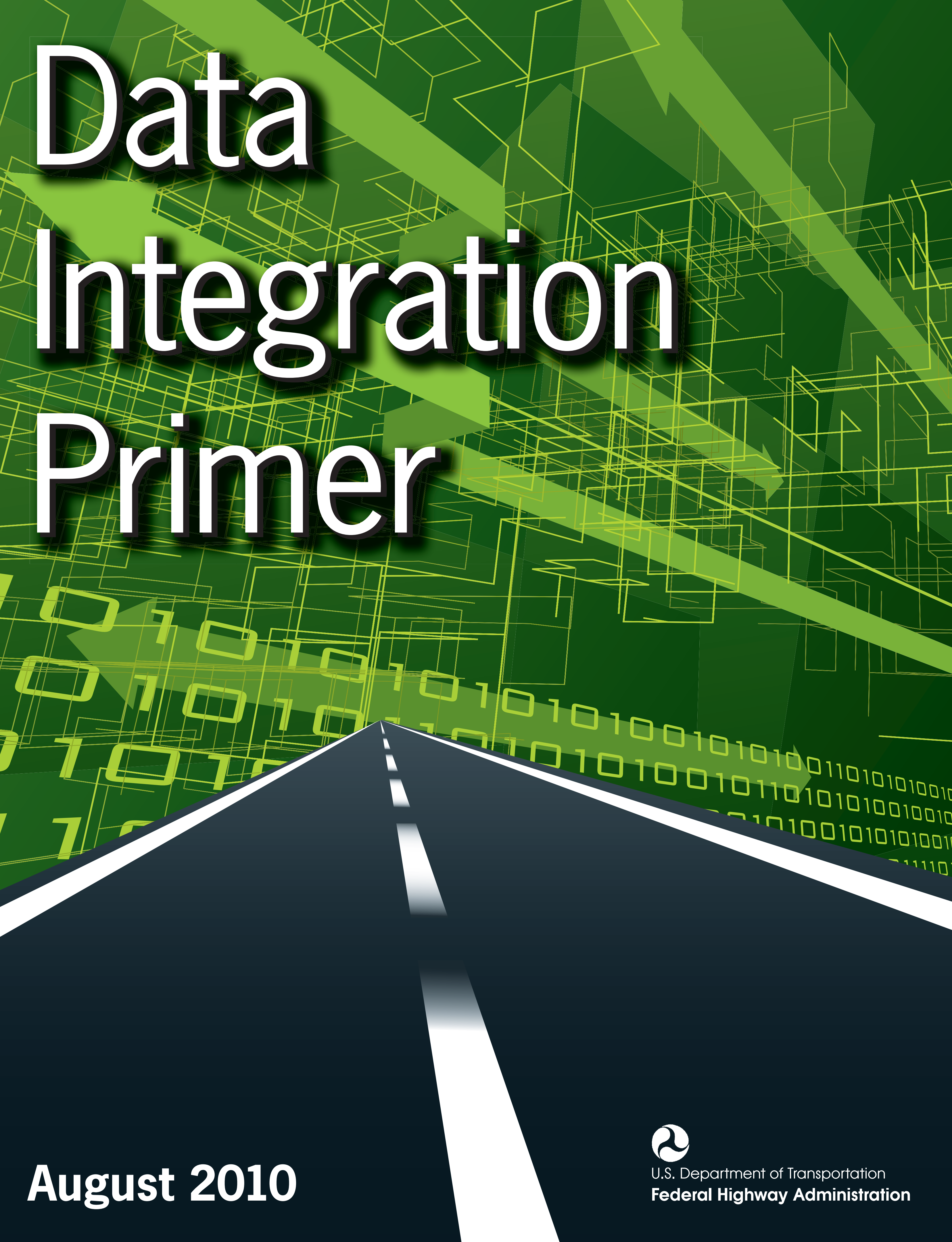
Data Integration Primer



August 2010

U.S. Department of Transportation
Federal Highway Administration

Data Integration Primer

OFFICE OF ASSET MANAGEMENT

August 2010

U.S. Department of Transportation
Federal Highway Administration
Office of Asset Management

U.S. Department of Transportation
Federal Highway Administration

TABLE OF CONTENTS

NOTE FROM THE DIRECTOR

Office of Asset Management, Office of Infrastructure
Federal Highway Administration

"Putting America to Work." As this publication goes to press, tens of billions of federal dollars are spurring forward thousands of roadway construction projects nationwide, stimulating America's economy. Despite such massive investments, the needs and the urgency of these historic infrastructure development initiatives stretch the resources, both human and financial, of transportation agencies large and small.

Transportation Asset Management (TAM) came into the common transportation vocabulary nearly a decade ago and, for those agencies that have been capable of incorporating TAM principles into their routine operations, times like these produce real pay-offs. However, even the most facile among them have probably discovered the challenges of comprehensive TAM built on a foundation of vital data that is dispersed and often disjointed.

This is a natural by-product of another historic era—the revolution in high technology capping the heralded "information age." Rapid advances in hardware and software left many with a patchwork of responses to an avalanche of available new, existing, and unforeseen information resources.

Revolution is messy. Along with remarkable advances, it tends to expose long-held practices that no longer serve. Maximum benefit from the very best TAM practices is nearly impossible without flexible, fluid, and fast access to accurate information. This Data Integration Primer aims to update agencies at various stages of Transportation Asset Management adoption and assist them in planning for the information-related decisions and investments involved in conquering that critical TAM frontier. This Primer is step one in evaluating how your agency can achieve cohesive access to and application of the information now housed in a variety of formats, locations, and silos—in other words, true "data integration."

Just five years ago, when FHWA's first Data Integration Primer appeared, many concepts that linked powerful distributed technology, such as desktop computers, were in relatively early stages of practice. Today, our baseline understanding of the tools that support data integration is somewhat more universal. However, the challenges remain steep: some are simply the province of experts.

The Data Integration Primer aims to provide the basic arguments for data integration, along with a framework for understanding and making the decisions necessary to select a data integration strategy that will work most effectively for your organization. From the "What is…?" to the "What if…?" questions, this Primer provides answers that will help stakeholders formulate their expectations, assess their existing resources, and choose appropriate levels of technical support for the data integration endeavor.

Data integration is not the "ultimate" solution for the inherent challenges of Transportation Asset Management. It is, however, a key component in the TAM toolbox. It will simplify the process of asset data integration and provide a valuable platform for asset evaluations, now and into the future.

I. Note From The Director

“How can data integration improve the Transportation Asset Management process in my agency?” “What important steps do we need to follow?” “What hardware, tools, and software options are available?” “What potential obstacles can hinder the process?” “How can we overcome the roadblocks?” Answers to these and many other questions are outlined in the following pages. Common issues, including location and other data referencing methods, database standards, staffing, and a variety of technical and institutional considerations are identified and discussed. The document also features recent experiences of transportation agencies that have integrated some or all of their related data.

The 2010 Data Integration Primer is designed as a stepping off point, helping agencies target resources to those areas of further investigation that will produce the best return on their data integration investment. From there, FHWA’s Office of Asset Management stands by to provide assistance in addressing the full range of data integration issues. Our workshops, publications, and relationships with strategic partners in the arena—from experienced peers to leading experts to State and private industry thought leaders—will help you access the resources you need to design an effective data integration strategy for Transportation Asset Management and promote the use of best practices within your agency. Call on us.

Butch Wlaschin
Director, Office of Asset Management

OVERVIEW

WHAT IS TRANSPORTATION ASSET MANAGEMENT?

Although considered one of the most effective innovations in the management of modern mobility, Transportation Asset Management (TAM) is not a software program, a database system, or even a specific package of professional tools. It is, in short, a decisionmaking process for allocating resources. TAM provides agencies with a strategic approach to managing transportation infrastructure and enables agency leadership to view the big picture before deciding how to deploy resources.[1]

TAM focuses on creating a strategic and systematic process for operating, maintaining, upgrading, and expanding physical assets effectively throughout their lifecycle. It relies on business and engineering practices for resource allocation and utilization. The goal is better decisionmaking based upon a high quality of information and well-defined objectives.[2]

TAM spotlights the whole transportation infrastructure, through an entire lifecycle, and makes possible decisions that produce the optimal performance of that infrastructure in relation to the resources required to operate and maintain it. Additionally, Transportation Asset Management examines investment timing, tools, and economic analyses to assure the most effective use of available funds.

Transportation Asset Management involves five core principles.[3] It is generally:

- **Policy-driven**—Resource allocation decisions are based on a well-defined set of policy goals and objectives.
- **Performance-based**—Policy objectives are translated into system performance measures that are used for both day-to-day and strategic management.
- **An Analysis of Options and Tradeoffs**—Decisions on how to allocate funds within and across different types of investments (e.g., preventive maintenance versus rehabilitation, pavements versus bridges) are based on an analysis of how different allocations will impact achievement of relevant policy objectives.
- **Dependent upon Quality Information for Decisions**—The merits of different options with respect to an agency's policy goals are evaluated using credible and current data.
- **Monitored to Provide Clear Accountability and Feedback**—Performance results are monitored and reported for both impact and effectiveness.

---

[1] Adapted from *Priority, Market-Ready Technologies and Innovations, Asset Management Guide*, FHWA Corporate Research and Technology, Updated 4/22/09, http://www.fhwa.dot.gov/crt/lifecycle/asset.cfm

[2] AASHTO Standing Committee on Highways. *Motion to Amend the Definition to Advocate the Principles of Transportation Asset Management.* May 6, 2006, http://www.transportation.org/sites/scoh/docs/Motion_Trans_Asset_Management.doc

[3] Adapted from NCHRP Report 551, *Performance Measures and Targets for Transportation Asset Management, Vol. I, Research Report*, 2006, p. ii, http://onlinepubs.trb.org/Onlinepubs/nchrp/nchrp_rpt_551.pdf

II. Overview

TAM principles apply to a number of functions within State transportation agencies, including:

- Planning
  - Initial goal setting
  - Long-range planning
  - TIP/STIP development
- Operations
- Preservation
- Maintenance
- Construction (including Quality)
- Safety
- Performance Measurement/Evaluation
- Financing

The Transportation Asset Management process is represented by the flowchart in Figure 1, Transportation Asset Management Framework. The flowchart shows not only the individual TAM components or business practices, but also their relationships and the order in which they will likely occur. The goals and policies of the agency consist of high-level, strategic statements that reflect the desired condition and performance of the transportation system from the perspectives of both the agency and its customers. These goals and policies therefore guide how the assets are managed at all levels of the organization.

An inventory of the assets and a means to assess their condition and model their performance is the first step, enabling the agency to identify investment requirements for improvement in the short and long term.

