

# INFORMATION MANAGEMENT AT STATE HIGHWAY DEPARTMENTS: ISSUES AND NEEDS

By William Rasdorf,<sup>1</sup> Ed Shuller,<sup>2</sup> Ron Poole,<sup>3</sup> Osama Abudayyeh,<sup>4</sup> and Forrest Robson<sup>5</sup>

(Reviewed by the Urban Transportation Division)

**ABSTRACT:** The North Carolina Department of Transportation, in its continuing efforts to more fully and effectively use advanced information technology tools to increase institutional productivity and effectiveness in managing its transportation infrastructure, embarked on a study that involved developing a comprehensive and unified enterprisewide information management system. This paper focuses on the issues and needs that emerged as the information system development study evolved. These issues and needs became so significant that they warranted a detailed analysis in and of themselves. It is these issues and needs, which have a degree of universality, that are reported herein. This paper first introduces and discusses the functions of various participating units at state departments of transportation (referred to as stakeholders). Then, the paper identifies the issues and needs that must be fully understood and considered in the development of a unified information management system.

## INTRODUCTION

To effectively plan, manage, and sustainably develop and maintain our physical transportation infrastructure requires that we first effectively model it using some form of an information system. Ideally we would utilize this information system in an automated manner using a range of advanced technologies (Sinha et al. 1988). Yet such a system is complex; it is large; it mandates a need for integration; it highlights incompatibility issues; and it mandates consideration of data sharing and security. Addressing these issues and needs is critical to building an information system to meet the ever increasing social, economic, and environmental demands placed upon our transportation infrastructure.

## Context

The North Carolina Department of Transportation (NCDOT) manages its transportation infrastructure using a number of large, complex systems and programs maintained by different units at NCDOT. Each program and/or system generates and maintains tremendous amounts of data and information in separate, independent databases. Many of these data are processed, analyzed, and shared by more than one unit in more than one format.

All state DOT data have a spatial component. That spatial component represents an area such as a highway corridor, a linear feature such as a road, or a point feature such as a sign or the location of an accident. Existing database technology and data management system design traditionally have not been effective at allowing units within DOTs to use or share data as extensively or as easily as should be the case. One

major reason for this lack of sharing has been that the spatial component of the data has not been effectively or correctly associated with the attribute data. This has led to problems in the quality, integrity, duplication, access, and availability of NCDOT's databases. Consequently, the decisions that are based on such data are significantly affected. This, in turn, has economic implications when it comes to ongoing operations or to the overall strategic planning of future transportation infrastructure investments.

Many DOTs are creating unified information management systems (Khan et al. 1994). That is, they have discovered that they possess widely distributed and vastly different databases, each of which contributes significantly to the mission of the organizational unit that owns and maintains it. DOTs have recognized these individual databases to be valuable components of an organizationwide information resource. They have thus recognized the need to combine them, to explore relationships among various units' data resources, and, quite often, to make them available to other agencies and to the public. Doing so will increase institutional effectiveness and productivity.

But these disparate databases also have widely varying levels of precision and accuracy in their data content. Additionally, a variety of data formats, ranging from sequential files to relational tables, are used. There are also no accepted standards or practices for dealing with spatial issues. Thus, diversity is predominant.

## Data Access and Information System Design

There are two ways to approach the development of a unified information resource in such an environment. The first is to maintain the data in the current diverse format used and to develop tools and methods that support data warehousing, access, and mining across multiple platforms and formats. Alternatively, one may redesign the entire information resource using a common format, thus rebuilding the data resource structure from scratch.

The first alternative allows each organizational unit to undergo the least change. However, complex tools and methods are required to link its data to that of other units. The management of the linkages and the formats becomes problematic. But an even greater problem is that data may be spatially represented differently in each such database. As a result, the actual spatial location of physical components cannot easily be determined. Thus, the data cannot effectively be used in combination with each other which allows robust graphical

<sup>1</sup>Prof., Dept. of Civ. Engrg., North Carolina State Univ., Box 7908, Raleigh, NC 27695.

<sup>2</sup>Former Mgr., Geographic Information Sys. Unit, North Carolina Dept. of Transp., P.O. Box 25201, Raleigh, NC 27611.

<sup>3</sup>Mgr., Statewide Plng. Branch, North Carolina Dept. of Transp., P.O. Box 25201, Raleigh, NC.

<sup>4</sup>Asst. Prof., Dept. of Constr. Engrg., Mat. Engrg. and Industrial Des., Western Michigan Univ., Kalamazoo, MI 49008.

<sup>5</sup>Mgr., Geographic Information Sys. Unit, North Carolina Dept. of Transp., P.O. Box 25201, Raleigh, NC.

Note. Discussion open until September 1, 2000. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on May 24, 1999. This paper is part of the *Journal of Transportation Engineering*, Vol. 126, No. 2, March/April, 2000. ©ASCE, ISSN 0733-947X/00/0002-0134-0142/\$8.00 + \$.50 per page. Paper No. 21020.

and spatial display and analysis, or with a geographic information system (GIS), for that matter.

The second alternative provides an opportunity to design a common spatial framework that alleviates the majority of the problems noted above. However, this approach introduces organizational impacts and requires significant restructuring and rebuilding of existing databases.

### **NCDOT Information Management System Design Project**

With the advent of advanced information technologies such as GISs and database management systems (DBMSs), and with their proliferation and increased use, the NCDOT has recognized the need to provide a degree of standardization in the representation, management, and use of data. The NCDOT's main objective is to create an environment where a wide range of transportation data can be universally retrieved, used, and shared in either a GIS format or in a nongraphical database format. In particular, NCDOT desires to efficiently combine tabular and spatial data into an integrated system to support queries, applications, data entry and maintenance, analysis, and report generation, to the extent possible, from both the GIS and the DBMS.

