CHAPTER TWO

FINDINGS

In an age of powerful and sophisticated electronic systems for recording, tracking, and manipulating data to help companies and organizations realize the asset value of their internal information pools, a poorly designed system will yield only nominal results. With the appropriate expertise, a database system can be constructed that sorts, organizes, manipulates, calculates, and displays information in nearly any manner conceivable. However, even the most complex and versatile information system can only give the desired results if the following criteria are met: the information is available, the information can be entered into the system, and the available information supports the desired outcome. In accordance with these criteria, a system methodology or design approach can be formulated in one of two ways, “bottom-up” (typical database design) or “top-down.” The bottom-up approach looks to the source information or raw data to direct the potential output, which in this case would be the state’s internal information. The “top-down” approach looks at the desired output and then attempts to verify the existence of source information that supports the output.

More specifically, the bottom-up approach takes the following form: (1) identify the information to be collected (data); (2) design the organizational and storage structure (tables); (3) create the input mechanism (entry forms); (4) populate the tables with data; (5) program statistical calculations and formulas, if needed; (6) create the output mechanism (report forms); and (7) test the application. The top-down approach includes the following: (1) identify the desired outcome; (2) discover the existence, location, and media state of the source data; (3) create the organizational and storage structure to accept the information; (4) program statistical calculations and formulas, if needed; (5) create the input mechanism, dependent on the media state (paper, electronic file, etc); (6) create the output mechanism; and (7) test the application.

The most significant difference in design approach is the locus of control and whether it lies with the source data or the outcome data. The construction of the model system required that both approaches be assimilated into one system to achieve the desired outcome, but without the use of a fixed uniform data standard to which all states would report. Most states expressing a willingness to participate in a national system indicated that their willingness was predicated on not having to significantly modify internal systems or enter data and information multiple times. In addition, requirements for the model data-management system included: compatibility with multiple software and hardware platforms used by the states, capability of processing dissimilar data items supplied by the states, ability to accommodate multiple formats of exported data, and cost-effectiveness of implementation and maintenance on a national level.

PILOT STATE SURVEY RESULTS

State Solicitation Process

Each of the five states selected initially responded favorably and without reservations to the prospect of participating as a pilot state in the feasibility study. The principal concerns were potential risk exposure and time constraints on state personnel. Two other concerns expressed but not emphasized during the initial contact phase were the question of value to states' internal data needs and decision-making processes and the aversion to making changes to state information systems in order to accommodate the data needs of the model system. With each of the five states, the research team spent considerable effort to elaborate clearly on the various responsibilities and requirements of pilot state participation, discussing each state's specific concerns in detail. Subsequent deliberations within the states resulted in two of the original five states withdrawing from participation in the project. Two other states replaced the two declining states, which largely met the selection criteria. The five pilot states thus were: California, Florida, West Virginia, Missouri, and Washington.

State Survey Process

The site survey consisted of a 3-day process of meeting with representatives from each department for presentation of the project goals, objectives, and needs; conducting the individual interviews with functional managers and technical staff, and conducting analysis of each department's information system. Interviews with the functional managers provided the content currently in the legal and risk management data fields within the model systems' horizontal table structure. The expected outcome of the interviews was to identify data resources with content related to the core data elements and to obtain a record layout that defined the structure and format of the states' internal data.

Interviews with functional and technical managers revealed that the management of requisite state data files was highly decentralized. A variety of state agencies maintain essential data components with little or no overlap in content. This fact made it impossible to link data electronically from one agency's file to those of another. All of the database applications used by state agencies had the capability to export data in a universally compatible format. There was not found among the pilot states a single agency that managed a majority share of the necessary data. The dispersal of data files and the omission of overlapping data content (in the form of key fields, reference fields, and docket numbers) were obstacles that required more on-site analysis to overcome than was budgeted. Tables 2.1 through 2.3 display the status of data content and system capability obtained during the interview process.

