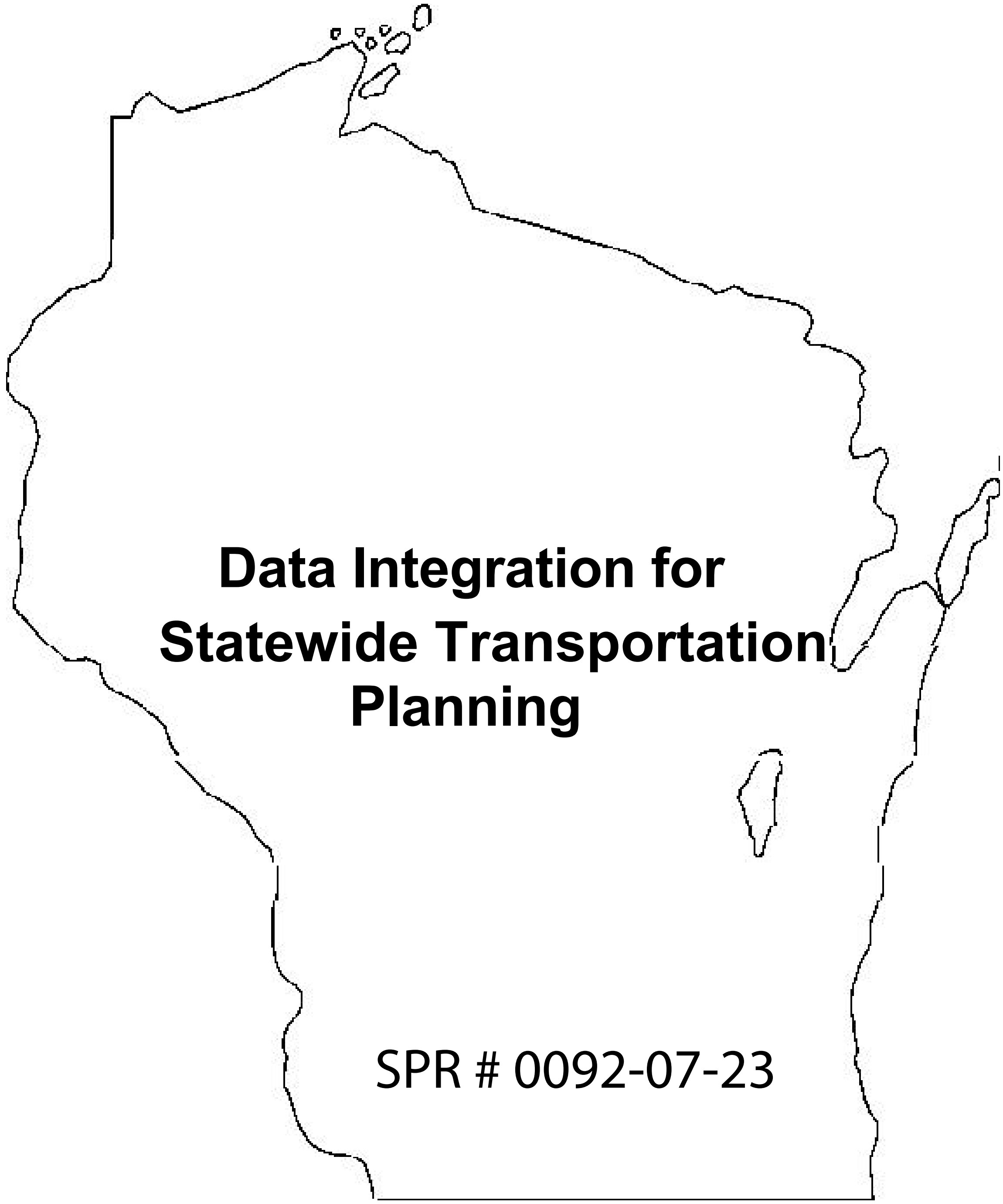
Wisconsin Highway Research Program



Data Integration for Statewide Transportation Planning

SPR # 0092-07-23

Jessica Y. Guo, Ph.D, Sasanka Gandavarapu

Transportation and Urban Systems Analysis Laboratory

Department of Civil and Environmental Engineering

University of Wisconsin - Madison

February 2010

WHRP 10-02

WISCONSIN HIGHWAY RESEARCH PROGRAM # 0092-07-23

DATA INTEGRATION FOR STATEWIDE TRANSPORTATION PLANNING

Final Report

Submitted to the Wisconsin Department of Transportation

by

Jessica Y. Guo, Ph.D.

Sasanka Gandavarapu

Transportation and Urban Systems Analysis Laboratory
Department of Civil and Environment Engineering
University of Wisconsin – Madison

February 2010

DISCLAIMER

This research was funded through the Wisconsin Highway Research Program by the Wisconsin Department of Transportation and the Federal Highway Administration under Project # 0092-07-23. The content of this report reflect the views of the authors, who are responsible for the facts and accuracy of the data published herein. The contents do not necessarily reflect the official views of the Wisconsin Department of Transportation or the Federal Highway Administration at the time of publication.

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its contents or use thereof. This report does not constitute a standard, specification or regulation.

The United States Government does not endorse products or manufacturers. Trade and manufacturers' names appear in this report only because they are considered essential to the object of the document.

ACKNOWLEDGEMENT

The TUSA Lab research team thanks the many staff members from WisDOT's Planning and Economic Development Bureau, State Highway Programs Bureau, and Information Technology Services Bureau who participated in the survey and interviews conducted for this research. The authors are also grateful to Ms. Aileen Switzer and Ms. Susie Forde for their assistance in providing contacts within WisDOT and valuable feedback on the research. Thanks are also due to all members of the project's Oversight Committee and the Data Integration Technical Oversight Committee of WHRP.

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No	3. Recipient's Catalog No
4. Title and Subtitle Data Integration and Partnership for Statewide Transportation Planning		5. Report Date August 2009 5. Performing Organization Code
7. Authors Jessica Y. Guo and Sasanka Gandavarapu		8. Performing Organization Report No.
9. Performing Organization Name and Address Department of Civil and Environment Engineering University of Wisconsin-Madison 1415 Engineering Drive, Madison, WI		10. Work Unit No. (TRAIS) 11. Contract or Grant No. WHRP# 0092-07-23
12. Sponsoring Agency Name and Address Wisconsin Department of Transportation Division of Business Services Research Coordination Section 4802 Sheboygan Ave. Rm 104 Madison, WI 53707		13. Type of Report and Period Covered Final Report, 2008-2009 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

The goal of this study was to investigate the data availability, accessibility, and interoperability issues arisen from the statewide transportation planning activities undertaken at WisDOT and to identify possible approaches for addressing these issues. A survey and a series of interviews were conducted with data users (planners) to identify current data use practices and the issues encountered when dealing with WisDOT-owned and non-WisDOT-owned datasets. Interviews were also conducted with data custodians and IT support staff to obtain information regarding the collection, maintenance and updating of frequently used planning data. The interview process confirmed some of the survey findings but also revealed differing views and data users' misconceptions of the availability and quality of planning data. The information collected about existing datasets was used to develop a catalog of planning data (with a focus on non-highway data) and a series of data flow diagrams to help documenting and informing data users of existing datasets.

Based on the findings through the various stages of this project, five recommendations were generated to help WisDOT in addressing the data challenges and opportunities relating to transportation planning: (1) information dissemination, (2) centralized data platform, (3) designated data coordinator, (4) data access tool for long range planning, and (5) data standardization. Information dissemination and centralized data platform are the low-hanging fruit that would address several data challenges with a relatively short timeframe. The remaining three recommendations require more financial and time investment and also stronger agency commitment to changing the current business practices. However, they are expected to yield high benefits in the long term.

17. Key Words Data integration; data sharing; transportation planning; data accessibility; database interoperability.		18. Distribution Statement No restriction. This document is available to the public through the National Technical Information Service 5285 Port Royal Road Springfield VA 22161	
19. Security Classif.(of this report) Unclassified	19. Security Classif. (of this page) Unclassified	20. No. of Pages 90	21. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

EXECUTIVE SUMMARY

Project Summary

The continuous and comprehensive nature of the statewide transportation planning process calls for diverse, reliable, and up-to-date data to inform decision-making. Data is required to identify the aspirations and concerns of the community, formulate alternative actions, evaluate potential impacts of alternatives, and monitor implemented solutions. Due to the wide range of planning factors to be considered, the planning process is a data-intensive one that involves the gathering, retrieval, storage, weaving, analysis, and communication of large quantities of transportation system, land use, passenger, freight, socio-demographic, economic, and environmental data. This requires accessing and integrating data from many different sources at federal, state, regional, and local levels. The need to review and revise transportation plans and programs over time also raises the added challenge of having to keep data up-to-date and to resolve temporal disparity among data sources.

The abovementioned data-related challenges do not suggest that all data for supporting the statewide transportation planning process should be statically integrated. Rather, they point to the need for data systems to be easily combined to inform decision-making. Establishing easy access and bringing consistency and compatibility to disparate multi-agency databases presents technical, cultural and institutional issues that have been largely under-explored. This project seeks to examine the data-related barriers experienced by the transportation planning staff at the Wisconsin Department of Transportation (WisDOT) and to identify data integration approaches for overcoming these barriers.

Background

Barriers relating to data awareness, availability, access, and interoperability have been found to impede the efficient flow of information and limit the full potential for comprehensive transportation analysis for many states. In Wisconsin, for example, planning for our statewide transportation services involves consolidating data from multiple business units within WisDOT, as well as federal agencies (such as Census Bureau), other state agencies (such as Department of Natural Resources, Department of Administration, and Department of Revenue) and local entities (such as MPOs and transit operators). Due to the diverse nature of potential data sources, identifying and assembling relevant data for planning studies is not a straight-forward task. While significant strides have been made to combine existing corporate databases within several functional areas of WisDOT, the process of statewide transportation planning requires access to many data sets beyond the existing integrated database systems.

Improved data integration and sharing can optimize available resources and enhance the quality of data brought to the transportation planning process. However, data integration and partnerships are often complicated by cultural and institutional factors, as well as how data are defined, collected, derived, stored, and managed. To date, little literature is available that addresses the many facets of database integration and partnership in relation to transportation planning. The goal of this research is to expand the knowledgebase about data-related issues and solutions to help state agencies take the next step toward more effective planning processes through data integration.