Next, the agency identifies the options or alternatives for addressing the investment requirement. These are then analyzed and evaluated on the basis of their cost effectiveness using a host of analytical and optimization tools. Budget and resource allocation constraints are incorporated into the alternatives evaluation criteria. Selected alternatives are then included in the list of projects that will go into the agency's short- and long-range plans. The final stages in the process consist of implementing the projects and monitoring the resulting performance of the assets.

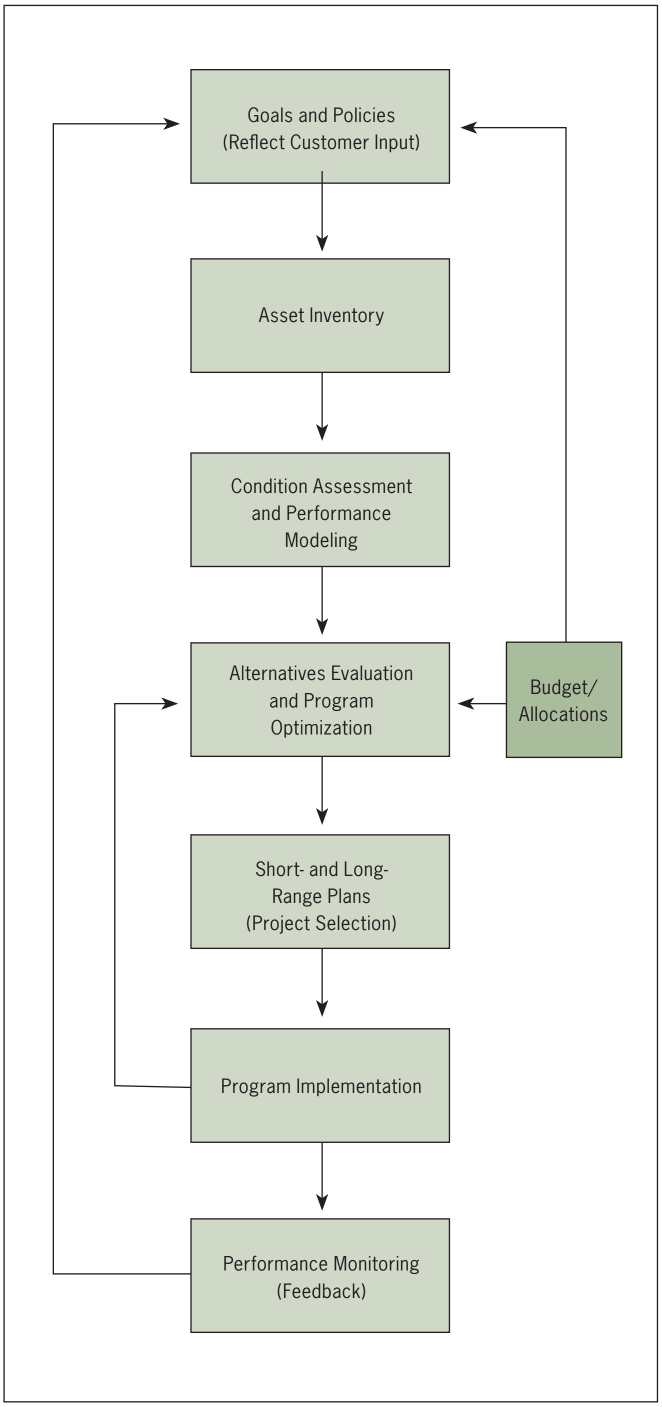


FIGURE 1. *TRANSPORTATION ASSET MANAGEMENT FRAMEWORK*

The knowledge gained from one cycle in the process is used to update and improve any or all of its components. Figure 1 depicts a generic Transportation Asset Management process in a transportation agency. However, the manner in which each component is carried out will vary from one agency to another.

WHY IS TRANSPORTATION ASSET MANAGEMENT IMPORTANT?

Since construction of the Nation's Interstate System is now functionally complete, the emphasis within the transportation community has shifted from "build it" to "improve system performance" and "preserve and maintain it." Thus, transportation agencies are under renewed pressure to demonstrate improvements in the performance of the transportation system and are being held increasingly accountable for funding decisions.

To help ease the transition from building new infrastructure to improving the performance of the existing system, many agencies are adopting TAM practices. TAM provides the tools and structure necessary to set goals, identify priorities, improve processes, and measure results that demonstrate improved performance.[4]

> *"Asset Management, in many ways, represents a 'revenge of the nerds,' . . . We are providing a rational basis for an investment process that can be inherently political."*
> *-Southeast Michigan Council of Governments Official*

Increased demands on the Nation's productivity and mobility present a significant challenge to many agencies. This is especially true where economic constraints necessitate more asset enhancement and preservation with fewer staff and less money. Transportation Asset Management helps agencies chart paths that leverage resources (including increasingly powerful computers, sophisticated analytical tools, and advances in information technology) to respond to rising system demands while maintaining adequate levels of service. TAM produces a "holistic" perspective that provides a means through which to prioritize requirements and allocate investment across different assets, over time, in the most cost-effective way.

Perhaps the most important argument for Transportation Asset Management, however, is that, without it, the underlying goal of any modern transportation agency, which is satisfying the customer, cannot be fully achieved. To fulfill that role, TAM must respond to the customer's fundamental values and priorities. With an ecosystem straining to support heavy demands, in a challenging economic environment, and at a time of mixed performance by societal leaders, words like "sustainability," "accountability," and "stewardship" rise in conversation and in public and individual consciousness. Transportation Asset Management allows agencies to strongly reflect, not only such shared values, but the solid, beneficial practices that serve them. With TAM, return on the taxpayer's dollar increases and the dialog between customers and the stewards of their investment moves toward greater understanding and partnership. Customers and agencies also benefit from a clear justification for the systematic application of funds that flow from coffers otherwise entangled by competing interests. No matter how large or small the agency, how challenging the operating realities, or how consistent the quality of customer interaction, the time is right for Transportation Asset Management.

---

[4] Adapted from *Priority, Market-Ready Technologies and Innovations, Asset Management Guide*, FHWA Corporate Research and Technology, Updated 4/22/09, http://www.fhwa.dot.gov/crt/lifecycle/asset.cfm

II. Overview

THE ROLE OF DATA IN TRANSPORTATION ASSET MANAGEMENT

Useful and reliable data are central to a fully functioning Transportation Asset Management process. TAM involves the gathering, retrieval, storage, analysis, and communication of large quantities of information. The guidance that is drawn from this data is essential to the cooperative and informed decisionmaking process underlying TAM.

Data inputs are required to evaluate and monitor the condition and performance of the asset inventory, develop performance objectives and measures, identify cost-effective investment strategies, and conduct asset value assessments. Information is also required to monitor the effectiveness of the Transportation Asset Management business process. Although it is not necessary to store all of the transportation system's data in a single repository, it is critical that the data be readily accessible and comparable. Data integration and data sharing, therefore, are vital components of TAM.

Peer Perspectives…
It is very easy for an agency to become "data rich and information poor." DOTs should strive to do the most rigorous analysis possible with current data resources. Early analysis, even with less sophisticated tools, may help agencies determine if they are collecting the "right" data.[5]

THE STATUS OF TRANSPORTATION ASSET MANAGEMENT IMPLEMENTATION

Since the publication of the American Association of State Highway and Transportation Officials' (AASHTO) *Transportation Asset Management Guide* in 2002, "many transportation agencies have begun, not only to understand the concepts and accept the principles of Asset Management, but also to implement its techniques and to incorporate these techniques in their day-to-day activities."[6]

FHWA's Office of Asset Management is actively engaged in helping transportation agencies successfully implement the Guide's principles, overcoming barriers to implementation as needed. This is important because 50 percent of State agency respondents in a recent National Cooperative Highway Research Program (NCHRP) survey noted that due, in part, to TAM implementation, they have been able to positively shift department policy and create legislative support for increased funding.

Implementation of Transportation Asset Management within individual transportation agency focus areas currently includes programs targeting:[7]

- Pavement Management
- Bridge Management
- Tunnel Management
- Roadway Safety (Hardware Management)

---

[5] Transportation Asset Management Case Studies/Data Integration, The Pennsylvania Experience, USDOT FHWA

[6] NCHRP 08-69, *Supplement to the AASHTO Transportation Asset Management Guide: Volume 2 - A Focus on Implementation.* http://www.trb.org/TRBNet/ProjectDisplay.asp?ProjectID=2190

[7] Adapted from *FY05 HIF-HIAM Strategic Program Roadmap*, FHWA, Office of Asset Management, Infrastructure, November 9, 2005, Updated April 23, 2008, http://www.fhwa.dot.gov/crt/roadmaps/hiamroadmap.cfm

- Transportation System Quality and Performance
- Construction and System Preservation
- Meeting Customer Needs
- Engineering Economic Analysis (EEA)
- Data Integration for TAM
- TAM Training and Support
- TAM Outreach and Implementation

TRANSPORTATION ASSET MANAGEMENT TOOLS DEVELOPMENT

An established body of tools is now in place to support the Transportation Asset Management process. Transportation agencies often deploy:

- Management Systems (for pavement, bridges, congestion, safety, maintenance, transit and intermodal management)
- Tools that evaluate investment levels and tradeoffs (e.g., NBIAS[8] and HERS-ST[9])
- Tools that help stakeholders analyze needs and solutions (e.g., QuickZone and EAROMAR)
- Tools that evaluate and compare options (e.g., MicroBENCOST, StratBENCOST, and LCCA[10])
- Tools that monitor results (e.g., BAMS/DSS and Estimator products in the AASHTO Trns•port suite)

In addition to these tools, a number of new technologies for collecting, processing, and managing safety data are entering the marketplace. The enactment of SAFETEA-LU in 2005 created a new emphasis on accountability for roadway safety. In response, FHWA undertook the development of tools to improve the quality of safety analysis and to prioritize safety problems in a more comprehensive fashion. These tools include:

- **Interactive Highway Safety Design Model (IHSDM)**—a recently released suite of software analysis tools for evaluating the safety and operational effects of geometric design decisions on two-lane rural highways.
- **SafetyAnalyst**—a set of software tools being developed for use by State and Local highway agencies to improve their programming of site-specific highway safety improvements.
- **Highway Safety Manual (HSM)**—a resource for use in quantifying and predicting the safety performance of various elements taken into account in road planning, design, maintenance, construction, and operation. *(Expected release: 2010)*

---

8 National Bridge Investment Analysis System

9 Highway Economic Requirements System - State Version

10 Life Cycle Cost Analysis

II. Overview

FHWA's Office of Asset Management (HIAM) recently added another item to the Transportation Asset Management "toolkit" now guiding transportation agencies nationwide:

- **Data Integration Workshop—**a comprehensive, one-day data integration class covering the basics of Transportation Asset Management and data integration techniques that includes a workshop providing an opportunity for participants to address issues related to pavement management, safety, and economic modeling, as well as to discuss data, data needs, and data integration efforts.