To support these objectives, and to provide a common spatial reference, NCDOT decided to redesign its entire information resource. Doing so involved determining and standardizing an effective base linear referencing system (LRS) to meet its needs, developing a comprehensive database design with focused attention given to the types of data analysis functions performed by NCDOT, standardizing data terminology, and determining the shared data needs of the participating business and engineering units. But before any work would be accomplished, an assessment was undertaken and a number of issues and needs was identified, considered, and resolved.

The first part of the project, which sought to determine the most effective base linear referencing method to meet NCDOT's needs, has been completed and reported in the literature (Geo Decisions 1997) and (Kiel et al. 1999). A brief summary of the results is included in this paper since it is intrinsically tied to the issues and needs described herein. This paper, however, is concerned primarily with the second task outlined above—the NCDOT database design (Sherk and Rasdorf 1998).

The NCDOT database project is aimed at the development of a unified relational database schema to meet the DOT's information management system needs. The relational database format was selected to provide a more manageable and expandable data environment than is currently in use, to standardize data formats, to promote data sharing, and to minimize training requirements over the organization. The unified database is initially designed to meet the needs of a specified set of individual DOT units including Traffic Surveys, Forecasting, and Planning; Traffic Safety Systems Management; Pavement Management; and Roadway Inventory. These DOT units will be referred to as the "stakeholders."

### **DATA MANAGEMENT STAKEHOLDERS**

Presently, there are many existing databases in use within NCDOT by the stakeholders. This section examines the scope, nature, and functions of each stakeholder database to gain an understanding of the data resource needs as currently embodied in existing data files. This is critical so that the design of a proposed new DBMS meets the existing needs of the stakeholders and accommodates all of their data. In other words, it must support the standard queries of interest to these stakeholders; it must support the data needs of their application programs; it must provide the reporting capabilities they uti-

lize; and it must support the data entry and management functionality required. These needs and requirements are echoed by others [e.g., Sadek et al. (1996)].

### **Traffic Surveys, Forecasting, and Planning Units**

The Traffic Surveys, Forecasting, and Planning Units are responsible for traffic counting, traffic analysis, and traffic projection operations. Their specific responsibilities are described below.

#### *Traffic Surveys Units*

The primary function of the Traffic Surveys Unit is to collect transportation data and to format it in a meaningful manner. This data can then be published for use by other units throughout NCDOT (Highway 1994). Approximately 90% of the data is specific to traffic volumes, vehicle types, and vehicle weights. The unit may be viewed as a transportation data warehouse. However, in addition to storing data, the unit also processes data to derive traffic and travel statistics that are of use to others.

The Traffic Surveys Unit is responsible for maintaining data that are collected in a series of "programs." In this context, the work program is defined as the data gathering and analysis activity within the unit. The Traffic Survey Programs include the following:

- **Continuous Count Program:** A program where traffic counts are continuously collected using automated traffic recorders.
- **Coverage Count Program:** A program where traffic counts are periodically collected at fixed locations using portable traffic counters.
- **Vehicle Classification Program:** There are approximately 300 sites devoted to counting and classifying 13 vehicle types.
- **Weigh-in-Motion Program:** There are 23 State Highway Research Program sites and 5 Weigh-in-Motion Program Monitoring Stations that are in continual operation.
- **Special Count Program:** By a special request from another unit within NCDOT, a wide variety of other data collection activities may be conducted.
- **Vehicle Occupancy:** Data related to the number of people occupying each vehicle.
- **Origin and Destination Survey:** Data related to the origin and destination of travelers.
- **Drawbridge Reports:** Data for drawbridges are maintained in a paper format and are mainly concerned with the number of different vehicle types that use drawbridges.
- **Ferry Utilization Reports:** Data for ferries contain counts and classifications of vehicles using the ferry system, accounting for vehicles towing trailers.
- **Historical Paper and Map Files:** The unit also stores old paper records such as past traffic counts and maps.

Key products of the Traffic Surveys Unit are maps that show traffic counts affixed to roads. These maps are developed from the processed data generated by the unit. They provide a graphical representation of traffic levels on roads throughout the state. These maps are sent to the GIS Unit's Mapping Section where they are digitized into a CAD package (Microstation).

#### *Traffic Forecasting Unit*

The Traffic Forecasting Unit is concerned with predicting future traffic volumes. It also predicts how the behavioral

trends of the population will impact the transportation system. Information related to future traffic volumes, turning movements, etc., are passed on to other business units throughout NCDOT. Traffic Forecasting uses the data collected and processed by the Traffic Surveys Unit to help accomplish its objectives. The unit has four basic objectives:

1. The development of future traffic volume estimates for possible road projects
2. Timely delivery of travel forecasts to different departments and units
3. Assessment of present and anticipated traffic and land development needs for the state
4. Ensuring that the transportation networks are adequate for future traffic and growth

The Traffic Forecasting Unit produces data that could be included in the unified database [e.g., future average daily traffic data, future truck percentages, and future turning movements]. Yet a word of caution must accompany this information. All traffic projections are highly subjective and are based on assumptions regarding the future configuration of the highway network. Their inclusion in a database that is widely available must, therefore, be carefully considered.

