COMPATIBILITY FILE: AN INFORMATION RETRIEVAL SYSTEM

Stuart D. Leland and James W. Burke, Connecticut Department of Transportation

The compatibility file serves as an interface between the transportation planning network files and other data files that are referenced by route number and cumulative mileage. The fundamental purpose of the file is to identify, by route number, each of the highway segments contained in the transportation study's link file (historical record) and also to identify by route number and cumulative mileage each node or highway intersection. The compatibility file would then be used to relate the computerized transportation study traffic networks to the physical inventory files maintained by the specialty divisions (traffic, highway needs, program planning, and accident) in a department of highway or transportation. The completed file not only provides a means of data exchange but, when used in conjunction with digitized networks and a data plotter, represents an excellent means of file editing and data presentation.

• HISTORICALLY, the physical inventory phase of a transportation study has presented the planner with a very costly task in terms of time and money. The data obtained during this important phase of a study are applied to the development of a historical record or link file that in turn serves as the base for computerized transportation networks.

Experience indicates that many portions of this inventory have been previously developed and are available but in a format that is not compatible with a transportation study's methods of storage and retrieval.

In the larger studies, such as those conducted by a department of highways or transportation, physical inventory data may be available from any one of several agencies that maintain separate inventories. Throughout the years, these agencies have been responsible for the development of portions of the existing transportation system and, therefore, data collection pertinent to the various specialty areas (such as traffic and needs) represents a normal function.

In recent years, the Transportation Planning Division in the Office of Planning, Bureau of Planning and Research of the Connecticut Department of Transportation, has been involved in several studies that require extensive use of data contained in the inventory files of other units. During this time, careful consideration was given to the possibility of developing a system that would allow the transportation planning data files to be compatible with the computerized inventory records of other units.

The Connecticut Department of Transportation has several such units or divisions that maintain inventory records of the existing highway system. Although the inventories have basic similarities, each is separately maintained and each serves separate functions. Typical data that may be obtained from the files are items such as number of lanes, pavement and shoulder width, geometric characteristics, and physical condition.

Several applications of the inventories were visualized. The data, however, were available only in printout form and could not be used in conjunction with the computerized networks except through manual conversion. To take full advantage of the data and,

Sponsored by Committee on Traffic Records.

where possible, to eliminate the manual process, we initiated an effort toward the development of the compatibility file.

TRANSPORTATION PLANNING NETWORKS

The Transportation Planning Division of the Office of Planning maintains a set of historical records (link files) that contain data relating to approximately 60 percent of Connecticut's total roadway system. These physical inventory data are used almost exclusively as input to the computerized networks. The largest of these networks describes some 9,100 miles of roadway, which are represented by approximately 11,700 links or segments. (For the purpose of this paper, the term roadway refers to all types of roads regardless of function or class, whereas the term highway is used to describe primary and secondary systems.) In addition to a town-node description, the file contains a length, operating speed, lane, and roadway class record for each segment.

A smaller network contains similar data but represents fewer total miles. This system has some 5,800 links that represent 6,900 miles of roadway. Aside from the roadway mileage, the major difference between the two systems is the number and, consequently, the size of traffic zones. Each zone system represents the same land area, and in each of the files the total state highway system is represented. The variation in represented miles is attributable to nonstate local and collector roadways that are recorded in the system because of their significance in the traffic assignment process.

The roadway segments described in these files are coded for use with computers. As such, there is no identification (route names or numbers) associated with the roadway segments.

The roadway segments (links) are identified at each end by a three-letter town designation and a three-digit intersection (node) number. For example, assume that a given segment of highway between towns ABC and XYZ was to be identified by this file. The roadway would be listed by the town designator and the intersection number at each end of the segment. The intersections are numbered separately for each town, from 101 to 999. The link between towns ABC and XYZ might therefore be described as ABC101—XYZ101. In the event that the roadway segment existed entirely within town ABC, the appropriate description might be ABC101—ABC102. For the transportation planning networks, no other identification is required. The relation of a particular link segment to the existing system is established manually through a visual map comparison.