HIAM is engaged in a major initiative to assess the data integration needs of transportation agencies and provide technical assistance to address those needs. This FHWA initiative underscores the critical nature of effective information management in supporting decisionmaking processes involved in Transportation Asset Management.[11]

---

[11] Adapted from *"Managing Infrastructure Data, Data Integration and Data Sharing for Transportation Asset Management"*, FHWA Office of Asset Management, http://www.fhwa.dot.gov/infrastructure/asstmgmt/diindex.dfm

WHAT IS DATA INTEGRATION?

The information revolution of the last half of the 20th century produced rich data collection resources that fueled process and technical improvements of all sorts. In the context of Transportation Asset Management, data integration is:

> *the method by which multiple data sets from a variety of sources can be combined or linked to provide a more unified picture of what the data mean and how they can be applied to solve problems and make informed decisions that relate to the stewardship of transportation infrastructure assets.*

In general, transportation agencies rely on large quantities of data to support routine operations as well as to respond to planning, design, construction, and other programmatic needs. Increasingly, divisions within a transportation agency need to exchange information with others. The speed of access to, as well as the consistency and accuracy of, that information across a variety of platforms and purposes enhances their ability to do so. Such elements also allow agencies to mine the collected data for its optimum potential value.

Data integration is a fundamental requirement for effective Transportation Asset Management. An agency's structures team, for example, may need access to information about the performance of advanced materials in use on an aging arterial bridge in order to evaluate the best technological design and construction method for upgrading the structure—perhaps offsite construction of a replacement span—while considering the impact of the decision on traffic and user costs along the corridor. Data required to analyze the most favorable scenario may exist separately in the agency's pavement, bridge, and traffic operations units' databases. The information itself is vital, but the way in which this information can be shared is critical to the success of a Transportation Asset Management program. TAM is dependent upon coordinated decisionmaking, which springs from an understanding of the interdependence of often highly technical aspects of a variety of disciplines.

In order to create that data integration process, sources of available data are considered in the context of likely needs of users. Applications are constructed that help "translate" the data into useable forms and formats, allowing for migration of the information into new channels that meet those needs.

Beyond Transportation Asset Management, the incentives for data integration are readily apparent to organizations that collect, store, and manage disparate databases. Agencies that combine or link their multiple databases can reduce data collection and management costs, improve the accuracy and timeliness of the information output, and support a variety of applications that draw data from various sources.

THE GOALS

As new technologies and business processes affecting data integration continue to advance, FHWA seeks to provide robust resources that help transportation agencies adopt and improve practices. Two early initiatives, the Data Integration Workshop and this Primer, are now assisting agencies. Also, a variety of publications, including several State case histories, are available at
http://www.fhwa.dot.gov/infrastructure/asstmgmt/didoc.cfm.

III. What Is Data Integration?

THE WORKSHOP

As referenced in Section II, a key resource to agencies considering or implementing programs is the Data Integration Workshop. This day-long course begins with the basics of Transportation Asset Management and the importance of data integration to the successful implementation and operation of a TAM program. The benefits of integrated data are detailed and participants receive an introduction to key concepts of data and process flow. Case histories are used to provide relevant peer experience, and the host agency shares its experiences, challenges, and lessons learned. Breakout groups allow participants to examine specific applications more closely and provide a chance to discuss a wide array of related issues and topics. Information on the Data Integration Workshop is available at http://www.fhwa.dot.gov/asset/dataintegration/workshop.cfm.

THE PRIMER

This Primer is another key educational resource. It offers transportation agencies an introduction to data integration and a comprehensive view of the principles and practices involved in applying it to Transportation Asset Management. It supplements FHWA's "Asset Management Primer," which identifies data integration as essential in supporting a state-of-the-art decisionmaking framework. The Data Integration Primer describes benefits from a general information management perspective as well as in the context of specific elements of Transportation Asset Management. It presents implementation options and describes the most common data integration challenges. Where possible, State agency examples are provided to illustrate peer success stories and lessons learned.

WHY INTEGRATE DATA?

As more and more transportation agencies implement successful Transportation Asset Management programs, the importance of data integration rises. Information systems used for TAM, including those for pavements, bridges, tunnels, hardware, and highway maintenance, commonly draw inputs from several data sources within an agency. Many transportation agencies have thus created effective databases and procedures for populating them. However, bringing the information from these disparate systems into a common decisionmaking framework exponentially increases the value of the information collected.

Transportation Asset Management relies heavily on highly organized and integrated databases to drive its many decision-support functions. With data integration, not only can individual departments within an agency access the information they need to make informed decisions about their own assets, but the impact of their decisions on other departments is clearer and the potential for synergistic decisionmaking increases.

The cost of effectively linking data agencywide can be high, but is generally far outweighed by the long-term benefits, both Transportation Asset Management specific and more broadly.

DATA INTEGRATION BENEFITS

In an age of information, the basic benefits of information sharing are easy to imagine. In a transportation era marked by increased demand for both mobility and accountability, certain benefits advance to the forefront.

Integrated Decisionmaking

Pavement decisions are a core consideration that impacts a variety of skill sets within every transportation agency. Integrating a full range of data relevant to the pavement arena results in significant improvements that can enhance the return on the taxpayer's investment. In one Data Integration Workshop hosted by FHWA, a conversation between an agency's maintenance management system (MMS) manager and pavement management system (PMS) manager offered a glimpse into this dynamic. The MMS manager explained how, with proper data integration, he can better use resources to address pavement sections in need of maintenance that will not be rehabilitated or reconstructed under the agency's PMS program. Likewise, the PMS manager explained that he needs good MMS data that considers improvements from maintenance activities so that he can develop effective pavement condition forecasting models. In this case, the benefit of the agency's data integration program did not stop at basic information sharing, but extended into sophisticated adjustments in process and practice within two separate, but related, areas of the agency resulting in an overall improvement in the results of each department's work.

Data Integration, HPMS and HERS-ST

The National Highway Performance Monitoring System (HPMS) is used extensively in the analysis of highway system condition, performance, and investment needs nationwide (see http://www.fhwa.dot.gov/policy/ohpi/hpms/abouthpms.cfm). HPMS data form the basis for using FHWA's Highway Economic Requirements System - State Version (HERS-ST). One agency manager responsible for submitting HPMS data noted in a recent FHWA-sponsored Data Integration Workshop that data integration allowed him to prepare such input in weeks rather than months, summarizing location information, traffic data, pavement management data, geometrics, and other factors utilizing transformation rules and processes like dynamic segmentation.

IV. Why Integrate Data?

Safety Analysis

Transportation agencies regularly sift through crash data to identify "black spots," which are areas of high risk to motorists. While important, this effort is considered reactive and many agencies now require a more proactive approach. Predicting where the next black spot might occur, and addressing its dangers before they become critical, is facilitated by the use of data integration. To accommodate proactive analysis, crash data need to be linked to various databases within a transportation agency. For instance, characteristics that contribute to risk might include the number of lanes (e.g., two versus four), vehicle speed, horizontal and vertical alignment, pavement condition, lighting, existing signs and their condition, and the presence and quality of pavement markings.

To isolate trends, an agency needs to identify contributing characteristics and develop a list of potential highway segments and locations in which multiples of those characteristics are present. At one FHWA Data Integration Workshop, agency representatives were asked how long it would take to implement a safety analysis using such a proactive approach. One agency responded that without properly integrated data, at least six months would be required to complete the analysis of just one corridor. With sufficient data integration, the building blocks of safety analysis can be accessed within days rather than months. The benefit of data integration in a high risk area grows with each day in which multifaceted information sources can drive the delivery of life-saving ameliorations.

Other Benefits

Additional benefits driving the adoption of data integration practices among transportation agencies are similarly compelling and numerous:

- **Availability/Accessibility**—Asset data that is easily retrieved, viewed, queried, and analyzed by anyone within an agency encourages the integration of such data into every area of an agency that can benefit from it, spurring both innovation and better decisionmaking.
- **Timeliness**—Well-organized data can be quickly updated; one input will often apply the data across a variety of linked systems, and the information can be time-stamped to reflect its currency.
- **Accuracy and Integrity**—Errors are greatly reduced because the integration environment drives a higher quality of input and can include automatic or convenient errorchecking and verification.
- **Consistency and Clarity**—Integration requires clear and unique definition of various types of data, avoiding confusion or conflict in the meaning of terms and usage.
- **Completeness**—All available information, including both historical and recent data, is accessible in an integrated database, with any missing records or fields identified and flagged via the integration process.
- **Reduced Duplication**—Identical data is eliminated reducing the need for multiple updates and ensuring everyone is working from the exact same information.
- **Faster Processing and Turnaround Time**—Less time is spent on consolidating and transmitting data to various users in the agency. The integrated data environment saves time by eliminating consolidation and transmittal to disparate users and allows many users to conduct separate analyses concurrently.

- **Lower Data Acquisition and Storage Cost**—Data are collected or processed only once, and the information is consolidated and stored at locations supporting optimal convenience and ease of maintenance.
- **Informed and Defensible Decisions**—Highly organized, comprehensive databases allow users to drill down through successive levels of detail for an asset, supplying more information to support decisions and supporting different types of analysis using various data combinations.
- **Enhanced Program Development**—Comprehensive and coordinated system information advances program development by providing timely data for high-priority actions, promoting efficient distribution of funding among competing programs, and improving consistency in programs from year to year and across departments, among other benefits.
- **Greater Accountability—**Data integration allows rapid and more accurate reporting of costs and accomplishments, including full attribution of results to relevant agency units and functions.

Peer Perspectives…
Arizona DOT recognizes that data integration will help it compensate for the loss of experienced personnel as valuable workforce leaders transition into retirement. When younger staff seek to meet the burgeoning demands of a growing population, the agency expects information and technology to offset traditional experience and precedent as the bases for important decisions.[12]

Supporting Core TAM Requirements

Five key components are required for any comprehensive Transportation Asset Management system:

1. An asset inventory
2. Methods of assessing current conditions and/or performance
3. A process to determine and evaluate future system needs
4. Tools to evaluate and select appropriate strategies to address current and future needs
5. Methods to evaluate the effectiveness of each strategy[13]

Table 1 delineates how data integration benefits each of these processes. Each process is carried out at different levels of a transportation agency and by a broad range of staff within the agency. Data integration, thus, gives each organizational level and relevant staff member access to consistent, high-quality information, enhancing the ability of each to contribute to an effective Transportation Asset Management program.