#### *Traffic Planning Unit*

The Traffic Planning Unit is responsible for conducting systems planning. This involves developing long-range thoroughfare studies for cities and towns. The primary purpose of these studies is to develop maps showing projected average daily traffic for the highway network surrounding a community. These maps are then used as a tool to assess future transportation needs. Although the Traffic Planning Unit does not generate any data to be included in the unified database, the unit will be a major data user. It should be noted that the Traffic Planning Unit is closely related to the Traffic Forecasting Unit. The reports generated by the Traffic Planning Unit in its thoroughfare studies utilize and include projections calculated by Traffic Forecasting. They also go one step further to include socioeconomic data such as the number of travel productions (residences) and attractions (businesses) within the limits of the studies.

#### **Traffic Safety Systems Management Unit (TSSMU)**

The TSSMU of the Traffic Engineering Branch is responsible for planning, designing, developing, implementing, and evaluating appropriate engineering response strategies that deal with traffic safety concerns throughout the state. TSSMU provides its information, expertise, and recommendations to other units within NCDOT as well as the Federal Highway Administration (FHWA), municipalities, and citizens of the state. TSSMU is also the primary stakeholder from the Traffic Engineering Branch that is involved in the NCDOT unified data model development effort.

The major programs or applications that are managed under the auspices of the TSSMU are listed below. Each major program can be broken into a specific set of functions that are carried out by unit personnel. The majority of these functions involve the collection and analysis of data. Those data help the TSSMU identify areas of concern and allow them to make informed decisions regarding the implementation of safety improvement programs.

- Highway Safety Management: This area of TSSMU is concerned with two major functions: (1) Safety policy and procedures; and (2) safety systems control, research, and development. The group works to achieve these functions

by investigating and evaluating new and emerging technologies that might improve public safety.

- Highway Safety Planning and Analysis: The Highway Safety Planning and Analysis Section is responsible for (1) identifying safety needs and analyzing potential enhancements; (2) providing safety and analysis information for feasibility studies, for thoroughfare plan studies, for planning and environmental studies, for highway design reports, and for bicycle and pedestrian safety reports; (3) monitoring all projects, highway systems, and special systems; and (4) identifying the safety needs of special user groups such as older drivers, pedestrians, bicyclists, motorcyclists, commercial motor carriers, and hazardous material carriers, in the planning, design, construction, and operation of highway systems.
- Highway Safety Improvement Program: The Highway Safety Improvement Program Section is concerned with identifying potentially hazardous highway locations throughout the state's highway system. The program examines crashes at several location types, including intersections, interchanges, and bridges. It also examines different kinds of crash situations, including pedestrians, wet pavement conditions, and nighttime crashes. The section's major responsibilities include (1) developing warranting criteria to identify hazardous locations; (2) analyzing potentially hazardous locations; (3) coordinating efforts with field personnel in establishing countermeasures and setting priorities; (4) monitoring and evaluating selected countermeasures; and (5) evaluating the overall program.
- Railroad Safety Management: The primary function of the Railroad-Highway Grade Crossing Safety Section is to identify and investigate potentially hazardous railway-highway grade crossings from among the 5,000 public at-grade crossings within the state. Some of the most important activities performed by the section are (1) authorizing and inspecting installation of active warning devices; (2) collection and maintenance of grade crossing data and response to external information requests; (3) monitoring and coordinating railroad/highway construction; and (4) performing final inspections of railroad installation.

#### **Pavement Management Unit (PMU)**

A coordinated effort by NCDOT is necessary to maintain the quality of the roadways through continuous rehabilitation, construction, and maintenance activities. The PMU is responsible for both the design and the maintenance of North Carolina's roadway pavements (*Pavement* 1998). Part of the PMU's mission is to ensure that the state's roadway pavements achieve a reasonable level of safety and performance at a reasonable cost. Effective and timely pavement rehabilitation and maintenance is important for the highway system to safely and adequately meet increasing traffic demands.

There are several major applications that are performed by the PMU. These applications include pavement design, pavement analysis, data collection and training, and the pavement management system. These applications are described in the following:

- Pavement Design: The Pavement Design Group's primary responsibility is to design pavement structures for any NCDOT Transportation Improvement Program that involves the construction of a new roadway or the expansion of an existing roadway.
- Pavement Analysis: The Pavement Analysis Group's primary function is selecting treatment types for existing pavement rehabilitation.
- Data Collection and Training: The Data Collection and

Training Group's primary activity is collecting pavement data and providing divisions with assistance and training related to the latest maintenance techniques, materials, and products. Data collected by this group are currently being integrated into the Pavement Management System (PMS) to provide a better evaluation method of the condition of the state's road network.

- **PMS:** The PMS Group is responsible for implementing and maintaining a PMS for the state that meets the FHWA's mandate for an operational PMS. This group's primary function is to analyze information provided to it from a pavement condition survey conducted by the Data Collection and Training Group (*Pavement* 1998).

## LRS

LRSs are used to manage and promote the analysis of spatial data elements on transportation networks. An LRS may incorporate one or more linear referencing methods (LRMs) designed to support a series of office, field management, and integration procedures. An LRS includes "procedures that relate all locations to each other . . . it includes techniques for storing, maintaining, and retrieving location information" position (Geo Decisions 1997; Vonderohe et al. 1997, 1998). An LRM is a way to identify a single location, that is, to reference a single position with respect to a known point (Geo Decisions 1997; Vonderohe et al. 1997, 1998).