COMPATIBILITY FILE PRINCIPLE AND DATA FILES

The basic concept of the compatibility file is to identify, by route number, each of the highway segments and also to identify by route number and cumulative mileage each highway intersection. The compatibility file then could be used to relate the transportation planning networks to the physical inventory files maintained by the Records and Inventory, Traffic, and Program Planning Sections. Each of the aforementioned sections maintains an inventory of the highway system and catalogs data by route and cumulative mileage.

Unlike the files maintained by all other units, the network files contain the length of the roadway segments to the hundredth of a mile. Each of the other inventory files involved in this project contains the length of a section to the nearest thousandth of a mile. A brief description of these files follows.

The state road inventory master file (RM) is a tape file that is maintained by the Records and Inventory Section of the Office of Planning. It contains the largest number of records pertaining to the existing highway system. The roadway length recorded in this file is the result of a field inventory utilizing a "fifth wheel." The file lists a highway mileage, for each route, to the nearest thousandth of a mile and indicates a cumulative mileage point for each intersecting roadway. Various inventory data are recorded for each highway segment designated by the cumulative mileage points. The intersections are not identified by route number or word description in this file. They are identified in a companion file, the highway master log, that serves primarily as a mileage record and was utilized (in this study) only to identify the intersecting roadways at the cumulative mile points.

The highway needs inventory file is a tape record maintained by the Program and Scheduling Division of the Bureau of Administration. The file contains extensive inventory data that are used to develop a set of priorities for funding the needs of the highway system. The file also logs each route by cumulative mileage, but it does not list intersecting roads as thoroughly as does the RM file. The highway sections between cumulative mile points are identified by a route and a sequence number. The sequence numbers are consecutive, but space is allowed for breaking existing sections into smaller sequences. The physical inventory and related needs data are recorded for each sequence.

The accident master file is a tape record of reported accidents that is maintained by the Traffic Division of the Bureau of Highways. The file records each accident by type and utilizes route number and cumulative mileage to identify locations. In both the highway needs inventory and the accident master files, the cumulative mile values are obtained indirectly from the master log file. Other data files, such as the straight line diagram, that are inventoried and prepared by the Division of Traffic are being converted to computer format and also will be compatible with the transportation planning traffic networks.

FILE DEVELOPMENT

All of the aforementioned files have much in common; however, each has its own route number and cumulative mileage record. Although the transportation files do not contain a cumulative mile record, they do have a progressive summation of the link distances that creates the potential for establishing that item.

The first step in the development of the compatibility file was the identification of the network links, which represent the state-numbered road system. This manual process required a visual comparison of networks and the state road system.

The coding format (Fig. 1) was developed for use in the link-route identification phase. Application of this format allowed the related items to be collected and recorded during a single review of the system. The data contained on this coding form represent the base of the compatibility file and are discussed in the following.

The individual nodes that represent state routes are identified on the coding form along with the corresponding route number. This is accomplished on a route-by-route basis with a link distance and route sequence number recorded for each node. The link distance is taken from the link file. It represents the distance from the node being coded to the next node having the same route number and the next sequence number. The sequence number is used in forming node strings, which are a computer representation of the various routes. The node sequence numbers, starting at the zero cumulative mile point, are assigned in consecutive multiples of five (5, 10, 15, etc.) (In Connecticut, the zero or starting cumulative mile point for any highway is considered to be the most southerly or westerly end.) This procedure is followed to allow coding space for future expansion of the road system.

The overlap route numbers serve the function of identifying multinumbered or overlap routes, e.g., US-6/US-202. This eliminates double records for single roadway or link segments. The lowest numbered route is always referenced during data retrieval.

Word description is the only nonfunctional item contained in the file. Its purposes are to allow identification of link segments that cross town boundaries and to provide any desired comment relating to a particular segment. There is no restriction on the description format because it serves only informational purposes.