[12] Transportation Asset Management Case Studies/Data Integration, The Arizona Experience, USDOT FHWA

[13] AASHTO-AGC-ARTBA Joint Committee, Asset Management Data Collection Guide, Task Force 45 Report, June 2006

IV. Why Integrate Data?

TABLE 1. DATA INTEGRATION BENEFITS OF SPECIFIC TAM REQUIREMENTS

Transportation Asset Management Business Process	Potential Benefits of Data Integration
Asset Inventory	• Acquire and upload data from a single source just once. • Update and process inventory records in a single transaction. • Determine more easily how much data exists and how much needs to be collected. • Reduce data handling and processing time with built-in data checking and verification.
Assessing Current Conditions and Performance	• Analyze historical and spatial conditions more conveniently. • Quickly identify assets that need immediate attention. • Standardize condition rating procedures and establish more uniform criteria for evaluation. • Store condition/investment analysis data and results more conveniently. • Support collective decisionmaking across various skills sets within an agency.
Determining and Evaluating Future System Needs	• Facilitate integrated decisionmaking. • Support investment trade-off analysis (across asset categories and modes). • Allow for more thorough and detailed assessment of investment requirements and help minimize the risk of flawed funding projections. • Effectively determine future funding needs. • Support the development of comprehensive improvement programs that cover multiple assets. • Enhance communication and improve the overall alignment of investment programs.
Evaluating and Selecting Strategies for Current and Future Needs	• Develop more effective management strategies by combining data about previous activities or decisions with existing condition data. • Reduce the risk of choosing inappropriate or ineffective action. • Prevent the inadvertent development of multiple strategies for a single asset. • Evaluate the economic viability of various alternatives. • Readily store and retrieve results of analyses. • Support fact-based strategies.
Evaluating the Effectiveness of Each Strategy	• Improve performance through immediate feedback. • Calculate performance measures and indicators with a higher level of confidence. • Promote more consistent performance measures agencywide. • Calculate and evaluate multiple performance measures for many assets in less time. • Conduct different types of analysis with data more flexibly. • Quickly compare assets, resources, personnel and activities.

HOW TO INTEGRATE DATA

Comprehensive Transportation Asset Management depends upon the availability of fully-integrated data. The process of integrating data is complex, and can be quite challenging. This is especially true when organizations are used to standalone records or database systems that rarely communicate with each other.

Where to begin? A thorough analysis of an agency's Transportation Asset Management activities is the ideal place to start. This helps an organization pinpoint needs, priorities, and existing capabilities for data integration. Before the analysis begins, however, it is wise to establish a data integration team consisting of all stakeholders in the TAM and data management processes.

THE TEAM

- Internal:
  - Data users
  - Asset managers
  - Decisionmakers
  - Information technology professionals
  - Database management professionals
  - Others? *Ask: Who else is critical to the Transportation Asset Management process?*
- External:
  - Data integration experts/consultants
  - Data collection vendors

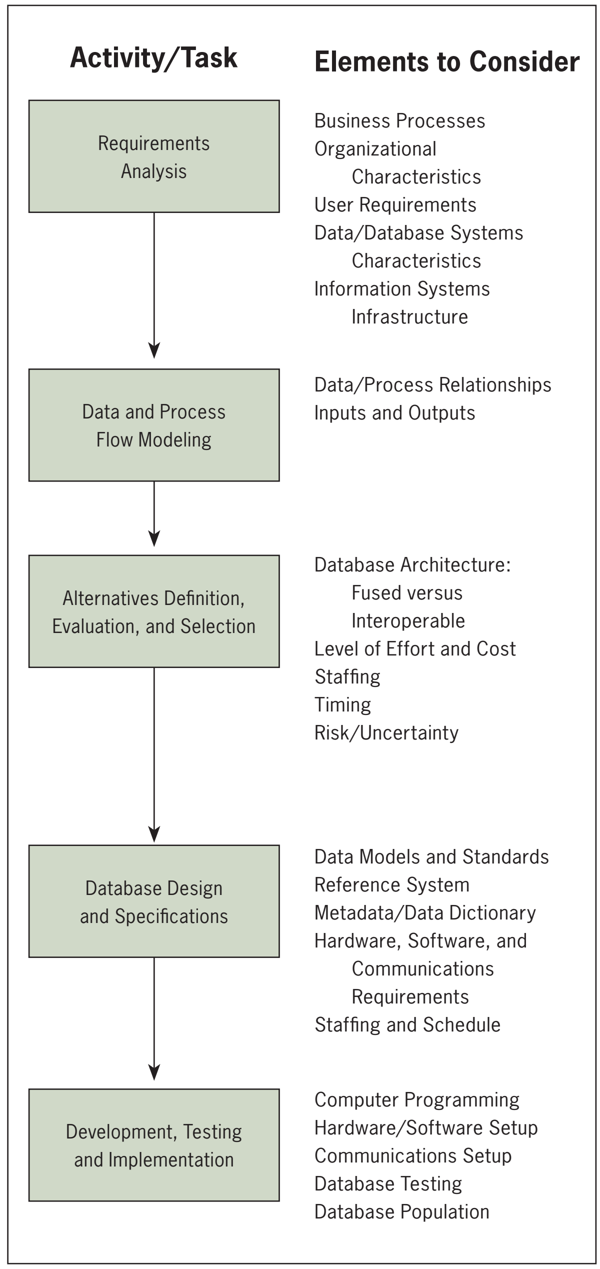


FIGURE 2. *THE DATA INTEGRATION PROCESS*

Figure 2 provides a general outline of the key activities in the data integration process, along with things to consider for each activity. Analyzing requirements is the first step in the process, followed by data and process flow modeling, then the definition, evaluation, and selection of alternatives. After this, database design and specification can be pursued. Finally, the development, testing, and implementation of the chosen database integration strategy can be implemented.

V. How to Integrate Data

Evaluating the cost of the data integration process depends upon several key factors, including the availability of existing:

- **Location referencing systems (LRS)**—Is a standard LRS being used by the agency?
- **Geographic information systems (GIS) tools**—Is the agency already using GIS databases and software?
- **Quantity and quality of data**—Do new data items need to be collected?
- **Management systems**—What systems are already in place for managing pavements, bridges, safety, signs, pavement markings, etc.?
- **Hardware and software**—Are legacy components sufficient to support the task of integrating large and complex sets of information?

Change is a constant in today's transportation agency. Maintaining momentum during organizational shifts is a key ingredient for the success of a long-term process such as data integration. In fact, agencies may wish to leverage windows of opportunity presented when top management changes or new requirements are mandated by the Federal government regarding Transportation Asset Management. Once established, the data integration system quickly becomes so integrated into business processes that future changes in the management or budget environment are unlikely to undermine support for it, or its value.

REQUIREMENTS ANALYSIS

Philosopher, Plato, famously noted that "the beginning is the most important part of the work." The most important stage of data integration is said to occur at the beginning: a requirements analysis. Depending upon the size and extent of integration, this can be a complex and time intensive step. Several areas must be examined to develop criteria for the best integration strategy.

Business Processes

Integrated databases can support a variety of functions, typically: inventory, data handling, decision-support processes, and systems for creating, acquiring, or maintaining pavements, bridges, tunnels, roadway hardware, equipment, and other physical transportation assets. To begin the requirements analysis, each business process is characterized according to the types of information it uses and produces, and the individuals who must be involved to do so. A system to support sign inventory and condition assessment, for instance, would identify key types of related information. These might include location, sign type and reflectivity, sign maintenance history, sign age, and the staff involved in assessing these data items (e.g., field crews, sign managers, and district and headquarters maintenance managers).

User Requirements

In any system supported by data integration, the requirements of data users such as field, technical, and management staff must be considered. Cooperation and involvement at all staff levels is critical to a successful integration strategy.

Requirements analysis includes ascertaining from a variety of staff where and how they obtain data, the business processes and information systems supported by that data, and any concerns they have about integrating databases. The very act of collecting this preliminary information helps to promote cooperation. Every data user must see that the strategy includes information relevant to his or her requirements and beneficial to the work at hand.

V. How to Integrate Data

Organizational Characteristics

Each transportation organization is unique and a requirements analysis should reflect the individual agency's characteristics. This includes recognition of the various groups that will be impacted by data integration. Each group's business process needs to be understood, along with factors such as the relationships between and within groups, staff skills and capabilities, availability of staff to collect additional data, and how receptive staff members are to data integration (how much they feel it will improve their effectiveness). The broad operational climate of the agency must also be taken into account. Is decisionmaking in the organization, by nature and practice, centralized or decentralized? How can an integrated data system best support either framework?

It is critical to recognize these realities and involve all stakeholders in, first, evaluating the optimum process for integrating the agency's data, then, migrating the data from traditional information structures to the integrated environment, and finally, testing and using the new data system. This creates the maximum level of trust, cooperation, and enthusiasm for the benefits of data integration. Full stakeholder involvement drives the highest possible return on the integration investment.

Information Systems Infrastructure

One critical area served by optimal stakeholder buy-in is the mapping of current information systems infrastructure. A clear view of the current picture helps the agency determine which software, hardware, and communications strategies will be required to integrate databases. From this analysis, the agency can then gauge its level of readiness for data integration. Most importantly, the analysis helps identify which potential data integration strategy can best marry the existing resources with the new infrastructure.

Useful information at this stage of the process includes an inventory of existing computer programming environments and database management or mapping software or servers, as well as computer hardware and operating systems.

Software systems are a particularly important component in planning the most efficient collection and reorganization of current data into the new structure. Many agencies use GIS software to manage a wide range of data inputs. It is important to ask what other software platforms contain information that must be identified, understood in context, and eventually harvested to build the integrated system.

Database and Database Management Characteristics

Key questions to ask when analyzing existing data and database systems within an agency might include:

- Where do the data come from and who collects it?
- How often, and how, are the data collected?
- What reference system or systems are used?
- What is the structure, format, and size of the data?
- How are the data currently transmitted, processed, and stored?
- What is the general quality of the data? Is it accurate? Complete? Recent? Unique or redundant?
- How are the data used—in what business processes?
- What applications draw data from the databases (e.g., bridge management system, pavement management system)?
- What types of reports are produced currently? What types are needed?

V. How to Integrate Data

DATA AND PROCESS FLOW MODELING

The objective of data and process flow modeling is to create a picture of the relationships between information and the business functions that the information supports. Data flow diagrams help database engineers and analysts determine the design specifications for the data integration system. All data and business processes identified in the requirements analysis can be captured in flow diagrams. A variety of software products exist to support this function.

To understand how data flows through an organization or agency division, analysts must know who collects the data, where it is stored, who uses it, and what levels of access users need (i.e., whether they need to modify, to view, or to update the data). It is also important to ascertain who "owns" the data, to provide guidelines or structure for its stewardship, and to establish a system of governance that protects the integrity of the data.