There are several types of LRMs employed by NCDOT and other transportation agencies. These include Route-Milepost, Link-Note, X-Y Coordinate, Address methods, etc. [The reader is referred to Geo Decisions (1997) for a complete definition of these methods as well as for a detailed description of the LRS.] In the NCDOT approach, a base LRM is used to relate all other LRMs to each other. A base LRM is an attempt to provide a common referencing platform in which may different data types can be represented spatially in a network structure in a standardized manner.

The base LRM is intended to be the spatial backbone of the relational database and the GIS. That is, it is the sole means to locate features that are a part of, or are located along, a particular road. At the same time, it supports translation among various other methods, thus allowing different stakeholders to access data using methods and procedures familiar to their personnel while at the same time providing a fixed underlying scheme for representing the topological and geometric aspects of the roadway network. The use of a base LRM also allows DOT units to incorporate additional databases into the organizationwide information framework at some future time without changing their existing spatial referencing systems. It is, therefore, highly expandable.

## INFORMATION MANAGEMENT ISSUES AND NEEDS

The information management issues and needs that emerged from the NCDOT experience will first be addressed in the context of the aforementioned stakeholders. A brief summary of the analysis of each stakeholder is provided and discussed. This is important to show the real issues and needs within a representative state DOT. Other issues and needs will be introduced in a later section that will further generalize from the specific NCDOT stakeholder issues and needs to more broadly address transportation information management system design considerations.

### Traffic Surveys, Forecasting, and Planning Units

The Traffic Surveys Unit has indicated that the current data management system is cumbersome and difficult to update. There are over five different data systems that are contained

in over five different database types (e.g., access on personal computers and flat files on the mainframe). Data downloaded from the mainframe and data acquired from other units' databases have led to replication of data and parallel databases. Often, data are collected in the field, transferred to digital format, printed in a paper report, and then reentered in a slightly different digital format. All of this results in outdated data that can no longer be updated and maintained properly. There is no mechanism or procedure for making orderly and consistent changes to these multiple copies of the databases.

Obviously, this is not the most efficient means of data management. Therefore, there is a need for revamping the data management process. To do this, software must be standardized by selecting DBMs that can be integrated with one another. Data processing must be streamlined so that needless efforts are not expended. Data queries that are both flexible and comprehensive must be possible so that specific information can be obtained when needed and in a proper format.

## TSSMU

The current NCDOT Crash Reporting System (CRS) resides on an IBM ES9000 computer at State Information Processing Services, an entirely separate state department. Monthly reports are generated that provide users with hard copies of the information they require for analysis. The data contained in these reports are often found to be in error, and corrections do not necessarily make their way back to the CRS. The process for making changes to the system and recreating previously extracted files is both difficult and time-consuming. Because of this, users often rely on extracted data that may have become out-of-date, which leads to different user groups relying on different versions of the same data.

Once a month the crash reports file is merged with mileposting information to link the crash data with the highway network. At the end of the year several standard reports are produced for the North Carolina Department of Motor Vehicles including the book *Traffic Crash Facts* (Traffic 1997). The Traffic Engineering Unit is provided with a file copy of the crash data. This file copy is then merged with roadway characteristics and traffic volume data from the division of highways. The resulting file is available for query, on a limited basis, using the statistical software package SAS. However, this procedure is cumbersome and leads to synchronization problems with the data. As a result, NCDOT is developing an entirely new CRS (Keane 1995).

Each of the stakeholders of the new CRS has its own needs, but the overall goal of each party is to improve the safety of the highways and reduce the number of traffic accidents. Some of these needs may be beyond the scope of the system but are still important considerations in the overall schema. The needs can be divided into two categories, the needs of the data providers and the needs of the data users. The needs of the data providers are

- Better data collection tools in the form of new hardware and software
- Ease of use
- Increased reporting efficiency
- Reduced errors during data collection
- Flexibility to adapt to changing requirements
- Integration with local efforts
- Electronic data transfer

The needs of the data users include

- Open access to data using standard relational database tools and techniques
- Improved quality and timeliness of data

- Integration with LRS for crash location identification
- Integration with GIS tools for expanded analysis
- Integration with the expanded data requirements of the CRS

Users may be interested in data that are organized in different ways. Some users may want access to specific crash reports, whereas others may want summary statistics from the data. Other users may want to integrate the crash data with roadway characteristics at the location of the accident. The system needs to be designed considering these needs.

## PMU

The PMU has developed a series of tables that are contained in a local ORACLE database. The PMU has developed its own spatial topology using the link-node method. Although the proposed NCDOT database has been designed using a form of the route-milepost method, the PMU would like to maintain its link-node method. For them, nodes are critical points that identify intersections, county boundaries, bridges, and certain culverts. The PMU feels this system of spatial management would allow for easier data collection and management for their purposes.

As a result of the PMU's desire to utilize a link-node method in the future, it would normally require a conversion program that would allow data contained in databases that utilize the base LRM to be interchangeable with data stored in the PMU link-node system. However, with a base LRM, the PMU database can simply be expanded to include the base LRM ID and mileposts, thus supporting full access to the PMU data via either the link-node method or base LRMs.

## Ownership and Responsibility

A key emerging concern regarding data and information in state DOTs is that of ownership and responsibility. Each unit's daily operation depends, in part, on the data it generates, uses, shares, and maintains. Therefore the security and accuracy of that data are of paramount concern to the "owning" unit. This organization's impact must thus be accounted for in any unified information management system design and development effort.