The intersecting route number and cumulative mile records represent the major control in the file. The intersecting route number is employed when a node is determined to represent a major route intersection. (The junction of two or more state-numbered routes is considered to be a major route intersection.) Where this occurs, a cumulative mile value, which is obtained from the master log file, is assigned to that node. It is important to note that this cumulative mile record is retained as a control value in the node string and is referred to elsewhere in the text as the checkpoint cumulative mileage.

Figure 2 shows the steps followed in arriving at the completed compatibility file. The more important aspects of this procedure are presented in the following text.

The node strings are formed by using a computer program that joins or strings the nodes into route sections. The program, as written, requires identical route numbers for nodes being added to the string and also checks the sequence number for compliance with the requirement that the sequence number of the node being processed be greater than the preceding sequence number.

The node string file was developed manually by transferring data from assorted maps and printouts. Because of this, the initial file contained many coding errors. The first consideration, therefore, in the editing process was to ensure that node strings do, in fact, represent valid links. This was accomplished by creating links from the node strings and matching these with a current link file that had been employed in the initial coding process.

The preliminary edited file is used to sum the link distances and to compare them with the checkpoint cumulative mileage. The summed link distance and the actual recorded distance, as computed from the control record cumulative mileage, must fall within a specified tolerance or the entire string is rejected. As each node string is processed, the preceding cumulative mile checkpoint becomes the zero or starting point for that string. During the editing of the string, only the difference between the newly designated starting point and next checkpoint and the summed link distances is compared. In this study's application, a 10 percent difference was allowed because the checkpoints rarely exceeded 3 miles and were generally less than 1 mile; therefore, any influence from compensating errors was minimal.

The rejected node string is recycled through a manual editing to determine the error and to correct discrepancies. It was found that rejections of node strings from the distance editing were due primarily to errors in coding and in transferring link distances. The distance editing also pointed out distance errors on links in the base network that had never been discovered because of the lack of an adequate editing program.

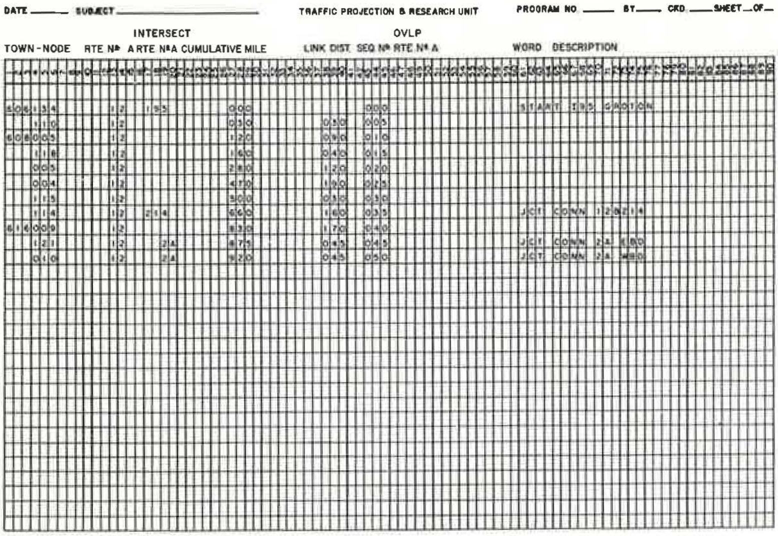
The final phase of the editing procedure was the compilation of the compatibility file. After it was established that the difference in cumulative mileage checkpoints and the sum of the link distances was no more than 10 percent, a cumulative mileage was computed for each node. This was accomplished through the use of a computer program that adjusts the link distances based on the error in the string segment, i.e., when a difference in distance existed between cumulative mile value and the summed link distances, and that difference was within the specified tolerance, the link distances were adjusted on a prorated basis to compensate for that difference. Then, by using the adjusted link distances, we computed cumulative mileages and assigned them to each node. The computed cumulative mileages then were added to the file, which completed the compatibility file.