The current type, status, form, location, and uses of this information are first examined to determine data integration needs and opportunities. A path is mapped that allows specific information within each location

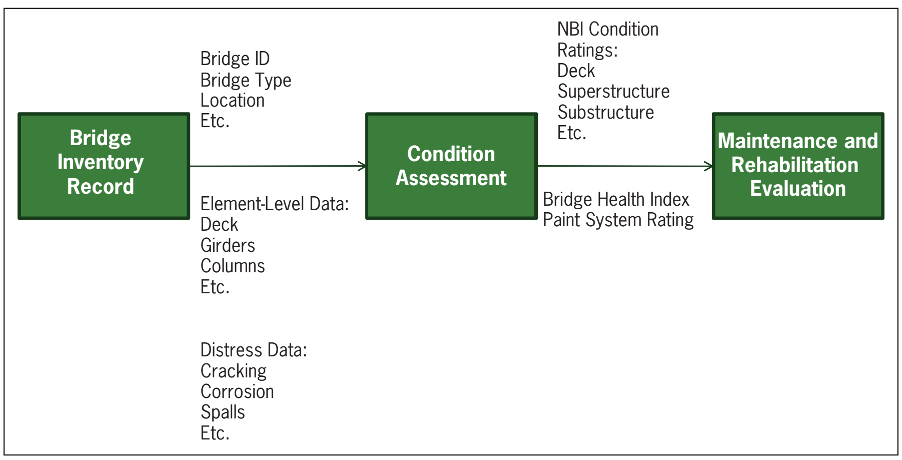


FIGURE 3. *DATA AND PROCESS FLOW: A BRIDGE EXAMPLE*

In the bridge example shown in Figure 3, basic information required to monitor the status of a bridge structure might reside in several general locations managed by a variety of departments.

or category to be accessed. Information extracted from a Bridge Inventory Record, for instance, will be maintained according to a protocol that allows it to relate to and be accessible across a wide platform of users for a broad set of purposes. In a well designed data integration system, inventory information might flow smoothly into a set of data that helps staff assess the structure's condition, leading to decisions about maintenance and rehabilitation. In a less cohesive environment, these categories of information are unlikely to be aligned in such a way that they can be readily synthesized to produce the most accurate picture from which to make sound decisions.

ALTERNATIVES DEFINITION, EVALUATION, AND SELECTION

Once requirements are analyzed and the flow of data is diagrammed, feasible integration alternatives can be identified. Two general approaches are available: fused databases and interoperable databases.

Data fusion (also known as data warehousing) combines information from multiple sources for the one-time use of making them accessible for data integration. The sources of fused data can be eliminated when the data is migrated to a centralized location. They can also continue to exist independently to serve various business processes. Ultimately, all fused data reside in a single database server with substantial processing and data storage capacity. All personal computers and terminals go through this server to access the data and perform the functions supported by the data "warehouse."

V. How to Integrate Data

Interoperable databases (also known as federated or distributed systems) consist of a series of data sources that communicate among themselves through a multi-database query. This requires a new interface through which a data source, such as an existing database, can be viewed and manipulated. In this environment, data reside in computers or database servers located in a variety of places, but each is linked through a computer network and viewed via the master interface. With interoperable databases, one computer can access or add to another's information.

Figure 4 shows how each of these options supplies access to data in response to a specific question. In this example, an integrated data system helps the agency conduct a proactive safety analysis.

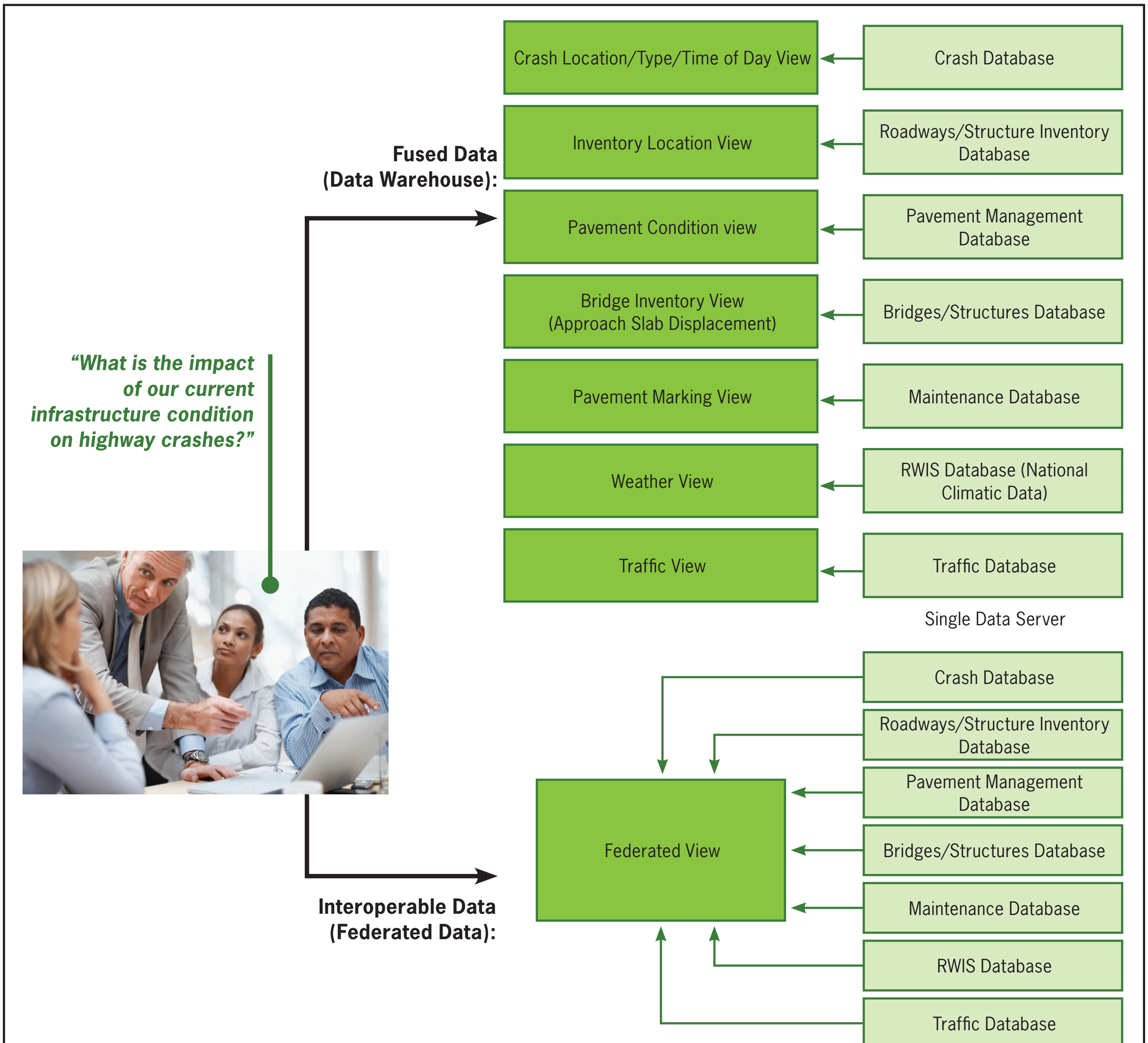


FIGURE 4. *DATA INTEGRATION ALTERNATIVES*

Fused Databases

In data fusion, the data is gathered, cleaned to remove inconsistencies, and exported to a centralized database. There it is stored in a format that replicates the way the data would be viewed in the source location. This allows users access to vast stores of data. The data fusion, or warehousing, program relies on a common user interface to organize the relevant subsets of all component databases from which it is fed, and specifies the rules for fusing the data it acquires.

Often, this requires converting a database and its applications from one format to another. Data are then shipped from the legacy system to the new one, using data reengineering or other integration methods. An agency can choose to continue to use the data in the old format after it is made available to the centralized data warehouse, or the old infrastructure can be abandoned when the warehouse is complete.

Regardless of the method used to achieve data fusion, the database management system is key—it must be able to handle the accumulation and management of a large amount of data while still ensuring that it can be accessed quickly and easily. In this sense, the interface with a fused database is something like using an Internet search engine to learn about a given topic—it delivers rapid access to information from a variety of sources, without requiring the user to have advance knowledge of each of those sources.

When fusing data, the variety of databases or formats, as well as sources and applications, can make it difficult to ensure the integrity of the information in each database. This complicates the task of mapping the movement of data from old systems to the new one and it is here that skilled database managers and information technology professionals, whether agency staff or consultants, provide critical solutions.

There is no single approach to data fusion that will meet all agencies' needs. The approach to data warehousing will continue to evolve as agency experience and technology advance.

Peer Perspectives…

Arizona DOT faced technical, cultural, and business process challenges in its data integration effort. The agency chose a data warehousing approach, and pulling data from many sources into a single repository exposed quality issues and data disconnects that had to be addressed at the source. As a result, the agency's strategy targeted cultural and process issues concurrently with technology changes.[14]

Interoperable Databases

As the name implies, an interoperable, or federated, database approach is one in which a variety of databases are linked through a communications network so that all appear to form a single source. Users can access and manipulate data from a variety of original sources, without harming the integrity of each source, and without having to learn the data model or write their transactions in the language of the source. Planners, for instance, can easily access and apply to their processes information from environmental engineers within their agency whose data might be maintained in a different format.

[14] Transportation Asset Management Case Studies/Data Integration, The Arizona Experience, USDOT FHWA

Interoperable databases allow a user to make a query without concern for where the data resides or how it is organized at its source location. A "federated view" is somewhat akin to shopping online at a department store website—it provides access to the product a user seeks without that person having to search for or be familiar with the universe of possible suppliers.

A federated view is created when an appropriate data interface is set up to link individual databases. The integrated format hides the complexity and distribution of the underlying component databases. Such a system can support different data models and execute transactions written in various data languages. It does not require the migration of all agency data into a single format, so it leaves intact the complexity and depth of data at its source.

Decentralized agencies are a natural fit for interoperable databases. While information may be structured consistently within a division, there is often great variation in data management formats across divisions. An interoperable approach works well when multiple well-organized subject databases exist (pavement, bridges, etc.), but there is a substantial need for access to the data for agencywide applications, such as maintenance management.

In short, the advantages of interoperable over fused (or centralized) database systems are that they provide:

- Easier access to resources on the network
- Improved database availability
- The ability to share data widely without relinquishing local database control

It is easy to imagine the myriad complications of developing a platform in which a wide variety of data sources converse in a universal language. In fact, the disadvantages of interoperable databases include:

- The difficulty of maintaining a functioning global model when thousands of source databases are involved (along with their query dialects, variations in supported functions, periodic updates, and differing data types and database versions)
- The expertise required to configure such a complex interface
- The ongoing "tuning" necessary to maintain acceptable performance of such a system

Federation does, however, allow agencies to preserve their investment in legacy systems while substantially improving data sharing capabilities. This, in turn, improves access to the information needed to maximize service to the highway customer.