The recommended information system design encompasses a conceptually centralized database but with distributed ownership and responsibility. That is, the nature of the relational database model is such that free exchange of relational database tables is easily achieved. This assumes, of course, a common relational database tool and the adoption of a common base LRM. Given that these two requirements are met, data exchange will be immensely facilitated as a result of adopting the unified database design. This may be the one most significant result of this investigation—global access, yet local ownership and responsibility. Furthermore, a suite of commercial database software products running on servers provides the functionality to implement such a design.

## Access and Availability

At the same time, it must be recognized that not everyone needs access to all of the data and that universal access need not necessarily be provided. The nature of the majority of the data is such that a single organizational unit generally works with its data extensively and holds significant organizational responsibility for it. Usually the needs of others for some of that same data are less immediate and less frequent. Some data may even be proprietary or protected from access by others outside the organizational unit. Thus, even though the distributed information system is universally accessible, there are

ways to limit access to certain data, particularly in a client/server environment.

## INFORMATION MANAGEMENT SYSTEM DESIGN CONSIDERATIONS

The items enumerated below represent some of the most critical issues and needs facing state DOTs related to enterprisewide data and information representation, management, and use (Hall 1991, Wang and Elliott 1995, Wang et al. 1998). It should be noted that no attempt is being made to suggest new fundamental information system frameworks or approaches. Rather we seek to identify those information system issues that are present in transportation data that significantly complicate the information system design, development, management, and use process. Upon identification and study, appropriate ways to deal with these issues in system design emerge. It should be noted that the spatial component of data certainly is the unique and critical integrator for transportation. At the same time, special information management design considerations also emerge from the nature of state DOTs themselves and are influenced by both their size and historical development as well as the size and historical development of their databases.

### Spatial Data

The key to data resource integration is the fundamental way in which information is referenced to the physical world spatial data. That is, data are joined together by a geographic reference of some sort. Residences, for example, are located using a street address; highways are located using posted route numbers; and items along a roadway may be located by specifying a distance from some landmark or otherwise recognizable starting point. The common thread among all these references is location. Yet location is described in vastly different ways, and the precision and accuracy of the description varies widely as well.

Location description must be sorted out to achieve a true integration among transportation information resources and to more fully and effectively use advanced information technology tools to increase institutional productivity and effectiveness in managing the transportation infrastructure. Location data must be accurately and universally represented in the spatial component of the framework of the database. Only then can true integration be achieved. Doing so will bear significant fruits in terms of facilitating the use of well understood data relationships and the identification of previously unknown (or simply unavailable) relationship, both between data and between DOT organizational units.

New interest is generated in the data of other units when those data are spatially referenced in such a way that one can spatially access the data. Thus, there is the possibility of having new data users as well as new data uses. Consider, for example, the common need to issue permits for routing over-size or overweight vehicles. This activity requires information about pavement design, bridge design, load ratings, lateral clearances, environmental sensitivity, etc. Clearly such data must be obtained from a diverse collection of possibly distributed organizational units and their databases.

Consider also a second example that illustrates the need to incorporate new data. Two new applications that have previously not been broadly supported (but are clearly useful) are worth mentioning. These are billboard and sign inventories and locations. Both have a variety of useful attribute data to be stored, and both need to utilize the common spatial representation shared by other applications. In doing so, interesting new relationships between signs and billboards, on one hand, and any other attribute(s) of the roadway network, on the other

hand, can be determined. We may thus, for example, investigate the relationship between sign locations and traffic accident locations to study possible correlations. Without a common and standardized spatial representation such linkages cannot be made without manual intervention, if at all. Thus, the criticality of spatial data and accurate representation of those data cannot be overemphasized.

### Existing Application Programs

It is essential to identify the scope, nature, content, field conditions, and use of the required input and output of the existing software application programs currently in use to effectively support their data needs. All data items that are required to support these application programs must be included in the database schema. This involves a comprehensive analysis of all of the uses (applications) of data. It involves a complete understanding of the technical business conducted by each unit and mandates that such business be supported. Additionally, consideration must be given to new applications being planned, designed, or developed.

### Data Migration

Consideration must be given to the migration of existing databases and data files to the newly structured and conceptually unified database. For each identified database and data item therein, a new location needs to be specified and illustrated. That is, it must be shown how and where each stakeholder's data are stored, and how the stakeholder gains access to the data. This alludes to the complexity of the whole information system problem. Not only must a new schema be investigated, designed, and developed, but a major effort of transforming all legacy data resources into the new schema must occur. The magnitude of this effort can be enormous.

### Future Data Needs

There should be communication with each of the stakeholders to identify new data item needs that can appropriately be incorporated into the overall database design. Thus, not only should the new database design take into consideration existing applications, but it should also take into consideration projected data needs to support new applications. At the same time, future data needs are often anticipated rather than concrete. The database design, therefore, should consider them, but should be grounded in and meet the data needs of current users.

### New Data Acquisition and Reporting

Appropriate consideration must also be given to the overall requisite data acquisition and reporting functions so that the resulting design efficiently supports those functions. Data acquisition must account for the emergence of new technologies [e.g., global positioning system (GPS)], which may alter the locational precision of the data and which might be significantly influenced by field methods, procedures, and training requirements. Furthermore, the FHWA Highway Performance Monitoring System (HPMS) reporting requirements must be met via the specified database design.

### Data Maintenance and Management Tools

A central and essential activity is the ongoing maintenance of the database. This activity must be considered so that the resulting relational database design efficiently supports its functions. A consistent set of procedures and methods must be developed to support this activity. Along with these organizational implementation tools must come a user-friendly in-

terface that supports efficient data manipulation and the location and identification of missing and incorrect data (Sadek et al. 1996).