Table 2.1 Legal Department Information

Legal Department Information	California	West Virginia	Florida	Missouri	Washington
Existence of "case status" database or flat file	Yes	No	Yes	Yes	Yes
Existence of electronic resources from which a "case status" might be derived		Yes. Currently maintained on the state's behalf by the AIG Insurance Company.	Yes	Yes	Yes
Existence of "case status" paper data sheet	Yes. Derived from database.	No	Yes. Derived from database.	No	Yes
Availability of file record layout	Yes		Yes	No	Yes
Database structure supports target content	Yes		No	No	No
Database is populated	No		Yes	No	Yes
Sample copy of electronic data in-hand	Yes		Yes	Yes	Yes
Paper representation of electronic data in-hand	Yes		Yes	Yes	Yes
Methodology for a single transfer of data (short-term scenario)	Yes		Yes	Yes	Yes
Methodology for periodic transfer of data (long-term scenario)	TBD		TBD	No	No
Expert witness compilation	Available through another resource.		Yes	Yes	No

Table 2.2 Claim/Financial/Administrative Information

Claim/Financial/Administrative Information	California	West Virginia	Florida	Missouri	Washington
Electronic file of case expense data	Legal Dept. – Some, but not enough for analysis. Account Dept. - will provide more complete data.	All such information is recorded and maintained by **AIG** on behalf of the state's risk management department.	Yes. Maintained by risk management.	Yes	Some
Existence of reserve and liability calculations in file data	Yes	Probably	Yes	Yes	Yes
Existence of settlement values in file data	Yes	Probably	Yes	Yes	Yes
Existence of resolution or judgment in file data	Yes	Probably	Yes	Yes	Yes
Existence of paper data file	Yes. Derived from database.	No. If it exists it is not readily accessible.		No	Yes
Availability of file record layout	Yes	Probably	Yes	No	No
Sample copy of record layout in-hand	Yes	No	Yes	No	No
Sample copy of electronic data in-hand	Yes	No	Yes	Yes	No
Paper representation of electronic data in-hand	Yes	No	Yes	No	No
Methodology for a single transfer of data (short-term scenario)	Yes	No	Yes	Yes	No
Methodology for periodic transfer of data (long-term scenario)	TBD	No	TBD	No	No

Table 2.3 Traffic and Highway Information

Traffic and Highway Information	California	West Virginia	Florida	Missouri	Washington
Existence of **central traffic events database** or flat file	Yes	Yes	Yes	Yes	Yes
Accessibility of file	Accessible	TBD	Accessible	Derivative accessible	Yes
Availability of file record layout	In-hand	TBD	In-hand	Yes	Yes
Sample copy of record layout in-hand	Yes	No	Yes	No	Yes
Sample copy of electronic data in-hand	Yes	No	Yes	No	Yes
Paper representation of electronic data in-hand	Yes	Yes	Yes	No	Yes
Methodology for a single transfer of data (short term scenario)	Yes	No	Yes	Yes	Yes
Methodology for periodic transfer of data (long term scenario)	TBD	No	TBD	No	Yes
Existence of **central highway environment** database or flat file	Yes	Yes	Yes	Yes	Yes
Accessibility of file	Accessible	Accessible	Accessible	Not available	Yes
Availability of file record layout	Available	Available	Available	Not available	Yes
Sample copy of record layout in-hand	Yes	Yes	Yes	No	Yes
Sample copy of electronic data in-hand	No	Yes	No	No	Yes
Paper representation of electronic data in-hand	Yes	Yes	Yes	No	Yes
Methodology for a single transfer of data (short term scenario)	Yes	Yes	Yes	No	Yes
Methodology for periodic transfer of data (long term scenario)	No	no	Yes	No	No

State Information Systems

A review of hardware and software technology in the states revealed various mainframe, mini-frame, and client-server environments with a range of operating systems and approximately 13 different databases being used for hosting and processing tort claims information. A result of the proliferation of computer-based technologies used to improve the administration of information has been a widening knowledge and communication gap between functional managers and system technicians. Though each understands his or her area of expertise, there is little or no common overlap that effectively melds the two together. Such was the case in the state agencies that were surveyed. The technicians did not fully comprehend the business process and the functional personnel had only a cursory knowledge of the technical processes. This circumstance manifested in technicians having complete access to data but limited knowledge of its content, and functional managers being aware of the existence and content of data but having no specific knowledge as to how they were generated, maintained, structured, or stored. The research team found that this knowledge gap was further exacerbated by the limited time available to conduct the on-site surveys and limitations of the personnel participating in the interview process. The intended objective of cross-referencing and verifying data labels, information sources, and functional authority was nearly impossible to achieve in the allotted time.