Evaluation and Selection of Alternatives

Table 2 provides a quick reference guide to the chief advantages and disadvantages of fused and interoperable databases. More generally, however, agencies will want to consider four key elements when evaluating integration alternatives:

1. What is the required level of effort to develop either approach?
2. How much time is involved in moving to this type of system?
3. What is the estimated cost of adopting this system—including the risks?
4. What are the benefits or improvements the agency anticipates from implementing the chosen system?

Additional evaluation factors might arise from the requirements analysis, including the identification of unique agency needs.

TABLE 2. COMPARISON OF FUSED AND INTEROPERABLE DATABASES

Characteristic	Fused Database (Data Warehouse)	Interoperable Database
Number of Data Servers	One (central)	Multiple (distributed)
Location of Data Server(s)	Single site	Multiple sites
Data Replication	Yes	No
Advantages	Easy to manage and control the databases. Maximum data processing power (quick access to the database). Able to handle large amounts of data and processing requests. Provides data security.	Can keep data in independent locations and file servers (autonomy of sites). No reliance on a single site that can become a point of failure. Changes made to data at one location can propagate quickly to become visible at other locations. Unified description of all data—no need to know database models. Allows access to resources in the computer network.
Disadvantages	Requires considerable time and resources to implement. Data is generally in read-only format and cannot be updated online. Storage requirements can become a major problem.	Hard to support and maintain integrated (global) data model. Need to rebuild the database system every time data export protocols change. Requires rigorous procedures for database access and updates.

Peer Perspectives…

Moving toward its integrated Transportation Asset Management process, Michigan DOT adopted four of the guiding principles identified by the National Performance Review for gathering data. Those principles are that data gathering must be:

- *Focused*
- *Flexible*
- *Meaningful*
- *Consistent*[15]

DATABASE DESIGN AND SPECIFICATIONS

Once requirements are analyzed, processes are diagrammed, and the most effective data integration alternative is selected, an agency is ready to develop a detailed implementation plan. This includes a schedule for database design, development, testing, and training, as well as plans for software and hardware purchases.

The process of developing the design and specifications for the database can help generate this plan. It can also help define the overall approach to the database development effort.

[15] Transportation Asset Management Case Studies/Data Integration, The Michigan Experience, USDOT FHWA

Database design for both fused and interoperable environments includes certain common components:

- Data models, standards, and reference systems
- Metadata and a data dictionary
- Network communication requirements
- Software and hardware specifications
- Staffing requirements
- Data management requirements

Data Models

A data model is the way a database is structured and configured to serve the needs of its users and the way data is organized. Data and process flow diagrams developed early in the integration process will help point up the appropriate models for agencies to use when integrating Transportation Asset Management data.

Typically, there are at least five different data structure options, or models.

- **Flat File—**A file structure for data records that have no structured interrelationship. A flat file takes up less computer space than other types but requires that the database application know how the information is organized within the file.
- **Hierarchical—**A structure that links records together like a family tree, but each record type can have only one owner (e.g., a purchase order "belongs" to one department, such as the one buying the product). Although common with early mainframe database management systems, these models do not tend to support the correlation of information structures with the real world applications of the data.
- **Network—**A special case of the hierarchical data model in which each record type can have multiple owners (i.e., a purchase order can be "owned" by both the buyer and the people producing the product or service.)
- **Relational—**A model in which data are organized in a set of formally described tables from which information can be accessed or reassembled in many different ways without changing the database tables themselves.
- **Object-Oriented—**A structure that defines a data "object" as containing computer code (sequences of instructions) and data (information on which the instructions operate). Traditionally these two are separated. When merged into a single, indivisible entity, they are considered an object.

The agency's existing models may be good candidates for use in data integration depending on the format. Relational databases (with spatial components) are the easiest to integrate given a standard location referencing system. Ideally, the data models chosen will be directly compatible with applications that will use the information. If not, a data conversion or transformation process can be developed to help applications use information that exists in a different format.

Data Standards

With many potential inputs, data entry personnel, and sources of data, standards must be developed to ensure that the information is optimized for use in the chosen system. Data standards are, in effect, a set of rules for representing, accessing, manipulating, transferring, and reporting information. Some of these standards may already exist within an agency (and, if so, will be identified in the requirements analysis stage) and some may need to be developed. New standards would be drawn from an analysis of the components of the data itself and of the processes that draw information for particular applications.

Three types of data standards are commonly used:

- How specific data will be stored in the database (i.e., its content and format)
- How data are accessed and manipulated (i.e., a protocol or convention for requesting information from the database as well as the query language that must be used)
- How data can be transferred and reported (i.e., the format in which data will be exported from the database into an application or another database that will use it)

Each is vital to a well-functioning integrated data system. Standards help to ensure that, as data is used and manipulated on an ongoing basis, the quality of the information stored and accessed is maintained at a consistent level and cannot be degraded by a patchwork of users and practices.

Peer Perspectives…
Alaska DOT's work on data governance and data business plans is on the cutting edge.[16]

Data Reference Systems

Reference systems are used by the database management system to link and relate separate data files. Location referencing systems, commonly used in the highway arena, can form a backbone for the organization of information in the data integration process. Route number and milepost, for instance, can be used as a common element through which to access and deliver information regarding processes as diverse as maintenance decisionmaking and pavement design. The coordinate reference system (involving latitude and longitude) might also be used.

A location referencing system can also be used to map and analyze data using GIS software. Administrative data, such as accounting, human resources, or inventory information, on the other hand, will likely be referenced by a means not related to location.

LOCATION REFERENCING SYSTEMS (LRS) V. LOCATION REFERENCING METHODS (LRM)

A highway location reference system is a set of office and field procedures that includes a highway location reference method.

Confused? A location reference *method* is a way to identify a specific location with respect to a known point. A location reference *system* is set of procedures that relate all locations to each other and includes techniques for storing, maintaining, and retrieving location information.

Typical highway location referencing methods include those that are:

- Sign-oriented (milepost signs or reference post signs along roadways)
- Document-oriented (saves the cost of installing signs in the field)
- Experimental (using coordinates and roadside land marks, for instance)[17]

[16] FHWA Data Integration Workshop, Notes, September 9, 2009

[17] NCHRP Highway Synthesis Report 21

Peer Perspectives…
Michigan DOT decided to abandon all existing linear referencing systems and adopt a single statewide system. This allowed consistency among key data components and enabled sharing within the agency as well as among county and city agencies and the State Police. To implement, MDOT worked with the Michigan Center for Geographic Information to fund the development of a statewide GIS capability built on a GIS basemap for use by all State agencies. The multi-agency partnership eliminated many process steps that occur between data capture, integration into appropriate shared databases, and final dissemination across State government.[18]

The Importance of Metadata

Metadata is a set of information that is needed to best access, understand, and use other information in a database or information environment. In other words, it is data about data. Exponential growth in the Internet and other communications channels, as well as improved access to all forms of information, has challenged government, business and others to manage effectively ever more complex sets of data. This has driven the need for standardized ways to manage information about such content, spawning the concept of metadata. The Dewey Decimal System, which is the way libraries traditionally organize books and reference material, is one very basic form of metadata.

> *"Metadata is key to ensuring that resources will survive and continue to be accessible into the future." --"Understanding Metadata," National Information Standards Organization (NISO)*

Metadata sometimes refers to information that a computer or program can read and understand in order to organize the location, delivery, or storage of data. Other times it refers to records that describe information available electronically. It can involve any level of information access, from a single record to a large, aggregated database.

A variety of metadata standards and models have evolved in the highway program, and some of these have their own sub-layer of standards (taxonomy, vocabulary, thesauri, etc.) to convey additional information. Sets of metadata, often called metadata schemes, might be expressed in a variety of different programming languages, and communicated in a variety of forms or syntaxes, such as in HTML or XML.

Metadata can help organize a set of data and it can also help facilitate the migration of existing blocks of information into an integrated environment, including both interoperable and fused databases. It also aids tremendously in the recovery of information from integrated databases.

[18] Transportation Asset Management Case Studies/Data Integration, The Michigan Experience, USDOT FHWA

TABLE 3. METADATA SAMPLE SCHEME

Field	Name	Type	Units	Comments/Description
1	Sec_num, Pave_ID	DECIMAL		Unique record identifier
2	Ipmp_rte	CHAR		IPMP GIS database/Iowa DOT road name
3	Loc_rte	CHAR		Local agency road name
4	Lit_desc	CHAR		Literal description of section location
5	Co	CHAR		County number for the county where city is located
6	City	CHAR		City number
7	Gen_surf_t	CHAR		Current (general) surface type from local agency (P-PCC, A-ACC, C-Combination)
8	Const_yr	DECIMAL		Year constructed from local agency
9	Fed_fc	DECIMAL		Federal functional classification
10	Avg_liri	DECIMAL	m/km	Average left IRI, 999 = invalid value
11	Avg_riri	DECIMAL	m/km	Average right IRI, 999 = invalid value
12	Avg_lrut	DECIMAL	mm	Average left rut
13	Avg_rrut	DECIMAL	mm	Average right rut
14	Allig_m	DECIMAL	m^2	Area of medium severity alligator cracking
15	Allig_h	DECIMAL	m^2	Area of high severity alligator cracking

The Iowa Pavement Management Program (IPMP) offers a good example of the elements that might be included in a metadata scheme for a transportation agency.

Generally, information technology professionals develop metadata appropriate to the data integration needs of a transportation agency. Sometimes, though, basic descriptions of the information sets are provided by the creators or collectors of the data. A variety of basic tools are commonly used in the development of metadata including templates, mark-up tools, extraction tools, and conversion tools.

Once a scheme is in place, “crosswalks” help map metadata elements—such as semantics and syntax—from one metadata scheme to another, fueling the effectiveness of even the most complex data integration strategies.

The information standards community as a whole continues to develop ever more effective metadata approaches and schemes, which are then applied to specific industry practices.

The Data Dictionary

As metadata has grown in importance, the use of data dictionaries has progressed as well. A data dictionary is a central repository—sometimes called a metadata repository—in which the meaning, origin, use, and format of information, along with its relationship to other metadata, are described and maintained. It is basically a register that lists all files contained in each database in an integrated data system, the number of records in each file, and the names and types of fields contained in each record.