In particular, this research investigates the data availability, accessibility, and interoperability issues arisen from the statewide transportation planning activities undertaken at WisDOT and identifies possible approaches for addressing these issues. The specific research objectives are to:

1. Review existing publications to identify best practices of data integration and partnership for transportation planning, thereby providing a synthesis of the motivation for integration/partnership, the problems encountered and lessons learnt, and the technologies/strategies used.
2. Identify the business and analytical processes constituting the current statewide transportation planning practice in Wisconsin, document the data currently used to support the processes; and catalog information regarding data sources useful for transportation planning.
3. Identify and document technical, financial, institutional, and other barriers to the access and integration of planning-related data. Additionally, assess the extent to which currently available data could support future transportation modeling and planning activities in Wisconsin.
4. Develop recommendations for overcoming the barriers to data integration and partnership within WisDOT, as well as between WisDOT and other agencies. Also, identify the next steps to further evaluate the cost effectiveness of the recommended methodologies.

This research project was conducted at the University of Wisconsin-Madison and funded by the Wisconsin Highway Research Program (WHRP) of WisDOT.

Process

The research was comprised of six tasks designed for collectively achieving the project objectives outlined above.

<u>Task 1: Synthesize existing literature on data integration and partnership.</u> This task corresponds directly to research objective (1). The research team reviewed published documents and web site materials pertaining to data integration and partnerships to aide statewide transportation planning and transportation decision making in general. The synthesis of literature is presented in chapter 2 of this report.

<u>Task 2.Understand the current practice</u> and <u>Task 3. Identify data opportunities, gaps and challenges.</u> These two tasks go hand-in-hand and together address research objectives (2) and (3). As part of task 2, the research team started by conducting a survey with the staff at the Bureau of Planning and Economic Development (BPED) of WisDOT to develop an inventory of their planning activities and the data sources used to support these activities. Further information about currently used data sources were collected through reviewing relevant documents and contacting data custodians within and outside of WisDOT. The information collected was used to compile a data catalog documenting the current status and use of planning-related data. Under the advice of the project oversight committee, the

data catalog adopted the structure of WisDOT's existing Highway Data Catalog (developed by the Data Management Section of BSHP) so that it can be viewed as an extension of the Highway Data Catalog. A total of 13 data sources were documented in addition to the many already described in the Highway Data Catalog. The data cataloging effort is described in section 4.1 of this report.

The survey designed for accomplishing task 2 also included questions for fulfilling task3. Specifically, the survey asked the respondents their perceived accessibility, quality, format and timeliness of planning data sources, both internal and external to WisDOT. The respondents were also asked to report any unmet data needs that they currently faced and that they envisage for the near future. Completed responses were received from nine planning staff. The survey process and findings are summarized in sections 3.1 and 3.2 of this report.

As part of Task 3, the research team also conducted face-to-face interviews with selected data users within BPED to ask clarifying, more in-depth questions regarding the data challenges reported in the survey, as well as for additional data needs and concerns not reported in the survey. Data custodians, database application managers, and agency-wide IT support staff were also contacted to obtain information regarding existing data management systems and any on-going or anticipated data access improvement efforts. The interview process not only confirmed many of the survey findings, but also revealed differing views and misconceptions of the availability and quality of planning data. The interview process and the findings are reported in sections 3.3 and 3.4 of this report.

Task 4. Data and process flow modeling. Based on information collected through the survey and interviews conducted as part of tasks 2 and 3, the research team documented the flow of frequently used planning data in the form of a series of data flow diagrams. The diagrams serve to help identify the organizational units and technology involved in the collection, maintenance and updating of each dataset; illustrate possible data redundancy and lack of centralized access; and help formulate recommendations to increase the efficiency of the current data flow mechanism. As such, this task contributes to meeting research objectives (2) and (3). The flow diagrams are presented in section 4.2 of the report.

Task 5. Develop recommendations for data integration and partnership. This task entailed first assessing the data needs and challenges identified through the survey and user interviews (tasks 2 and 3) against the current data access and flow mechanisms revealed through the custodian interviews and process flow modeling (tasks 3 and 4). Based on the assessment, the research team developed a set of recommendations for addressing existing needs and challenges and for helping achieve the desired level of data integration and partnership. The recommendations incorporated lessons learnt from past data integration practices (as reviewed in task 1), inputs from WisDOT staff, and knowledgebase of the research team. The recommendations are documented in chapter 5 of this report. Also included in the chapter are the respective factors for consideration, ballpark estimates of the relative cost requirement, the implementation time requirement, and the anticipated scope of impact on planning practices associated with each recommendation. This task fulfills research objective (4).

Task 6. Prepare final report. This entailed summarizing all study elements in this report.

Findings and Conclusions

Compared to the amount of literature on the topic of data integration for asset management, there are much fewer publications regarding data integration for transportation planning, indicating the need for further research on this latter topic. Among the data integration efforts reviewed in this project, developing and disseminating data dictionaries or catalogues was found to be a common and useful approach. More advanced approaches used to enhance data access include the use of metadata publishing tools and content management systems. There are examples of state DOTs using either the centralized data warehousing approach or the interoperable database approach to integrate databases within their agencies. Geographic information systems (GIS) typically play a central role in managing the consistency across transportation-related spatial data and providing data in a unified environment. The literature also pointed to organizational issues – such as resistance to change from staff – as often being a major impediment to data integration efforts. The fact that the benefits of such efforts often lag its implementation and are not very easily quantifiable in short run also makes gaining agency-wide support in such efforts a challenge.

Our survey to BPED staff resulted in a response rate of about 30% (9 responses). Thus, caution needs to be exercised when interpreting these individual responses. The general themes that emerged from the survey include:

(a) Data access appears to be hindered by the perceived lack of up-to-date information regarding what data are available and where. This perception is a result of having no clear documentation of data or users not aware of data documentation that exists.

(b) Planning staff generally found the data quality to be satisfactory. An exception is when multiple sources appear to be available for the same data item.

(c) Since the wide range of planning-related data often comes in differing formats and level of spatial scale, most respondents reported that they often need to perform minor manipulation and sometimes need major manipulation to use planning data.

(d) The differing, and sometimes irregular, cycles by which datasets are updated were reported to hinder users' ability to locate the latest version of a dataset.

(e) Regarding data shared between WisDOT and other agencies, two major issues were reported. First, as data from external agencies are sometimes in hard copy and represent a snapshot in time, the timeliness of such data is of concern. Second, the data models used by WisDOT and other agencies are often different, making data interoperability an issue.

The data cataloging effort and the data flow modeling process provided further insights into the overall process of the data sharing mechanism within WisDOT. The primary findings are:

(a) Much of the existing, well-established data sharing mechanism is placed on collecting, maintaining and reporting highway related data. Many of the non-highway data are not available as an enterprise resource that is easily accessible and personal data requests have to go to designated individuals. The seeming emphasis on highways in the current data management systems is largely driven by the business processes and reporting requirements. This has led to the implementation of effective highway data management strategies in Wisconsin and other states. However, the statewide transportation planning takes place in a multimodal context and requires a much broader range of data beyond highways. Currently, a streamlined process for maintaining multimodal data that parallels the existing process for managing highway data remains lacking in WisDOT. As revealed in our literature review, this is a common challenge also being faced by other state DOTs in the country.

(b) The data flow modeling process revealed that more than one access channel exists for some datasets. In most cases, there is not a duplication of data, but merely multiple access methods that offer greater convenience to the users. In these cases, the end users need to be aware that the data obtained through the alternative access channels is the same and there is no concern regarding data consistency across differing channels. In other cases, however, duplicate copies of data or variants of the same data do exist, particularly on the Geo-base system.

(c) The disparity across the planning-related data in their sources and flow paths – as reflected in the data flow diagram – suggest that a well structured and managed central clearinghouse for planning data would be a valuable asset to BPED.

(d) The disparate format, quality, and update cycle of data sources from other state and federal agencies present a data challenge that is beyond WisDOT's administrative boundary and requires federal leadership to achieve standardization across planning data providers and users.

Recommendations for Further Action

Based on the findings through the various stages of this project, five recommendations were generated to help WisDOT in addressing the data challenges and opportunities relating to transportation planning.

<u>Recommendation #1: Information dissemination.</u> Raise awareness of where data are by providing an up-to-date data catalog (this has been implemented with the Highway Data Catalog). Adopt technologies such as web-based metadata application or content management system to provide more detailed and dynamically maintained metadata.

<u>Recommendation #2: Centralized data platform.</u> To address the concerns of redundant or duplicate data, move the data items currently residing on the Geo-base platform that are worth keeping to the SDE under the management of BITS.

<u>Recommendation #3: Designated data coordinator.</u> Designate a data coordinator, whose job function would be to streamline the data business process for BPED and provide the needed data services and products to support data applications, planning processes, and business objectives. The data coordinator would structure these services, coordinate among organizational units that provide the original data, coordinate with the data

management units in other Bureaus (e.g. BSHP and BITS), and ensure the services are maintained and running.

Recommendation #4: Data access tool for long range planning. To meet the strong business need for working with and integrating multimodal data, house all planning required data items in SDE and develop a data access tool based on the 'layer file' feature in ArcGIS. The tool essentially provides pointers to the planning related data items on SDE with predefined display settings, queries, and labels. An alternative to implementing the data access tool using layer files is to build a graphical user interface (GUI) application on SDE. Such a GUI could allow users to interactively select the data layers needed and run customized queries. However, this development of such a GUI would be more expensive and involved than that of the layer-file based tool.

Recommendation #5: Data standardization. A solution that would have the most long term impact on enhancing data exchange and interoperability between a state DOT and other agencies is to establish data standards. There has been ongoing research effort in developing standards for transportation data, e.g. TransXML, using XML schemas to provide a common vocabulary and information structure for transportation agency activities and assets. This is not an approach that is ready for adoption, but a direction for future practice that warrants planning agencies' attention.

The implementation of our five recommendations would entail differing levels of costs and time requirements. Information dissemination and centralized data platform are the low-hanging fruit that would address several data challenges with a relatively short timeframe. The remaining three recommendations require more financial and time investment and also stronger agency commitment to changing the current business practices. However, they are expected to yield high benefits in the long term.