The database software used for the information systems in each state varied greatly in complexity, functionality, and capability. The technical expertise of the personnel responsible for managing the information systems ranged from secretarial staff with limited input/output knowledge to systems administrators with a thorough knowledge of the system. The level of expertise that was available to assist the project team in identifying the content of each system greatly impacted the success of identifying the appropriate data items for export. Systems built on standard, off-the-shelf brands posed no real problem for accessing the table structure and exporting the content. Since the applications of most of these systems were developed internally, accessing the record structure to verify content was a straightforward exercise.

In contrast, accessing the record structure of proprietary database software, such as Dorn Risk Master that is used in Missouri, involved insurmountable obstacles for this project. The Risk Master software is constructed from a relational database that uses a parent/child table structure. The software offers optimal versatility and functionality but has no means of providing a visual display or tree view of the table structure, which would have allowed the team to easily identify and verify data flow and content. A search of the child tables revealed that data fields with the same label were found in other child tables with no means of determining field duplication or reference. The software implemented an expansive export function via a proprietary reporting mechanism called Report Master. However, creating a report with the necessary data fields required expert knowledge of the field labels and content within each field. The state's central technology personnel were not familiar enough with the internal structure of the program and had to defer to the clerical person in charge of data entry and reporting. The clerical person could not retrieve a record or table structure and contacted the technical support representative for the software vendor, Dorn. The vendor representative provided assistance for creating a report from the selected data fields, but the software licensing agreement protected access to the internal structure of the program. This made identifying the specific content of data fields related to tort claims information a hit-or-miss proposition based on foreknowledge of the data field labels and data structure set forth by the department. The ability to identify and verify the exact data fields without a full awareness of the department's data labeling and reference terminology and cooperation from the software vendor would require a timely process of on-site analysis. The same issues would probably apply when dealing with any private, proprietary software programs.

A significant issue expressed in the Interim Report impacting the design approach for constructing the model data-management system was that states did not want to change their internal information systems. However, of the four states using computer-based information systems, one or more of the departments interviewed in three of states reported that they had either recently converted, were currently involved in a conversion, or were planning to convert to a newer version or completely different software program. The majority were inclined toward adopting a completely different software program that offered integration and scalability to include access by multiple departments statewide.

States' Relevant Data Structures

The states' data objects required to populate the data-management system consisted of select information from tort claims records, risk management records, highway deficiency data, expert witness information, and injury/accident statistics. With the exception of Florida and California, legal agencies did not maintain accident, injury, highway, or judicial statistics as a part of their case file. These agencies generally managed case information in word processing documents rather than databases. The rest of the case (legal) statistics were available from risk management in all states except West Virginia.

Accident, driver, injury, and highway statistics were available from a variety of bureaus within transportation agencies. However, none of them maintained a reference field that linked accident records with legal records, or risk management records, or highway event (highway maintenance) records. Through the course of the investigation, it was discovered that linked information was available through the department of motor vehicles or the states' electronic information offices. Access to relevant files in these resources would have required prior knowledge of their existence. Unfortunately, such information was not available prior to the scheduling of interviews.

The research team was able to collect record layouts of existing data files when available, and personally inspect the files of target agencies to verify the applicability of data content for this project. The core data elements that refer to policy issues required multiple-choice or short-answer responses. Statistics required the compilation of values from legal files and claim files. The content of database files in the legal and risk management agencies visited were function specific. They did not maintain information from which the target data could be derived. It was obvious that managers could not have relied on those files exclusively to generate the statistical data requested in the AASHTO study. None of the states would have been able to complete the AASHTO survey by exclusive use of the data files that were made accessible to the interview team.

