - 52	Surface Width	Measured/Descriptive
53 - 54	Shoulder Width	Measured/Descriptive
55	Surface Type	Descriptive
56	Base Type	Descriptive
57 - 58	Kind of Highway	
	# Lanes	Measured/Descriptive
	Type of facility & access control	Descriptive
59 - 60	Rural & urban public at grade crossings	Measured
61 - 66	AADT	Measured/Descriptive
67	Type of Section Identification	Descriptive
68 - 72	Milepoint	Measured
73	Route signing	Descriptive
74 - 79	System	
	Governmental level of control	Descriptive
	Location	Descriptive
	Federal aid system	Descriptive
	Federal, state & local domain	Descriptive
	Special system	Descriptive
	Public road	Descriptive
80 - 83	Operation	
	Reversible lanes	Descriptive
	Trucks/commercial vehicles	Descriptive
	HOV lanes	Descriptive
	Toll	Descriptive
84	Functional class	Descriptive
85	Special code	Descriptive

Once a link has been selected for conversion, the immediate network surrounding the link is displayed, including node numbers. This link is placed in a box and highlighted on the regional map. Information on route number of the link as well as route numbers of intersecting links is provided (see Figure 3). Link length, in miles, is calculated from digital coordinates and scaling information. Error in determining distance from digital coordinates is introduced by the digitizing process and compounded by line generalization techniques as described by McMaster (*6*). Thus, it is necessary for the conversion procedure to rely on additional information, when available, and estimation routines to match milepoint segments to links. This interactive system allows the user to have the final say in piecing together the network data base. A county code associated with each link is used to determine the appropriate file to search for road attribute data. Within a county file, all records associated with the highway route of interest are identified.

Primary information in road inventory files used in the search algorithm is beginning and end termini labels. If a sequential data conversion technique is used, all links in a route identify a path from reference point zero through the county to a second zero reference point. Links are processed along this path by matching record end termini labels to cross-

TABLE 3 ROAD INVENTORY HPMS FORMAT

Column(s)	Variable	Classification
97 - 98	Surface width	Measured/Descriptive
99 - 101	Approach width	Measured/Descriptive
102	Shoulder type	Descriptive
103 - 106	Shoulder width	Measured/Descriptive
107	Median type	Descriptive
108 - 109	Median width	Measured/Descriptive
110	Widening feasibility	Descriptive
111	Horizontal alignment adequacy	Descriptive
112	Vertical alignment adequacy	Descriptive
113 - 115	Percent passing sight distance	Measured/Descriptive
116 - 117	Speed limit	Measured/Descriptive
118 - 119	Average highway speed	Measured/Descriptive
120	Signal type	Descriptive
121 - 122	Percent green time	Measured
123 - 124	Parking	Descriptive
125	Terrain type	Descriptive
126	Type of development	Descriptive
127	Urban location	Descriptive
128 - 129	Grade separated interchanges	Measured
	At grade intersections with	
130 - 131	signals	Measured
132 - 133	stop signs	Measured
134 - 135	other or no control	Measured
136 - 137	Commercial access points	Measured
138 - 139	Structures	Measured
140 - 141	At grade railroad crossings	Measured
142 - 143	Type of Improvement	Descriptive
144 - 155	Sample number identification	Descriptive
156	Sample section subdivision	Descriptive
157 - 161	State HPMS sample number	Descriptive
162 - 163	AADT volume group identifier	Descriptive
164 - 168	Expansion factor	Measured
169	Pavement section	Descriptive
170 - 171	Structural number or slab thickness	Measured/Descriptive
172 - 174	Pavement condition	Measured/Descriptive
175 - 176	Skid resistance	Measured/Descriptive
177 - 179	Existing right-of-way width	Measured/Descriptive
180 - 181	Percent trucks peak period	Measured

TABLE 3 (*continued on next page*)

TABLE 3 (*continued*)

Column(s)	Variable	Classification
182 - 183	Percent trucks off-peak period	Measured
184 - 185	K-factor	Measured
186 - 188	Direction factor	Measured
189 - 193	Urban peak capacity	Measured/Descriptive
194 - 198	Urban off-peak capacity	Measured/Descriptive
199 - 204	Future AADT	Measured/Descriptive
205	Drainage adequacy	Measured/Descriptive