The term “database dictionary” can be used broadly to refer to either a document, a component of a database management system (DBMS), or a piece of “middleware” that extends or replaces an existing data dictionary within a DBMS. Depending upon their design, data dictionaries can help facilitate various levels of access to information in an integrated data system. Examples of these functions include simplifying the input required to generate a report that results from a query, and automating the assembly of code required to use data from or enter information into a system. While a data dictionary does not contain any actual source data, it is used by a DBMS to find and use such information.

Software and Hardware Requirements

Database design and specifications drive decisions about which software and hardware will best serve the integrated data strategy. Such elements as servers, network communications, data mapping, user interfaces, computer operating systems, and programming environments will be chosen to support the integrated system.

Database Server. Both software and hardware choices are involved here and they are critical because this is how and where the system's databases are stored and manipulated. In making decisions about the database server, consider the following:

- Maximum number of users expected to access the database at any one time
- Level of "uptime" needed
- Types of programs that will be used to access information from the database
- Hardware and operating system the server will use
- Level of familiarity the organization already has with a particular server environment
- Level of speed and storage that will be required to handle large and complex data processing tasks

A variety of highly secure, customizable, Web-enabled commercial products now exist to support thousands of distributed users. Both software and hardware are becoming better, faster, and cheaper as use and technology advances.

Database Management Systems. The data integration marketplace now includes a variety of commercial software packages for database management. An agency can either purchase a commercial product and customize it to suit its needs, or build its DBMS from scratch.

Geographic Information Systems. Geographic information systems lend themselves beautifully to the organization of data within transportation agencies. Beyond a tool to pinpoint a specific location on the map, in an integrated data environment, GIS software analyzes and queries information. It aids in the construction of spatial databases of transportation networks and features. Additionally, GIS software allows users to conduct a variety of analyses and applications on that data, while integrating management and decisionmaking information and processes.

GIS provide the common information threads that allow a variety of databases and DBMS to communicate with each other.

In the simplest terms, GIS is the merging of cartography (graphic elements) and data (raw information) through the use of database technology. In a generic sense, GIS applications are tools that allow users to perform a variety of functions, including creating interactive queries (such as user-created searches) analyzing spatial information, editing data, creating maps, and presenting the results of all these operations in a manner that can be easily communicated to the end user. In specific terms, GIS integrates, stores, edits, analyzes, queries, shares, and displays geographic information graphically.

GIS tools can play an important role in the process of data integration for any transportation agency because they link disparate sets of information to facilitate decisionmaking. For instance, GIS can be used to overlay pavement condition data and vehicle crashes to determine if there is a relationship between infrastructure condition (highway roughness and friction) and safety. Other information such as traffic volumes and roadway function class can be used to determine the most appropriate pavement marking material.

GIS also allows transportation agencies to leverage the information available through other State or local agencies, such as departments of natural resources, to inform decisions about roadway maintenance and construction. One example is a cutting edge practice in which an agency overlays transportation and environmental spatial data to make decisions about corridor-wide mitigation options that maximize the efficiency and impact of roadway-related environmental quality investments.

Peer Perspectives…

Pennsylvania DOT's GIS work spanned several years and a wide range of issues. Critical success factors identified by staff include:[19]

- *Adherence to and periodic review of a GIS strategic plan*
- *Development of a GIS plan that addresses problems but does not constrain solutions*
- *Contractor relationships that promote training and technology transfer*
- *Emphasis on outreach and public relations efforts*
- *Documentation of the data structure and available applications to aid use and understanding of the system by data customers*
- *A balance between strategic planning, practical applications, and future maintenance and operations requirements*

Commercial Off-The-Shelf Packages (COTS). Agencies have the option of choosing COTS software that supports the basic, generic functions of Transportation Asset Management processes and can be tailored to serve specific agency data integration routines and applications. Typically, this software ranges from enterprise-wide suites of applications (known as enterprise resource planning—or ERP—software) to products that can be used for several asset types or TAM processes. Where applicable, COTS can save time and money in the data integration process.

Enterprise Resource Planning (ERP) Software. An ERP system is a multi-module application software that supports a broad set of activities to help an agency or business manage important parts of its business. Typically, ERP solutions operate in a client-server structure, but some work with midrange and mainframe computers as well. ERP systems are typically Windows-based, grouping functions that serve such areas as accounting, financing, manufacturing, and human resources. The systems work together to control and analyze related data, often in real time—a tremendous benefit to Transportation Asset Management. A variety of COTS now exist to support specifically transportation data management and integration.

STAFFING THE INTEGRATED DATA SYSTEM

A final step in the database design and specification stage is the setting of responsibilities for management and administration of the integrated databases. This includes identifying those who will manage database development (programming, prototyping, and testing), procurement (software and hardware purchases), systems administration (computer network setup), and database general management (maintenance and upkeep).

DEVELOPMENT, TESTING, TRAINING, AND IMPLEMENTATION

The final phase in data integration involves software development and system implementation. In this stage, prototype software and use case applications are developed, computer systems and network communications are set up, and the database is populated with information.

Database models, data management applications, and communications interfaces are tested, evaluated, and modified as needed. It is now time for the appropriate personnel to be trained in the requirements and uses of the new system as well. It is wise to approach this stage in as modular and incremental a fashion as feasible since an agency will almost certainly need to make future, ongoing additions or changes to the database and

[19] Transportation Asset Management Case Studies/Data Integration, The Pennsylvania Experience, USDOT FHWA

many components of the integrated environment.
When creating software, developers ideally create prototypes and use case applications in tandem so that they can focus their attention on the data users and how they actually work with the system. Initially, a prototype will likely consist of major program modules designed to move data back and forth between screens, databases, reports, and inputs and outputs used to communicate with other data systems. At this stage, the software may perform only limited amounts of actual data processing. Subsequent versions of the programs that perform complete data processing functions will eventually replace the prototypes.

Peer Perspectives…
Pilot-level rollout programs provide the opportunity to test the validity of a system before undertaking a large-scale implementation. Virginia DOT initially implemented its system in three counties—rural, urban, and mixed rural/urban—then fine tuned data collection processes and technologies for use throughout the rest of the State.[20]

Network communications components of the integrated data system are put in place during the programming and software development stage, once the interface between databases is ready for setup and testing. Significant effort and agency resources may be expended at this stage, depending upon the scope of integration and the type of network configuration an agency chooses. Generally, the more complex the communications requirements of the integrated database system in terms of connectivity, data access rates, data retrieval and processing, or system flexibility and reliability, the more time and money will be required.

As the new system nears its operational launch, a variety of agency staff can be educated as to the design and uses of the system and trained to operate or maintain its various functions. Training might include, at a minimum, the following:

- The capabilities of and instructions for using new software tools
- The means by which to access data from the variety of sources that have been integrated into the system
- In-depth instruction for more advanced users on how integrated data will help them analyze transportation-related issues

The final—and arguably, most important—step in the development and implementation phase is, of course, the actual population of the integrated database system with data, whether new information, or material that is a legacy of the previous systems.

SUMMARY

This Primer describes requirements analysis, data and process flow modeling, alternatives evaluation, detailed database design, software development, and implementation. Application of each of these methodologies help ensure the most successful possible outcome for a data integration initiative. Within various stages of this complex process, a few general guidelines serve an agency well:

- Use a data environment that best supports making changes in database functions or adding new data sets.
- Adopt an incremental development approach to assure the maximum flexibility for inevitable change and

[20] Transportation Asset Management Case Studies/Data Integration, The Virginia Experience, USDOT FHWA

evolution, and to provide sufficient time to test and upgrade integrated databases.

- Involve database users in every stage of design and development to benefit from their input and to assure their cooperation and buy-in.
- Select hardware and software that meet the goals of data users, the agency, *and* the database management system.

CHALLENGES TO DATA INTEGRATION

One of the most fundamental challenges in the process of data integration is setting realistic expectations. The term data integration conjures a perfect coordination of diversified databases, software, equipment, and personnel into a smoothly functioning alliance, free of the persistent headaches that mark less comprehensive systems of information management. Think again.

The requirements analysis stage offers one of the best opportunities in the process to recognize and digest the full scope of complexity of the data integration task. Thorough attention to this analysis is possibly the most important ingredient in creating a system that will live to see adoption and maximum use.

As the field of data integration progresses, however, other common impediments and compensatory solutions will be easily identified. Current integration practices have already highlighted a few familiar challenges as well as strategies to address them, as outlined below.

HETEROGENEOUS DATA

Challenges:
For most transportation agencies, data integration involves synchronizing huge quantities of variable, heterogeneous data resulting from internal legacy systems that vary in data format. Legacy systems may have been created around flat file, network, or hierarchical databases, unlike newer generations of databases that use relational data. Data in different formats from external sources continue to be added to the legacy databases to improve the value of the information. Each generation, product, and home-grown system has unique demands to fulfill in order to store or extract data. So data integration can involve various strategies for coping with heterogeneity. In some cases, the effort becomes a major exercise in data homogenization, which may not enhance the quality of the data offered.

Strategies:

- A detailed analysis of the characteristics and uses of data is necessary to mitigate issues with heterogeneous data. First, a model is chosen—either a federated or data warehouse environment—that serves the requirements of the business applications and other uses of the data. Then the database developer will need to ensure that various applications can use this format or, alternatively, that standard operating procedures are adopted to convert the data to another format.
- Bringing disparate data together in a database system or migrating and fusing highly incompatible databases is painstaking work that can sometimes feel like an overwhelming challenge. Thankfully, software technology has advanced to minimize obstacles through a series of data access routines that allow structured query languages to access nearly all database management and data file systems—relational or non-relational.

BAD DATA

Challenges:
Data quality is a top concern in any data integration strategy. Legacy data must be cleaned up prior to conversion and integration, or an agency will almost certainly face serious data problems later. Legacy data impurities have a compounding effect; by nature, they tend to concentrate around high volume data users.

If this information is corrupt, so, too, will be the decisions made from it. It is not unusual for undiscovered data quality problems to emerge in the process of cleaning information for use by the integrated system. The issue of bad data leads to procedures for regularly auditing the quality of information used. But who holds the ultimate responsibility for this job is not always clear.

Strategies:

- The issue of data quality exists throughout the life of any data integration system. So it is best to establish both practices and responsibilities right from the start, and make provisions for each to continue in perpetuity.
- The best processes result when developers and users work together to determine the quality controls that will be put in place in both the development phase and the ongoing use of the system.

LACK OF STORAGE CAPACITY

Challenges:
The unanticipated need for additional performance and capacity is one of the most common challenges to data integration, particularly in data warehousing. Two storage-related requirements generally come into play: extensibility and scalability. Anticipating the extent of growth in an environment in which the need for storage can increase exponentially once a system is initiated drives fears that the storage cost will exceed the benefit of data integration. Introducing such massive quantities of data can push the limits of hardware and software. This may force developers to instigate costly fixes if an architecture for processing much larger amounts of data must be retrofitted into the planned system.