TABLE OF CONTENTS

DISCLAIMER .... **I**
ACKNOWLEDGEMENT .... **II**
TECHNICAL REPORT DOCUMENTATION PAGE .... **III**
EXECUTIVE SUMMARY .... **V**
TABLE OF CONTENTS .... **XI**
LIST OF TABLES .... **XII**
LIST OF FIGURES .... **XIII**
1. INTRODUCTION **1**
1.1. STATEWIDE TRANSPORTATION PLANNING AND THE ROLE OF DATA 1
1.2. DATA INTEGRATION 2
1.3. RESEARCH OBJECTIVES 3
1.4. SCOPE, METHODOLOGY, AND REPORT OUTLINE 3
2. LITERATURE REVIEW **5**
2.1. DATA CHALLENGES 5
2.2. DATA INTEGRATION STRATEGIES 8
2.3. IMPLEMENTATION CONSIDERATIONS 22
2.4. SUMMARY 23
3. CURRENT DATA PRACTICES AND CHALLENGES **25**
3.1. SURVEY PROCESS AND INSTRUMENT 25
3.2. SURVEY RESPONSES 26
3.3. INTERVIEWS 30
3.4. SUMMARY 31
4. DATA CATALOG AND FLOW MODELS **33**
4.1. DATA CATALOG 33
4.2. DATA FLOW MODELING 38
4.3. SUMMARY 44
5. RECOMMENDATIONS AND CONCLUSIONS **46**
5.1. INFORMATION DISSEMINATION 46
5.2. CENTRALIZED DATA PLATFORM 49
5.3. DESIGNATED DATA COORDINATOR 50
5.4. DATA ACCESS TOOL FOR LONG RANGE PLANNING 50
5.5. DATA STANDARDIZATION 53
5.6. SUMMARY 53
6. REFERENCES **55**
APPENDIX A. SURVEY QUESTIONNAIRE **59**
APPENDIX B. DISCUSSION **71**

LIST OF TABLES

Table 1 Questionnaire structure 26
Table 2 Summary of recommendations 47
Table 3 Geo-base Vs. SDE 49

LIST OF FIGURES

Figure 1 Overview of study components and interdependency among components.......... 4
Figure 2 Data quality and ambiguity faced by the Pend Oreille County (10) 6
Figure 3 Translations from one LRS to other (direct transformation) (22) 11
Figure 4 Translating Different LRS into one system (indirect transformation) (22)........ 11
Figure 5 Pend Oreille County's solution to their data ambiguity issue (10) 13
Figure 6 Data warehouse (33)... 14
Figure 7 IDOT Spatial Data warehouse architecture (34) 14
Figure 8 Arizona DOT's data architecture (31).. 15
Figure 9 Interoperable database System (34)... 16
Figure 10 Comparison between fused and interoperable database (7) 17
Figure 11 GUI of WSDOT's search tool (37) 20
Figure 12 Example of a search result returned by WSDOT's online tool (37) 20
Figure 13 Phased implementation of MeDOT's TIDE data warehouse, ... 23
Figure 14 Database schema describing the structure of WisDOT's existing Highway Data Catalog 34
Figure 15 Proposed extension to the Systems table of WisDOT's data catalog............... 35
Figure 16 Proposed extension to the Attributes table of WisDOT's data catalog 36
Figure 17 Proposed extension to the Descriptions table of WisDOT's data catalog........ 37
Figure 18 Data flow diagram depicting the flow of key data used to support planning activities 39
Figure 19 STN, WISLR, FIIIPS 40
Figure 20 Photolog, railroads, ports and harbors 41
Figure 21 Airports, TRADAS, and economic data 42
Figure 22 Reference USA, Rail passenger, WorkNET, and environmental data 43
Figure 23 A sample SDE web-based catalog (source: 49) 48
Figure 24 Proposed data access tool to support long range planning activities................ 52

1. INTRODUCTION

1.1. Statewide Transportation Planning and the Role of Data

Statewide transportation planning is the *cooperative*, *continuous*, and *comprehensive* process through which investment decisions are made to provide safe and efficient transportation services throughout a state (1). It plays a fundamental role in a state's vision for its future. *Cooperative* suggests that planning decisions are made with the inputs from various community sectors, interest groups, agencies, and stakeholders. Everyone has a right to voice their opinion on proposed transportation programs and projects. *Continuous* refers to the fact that planning decisions are made and revisited over time to ensure the decisions remain the best courses of action. *Comprehensive* means that the decision-making process needs to account for all relevant factors. As outlined in the Safe, Accountable, Flexible, Transportation Equity Act – A Legacy for Users (SAFETEA-LU), the planning factors that constitute the statewide transportation planning process include:

1. *Support the economic vitality of the metropolitan area, especially by enabling global competitiveness, productivity and efficiency.*
2. *Increase the safety of the transportation system for motorized and non-motorized users.*
3. *Increase the security of the transportation system for motorized and non-motorized users.*
4. *Increase the accessibility and mobility of people and for freight.*
5. *Protect and enhance the environment, promote energy conservation, and improve quality of life, and promote consistency between transportation improvements and State and local planned growth and economic development patterns.*
6. *Enhance the integration and connectivity of the transportation system, across and between modes, for people and freight.*
7. *Promote efficient system management and operation.*
8. *Emphasize the preservation of the existing transportation system.*

Due to the wide range of planning factors to be considered, the planning process is a data-intensive one that involves the gathering, retrieval, storage, weaving, analysis, and communication of large quantities of transportation system, land use, passenger, freight, socio-demographic, economic, and environmental data. As these data may come from many different sources at federal, state, regional, and local levels, the ability to access and integrate data is vital to successful planning practices. Moreover, since transportation plans and programs need to be reviewed and revised over time, the need to keep data up-to-date and to resolve temporal disparity among data sources places added challenge upon the planning process. Without comprehensive and timely data, the various analyses and modeling exercises undertaken as part of the planning process cannot provide reliable results to inform transportation investment decisions.

Identifying and assembling the broad range of relevant data for planning studies often requires pulling together databases from multiple business units within and outside of the state's Department of Transportation. Although it is not necessary to store all the relevant data in a single repository, it is critical that the data can be easily combined to inform decision-making. Yet, linking disparate and distributed datasets is a process often complicated by cultural and institutional factors, as well as how data are defined, collected, derived, stored, and managed (2). Barriers relating to data awareness, availability, access, and interoperability have been found to impede the efficient flow of information and limit the full potential for comprehensive transportation analysis for many states. Data integration and data sharing, therefore, are vital components of statewide transportation planning.

1.2. Data Integration

From an information management perspective, data integration can be simply defined as the problem of combining data residing at different sources, and providing the user with a unified view of these data (3, 4). The Federal Highway Administration (FHWA) adopts a similar but more expanded definition of data integration: "*the process of combining or linking two or more data sets from various sources to facilitate data sharing, promote effective data gathering and communication, and support overall management activities in an organization*" (5). At the 2001 Peer Exchange on Highway Information Integration organized by the TRB Statewide Data and Information Systems Committee, the participants reviewed data integration as "*coordination and sharing of inputs, processes and outputs of systems; dynamically linking systems where data is consistent and easily accessed, displayed and transferred between systems thereby creating valuable information for stakeholders and business decision making process*" (6). This latter definition is considered to be most suitable for the context of this project. The definition highlights the point that not all data needs to be statically integrated. Instead, ensuring that data systems are consistent and interoperable is critical to successful data integration and effective decision making.

The incentives for data integration are many. Agencies that combine or link their multiple databases can reduce data collection and management costs, improve the accuracy and timeliness of the information, and support a variety of applications that draw data from various sources (7). Other benefits of data integration include (8):

- *Long term operational savings through reduced redundancy of effort*
- *Long term cost avoidance through identification of opportunities to build more efficient, integrated systems from the ground up*
- *Improved customer service*
- *Improved data consistency / quality / accuracy / timeliness of data, driven by expanded use*
- *Development of cross-agency trust and communication driven by inter-dependence*

- *An enterprise-wide awareness of the broader business process, which will foster identification of future integration opportunities, and better, more efficient overall operation*

1.3. Research Objectives

In Wisconsin, planning for our statewide transportation services often involve consolidating data from multiple business units within WisDOT, as well as federal agencies (such as Census Bureau), other state agencies (such as Department of Natural Resources, Department of Administration, and Department of Revenue) and local entities (such as MPOs and transit operators). Due to the diverse nature of potential data sources, identifying and assembling relevant data for planning studies is not a straight-forward task. While significant strides have been made to combine existing corporate databases within several functional areas of WisDOT, the process of statewide transportation planning requires access to many data sets beyond the existing integrated database systems. There remains a lack of knowledge about data availability and of interoperability and integration among disparate data sources, especially for non-WisDOT data sources. These issues impede the efficient flow of information and limit the full potential for comprehensive transportation analysis.

This project aims to examine the data-related barriers experienced by the transportation planning staff in WisDOT and identify data integration approaches for overcoming these barriers. The primary objectives are:

1. To review existing publications to identify best practices of data integration and partnership for transportation planning, thereby providing a synthesis of the motivation for integration/partnership, the problems encountered and lessons learnt, and the technologies/strategies used.

2. To identify the business and analytical processes constituting the current statewide transportation planning practice in Wisconsin, document the data currently used to support the processes, and catalog information regarding data sources useful for transportation planning, including the current data custodians, formats, contents, quality, collection and maintenance costs, user requirements, and access methods.

3. To identify and document technical, financial, institutional, and other barriers to the access and integration of planning-related data. Additionally, assess the extent to which currently available data could support future transportation modeling and planning activities in Wisconsin.

4. To develop recommendations for overcoming the barriers to data integration and partnership within WisDOT, as well as between WisDOT and other agencies. Also, identify the next steps to further evaluate the cost effectiveness of the recommended methodologies.

1.4. Scope, Methodology, and Report Outline

This project focuses on the data needs and challenges faced by the personnel of the Bureau of Planning and Economic Development (BPED), which is consisted of the

Urban planning Division, Economic Development Division, and Travel Forecasting Division. Emphasis is placed on the multimodal planning context in which data describing multiple modes for both passenger and freight transportation are needed.

As shown in Figure 1, the study involves several components: literature review, online survey to BPED staff, interviews with several business units within and outside of WisDOT, development of a data catalog and a data flow model. Finally, all components feed into the development of our data integration recommendations.