Table 2.4 provides a complete breakdown of core data elements for the model system by category and the availability of supporting data from within the state's internal information systems. A closer look at the data revealed that a more complete picture of the highway safety and condition environment in support of the core data elements might be compiled by including data from actual accident reports and information retrieved from traffic operations departments. However, this option could not be pursued within the time and monetary constraints of this project.

Table 2.4 Distributions of Data Elements and Target Data

Subject Category	Total Responses	Short Answer	Statistics	Response derived from accessible legal / claim files
Sovereign Immunity	9	9		0
Claims Procedures	7	7		0
Claim Statistics	24		24	4
Attorney Statistics	11		11	0
Employment Liability	4		4	NA
Contractor Indemnification	4	4		NA
Insurance	14	14		NA
Training Policy	25	25		NA
Risk Management	2	2		NA
Expert Witness	6	6		0
Totals	**106**	**67**	**39**	
Highway Characteristics	320		320	
Injury Characteristics	22		22	

Impact of States' Data Structures on Model System Development

The process of data analysis began with the first set of exported data from the state systems and implementing a mapping or translation process to link states' data items to the associated data elements in the model system. The mapping process applies a common structure to the cumulative data by identifying and routing the information to the appropriate data fields within the model system's data tables, where statistical functions would calculate the aggregate results for display in the reporting mechanisms. A traditional database design would have made this process an effective and efficient method of collecting and processing the states' information; however, the magnitude of calculated data variables in claims and highway data records, along with the software development issues, hindered the use of a traditional horizontal table structure and required an alternate design approach. A vertical table structure was developed that would allow the aggregate totals for each of the core data elements to be displayed to the user community in a familiar format. A utility program within the application collects data from reference data tables and the data table that supports the entry form, performs all necessary calculations, and performs formatting that is required to prepare the data in the horizontal tables for conversion to the vertical tables. Displaying the results of this process could not be completed for two reasons: (1) a majority of the data fields retrieved from the state's internal systems did not support the core data elements, and (2) data fields that did appear to support the core data elements were not populated sufficiently to verify the content.

There were a few additional issues that hindered the use of raw data from the states to calculate the aggregate totals required by the central database. Maintenance of data required to calculate aggregate totals was distributed among many more departments than the initial survey responses indicated. The time lapses for data entry on a case and the actual progress for a case are very broad in some states. The reasons for this issue varied from state to state, but all states experienced some form of this problem, such as delays in receiving information from their business or accounting departments to provide certain cost figures, awaiting information from subordinate database systems that had not been updated, and failure to obtain the information available for input. There also appears to be some duplication of information between departments within a state that would make the totals unreliable.

The most significant issue that surfaced was that existing state legal and risk management information systems contain too many data fields that are not populated with data. This situation severely constrained the ability to construct a model data-management system. Further hampering the model system's development were the difficulties encountered in attempting to evaluate the data content of the existing state systems. As indicated previously, the data field references/labels used to identify data items during the site survey were often not enough to make an accurate evaluation of the data content of these systems. Consequently, the data fields of existing systems could not be accurately mapped into a common coding structure for this project. It became apparent that since the field identifiers used in each state are often unique to their own organizational, cultural, or legal environment, a more significant amount of time will be required at each of the states and with state personnel that are proficient in all aspects of the existing state database systems.

Given the general state of existing legal and risk management electronic data, attempts were made to find supplemental state and federal transportation information sources to use in the model data management system. It was determined that highway engineering and technical data (accident files, highway maintenance files, highway characteristics files) and injury characteristics data maintained by all of the states are comprehensive and available in database structures. However, these databases are very large, with complex record layouts (including a description of values) that require much more time to process than the budget allowed. For these reasons, the core horizontal tables remain populated with data retrieved from legal and risk management departments during the interview process and readily available highway and fatality statistics retrieved from the Bureau of Transportation Statistics (BTS) Fatal Accident Reporting System.