Strategies:

- Alternative storage is becoming routine for data warehouses that are likely to grow in size. Planning for such options helps keep expanding databases affordable.
- The cost per gigabyte of storage on disk drives continues to decline as technology improves. From 2000 to 2004, for instance, the cost of data storage declined ten-fold. High-performance storage disks are expected to follow the downward pricing spiral.

UNANTICIPATED COSTS

Challenges:
Data integration costs are fueled largely by items that are difficult for the uninitiated to quantify, and thus predict. These might include:

- Labor costs for initial planning, evaluation, programming, and additional data acquisition
- Software and hardware purchases
- Unanticipated technology changes/advances
- Both labor and the direct costs of data storage and maintenance

It is important to note that, regardless of efforts to streamline maintenance, the realities of a fully functioning data integration system may demand a great deal more maintenance than could be anticipated.

Unrealistic estimating can be driven by an overly optimistic budget, particularly in these times of budget shortfalls and doing more with less. More users, more analysis needs, and more complex requirements may drive performance and capacity problems. Limited resources may cause project timelines to be extended, without commensurate funding. Unanticipated issues, or new issues, may call for expensive consulting help. And the dynamic atmosphere of today's transportation agency must be taken into account, in which lack of staff, changes in business processes, problems with hardware and software, and shifting leadership can drive additional expense.

The investment in time and labor required to extract, clean, load, and maintain data can creep if the quality of the data presented is weak. It is not unusual for this to produce unanticipated labor costs that are rather alarmingly out of proportion to the total project budget.

Strategies:

- The approach to estimating project costs must be both far-sighted and realistic. This requires an investment in experienced analysts, as well as cooperation, where possible, among sister agencies on lessons learned.
- Special effort should be made to identify items that may seem unlikely but could dramatically impact total project cost.
- Extraordinary care in planning, investing in expertise, obtaining stakeholder buy-in and participation, and managing the process will each help ensure that cost overruns are minimized and, when encountered, can be most effectively resolved. Data integration is a fluid process in which such overruns may occur at each step along the way, so trained personnel with vigilant oversight are likely to return dividends instead of adding to cost.
- A viable data integration approach must recognize that the better data integration works for users, the more fundamental it will become to business processes. This level of use must be supported by consistent maintenance. It might be tempting to think that a well designed system will, by nature, function without much upkeep or tweaking. In fact, the best systems and processes tend to thrive on the routine care and support of well-trained personnel, a fact that wise managers generously anticipate in the data integration plan and budget.

LACK OF COOPERATION FROM STAFF

Challenges:

User groups within an agency may have developed databases on their own, sometimes independently from information systems staff, that are highly responsive to the users' particular needs. It is natural that owners of these functioning standalone units might be skeptical that the new system would support their needs as effectively.

Other proprietary interests may come into play. For example, division staff may not want the data they collect and track to be at all times transparently visible to headquarters staff without the opportunity to address the nuances of what the data appear to show. Owners or users may fear that higher ups without appreciation of the peculiarities of a given method of operation will gain more control over how data is collected and accessed organization-wide.

In some agencies, the level of personnel, consultants, and financial support emanating from the highest echelons of management may be insufficient to dispel these fears and gain cooperation. Top management must be fully invested in the project. Otherwise, the likelihood is smaller that the strategic data integration plan and the resources associated with it will be approved. The additional support required to engage and convey to everyone in the agency the need for and benefits of data integration is unlikely to flow from leaders who lack awareness of or commitment to the benefits of data integration.

Strategies:

- Any large-scale data integration project, regardless of model, demands that executive management be fully on board. Without it, the initiative is, quite simply, likely to fail.
- Informing and involving the diversity of players during the crucial requirements analysis stage, and then in each subsequent phase and step, is probably the single most effective way to gain buy-in, trust, and cooperation. Collecting and addressing each user's concerns may be a daunting proposition, particularly for knowledgeable information professionals who prefer to "cut to the chase." However, without a personal stake in the process and a sense of ownership of the final product, the long-term health of this major investment is likely to be compromised by users who feel that change has been enforced upon them rather than designed to advance their interests.
- Incremental education, another benefit of stakeholder involvement, is easier to impart than after-the-fact training, particularly since it addresses both the capabilities and limitations of the system, helping to calibrate appropriate expectations along the way.
- Since so much of the project's success is dependent upon understanding and conveying both human and technical issues, skilled communicators are a logical component of any data integration team. Whether staff or consultants, professional communications personnel are most effective as core participants, rather than occasional or outside contributors. They are trained to recognize and ameliorate gaps in understanding and motivation. Their skills also help maximize the conditions for cooperation and enthusiastic adoption. In many transportation agencies, public information personnel actually focus a significant amount of their time and budget on internal audiences rather than external customers. This makes them well attuned to the operational realities of a variety of internal stakeholders.

Peer Perspectives...

At least three conditions were required for the success of Virginia DOT's development effort:

- *Upper management had to support the business objectives of the project and the creation of a new system to meet the objectives.*
- *Project managers had to receive the budget, staff, and IT resources necessary to initiate and complete the process.*
- *All stakeholders and eventual system users from the agency's districts and headquarters had to cooperate with the project team throughout the process.*[21]

[21] Transportation Asset Management Case Studies/Data Integration, The Virginia Experience, USDOT FHWA

VI. Challenges to Data Integration

LACK OF DATA MANAGEMENT EXPERTISE

Challenges:
As more transportation agencies nationwide undertake the integration of data, the availability of experienced personnel increases. However, since data integration is a multi-year, highly complex proposition, even these leaders may not have the kind of expertise that evolves over a full project lifecycle. Common problems develop at different stages of the process and these can better be anticipated and addressed when key personnel have managed the typical variables of each project phase.

Also, the process of transferring historical data from its independent source to the integrated system may benefit from the knowledge of the manager who originally captured and stored the information. High turnover in such positions, along with early retirements and other personnel shifts driven by an historically tight budget environment, may complicate the mining and preparation of this data for convergence with the new system.

Strategies:

- A seasoned and highly knowledgeable data integration project leader and a data manager with state-of-the-practice experience are the minimum required to design a viable approach to integration. Choosing this expertise very carefully can help ensure that the resulting architecture is sufficiently modular, can be maintained, and is robust enough to support a wide range of owner and user needs while remaining flexible enough to accommodate changing transportation decision-support requirements over a period of years.

PERCEPTION OF DATA INTEGRATION AS AN OVERWHELMING EFFORT

Challenges:
When transportation agencies consider data integration, one pervasive notion is that the analysis of existing information needs and infrastructure, much less the organization of data into viable channels for integration, requires a monumental initial commitment of resources and staff. Resource-scarce agencies identify this perceived major upfront overhaul as “unachievable” and “disruptive.” In addition, uncertainties about funding priorities and potential shortfalls can exacerbate efforts to move forward.

Strategies:

- Methodical planning is essential in data integration. Setting incremental (or phased) goals helps ensure that each phase can be understood, achieved, and funded adequately. This approach also allows the integration process to be flexible and agile, minimizing risks associated with funding and other resource uncertainties and priority shifts. In addition, the smaller, more accurate goals will help sustain the integration effort and make it less disruptive to those using and providing data.

CONCLUSION

Transportation Asset Management helps agencies make the best use of resources to create the maximum return on the roadway investment. TAM is dependent, however, upon access to a range of data that too often resides in "silos" or other pockets dispersed within an agency. Data integration (DI) helps align all relevant information, regardless of its legacy format or source, so that it is more readily available for use in decisionmaking for the stewardship of transportation assets.

Thorough planning is essential to the process of data integration. An understanding of the requirements that an integrated data environment will be expected to fulfill is the first step. The agency's business processes, user needs, organizational characteristics, existing information systems infrastructure and database and database management characteristics must be assessed. Then data and process flow models must be established. From there, alternatives—such as aligning and connecting similar data systems (fused databases) or helping disparate information sources better relate to each other (interoperable databases)—are evaluated and selected.

The design and specification of the resulting database are then addressed. A series of decisions ensues, including recommendations regarding data models and standards. Data reference systems are explored and data dictionaries are assessed. Finally, software and hardware choices that best support the envisioned integrated database are made.

Development, testing, and implementation of the processes and solutions chosen is the final step toward full data integration. Well after the new system is in common use, however, it is important to understand that the more successful the result of the data integration process, the more robust the demands on the infrastructure are likely to be. This generally creates a need for vigorous ongoing maintenance to ensure it sustains peak performance, a phenomenon that must be anticipated in the planning and cost evaluation stages.

On the road to data integration, challenges abound. The quality of existing data varies and must be addressed. Determining a sufficient level of storage capacity can be difficult, yet is critical for a highly functioning end product. Buy-in from top management to end users is critical and can create a hurdle, particularly if the effort is undertaken at a time in which resources are stretched to capacity. Managing the cost and level of effort associated with data integration calls for considerable knowledge and foresight in the planning and analysis stages of the project. Finding project managers and other staff with relevant experience is becoming easier as more agencies achieve data integration, but the number of DI veterans remain small nationwide. And a process that can sometimes seem analogous to design-build construction practices is difficult to cost accurately, requiring a quality of commitment that allows for some flexibility as it moves toward and into practice.

Yet the dividends paid by a fully functioning integrated data environment are well worth the investment. Access to information drives innovation, efficiency, safety, and maximum return on investment. Empowering talented professionals with the information they need to make their most insightful contributions harvests from the transportation community the best possible options for addressing immediate priorities such as "Putting America to Work." It also helps ensure consistent action in the long-term best interests of the highway customer and the Nation.

Peer Perspectives...

Colorado DOT's enterprise data system is expected to grow as more data systems are converted, helping support the agency's five-year programming and twenty-year planning cycles. CDOT also recognizes that data integration becomes increasingly critical to providing stakeholders, such as the legislature and elected officials, with reliable information, including accomplishments involving performance measures and Transportation Asset Management, to ensure efficient resource allocation. CDOT is constantly nurturing and reevaluating its business processes and information management tools to ensure this goal is met. In the words of agency personnel, "We are never done."[22]

[22] Transportation Asset Management Case Studies/Data Integration, The Colorado Experience, USDOT FHWA

For further information and additional copies of this document, contact:

Office of Asset Management
Federal Highway Administration
U.S. Department of Transportation
1200 New Jersey Ave SE—HIAM 1
Washington, DC 20590

Telephone: 202-366-0392
Fax: 202-366-9981

U.S. Department of Transportation
Federal Highway Administration

This document updates FHWA-IF-01-016 - "Data Integration Primer," August 2005.

FHWA-IF-10-019