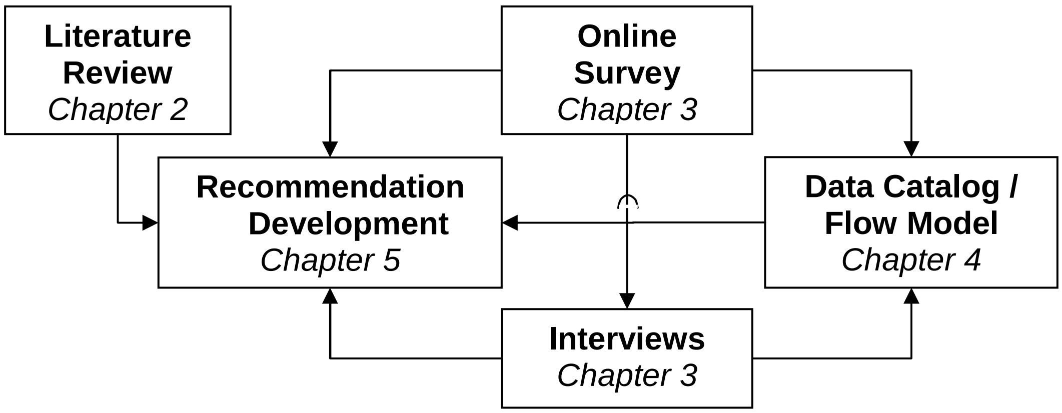


Figure 1 Overview of study components and interdependency among components

This report documents our research effort, findings, products, and recommendations. Specifically, Chapter 2 summarizes past data integration efforts and presents two case studies to illustrate successful data integration and partnership initiatives specifically for statewide transportation planning. Chapter 3 describes the data-related challenges and future data needs as identified by a small sample of WisDOT transportation planners through the survey and interviews. Chapter 4 presents our effort of inventorying WisDOT's planning-related data and the data catalog and flow model produced from this effort. Finally, chapter 5 provides a set of data integration recommendations to help streamline the statewide, multimodal transportation planning activities in Wisconsin.

2. LITERATURE REVIEW

Statewide transportation planning requires a variety of data from many different federal, state, regional, and local sources. State DOTs across the country have inevitably encountered data-related challenges during their planning process and have utilized various strategies to address these challenges. This chapter presents a synthesis of DOTs' experience with these data-related challenges and solution strategies.

The original intent of this literature review was to summarize published documents and web site materials pertaining to data integration and partnership to aide statewide transportation planning. However, our literature search process revealed that little published material has discussed data integration issues and solutions related specifically to statewide transportation planning. Therefore, the scope of our review was expanded to include discussions of these issues and solution options related to the broader domain area of transportation decision making. It should also be noted that, due to the dynamic nature of the planning process and practice and the limited amount of publicly available information on the subject, this synthesis is not intended as a comprehensive review of the most current state of practice. Rather, our focus is on surveying the range of data-related challenges experienced by DOTs, providing insights into the nature of these challenges through documented DOT experiences, and identifying the best practices for addressing these challenges.

This chapter is organized as follows. Section 2.1 outlined the many issues arisen from the transportation decision-making process that call for data integration initiatives. Section 2.2 describes data integration tools and strategies that have been adopted in various states across the country. Section 2.3 raises a few points for consideration when implementing data integration strategies. Finally, Section 2.4 presents a summary of the literature scane.

2.1. Data Challenges

Our literature review reveals that the statewide transportation planning process is often subject to the following five types of data-related challenges:

- Data accessibility
- Data quality and timeliness
- Data interoperability
- Data redundancy
- Staff delegation and cooperation

Each of these challenges is described in more details below.

2.1.1. Data Accessibility

Data accessibility refers to the ease with which data can be reached. Having quick and easy access to the data that is frequently needed to support decision-making – as opposed to having to go through a complex or tedious procedure to access data – is high on probably every planer's wish list.

The lack of data accessibility has often been identified as the main factor driving data integration efforts. For example, at the *Operations Data for Planning Applications: Identifying Needs, Opportunities, and Best Practices Peer Exchange* held in Washington, D.C., in May 2005, state DOT representatives concurred that data accessibility was a major concern when using operations data in planning (9). The inaccessibility was at times manifested from the lack of knowledge of data availability. It may also be attributed to institutional controls over data access.

2.1.2. Data Quality and Timeliness

Quality refers to the accuracy, precision and the level of detail in the data. For example, the quality attributes of spatial data would be accuracy of representing a physical location or resolution of the map etc. In the spatial context, having bad quality data refers to a missing node or an extra link or misreported geometry etc. Dealing with these kinds of network inconsistencies is an arduous task as it involves a lot of human intervention. In the peer exchange (9) cited above, the Center of Transportation Studies at university of Virginia reported that "*Metadata and data quality are two big challenges, after obtaining the access itself. In our experience, getting everyone to the table and sorting out data issues takes time. Fast prototyping allowed us to provide data in an incremental manner, instead of one final product at the end of the project.*"

Similar issues were faced during integrating distinct transportation data into a single Geo database Framework in Pend Oreille County, Washington. In 2006, Pend Oreille County was awarded funding to work in cooperation with Spokane County and integrate and consolidate the existing spatial data and make it available through the web (10). To accomplish this project goal, datasets from different sources had to be acquired. First, the County Sheriff/911 data is used as the network file containing all the roads. For county road number and milepost data, the Pend Oreille County engineering roads data is used. Also, Seattle city light roads dataset collected using the GPS is used to check the spatial accuracy of the project work (11). Later, the Washington Department of Natural Resources Othoimagery was used to alter spatial imagery of roads/intersections in river corridors. Figure 2 below shows the inconsistency in the roadway alignment represented by the different data sources.

Figure 2 Data quality and ambiguity faced by the Pend Oreille County (10)

Also related to data quality is the timeliness of data, i.e. how accurate the data is in representing the most current situation. Often in transportation planning, having frequently updated data is critical to support timely analysis. It is also important to maintain historic data for trend analysis.

2.1.3. Data Interoperability

Sometimes, even if a dataset is accessible and of good quality, the format in which it is made available could make it difficult to use the data. For example, the format might not be compatible to work with the planner's typical choice of analysis tools. This may result in additional effort and time to manipulate data and delay planning activities. Planners often need to manipulate data formats to suit the specific needs of a task. This could mean a minor manipulation such as converting data from Microsoft Excel spreadsheet to a more advanced application like SAS or SPSS. It could also be a major manipulation such as resolving different linear referencing systems. For each planner, whether a particular kind of manipulation is major or minor depends on the level of experience he/she has with the software tools.

Complete and consistent transportation data that have the same reference frames in spatial and temporal aspects are often desired to support transportation planning. Yet, since data are often collected and compiled originally for different projects and applications, such consistency is hard to achieve (12). For example, different roadway network representation models exist for different applications. Pavement management and engineering/construction applications have traditional CAD and aerial representation needs, but routing and transportation logistics require network connectivity models with road types and route numbers that can overlap. For some applications, street centerline could be used to represent transportation networks. For other applications, however, physical characteristics such as lanes and width would also be needed. Other issues involve 3-D problems in situations where the data needs to represent overpasses, tunnels, and elevated roads. Since one data set may not suit all purposes, ensuring the interoperability across differing data sources that describe similar entity would greatly enhance the ease of data use to support planning exercises.

2.1.4. Data Redundancy

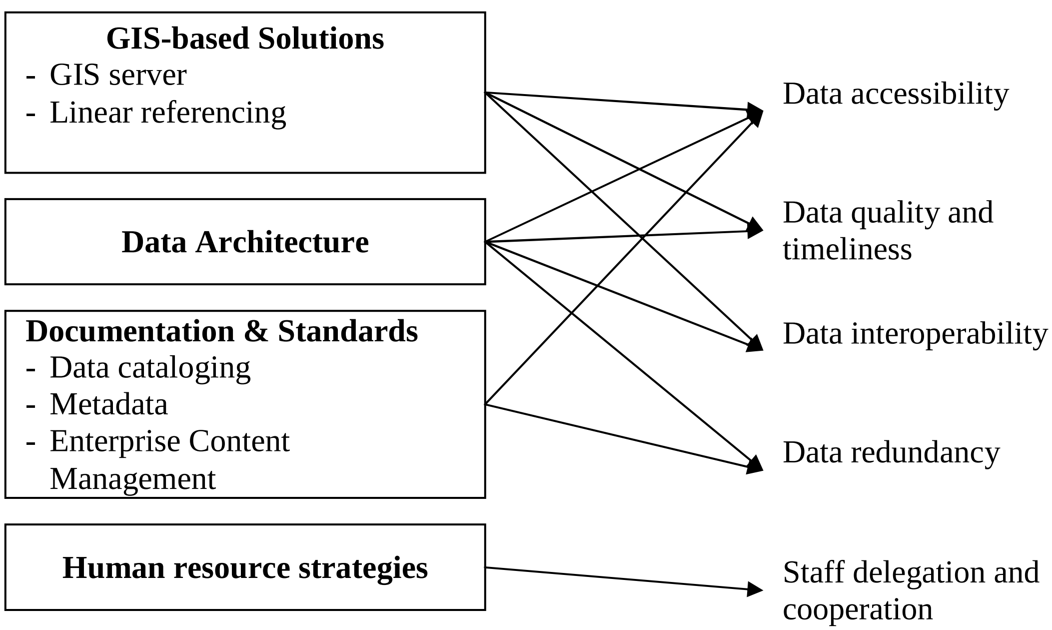
As pointed out earlier, with the number of organizations involved in collecting and maintaining data, there is bound to be some degree of redundancy. Our literature scan revealed that data redundancy is a problem faced by many state DOTs. For example, Kansas DOT has learnt that there are more systems (88+), databases (56), and technologies (250+) than expected and there is a great deal of redundancy in data and technology (13, 14). Also, many state DOT's already have a data sharing partnerships with other organizations or within the business units. But, the effectiveness of the sharing mechanism would determine the extent of redundancy. For example, prior to setting standards for its GIS application in 2001, Wyoming DOT's database connection allowed software to connect to corporate sources and to download a snapshot copy of selected data. "*The result of this process was that there were multiple copies of data on local machines, with the risk of redundancy and low integrity, due to changing values on the main database*"(15).

2.1.5. Staff Delegation and Cooperation

Typically, each dataset has a custodian who is responsible for collecting, maintaining, updating and sharing the data with the rest of the organization. The custodians are also responsible for communicating and cooperating with personnel in other divisions to ensure efficient data sharing. In practice, however, there could be cultural and organizational issues that inhibit healthy interaction between business units within an organization, thus presenting barriers to effective data flow and integration. For example, operations and planning staff typically reside in different agency silos. Therefore their priorities (e.g., short term vs. long term) often do not match (9). Also, many agencies are "getting by" with existing data and hence there is little motivation to adopt to change. Staff cooperation, knowledge transfer among the IT and other divisions are key to a successful data partnership.