There needs to be a clear identification of management responsibility for different data items. For example, the data representing the base roadway spatial network, which is used organizationwide, have such an impact that great care must be taken to ensure integrity and utility of these data to all users. Therefore, a clear assignment of the responsibility for this aspect of the database must be made and made known throughout the organization.

### Regeneration of Existing Databases

Existing applications are tightly coupled to existing data files. During the actual development of the new database, and for some time after its implementation, the application programs will require access to data in its original format. This need will persist until those application interfaces are rewritten to accommodate the new data format. Until that time, it will be necessary to regenerate the existing data files from the new database schema so that existing applications can continue to run. This is essentially a transition cost. That is, it would not need to be borne if a new system were being completely developed from scratch but is essential for a system being transformed.

The most obvious example of this situation is the HPMS report. Historically, at NCDOT, this report has been generated using the universe file and a mainframe application. If no other changes took place, we would have to be able to reconstitute the current universe file from the new universe database so that the existing HPMS application could run off of that file and generate the HPMS reporting data as has always been done in the past.

However, under development is a computer program to generate this report from an Oracle universe "table" instead of from the universe file. This is a single multiattribute Oracle relational database table with exactly the same columns and content of the universe "file." Thus, the new universe database schema will have to be able to reconstitute this universe relational table so that future HPMS reporting requirements can be met. At some time in the future, the HPMS reporting application can be rewritten to run directly from the new universe database schema. Until that time we will need to be able to "regenerate existing databases."

### Data Dictionary

As part of the review of existing databases, applications, and files and new data needs, a comprehensive data dictionary must be compiled. This data dictionary should include at least three major categorizations of data: spatial, tabular, and temporal data.

Spatial data provide information about topology and geometry. Topology describes the connectivity between components or elements of the system. For example, a section of roadway may be said to be bounded by two intersections. Geometry describes the precise location in space of a component or element. For example, an intersection may have an NEZ (Northing, Easting, Elevation) state plane grid coordinate location.

Tabular data describe the individual attributes or characteristics of a component or element. A roadway will have some width and number of lanes, for example. An intersection will have attributes to describe the signaling. Finally, temporal data describe, in some way, the relationship that components or elements have with time.

The data dictionary should provide complete definitions for each data item; identify both the organizational unit(s) and

application(s) using (or being responsible for) that data item, data type, data domain and range; and provide either typical or comprehensive value sets for the data item. The data dictionary should be organized and categorized into groupings related to needs of stakeholder and applications for locally used data. It should support data dictionary analysis and standard queries.

### Data Dictionary Analysis

One purpose of a critical analysis of all data items is to identify those data items that could be omitted from the new database. A second purpose is to resolve conflicts, formalize naming conventions, and standardize definitions, thus obtaining a more concise, robust set of data items to meet all previously determined needs. A third purpose is to appropriately assign responsibility for, and ownership of, data (i.e., to ensure that the data is properly positioned in the organization).

Three conditions contribute to the need for a critical analysis of the new comprehensive data dictionary. First, it is anticipated that over time a mismatch has occurred between the supply of data on one hand and the anticipated program demand for, or use of data, on the other hand. It is expected that some existing data items may no longer be needed. These items should be identified and removed. Maintaining unnecessary data is extremely costly and is to be avoided.

Second, it is known that secondary data items have been created from combinations of other more fundamental (or base) data, thus proliferating a degree of redundancy of existing data. Furthermore, different ways of referring to and using data also result in redundant data items. These data items should be identified and removed if not needed in recreating data sets in previous formats.

Finally, it is known that various organizational units maintain data for other units unnecessarily. That is, some units could assume ownership of, and responsibility for, data items within their domain or sphere of activities. Regardless of the original reasons for data maintenance and management by any unit, all items should be examined to determine their rightful or most logical home. For example, in North Carolina the GIS Unit maintains data that are of primary interest to the PMU (number of travel lanes, surface width, etc.), the Traffic Surveys Unit (design hour speed, PTC count station, etc.), and the Forecasting Unit (future annual average daily traffic, future annual average daily traffic year). Both the PMU and the Traffic Surveys Unit possess the expertise, knowledge, interest, and ability to assume control of these data items. The Traffic Forecasting Unit, however, may not currently possess all of the requisite resources to maintain their own data and, therefore, it may be more appropriately maintained by the Roadway Inventory Unit.

### GIS Interface

The unified relational database should support direct and seamless integration with the GIS so that data entry, searching, and queries can be issued spatially via the GIS directly to the relational database. That is, the database should support traditional GIS spatial queries and analysis in addition to the broad range of data analysis capabilities normally provided by a relational DBMS.

With the ready availability of ever more powerful GIS software it is critical that DOTs make use of it to perform complex spatial analysis. However, to do so requires its availability. In a carefully designed transportation information system that database will provide it. But for the data to be effectively used requires a seamless coupling with the GIS. The DBMS and GIS must work hand in hand. Each must be used when its particular strengths are required. Yet to do so requires a seamless interface between them.

### GPS

GPS can be used in one of two ways: (1) To determine the location of a given object; or (2) to find an object given a location. The former is called measurement (used in surveying), and the latter is called navigation (used in ships, cars, and cargo trucks). Today, more and more organizations are using GPS for collecting location data. As such, GPS must be accounted for in the database design.