IMPLEMENTATION

The actual data exchange is accomplished by using a generalized computer program written for the specific purpose of implementing the compatibility file concept. This program reads the compatibility file plus a data tape referenced by route and cumulative mileage and, according to user-specified parameters, performs certain data manipulation. The program uses the compatibility file as an interface between the networks and inventory files and produces a data tape referenced for use with the appropriate planning files.

The user may specify a maximum of seven data fields for manipulation. Allowable arithmetic operations include finding the minimum value, the minimum nonzero value, the maximum value and the mean, and the sum of the applicable input data. The user must supply the location of the data field on the input and the location for the output. He must also supply the tape blocking factor, the item size, and the field location of route and cumulative mileage for all files. The following example will best illustrate the computer program process. Let us assume that a user wants to update a highway

Figure 1. File coding format.



DATE ____ SUBJECT ____ TRAFFIC PROJECTION & RESEARCH UNIT PROGRAM NO ____ BY ____ CKD ____ SHEET __ OF __

TOWN-NODE	RTE Nº A	INTERSECT RTE Nº A	CUMULATIVE MILE	LINK DIST	OVLP SEQ Nº	RTE Nº A	WORD DESCRIPTION
506134	12	195	000		000		START I95 GROTON
110	12		030	030	005		
608005	12		120	090	010		
118	12		160	040	015		
005	12		280	120	020		
004	12		470	190	025		
115	12		500	030	030		
114	12	214	660	160	035		JCT CONN 12&214
616009	12		830	170	040		
121	12	2A	875	045	045		JCT CONN 2A EBD
010	12	2A	920	045	050		JCT CONN 2A WBD

Figure 2. Development of compatibility file.

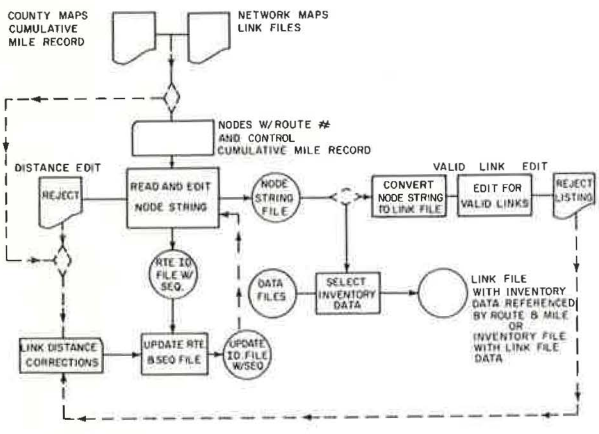
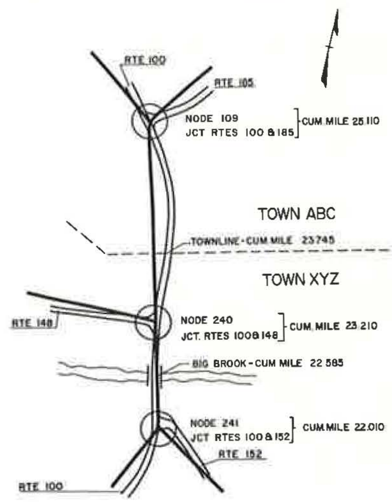


Figure 3. Data overlap section.



network link file with accurate street widths and that he wants the minimum street width for the length of each link recorded.

The computer program, after initial validation of the input parameters and tape files, reads the data tape and stores the applicable data (street width and route and cumulative mileage) for each inventoried segment contained in the data master file. The compatibility file is processed next. As each link is read, the data array is referenced by using route and cumulative mileage. If any part of an inventory segment is found to overlap the link in question, it is included in the development of data for that link. Any link for which there are inadequate data on the data master file is reported in an on-line printed output and included on the tape output with partial or no data.

The primary output is a data tape referenced by route, cumulative mileage, and link number that is used to update the file.

PROBLEM AREAS AND LIMITATIONS

The overlap of data segments is a problem that is difficult to eliminate from the data exchange system.