2.2. Data Integration Strategies

In view of the various data-related challenges described in section 2.1, we turn our focus to the tools and strategies that have been adopted by different organizations to address these challenges. This section presents a list of common strategies across different data integration projects, citing appropriate examples of specific projects. The strategies are grouped into four categories: GIS-based solutions, database architecture, data documentation, and human resources.



2.2.1. GIS-based Solutions

As discussed in chapter 1, transportation planning data is often spatial in nature. Hence, Geographic Information Systems (GIS) are becoming more widely used among transportation planning agencies. The extension of GIS technology into transportation (GIS-T) has been driven largely by the requirements of the Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA), the Clean Air Act Amendments of 1990, and legislation by states that mandate the development of transportation programs to reduce traffic impacts (18).

Enterprise GIS solutions are fitting data integration tools for addressing data access, quality and interoperability issues. For example, sharing data through GIS server and publishing through web-based applications allow data to be readily accessible. Location referencing provides a means of attaining interoperability between disparate referencing systems. The use of metadata enhances data accessibility by providing users accurate information about the data content and format. These three GIS-based strategies are described in detail below.

<u>GIS Server</u>

With a GIS Server, organizations can (19):

- Take control of their spatial data through a centralized management of data, applications, and services.
- Provide fast access to large volumes of imagery using image services, thereby reducing storage costs and data processing overhead.
- Improve decisions and productivity with Web mapping services and applications that can be delivered to Web, desktop, and mobile workforces.
- Leverage existing IT architecture by integrating a GIS server and spatial data with other enterprise systems, such as customer relationship management (CRM) or enterprise resource planning (ERP) systems.
- Rapidly meet specialized demand for focused applications by mashing up geographic content with GIS functionality.

For example, San Bernardino County, California, needed an automated method for locating and displaying county road data that could be used by any planning department across the county (20). After the successful implementation of an online querying tool using ESRI's ArcGIS server web mapping application, staff could now easily locate and display individual county roads. The mapping application also helped integrate existing legacy databases and procedures. Thus, GIS server was said to have enhanced the data accessibility of staff at the San Bernardino County.

Prior to 2006, the transportation networks and databases required to support planning and modeling studies in the state of Florida – including the Florida Department of Transportation (FDOT) Work Program (WPA), Roads Characteristics Inventory (RCI), Florida Intrastate Highway System – Decision Support System (FIHS-DSS), Florida Standard Urban Transportation Model Structure (FSUTMS), Florida Geographic Data Library (FGDL), Efficient Transportation Decision Making (ETDM), and Strategic Intermodal System (SIS) – existed as independent databases with disparity in format and nature. As the amount of data and the size and complexity of multimodal transportation networks continue to rise and the number of data sources involved in planning continues to increase, FDOT recognized the need for a set of consistent, easy-to-use, and flexible data integration procedures to support the modeling and planning processes in the state (27). The solution, as recommended by researchers from the University of Florida, included a GIS platform through which the disparate databases were integrated and a user-friendly data access tool as a front end of the GIS.

Similarly in Colorado, GIS data collection standards have been developed so that any GIS data collected would be consistent and easily integrated to other corridor data (23). The GIS displays information on highway and aviation transportation assets and is updated through links to the databases. GIS is now applied to applications ranging from environmental impact analyses, project scope studies, identification of maintenance and bridge needs, and planning studies.

Location Referencing

Location referencing is one of the methods for bringing data onto a unified GIS platform. The reason for the special mention of location referencing is that most DOTs' use this method to resolve the issue of ambiguity in referencing the physical location. Each spatial feature can be attributed to the physical location using a particular referencing scheme. Typically, the location referencing problem pertains to the representation of streets/road network. For example, it is often found that organizations at different levels maintain a different way of representing spatial features of the network. Some key concepts related to location referencing are defined below (18, 24).

A *location referencing system* is a set of data and procedures that embodies the management of (one or multiple) location referencing methods. A *location referencing method* is a way of describing the location of an object or event relative to some known point (on the earth). A *Linear Location referencing method* (LLRM) allows the location of an object along a linear feature to be determined by specifying the direction and distance (along the linear feature) from any known (reference) point to the object. There are three general approaches to integrating data with different linear referencing systems (LRS) (13):

- Standardize on a single LRS to be used throughout the agency. For example, data collection, storage in all legacy systems, reporting can be performed on a single LRS.
- Maintain legacy data in multiple LRS's and create routines to translate these data into a common LRS for the purpose of integration; and
- Develop an exchange engine that translates data between various LRS's. If this approach is chosen, there is no common or standard LRS for the organization.

The use of multiple location referencing methods is a common data integration challenge. Figure 3 and Figure 4 below illustrate the concept of transferring the coordinate systems through direct and indirect transformations, respectively. Each circle represents a particular LRS. As shown in Figure 3, if we pick a methodology to translate the data from one system to other (often called direct transformation), the number of operations to be performed increases exponentially. The purpose of having such interoperability is to maintain relative independence among systems, with the only 'dependency' requirement being the interface to share data between them (18, 24). Adams, Koncz and Vonderohe (25) stated that three types of interoperability are possible: *Procedural Interoperability* through data and procedures that exchange information, *technical interoperability* through heterogeneous software and hardware component communications; and *institutional interoperability* through formal relationships between transportation agencies.

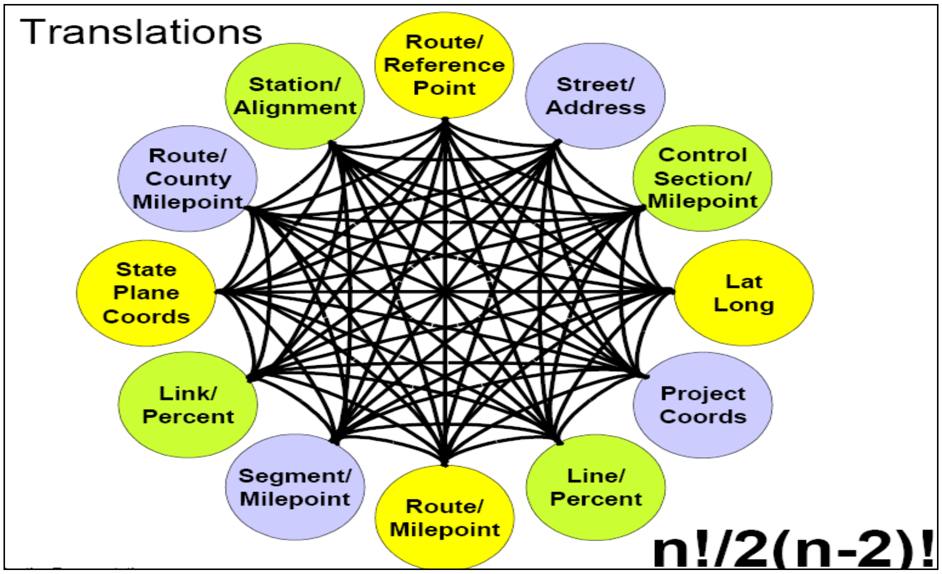


Figure 3 Translations from one LRS to other (direct transformation) (22)

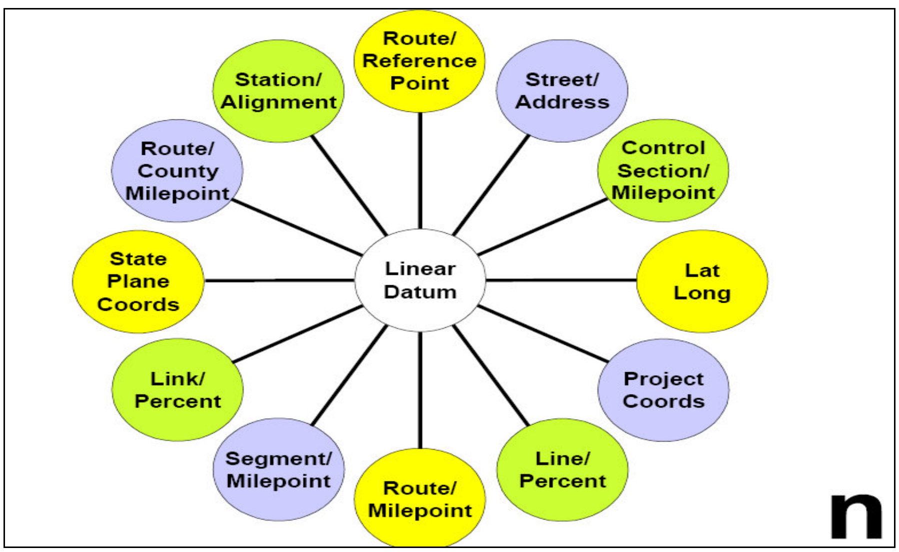


Figure 4 Translating Different LRS into one system (indirect transformation) (22)

When dealing with a large number of coordinate systems, however, it might be effective to convert them to a single linear referencing system (also known as indirect transformation) as shown in Figure 4. According to Ries (26), there are several advantages of using this indirect transformation apart from the less number of transformations required. Adding a new LRM would typically require adding only two transforms. Also, this approach would simplify the interoperable interface development effort. Depending on the number of LRS's that the DOT handles, direct or indirect transformation is chosen. Below are a few examples of how planning agencies addressed the issues of location referencing.

According to the North Carolina State DOT, modeling the road network to integrate all the attributes of road data corresponding to a specific physical location presented a major challenge: "*The challenge is to capture attributes that vary by length without duplicating geometry. Linear referencing efficiently meets the representational challenge. It allows virtual networks to be constructed upon a single, stable base network. Attributes are attached (referenced) to linear features along virtual networks using relative (known locations and offset distances) rather than absolute (x and y coordinates) measures. Linear referencing simplifies spatial data management by eliminating the need to maintain separate coordinate geometries for each portion of a road where attribute values may vary independently*" (23). NCDOT's data integration effort focused on integrating their four most commonly used linear referencing methods with minimum loss of accuracy. The result is a LRS composed of two main components: a road line-work geodatabase and the LRS Database Core Module. The geodatabase stores all explicitly spatial data for the state-maintained road network. The LRS Database Core Module represents the road network in a non-spatial, tabular format utilizing an Oracle9*i* database running on Red Hat Linux servers. It has a set of related subcomponents that model the transportation network, the NCDOT route number system, and change histories for both.