Posted routes and mileposts are most often used to locate linear attributes. That is, they determine location by measurement along a line from a starting point regardless of the path that the line takes. GPS coordinates, on the other hand, are not mileposts or linear measures. Rather, GPS determines location as a precise position in space and its measurement system is a coordinate system. However, GPS can be incorporated into a linear database by providing accurate coordinate positions for nodal entities such as intersections and country boundaries. These coordinate positions can be converted to mileposts through the application of filter programs. Or, they can simply be stored as coordinates (as attributes in the database) and programs can be written to use them as needed by a particular organizational unit.

One possible organizational dilemma is that GPS can be more accurate than digitized GIS maps or than the spatial database representation of the highway system maintained by the state DOT. This could result in the incompatibility of location data. That is, an existing point, as stored in GIS, may be positioned differently than the GPS coordinates of that point. Put another way, the actual positional data (that was field measured with highly accurate instruments and stored in the database) can be more accurate than the positional data of the line-work that was digitized into the CAD or GIS systems. For example, consider the GIS lines representing our most accurate positioning of a highway. A highly accurate GPS point measurement of the highway centerline might yield a location that does not even land closely on the GIS line. If not, how is the GIS line-work adjusted? This is a future issue that must be resolved as GPS and other positional measurement instruments become more widely used.

### Positional Accuracy

Current GISs and DBMSs do not directly provide information on the quality of the data that they present to the user. This shortcoming is becoming increasingly recognized as critical. Leica Geosystems AG, for example, is working with the Environmental Systems Research Institute to design core GIS technology, including data structures and software tools to operate on those data, to service the unique needs of survey data and land record data. Such efforts are needed to enable DOTs to bring to bear the full power of DBMSs and GISs on transportation data.

The previous discussion regarding GPS highlights the issue of positional accuracy and precision. Accuracy is the relationship between the value of a measurement and the "true" value of the dimension being measured, that is, the correctness of the result. Precision describes the degree of refinement with which the measurement is made, for example, to additional decimal places. On one level, we are concerned with the accuracy and precision of reported data. First, was the GPS receiver accurately positioned (accuracy), and, second, what is the range of variability in its reading (precision)?

It is critical to know, for each field data gathering method or procedure, the degree of accuracy that can be obtained by those using the equipment. It is also critical to know the precision of the instrument, procedure, or method. This information needs to be known and used in downstream applications. It must be reported to users as well. Otherwise, users

tend to assume that the value given is the actual value. But if the value is actually 10 ft from where the number tells the user it is, the user needs to know and account for that.

Positional data are a key ingredient in transportation information management systems. Factors that contribute to the existence of positional data errors include the sources of positional data; the nature and volume of these data; the methods for acquiring, inputting, and modeling them; the processes they can undergo; and the ways in which they can be presented. In turn, these errors affect the reliability of the information derived in and by applications. Tools must be made available to enable the user to fully understand the accuracy and precision of positional data (Goodchild 1998) as it is obtained from the field.

Finally, the accuracy needs of the user must be considered so that gross mismatches between what a user receives and what a user needs in the way of positional accuracy do not occur. Related to this are the differing accuracy requirements between organizational units, which can also result in mismatches. Traffic Survey Units, for example, require less precision and accuracy than PMUs. Accident location reporting requires a high degree of accuracy and precision. This is an issue that must be carefully studied and documented to ensure that information system design reflects accurately the needs of the organizations.

### Spatial Database Interface

Any linear spatial data access or capabilities currently in use by the stakeholders must be able to be performed directly on the relational database itself (through the use of the DBMS) in addition to being able to be performed by the GIS. These include, in particular, linear spatial search and spatial queries via the DBMS. That is, when GIS is not available, spatial analysis should still be able to be performed using the DBMS alone.

Linear queries using a GIS are well understood. Issuing linear spatial queries to traditional transportation databases, however, is a rather novel concept. A multitude of legacy databases presently exist. Enabling those databases to support spatial queries significantly enhances their value, thus significantly enhancing the utility of data already in place at most DOTs. Thus, a limited spatial analysis capability can be achieved without the wide distribution of costly GIS software. The value of existing data is significantly enhanced.

### Temporal Data

Time comes into play in many situations and must be considered and dealt with in the information management system design. One of the unique aspects of a state DOT database design is that the physical infrastructure can change over time; new roadways can be built, and others can be abandoned or moved. These infrastructure changes are time-critical spatial network changes, and they impart unique situations on the database design.

Consider accident locations, for example. Accident data may have been collected for many years at a particular intersection. If that intersection was later relocated, problems can arise in interpreting accident location in the vicinity or region. A discontinuity in time and location can have an adverse effect on historical analysis of data that has fundamentally changed.

Time thus influences the interpretation of historical data. A change in roadway width from two to four lanes is a significant change in the physical infrastructure. When that change occurs it is important because it influences how we interpret other data. Traffic survey counts, for example, would significantly change in such a scenario. But if there is no way to establish a correlation between the roadway widening ac-

tivity and the counts, this change in count data could cause concern. Temporal data, therefore, are a consideration in transportation database design.

### Data Exchange

A complete understanding of the scope and extent of data sharing is an important consideration in database design as well. Historically, many state DOT organizational units have operated in isolation, and data sharing was at a minimum. However, the proliferation of standardized software tools, the need to access data owned or maintained by others, and external demands for data all mandate that more and more will be shared.

Individual users within an organizational unit are operating in a much more automated office working environment. This environment itself promotes the desire to enhance data sharing within the unit. Increasingly complex design situations are mandating cross unit data sharing. Relationships between state and federal agencies mandate data sharing; the most prominent example being the HPMS data reporting requirements. A critical need may also exist, for example, between a state highway department and its division of motor vehicles that would significantly benefit the ability to readily share and analyze data.