Because each unit that maintains an inventory file collects and records data primarily for use in its own specialty area, the beginning and ending points of data sections do not necessarily coincide. This is particularly true of transportation planning networks and needs files. Because the networks emphasize traffic projection and simulation techniques, they terminate links at nodes or intersections. The needs inventory file is concerned with the actual physical condition and consequent needs of the various facilities. In this file, therefore, data sections are referenced to needs projects and may begin or end at any point along the highway.

When a cumulative mile value is assigned to the nodes, the link sections may overlap the data sections represented by the other files and vice versa. Where these conditions occur, more than one data set may be available for a specified route segment (link).

This situation is less likely to exist in relating the networks to the RM file because the length of data sections contained in the RM is relatively short.

It is not a serious problem so long as the user is aware of the potential situation, and the degree to which any selected output is affected is dependent on the data involved.

A typical overlap situation is shown in Figure 3. If such a condition did exist, several sets of data might be available for that segment of Route 100 described as link ABC 109—XYZ 240. Whether a data retrieval problem would exist for this segment would depend on the data requested. If the user parameters specified a minimum or maximum value, then a single record, such as average daily traffic (ADT) or width, would be selected from the several records available.

A problem may occur when a listing of segments falling within a certain value range is requested. Under these conditions, and depending on the tolerances specified, a roadway segment may be identified as possessing the characteristics of an adjoining roadway section. In these instances, both sections would appear in the listing.

Assume that the user parameters specified a listing of all two-lane facilities carrying an ADT of 13,000 or greater. If the segment of Route 100 (Fig. 3) between cumulative miles 22.585 and 23.745, which overlaps link segments ABC 109—XYZ 240 and XYZ 241, has a recorded ADT equal to or greater than the specified minimum (13,000), then both link segments would appear on the listing. This, of course, assumes that the file being used records the sequence between the aforementioned cumulative mile points as one section. Some files, such as the RM file, might very well consider the same sequence as being several smaller sections, in which case the link ABC 109—XYZ 240 might represent two data sections. Those data sections would be from cumulative mile points 23.210 to 23.745 and 23.745 to 25.110. In the event that either of those sections recorded a value within the user's specified tolerance range, the link would be entered on the listing.

ADAPTATION TO OTHER SYSTEMS

Although the system described in this paper was designed for the inventory referencing scheme and transportation planning networks that are used in Connecticut, it is felt that the basic elements of the compatibility system can be used by every state and most transportation planning study groups. It is recognized by the authors that a majority of these planning study groups utilize the available Federal Highway Administration Urban Transportation Planning Program battery. The historical record format (Appendix), represents the basic inventory file of the computerized network used in this system. It contains items such as street width, parking, pavement type, and roadway classifications. These inventory items are allotted space for a rigorous analysis of network adequacy and, therefore, they should be included in the file. The problem, as was true in Connecticut, is that these data must be transferred to the historical record manually.

Before we discuss the system's applicability to other jurisdictions, let us review the system elements. First, a detailed transportation planning highway network must be available; this file should include route number (or similar identification of the link segments) and a record of the distance between nodes. If automatic data plotting is desired, geographic coordinates must also be available. Second, a compatibility file is prepared by using this coded network. The mandatory items contained in this file include route number, node number, and a cumulative mileage value for each node. The third and final consideration is the availability of data files referenced by route and cumulative mileage.

All states are required by law to maintain an inventory of the physical aspects of the state-maintained highway system. In addition, most states have accident information, needs, traffic control, and traffic count files. Usually, a computerized system does not exist to transfer the data from one file to another or from a file to the historical record. In most states, the files are referenced by route and cumulative mileage, and a few are referenced from mileage posts or some other variation of the two. Regardless of the referencing scheme, the compatibility system described in this paper can be applied.

The historical record format used in the FHWA Urban Transportation Planning Program battery readily lends itself to the development of a compatibility file. All of the previously mentioned items, node number, route number, and distance between nodes, are found in the historical record. The application of the techniques of progressive link distance summation combined with proper editing procedures will yield a cumulative mileage value for each node that can be further adjusted to the user's tolerances.