As previously indicated, the states have a keen interest in including alleged highway deficiencies within the data-management system. A four-level classification and coding system for alleged deficiencies was developed and proposed for the model system. The data matrix alone for Levels I, II, and III is approximately 360 discrete fields, which is not unreasonable for mid-level analysis. However, if Level IV were to be included, the matrix would expand to 20,544 discrete combinations. Construction of a presentation scheme to deliver a quality search function for dimensions of this magnitude would require a research initiative of its own, as evidenced by the Public Risk Database Project. The project team decided that a database constructed from Levels I, II, and III would be sufficient to demonstrate the value and function of this information. The decision was supported by the findings in the Phase I report, which noted that most states could not provide this level of detail for a national perspective due to varied geographical and climate features, unique organizational structures, and diverse tort laws.

The data entry and reporting functions for the highway deficiency database pose the most challenge of all the components in the data-management system due to the linking process between the levels. The design and construction would be simple for “just” a mutually exclusive perspective or “just” a relational perspective, but the design and construction of a system that offers a mutually exclusive perspective “and” a relational perspective within existing time and budget constraints was beyond the scope of the project.

MODEL SYSTEM DESIGN AND CONSTRUCTION

Database Application Environment

Two principal factors dictated the design and development of the central information system—the data structures and software programs used to create the model system. The type (state’s data or aggregate) of data and output needs dictate the design of a table. The horizontal table is a common data matrix structure that allows for a highly versatile search capability. The vertical table structure has a more fixed data structure that has less capability for searching the data content of the table. The traditional horizontal layout and the less common vertical layout dramatically dictate the way information can be displayed and manipulated.

The central database environment for this project is comprised of both horizontal and vertical tables due to the incongruence of the data types. The horizontal tables contain the raw data records being retrieved from the state’s information systems and the vertical tables contain the aggregate information being collected via the entry forms. The construction of this environment had to be altered dramatically with the discovery of system design flaws due to the misinterpretation of the database program’s integration capability with the web application software. A key component to the system that ties the state’s data to the core data elements is the translation program. The translation program converts states’ data structures to meet the requirements of compiling the output dictated by the core data elements and back to the states' data structures.

The shared access to a database within a local area network and shared access to a database within a wide area network that utilizes the World Wide Web for connecting computers pose fundamental differences. The most significant difference between them is the client capability or functionality. A shared database application in a local area network splits the responsibility of data processing between the client and server, with the majority of processing work assigned to the server. A shared database application via the World Wide Web relies completely on the server for all data processing needs and the client is programmed to merely transmit requests and receive the output of the request. Therefore, a significant amount of programming is required to get the functionality of a locally shared database within a remotely shared environment. The Microsoft Visual Studio Suite was chosen with the understanding that the functionality of a locally shared database could be achieved without the extensive programming requirements. This was not the case and the entire design had to be adjusted to meet the requirements of a traditional web-based database environment. The central database environment became a collection of segregated data tables that store the content of information received from the state's information systems and entry forms. The vertical table structure was implemented, in part, due to the problems and limitations encountered with the development software. The project team needed a quick work around to get a reporting component that would display the information in the desired format. A significant amount of functionality for performing custom data searches was lost in the vertical table structure.

The state's exported data are entered into the system via the data input function designed by the database developer. The initial work scope called for a "fully automated" process for the state's data to be retrieved and entered into the model system. The process proved to be a liability to both the states and the project supporters due to potential security vulnerabilities of linking to a remote network and system stability issues related to introducing a foreign program into a state's computer system. The project team had envisioned a direct link to the state's computer network with a scripted program that would extract the required information from the system automatically and store it in an accessible location on the network for retrieval by the project team. This process was rejected based on department policy and the legal liability of introducing a non-departmental script program into the system. Therefore, the information was extracted by the state's technical staff and sent to an FTP site (shared folder on the Internet that is password protected) or copied to removable media (disk, CD-ROM) and sent via the U.S. mail. The data were to be extracted and sent to the database developer on a monthly basis. The database developer was then to verify the data content, map the data to the appropriate fields, and import the data into the data-management system. The data import process would eventually become more automated after establishment of the consistency of the data, reliability of the extraction method, and integrity of the data upon arrival.

The data import process never reached an adequate flow level to test its effectiveness due to delays in system development. Based on the findings of the site surveys, the schematic in Figure 2.1 diagrams the model for information flow and process that would meet the expressed needs for the model data-management system and a national implementation. This structure was not completely realized in the pilot system.