Similar challenge was faced by the FDOT, which found the discrepancies among street data used at different levels of transportation planning. Such discrepancies were said to "hinder efficient exchange of information among related transportation planning applications" (27).The solution recommended to FDOT by researchers at the University of Florida entails adopting the commercial product Dynamap/Transportation Streets, which includes local streets and is updated every 6 months. The solution also includes a GIS data association tool that help transfer attributes between different reference data.

The Pend Oreille County addressed their street data discrepancy issues (as discussed in section 2.1.2) by adopting a GIS platform and the direct transportation method. As shown in Figure 5, after the integration of data and conversion into a single LRS, the network is now represented with a single line. It should be noted that, in cases where automatic conversion through computerized algorithms is not possible, this exercise of correcting the referencing system is very time consuming and hence should be implemented in a phased manner to see better results.

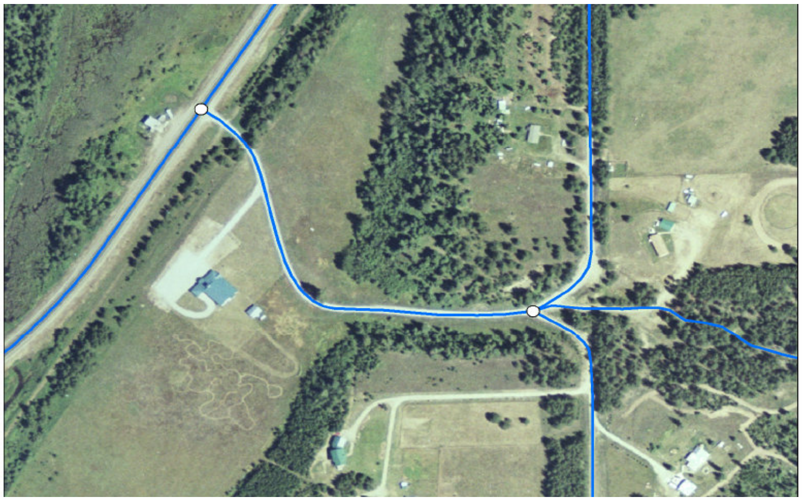


Figure 5 Pend Oreille County's solution to their data ambiguity issue (10)

2.2.2. Data Architecture

The data flow process within an organization is dominated by the design of the data architecture of the organization. The integration of disparate existing databases to enhance the flow of information thus requires agencies to examine and rethink about their existing database architecture. The process would entail carefully studying the existing data flow mechanism within an agency, identifying the pitfalls, and developing a new data flow process. This process would help identify the databases to be integrated and provide a blueprint for the data integration effort. A good data architecture would therefore address data challenges pertaining to maintenance, accessibility and redundancy.

Typically, there are two approaches for designing data system architecture: (a) centralized data warehouse, and (b) interoperable databases. These approaches are discussed below.

Data warehousing

A centralized data warehouse is typically designed to span multiple divisions, departments and functions throughout a complex organization. It serves to consolidate and integrate data from multiple operational systems. A data warehouse database is typically created with a database management system such ass Oracle, Sybase or SQL Server. Figure 6 depicts a typical data warehouse architecture. Described below are a few projects that have recently adopted the data warehousing strategy.

IDOT employed a methodical and incremental approach to spatial data warehouse design (34). Figure 7 shows the data architecture that is implemented. One notable feature of Iowa's effort is the embedding of the underlying link node structure into roadway inventory databases. This enabled direct linkage of data using other linear referencing methods. As can be seen from the figure, the central database is the spatial data warehouse which is linked to various other link/node referencing scheme. It is similar to the indirect transformation of the linear referencing systems described earlier. The spatial data warehouse acts as a data storage location consisting of data regarding political boundaries, streams, lakes, aerial photography, census, environmental, traffic, pavement conditions, railroad data etc.

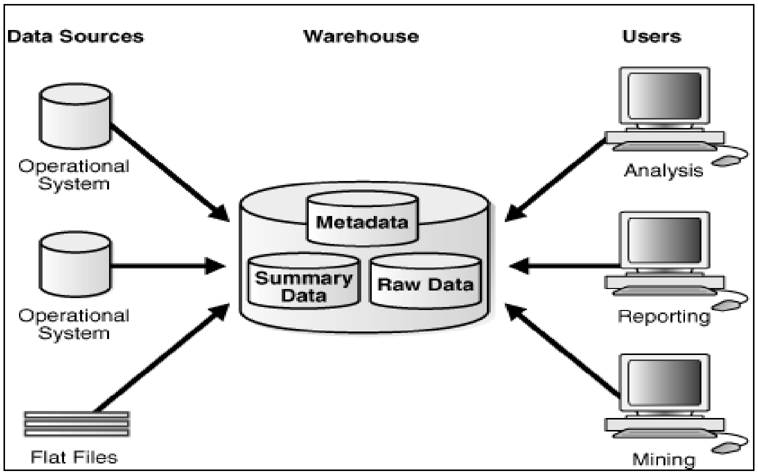


Figure 6 Data warehouse (33)

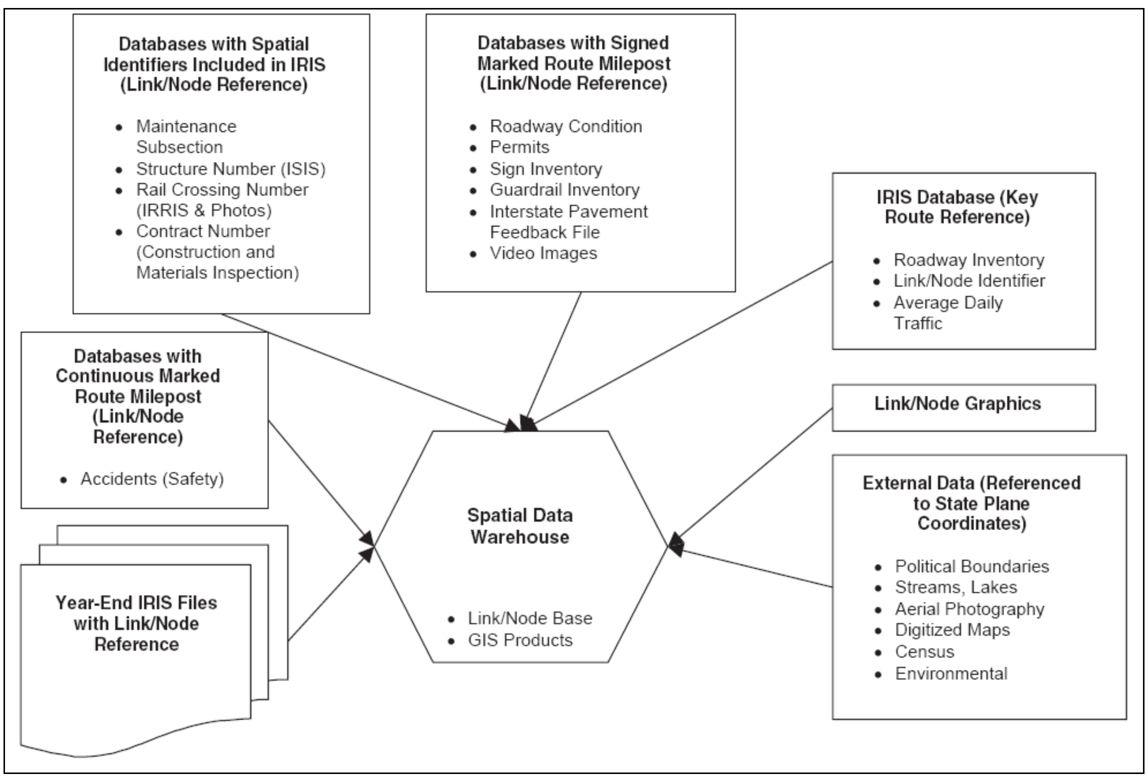


Figure 7 IDOT Spatial Data warehouse architecture (34)

The Arizona Information Data Warehouse (AIDW) is an online analytical processing system that serves as a read-only repository of information (31). The basic architecture is shown in Figure 8. The system uses a Windows 2000 Server and SQL Server 2000. Microsoft's Analysis Services and Data Transformation Services (both of which are components of SQL Server 2000) are used to provide online analytical processing capabilities. ESRI's ArcIMS and ArcSDE (Spatial Database Engine) software are used to provide an agency-wide GIS solution for performing queries and viewing results without requiring users to store large shape files locally. ADOT is also developing a series of business intelligence tools to enable users with little or no technical training to readily access integrated data. Users will be able to query data with an online interface or GIS with point-and-click and drag-and-drop query capabilities. Users will be able to retrieve summary data or drill across the data marts to get information on projects, traffic, accidents, features, maintenance history, and other items at any given milepost.

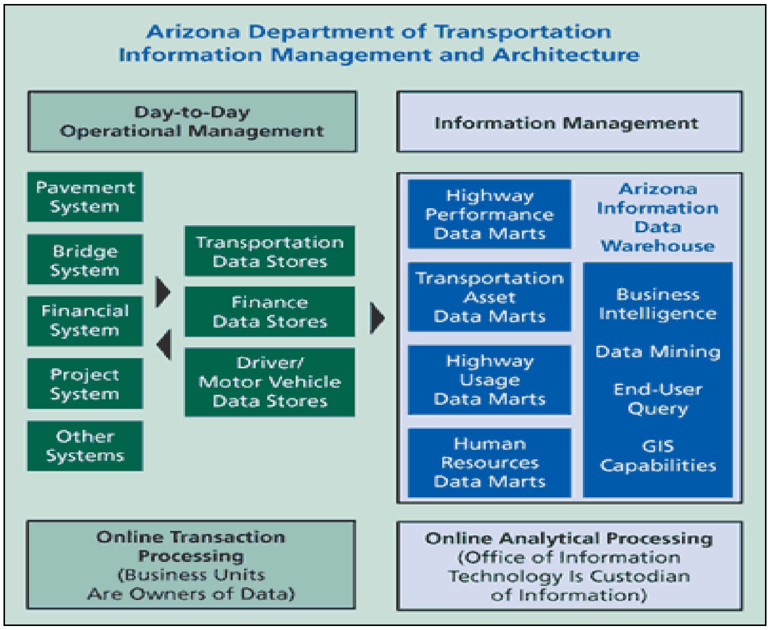


Figure 8 Arizona DOT's data architecture (31)

Delaware DOT recently developed an integrated, multimodal transportation management system called Deltrac. A key feature of the Deltrac system is an Oracle database with a web interface. The database stores transportation data from both legacy systems and new components, and allows DelDOT to perform both planning and real-time operations using a single database. All users can access the system through a uniform (Extensible Markup Language) XML-based interface from any computer with access to the web.