The development of the compatibility file provides the user with a new and powerful data-processing tool that, as the following application descriptions relate, allows previously inaccessible data to be transferred, sorted, plotted, and edited.

APPLICATION

The first data exchange accomplished through the use of a compatibility file was directed at the highway needs inventory files.

In this application, a program was written to utilize the geometric inventory data contained in the needs file as input to highway capacity programs. The resulting capacity values, based on procedures given in the 1965 Highway Capacity Manual were computed for each highway section listed in the needs file. Through use of the compatibility file, the highway sections and corresponding link segments were identified and the capacity values were transferred to the network files.

The capacity values also were used to add volume-capacity ratios to the needs file, which provided some measure of traffic demand. This was accomplished by utilizing the ADT value recorded in the needs file to obtain a service volume that was then compared with the computed capacity value to obtain the ratios. It is also possible to obtain a service volume by using projected traffic assignments to measure future traffic demand on existing systems.

Perhaps the single application that has proved most beneficial to the transportation study is the ability to easily relate links to routes and to add current ADT values to the link files.

The Connecticut Department of Transportation produces an annual ADT log. As this tape file (ADT log master file) is updated, it is a simple mechanical operation to transfer the current data to the link files. A recent application of this type has resulted in a 20 percent increase in the total number of ADT links in the network. This is a tremendous asset to the network calibration process because it represents a substantial increase in the number of links with ADTs and also updates existing records with recent counts.

Although initial implementation of the compatibility file was intended as a transportation planning tool, its benefits are easily extended to each of the agencies whose files are involved. Data editing and manipulations may be performed by working from the network and compatibility files. Many programs that are developed primarily for a transportation study may be used to edit or update files of other agencies. This type of application has recently been used to transfer current ADTs from the ADT master log to the RM file. Prior to this application, such a process was a time-consuming and lengthy manual project that created a difference of up to 2 years in the recorded ADT values.

One outstanding example of multiple-agency use in applying the file is in file editing. The transportation planning networks have been digitized for use with a data plotter, and, therefore, all data compatible with the network files may be plotted. Experience has shown that plotting data in this manner is an excellent form of file editing. It allows a rapid visual examination and very quickly presents a broad overall view of the data.

By using available computer programs in conjunction with the plotter, various combinations of data may be sorted, grouped, and presented for visual review. Data may be selected from individual files and combined and plotted to present relationships such as accident areas versus needs projects or accidents versus safety projects. Other items such as needs, railroad crossings, bridge repair projects, capacities, and ADTs may be plotted by individual route or area as well as on a statewide basis.

Recent applications have included the plotting of a statewide traffic flow map by using current ADTs. Plots of this type are also serving as drafting aids that provide the cartographer with scale values for use in producing published traffic flow maps.

SUMMARY

A procedure has been presented that provides the transportation planner with the capability to easily access typical inventory files such as highway needs, physical inventory, and accident and traffic files.

By using as an interface a node-to-node description of the state highway system, a planner can transfer data between transportation planning network files and inventory files. Connecticut Department of Transportation planners can perform the following by using this system:

1. Analyze the effects of the physical characteristics of the existing highway network in conjunction with developing a future transportation system;
2. Increase the data base of the transportation planning study into areas such as accident analysis, highway needs, and traffic inventory that heretofore were inaccessible, except on a small scale, because of the difficulty in manually transferring the data;
3. Graphically present any aspect of the transportation network and inventory files by using a data plotter; and
4. Aid other departmental units by making available to those units that maintain inventory files the various sorting, editing, and data plotter computer programs that have been written for transportation studies.

The system has been implemented to transfer current ADTs to the transportation planning network files. Also, capacities that were calculated by using procedures outlined elsewhere (1) and physical inventory data have been transferred to all state road links in the network files. Volume-to-capacity ratios have been calculated and transferred to the highway needs inventory file. All of the preceding items plus many aspects of the highway needs file have been plotted. Pictorial representation of many of these data is available for the first time.

It is felt that this procedure represents a real asset to the transportation planner by adding further to his ability to provide an effective transportation plan.