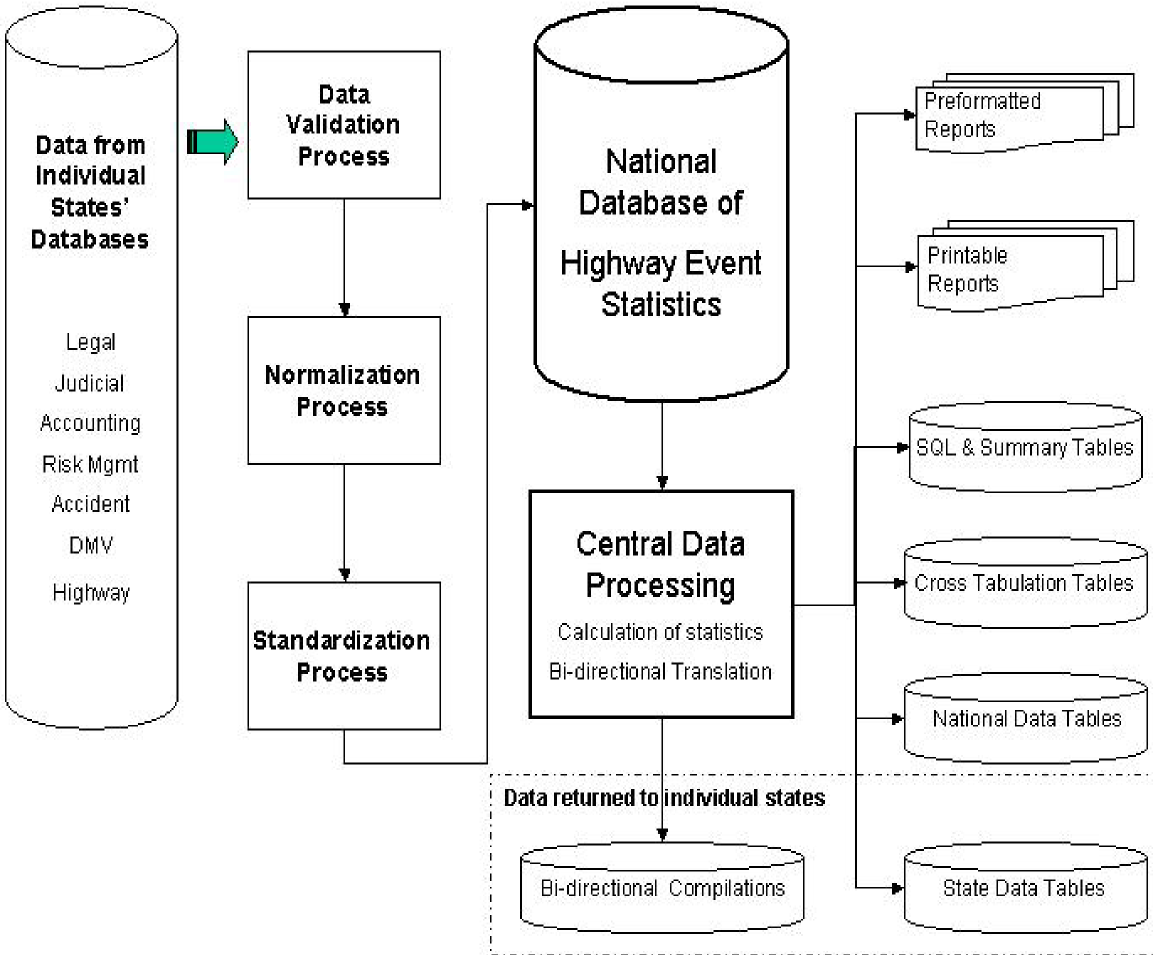

Figure 2.1 Proposed Data Environment for Model Data-Management System

Web Application Environment

The Macromedia Cold Fusion 4.5 web application server and Studio 4.5 web development software proved to be an invaluable asset to the production of the model system, especially in view of the issues encountered with the Microsoft web development software. As discussed in the Database Environment portion of this section, integration of the horizontal data tables into the website proved to be problematic. The original approach to the design and development of the model system was abandoned for a complete web-based interface that utilized only the data tables and relied on the web application software to design the user interface. The user interface is comprised of the data entry and reporting mechanisms. The data entry mechanism is in the form of web-based data entry forms within the data-management system. The data entry forms are used for entering aggregate totals similar to those prepared for the AASHTO paper survey. The entry forms are organized according to function and department: Legal, Risk Management, and Insurance and Liability using the core data elements selected to construct the model system. The table structure contains a record for every state for the years 1992 and 1997-2004. The information is entered into the system by indicating the state, calendar year of information entered, and entry date. This information is used to track the last time each state record is updated. A batch program is used to transfer the table content to the vertical tables for access by the reporting mechanism. Depending on the search criteria, the reporting mechanism pulls the specific table statistics into the appropriate item field for display and review by the user.

Server and Network Environment

The software and hardware components used to construct the data-management system from client to server are basic, midrange mechanisms for creating an environment of this type. The server computer is a mid- to low-range, dual-processor product configured with a single Pentium III (500 MHz) processor and 256 MB of Random Access Memory that provides sufficient processing power to implement a project of this limited scale. The system was purchased with six 9-GB hard drives that are integrated in a redundant array of internal drives (hardware-based RAID 5) for data protection and provides maximum storage space to meet the potential storage needs envisioned in the Interim Report. The computer has two network interface cards that will allow any entity that might assume temporary or permanent custody of the server to connect it to a network without creating a security breach by assigning a public IP address for the website and a private IP address for the local network. The server was upgraded from Windows NT4 to the Windows 2000 operating system for improved stability, administrative features, and enhanced security features. The Windows 2000 operating system has proven to be a more stable operating and processing environment than the Windows NT4. The expanded administrative functions boast improved integration with web applications, the ability to control hard-drive space per user, and added components for configuring