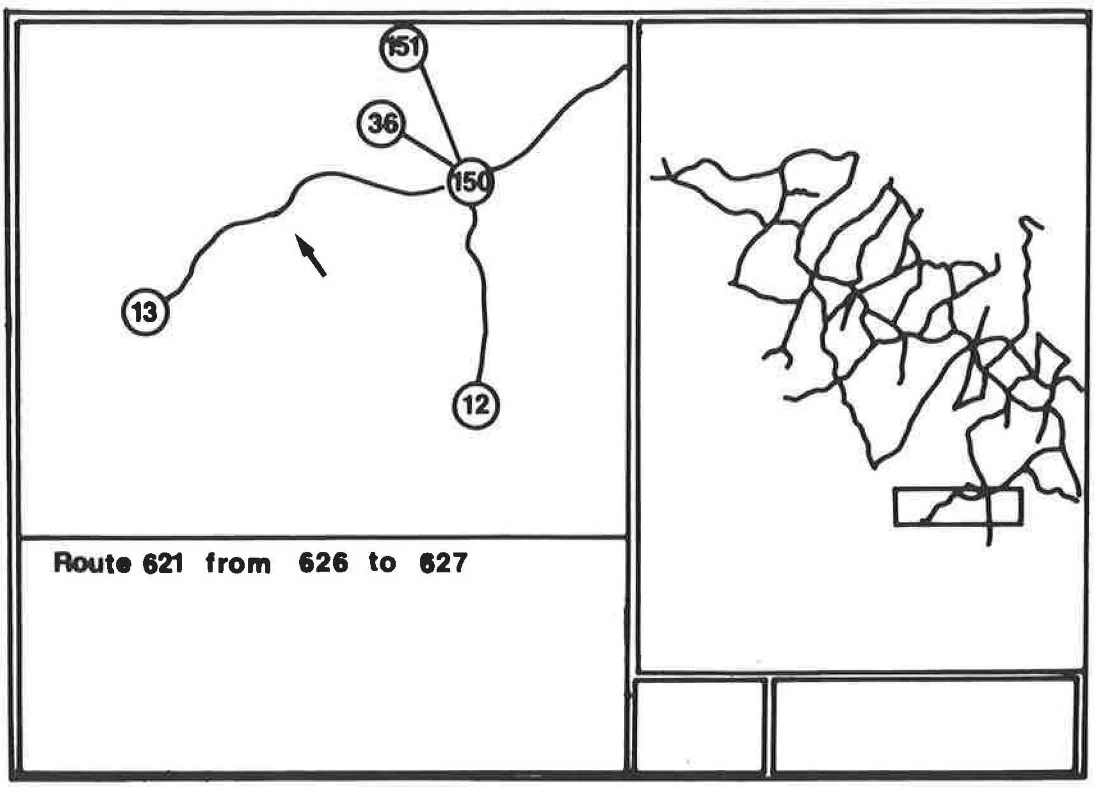

FIGURE 3 Screen representation of data conversion program.

route labels at intersections. If an intersection occurs between the endpoints of a road segment, distance information is used to disaggregate data in this segment and assign converted data to the two links. When multiple road segments form a link, selected variables are aggregated using rules defined by users. As records in road inventory files are being assigned to network links, they are displayed on the terminal to facilitate user intervention.

SUMMARY AND CONCLUSIONS

The model described is still in a design and development stage. On completion of programming, tests will be run on several networks. Emphasis will be placed on using available digital files, such as TIGER files, to reduce the effort and cost associated with digitizing network maps. Investigations will concentrate on sensitivity of network analysis to aggregation and disaggregation of supply variables, ability of the model to improve, in terms of accuracy, speed, and flexibility, the manual process of building a network using incompatible data, and user-friendliness of the procedure. As tests are run, improvements will be made to the model as required in order to develop a system that facilitates sharing of transportation data bases.

Issues of data accuracy and the impact of aggregation and disaggregation of network supply attributes for subregion analysis and data structure conversion are topics for continued

research. The model presented here is intended to facilitate the data conversion process. However, further investigation is required into appropriate models for network data aggregation.

REFERENCES

1. T. Nyerges. Locational Referencing and Highway Segmentation in a Geographic Information System. *ITE Journal*, Vol. 60, No. 3, 1990, pp. 27–31.
2. T. Henderson. Position Paper: A Comprehensive Proposal for the Acquisition, Use, and Maintenance of Geographically Related Information in the New Mexico State Highway and Transportation Department. In *Proc., GIS for Transportation Symposium*, AASHTO, Washington, D.C., 1989, pp. 219–268.
3. H. J. Simkowitz. Geographic Information Systems: An Important Technology for Transportation Planning and Operations. In *Proc., GIS for Transportation Symposium*, AASHTO, Washington, D.C., 1989, pp. 161–192.
4. F. L. Mannering and W. P. Kilareski. The Common Structure of Geo-Based Data for Roadway Information Systems. *ITE Journal*, Vol. 56, No. 7, 1986, pp. 43–49.
5. K. Kuykendall. *Representing a Highway Location Referencing System with a GIS in a Relational Environment.* Environmental Systems Research Institute, Redlands, Calif., 1989.
6. R. B. McMaster. The Integration of Simplification and Smoothing Algorithms in Line Generalization. *Cartographica*, Vol. 26, 1989, pp. 101–121.

Publication of this paper sponsored by Committee on Photogrammetry, Remote Sensing, and Related Automated Systems.