Finally, the general public is developing a larger and larger appetite for a variety of data. As transportation significantly impacts our environment and other aspects of our infrastructure, increasing demands will be made to share transportation data. Consideration thus must be given in the transportation information management system design as to how much sharing will occur, with whom, how often, and in what form. The delivery vehicle(s) for that sharing must be considered as well.

### Expanding the Organizational Scope

How is a DOT to proceed with an information system development effort? The successful experience at NCDOT began with a core stakeholder group. That group was of sufficient size to involve enough participants to get the work done and encompassed a wide enough variety of data to support a generalized and expandable solution. Other new organizational units can then be incrementally included in such an information system development effort if they adopt the methods and procedures outlined in the original information system design.

### Training

Because of their large size, training is an issue to be seriously considered by state DOTs. The adoption of any new methods, procedures, or software requires training consideration. The adoption of a relational database format, for example, will necessitate some employee training in this method. The adoption of a specific relational database tool will also necessitate employee training to use that software. The amount of this training will largely be determined by the level of standardization that can be achieved and by the organizational structure. For NCDOT we found that standardization on a base LRM, and on the relational database format, provided an ideal level of standardization that could readily be supported via training.

### SUMMARY

This paper has provided a detailed examination of many of the issues and needs in data management at state DOTs. The data structures currently used by NCDOT as well as the existing databases have been analyzed. A set of database and information system design considerations unique to NCDOT (to be used in the development of the unified database) and common to all state DOTs have been identified and discussed.



Key to the development of the unified highway database is the adoption of a base LRM and a common database format that effectively incorporates spatial data and correlates well with the physical roadway network and how it is described.

State DOTs have a unique database problem—much of the enterprisewide data is related to a linear roadway network. Thus, a common way of describing the topology and geometry of the network must be determined and incorporated into the database. Next, a set of conversion capabilities must be provided to support data access via other referencing methods than the base method. Finally, a whole series of database issues and needs emerge because of the diverse and highly distributed nature of state DOT data.

Resolution of the issues and needs identified herein should enable DOTs to meet their information needs into the next century and should more than adequately serve other users of the information. All stakeholders participating in the unified database need to adopt a standard roadway topological structure and convert their data to fit both the relational model and the proposed LRM structure. The design considerations outlined herein can be followed as a guideline for state DOT enterprisewide information system design.

## APPENDIX. REFERENCES

- Geo Decisions (a Division of Gannett Fleming, Inc.). (1997). "Linear referencing systems—Review and recommendations for GIS and database management systems." *Tech. Rep.*, North Carolina DOT, Raleigh, N.C.
- Goodchild, M. F. (1998). "Uncertainty: The Achilles heel of GIS?" *Geo Information Sys.*, November, 50–52.
- Hall, J. P. (1991). "Guidelines for the development of a strategic information process for state departments of transportation." *Transp. Res. Rec. 1305*, Transportation Research Board, National Research Council, Washington, D.C., 141–146.
- Highway traffic statistics (1994). Statewide Planning Branch, North Carolina DOT, Raleigh, N.C.
- Khan, S. I., Ritchie, G. S., and Kampe, K. (1994). "Integrated system to develop highway rehabilitation projects." *J. Transp. Engrg.*, ASCE, 120(1), 1–20.
- Keane, Inc. (1995). "A strategic plan for traffic records improvement." *Tech. Rep.*, Revision 4, November, Durham, N.C.
- Kiel, D., Rasdorf, W., Shuller, E., and Poole, R. (1999). "An LRS model to support GIS and DB analysis." *Transp. Res. Rec. 1660*, Transportation Research Board, National Research Council, Washington, D.C., 108–113.
- Pavement management unit—pavement condition survey manual. (1998). North Carolina DOT, Raleigh, N.C.
- Sadek, A. W., Freeman, T. E., and Demetsky, K. J. (1996). "Database preparation for pavement modeling—Virginia's experience." *J. Transp. Engrg.*, ASCE, 122(6), 454–461.
- Sherk, S. A., and Rasdorf, W. J. (1998). "Unified database schema for North Carolina Department of Transportation." *Tech. Rep.*, North Carolina DOT, Raleigh, N.C.
- Sinha, K. C., Cohn, L. F., Hendrickson, C. T., and Stephanedes, Y. (1988). "Role of advanced technologies in transportation engineering." *J. Transp. Engrg.*, ASCE, 114(4), 383–392.
- Traffic crash facts. (1997). Dept. of Motor Vehicles, North Carolina DOT, Raleigh, N.C.
- Vonderohe, A., Adams, T., Chou, C., Bacon, M., Sun, F., and Smith, R. (1998). "Development of system and application architectures for geographic information systems in transportation." *NCHRP Res. Results Dig., No. 221*, Transportation Research Board, National Research Council, Washington, D.C.
- Vonderohe, A., Chou, C., Sun, F., and Adams, T. (1997). "A generic data model for linear referencing systems." *NCHRP Res. Results Dig., No. 218*, Transportation of Research Board, National Research Council, Washington, D.C.
- Wang, K. C. P., and Elliott, R. P. (1995). "Design considerations for distributed, multimedia-based highway information system." *Transp. Res. Rec. 1505*, Transportation Research Board, National Research Council, Washington, D.C., 95–102.
- Wang, K. C. P., Li, X., and Elliott, R. P. (1998). "Generic multimedia database for highway infrastructure management." *Transp. Res. Rec. 1615*, Transportation Research Board, National Research Council, Washington, D.C., 56–64.