various connection types via the World Wide Web, to name a few. A security feature enhancement with this operating system permits data encryption for transmission (IPSec) and storage (Stored Data Encryption) that offers a built-in capability for secure implementation of this site by an entity with minimal budgetary resources. The Penn State network provided the level of bandwidth for a large-capacity data transmission rate necessary for multiple and simultaneous connections to a database-driven website.

Beta Test and Monitoring

The beta testing phase of the model system was impacted greatly by the numerous delays and technical problems encountered during the development of the project site. The testing period was to begin in June 2001 and last 6 months, during which time the panel members and participant states were to explore and experiment with different components within the site. However, the test period did not begin until October 2001. A site manual that explains the different components was to be provided to the user community prior to the site coming on-line, but due to insufficient time and resources, it was replaced with the built-in User's Guide that is accessible from within the site.

A brief e-mail was forwarded to each panel member with instructions on accessing the site and user account and password information for secure logon. The logon process did offer some problems that were easily resolved. The site components are fairly intuitive and each offers some direction and instruction at the point of interaction with each component. The Discussion Forum contains its own user's guide that is accessible once the user is logged into the Discussion Forum component. In addition to testing the user functions, the beta test period was to monitor the data export and import process of the state's data into the model system. The monitoring process was to analyze and verify that the data content was correct and consistent, the data mappings for the state's data to the central data elements were correct, and the integrity of data being sent to the model system was sound. Unfortunately, the amount of external user activity was not sufficient to allow a conclusive beta test.

The in-house testing process was ongoing throughout the development of the data-management system. The different components were tested on a non-production web server prior to being programmed into the central data-management system. This process still did not prevent technical problems and human error from leading to issues that required significant debugging and, at times, redesign of entire functions or site components. The following issues remain to be addressed:

- Editing for spelling, grammar, use of terminology, and general content;

- Editing for format, color, and overall appearance;
- Correction of alignment problems in report displays;
- Correction of item descriptive/label in report displays;
- Correction of numeric character display for dollar values in report display;
- Completion of the Highway Deficiency component;
- Programming of daily interval for automatic update of entry data to report tables;
- Creation of annual reporting component with Crystal Report software; and
- Verification that the report displays contain the correct data fields.