Geo-Referenced Information Portal (GRIP) is a geo-referenced information system that integrates diverse locational data from a variety of data sources within Florida DOT (13). It provides users an interface for access to the centralized database (including legacy systems) via FDOT's intranet. It consists of an integrated multi-layered GIS spatial database as well as attribute and image servers residing in each of the FDOT districts. Personal computers connect to the server via the intranet through Local Area Network (LAN), Wide Area Network (WAN), Remote Access Server (RAS), or Virtual Private Network (VPN) connections. Users access GRIP through Internet Explorer. The application is designed to meet the needs of heavy as well as casual users. Oracle 8.17 is used with the full spatial capabilities to maintain the GIS data. In the initial version of

GRIP, Formida Fire, an advanced application which links complex data including spatial, image, video, audio, and time series, was used to serve it to the web. GRIP is an application that allows centralized management and supports an unlimited number of users.

Interoperable databases

An interoperable database system – also referred to as a federated database system – is a database front-end which communicates with other databases of different formats so that they all work together, and data from one system can be accessed from another system. Figure 9 depicts a typical interoperable database system, in which many databases can interact with each other and also interact with a main database of the organization.

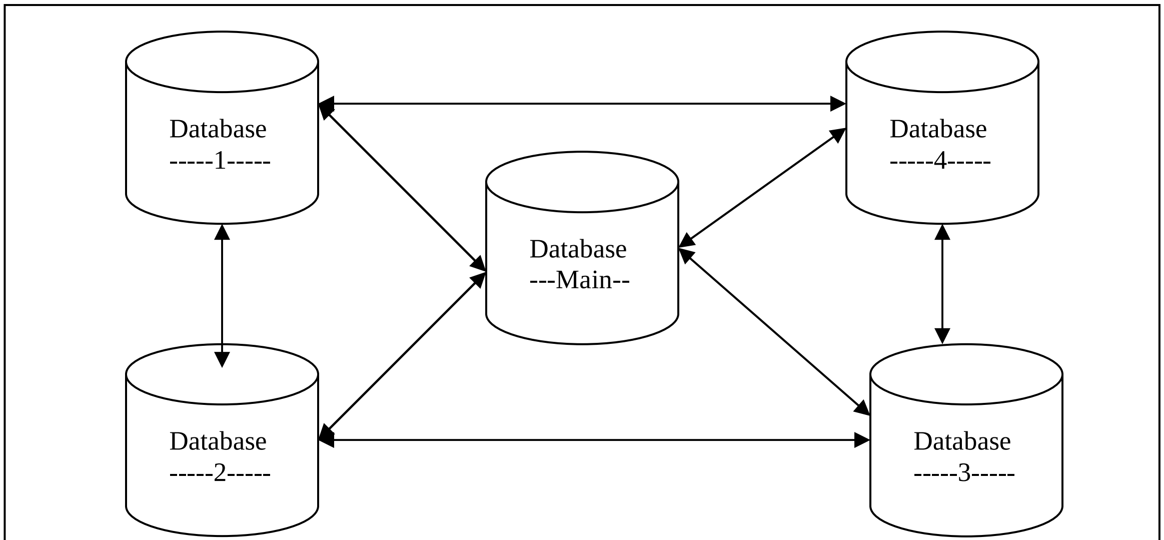


Figure 9 Interoperable database System (34)

An example of interoperable database approach was found in Vermont DOT, who identified more than 200 concurrent users over a wide area network over two terabytes of imagery. Consequently, the system requirements specified a large number of hard drives on data servers, massive data backup, high bandwidth, and significant performance tuning. The solution was to develop an interoperable database system that separates the business/spatial data from the image data. The system configuration includes two data servers, one for the image data and one for the business/spatial data, and one application server for ArcIMS. The data servers also host the GIS software (ArcSDE) and Oracle DBMS. Both sets of servers use Windows 2000 as the operating system.

On the one hand, a data warehouse can handle more data and control the whole system easily while requiring considerable time and resources to implement. On the other hand, an interoperable database would keep the files in independent locations and help preserving the autonomy of the files while requiring rigorous procedures for database access and updates. The advantages and disadvantages of both approaches are summarized in Figure 10 below. The choice between the two approach depends on the number of databases involved, storage capacity, and other resource availability.

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.

Figure 10 Comparison between fused and interoperable database (7)

2.2.3. Documentation and Standards

Data Cataloging

Not all data needs to be integrated. In fact, prior to start considering the previously discussed data integration strategies, most agencies recognize the need to carefully document the available data resources in the form of a data dictionary or data catalog. For example, the documenting process allowed the Colorado DOT to identify, and subsequently eliminate, duplicate data items (31). The development of data dictionary and catalog, however, could be a time-consuming undertaking that in turn delay achievement of visible results (13).

Metadata

A formal way of documenting data is through metadata. Metadata are data about data. A metadata record typically captures the who, what, when, where, why and how of a data or information resource (30). Thus, developing metadata could help eliminate data redundancy and enhance data accessibility.

According to the TRB Metadata working group (29), metadata are important to the transportation community for several reasons:

- *Metadata maintains a transportation organization's investment in information resources -- Metadata provides automated, searchable access to information resources so that employees or clients can find the information they need with minimal time and effort. (Example: search for GIS coverages in a particular format for a particular area)*

- *Metadata provides information necessary for data to be understood and interpreted by a wide range of users -- Thus, metadata is particularly important when the data users are physically or administratively separated from the data producers. Metadata also reduces the workload associated with answering the same questions from different users about the origin, transformation, and character of the data. (Example: traffic datasets posted on a state DOT website)*
- *Metadata enables data to be discovered and used to its full potential – Metadata may also provide information about intended or planned uses (as well as limitations), which can assist the data users in realizing the full potential of the data. (Example: crash records datasets for different time periods, and using different criteria for what crashes are included)*
- *Metadata facilitates the operation of federated database systems (i.e., distributed holdings) -- Metadata enables data to be centrally stored with searchable interfaces; thereby, providing a single access point for varied types of data resources. (Example: metadata repository that enables access and content details for databases held by data owners who agree to have their data listed).*
- *Metadata extends the efficiency and reliability of discovery and utilization processes. By embedding metadata in statistical and modeling elements and clarifying between raw and derived or estimated data elements, the outputs from hypothesis tests, model tests, or other statistical analyses can be more readily compared, evaluated and disseminated (Example: compilation of information which enables efficient and accurate use of travel survey datasets, e.g., source agency of data, data collected, data quality, caveats, and other descriptive information)*

Metadata in conjunction with XML-based specifications, schemas, and tools allow a high level of automated and validated inter working between different types and sources of data (32). There is a need for a formal approach to specify metadata, that is the vocabulary that is used and the scope it covers. For example, metadata also should include the details about the source of the data, that is, who collected it and where was it processed and edited. The Federal Geographic Database Committee (FGDC) selected standards for the role of metadata in GIS applications based on availability, fitness to use, access and transferring of data. Geospatial metadata are used to document geographic digital resources such as GIS files, geospatial databases, and earth imagery. A geospatial metadata record includes core library catalog elements such as title, abstract, and publication date; geographic elements such as geographic extent and projection information; and database elements such as attribute label definitions and attribute domain values (30).

In addition to developing metadata, several planning agencies across the country have spent time and effort in designing tools to help user search through metadata to locate the dataset of interest. Most of these efforts are limited to the GIS data in the organization and do not include non-GIS data. For example, King county GIS center knowledge base contains a searchable Spatial Data Catalog (SDC) (35). The data is grouped by agency (eg. DOT), subject (e.g. transportation) and format (e.g. vector vs. raster). Similarly, New Hampshire DOT's GIS section data catalog (36) also provides GIS maps grouped by subject such as functional class, bridges, and facilities.

Enterprise content management

Enterprise content management (ECM) is a set of technologies used to capture, store, preserve and deliver content and documents and content related to organizational processes. ECM tools and strategies allow the management of an organization's unstructured information, wherever that information exists. Therefore, ECM provides a way of developing a 'smart' data catalog that also serves as a platform to search through metadata.

Washington DOT recently made the transition from a simple metadata repository to a comprehensive ECM system (37). The WSDOT GeoData Distribution Catalog, maintained by the Office of Information Technology, is a centralized distribution site for geographic information system data produced at the Washington State Department of Transportation. The agency management supported the need and request for funding to accomplish ECM. Washington DOT had its first metadata repository produced in 2001 and it was further developed with a user interface that served as a repository and provided both a list of data sets and user search tools. The data catalog was developed in-house using methods developed from the engineering domain. This resulted in an inflexible design that did not meet the information search and retrieval needs of our users. The agency management understood the need for a more organized data resource and obtained funding through the Washington State Legislature.

The initial phase of the project dealt with educating the WSDOT Information Resource Management group on controlled vocabularies, content oriented metadata as opposed to the technical metadata found in databases. Next step was to develop a classification scheme geared to the WSDOT knowledge worker. The classification scheme developed was a strict mono-hierarchical taxonomy that described the data in precise technical terms. The classification scheme relies on the grouping the metadata into business topics. The business topics were: assets, construction, environment, geospatial, management and organization, operations, persons, planning and design, projects and development, structures, traffic way, uncategorized, utilities, weights and measures. The GUI of the search tool is shown in Figure 11. In the search tool, business topics and data subjects are linked. This ensures that the search from the home page includes all the synonyms in the results searched. For example, searching for the term 'crash' would lead us to a screen as shown in Figure 12. The related business topic is highway safety and 'Collision' is one of the synonyms for the word 'crash'. By selecting collision, one can view all the metadata about collision such as what type of data is collected and data stewards. WSDOT's data repository contains 500 databases, 18000 tables, 240,000 data elements. Currently, bi-weekly automatic update maintains an inventory of databases (relational and mainframe). One drawback of the current design of the catalog is that it does not allow for separate navigational taxonomies for different user communities (developers, knowledge workers, data stewards).

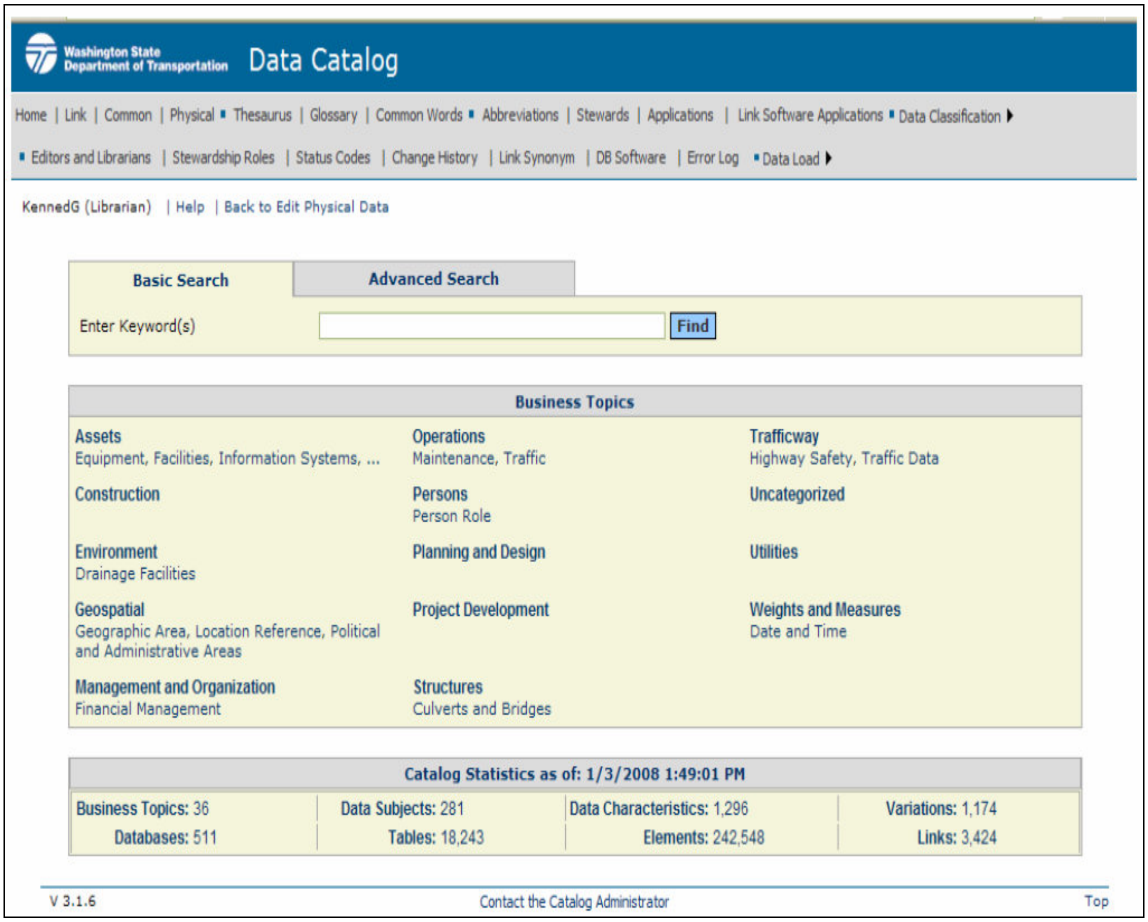


Figure 11 GUI of WSDOT's search tool (37)

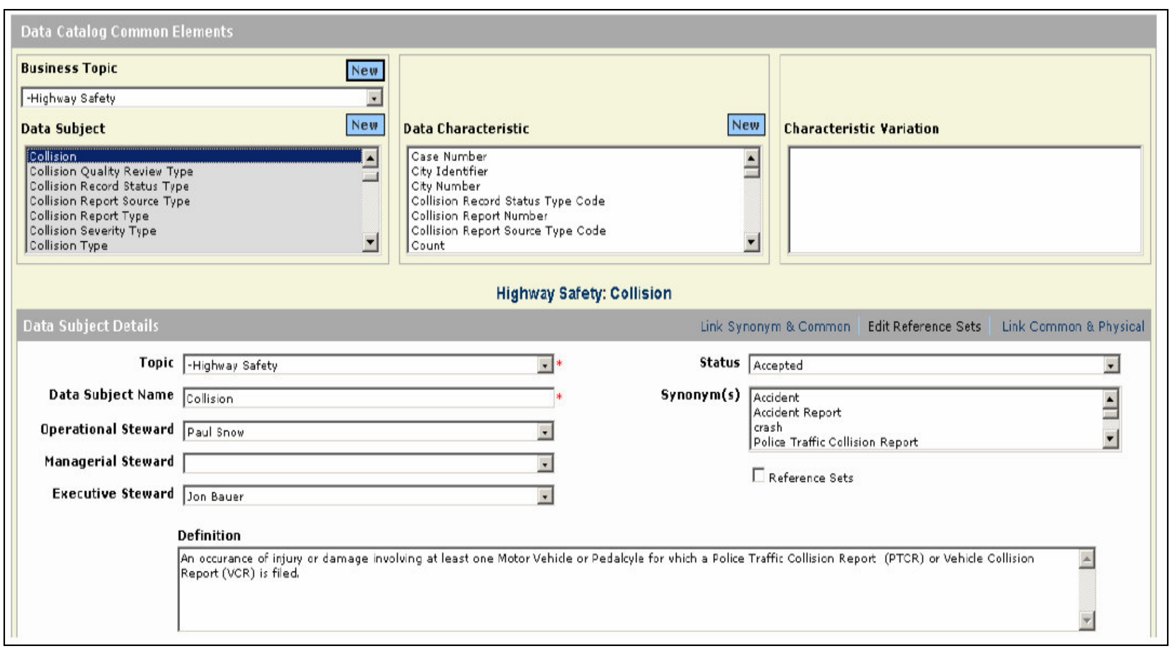


Figure 12 Example of a search result returned by WSDOT's online tool (37)

2.2.4. Human Resources

Data integration, as technical as it sounds, also has another critical angle to its success. It is the way the human resources are trained and equipped to maintain the integrated system. More often than not, many integration projects require an active interaction among staff from many departments within the organization and also other organizations. Also, there will be basic cultural and organizational differences in the functioning and hierarchal classification among various organizations.

Strong partnerships between IT staff and agency practitioners are required for a successful agency-wide IT initiative. In some cases, IT efforts motivate reluctant practitioners to change. In other cases, the practitioners want to move forward quickly and need to be restrained by IT staff to insure consistency with an overall strategy.

A strong mandate for a comprehensive data integration initiative from above is unlikely to happen. Bottom-up desire is usually strong but can be uncoordinated, particularly across divisions, and by itself cannot provide the impetus for moving forward. A critical success factor is a carefully blended mix. Management support is required to ensure that the appropriate tools and resources are available. Bottom-up support is required so that a regression does not occur when the current management leaves.

As part of Arizona DOT's data integration efforts, Arizona Information Data Warehouse (AIDW) staff worked to help operational units (e.g., planners and engineers) understand that they own the data they collect and are responsible for its integrity (31). It was considered that IT staff were merely custodians of the data, and the data warehouse merely a tool with which to access the data. By adopting the principles of information resource management, ADOT is working to establish an agency-wide data culture. This cultural view is that data are valuable and that strategic resources need to be managed in the same fashion as human resources and capital assets. It stresses that data needs must be driven by business process needs. Clear ownership of data items must be assigned, and owners must be fully accountable for meeting data standards. Adequate financial and human resources must be allocated for data collection and management. In this context, the AIDW is viewed as one tool under the information resource management umbrella.

The New York State Department of Transportation (NYSDOT) has undertaken an effort to design and implement an asset management system for overseeing the state's diverse and complex transportation system (37). The department has built an appropriate organizational and business foundation for the effective use of sound, integrated databases and technical modeling tools. One of their primary focus areas was developing well-defined organizational roles within a highly decentralized department. When identifying and developing additional technical elements, state departments of transportation should heed the importance of building and maintaining an effective organizational and business foundation.

In a study done in nine states about organizational location of data in DOT’s, it was found that most interviewees believed that the organizational arrangement was in itself not a significant factor in successfully managing planning data (37). Two important factors reported from this study, though not directly related to organizational structure are leadership and involvement of IT Staff. The report notes that “*effective data group managers should have experience in, or at least a solid understanding of, information* *technology and should also understand how data is ultimately used*". Needless to say, such DOT's with high priority on data management, along with supportive technical expertise and capable executive managers would have a better decision making process.

These above mentioned examples suggest that having a HR strategy is a key element towards a move to integrated data. Ultimately, it is the staff of DOT that manage data. Data integration effort that does not consider HR issues such as management expertise and data stewardship would be at high risk of failing.

2.3. Implementation Considerations

Integrating an existing system of data structure for future benefit might affect the current planning process. Care must be taken to implement the integration process without affecting the day-to-day planning procedures in the organization. A Key issue with most data integration projects is the time consuming, data-intensive and voluminous work involved. Hence, it would often put off clients/organizations because the results of this effort are not readily available. In this case, a phased approach may be desirable.

For example, the Maine DOT's effort in developing its Transportation Information for Decision Enhancement (TIDE) data warehouse was unique in its kind at the time (13, 31, 40). Not only was a phased implementation adopted, but also there was an iterative process involved which took regular inputs from the end users of the system. In phase 1 (as shown in Figure 13(a)), the scope of the project was limited and focused on integrating two databases TINIS and PMS data. Also, a platform for integrating the data and sharing between different users was created using GQL queries and other tools mentioned above. In phase 2 (see Figure 13(b), the entire backlog data is linked on a GIS platform and a location referencing management system was implemented. Other advanced functionalities like viewer for videolog images, more layout choices for ArcView, SQL queries directly from ArcView were added. Also, the scope was extended to sharing data with other agencies like MPO's and other state agencies.

Data integration efforts often incur unanticipated costs. Hence, a careful examination of the cost of project should be done at the time of proposing a new data flow model. At this stage, alternate strategies for data architecture should be discussed and the one with least cost and relative ease of practical implementation should be picked.

If the proposed model includes collecting and maintaining data at the same physical location, storage capacity would also be in an important factor of consideration. Provision for adding new data fields in future should also be made. This issue of storage capacity is interrelated with data quality. Sometimes, if huge amount of data has to stored, trade off should be made with the level of detail to which we require data. Hence, there is a need for one system which encompasses all the data available but makes only the required part available during specific operations. Storage capacity may not directly influence the implementation phases but it could affect the data architecture (fused vs. interoperable) which would play a key role in the implementation stages.

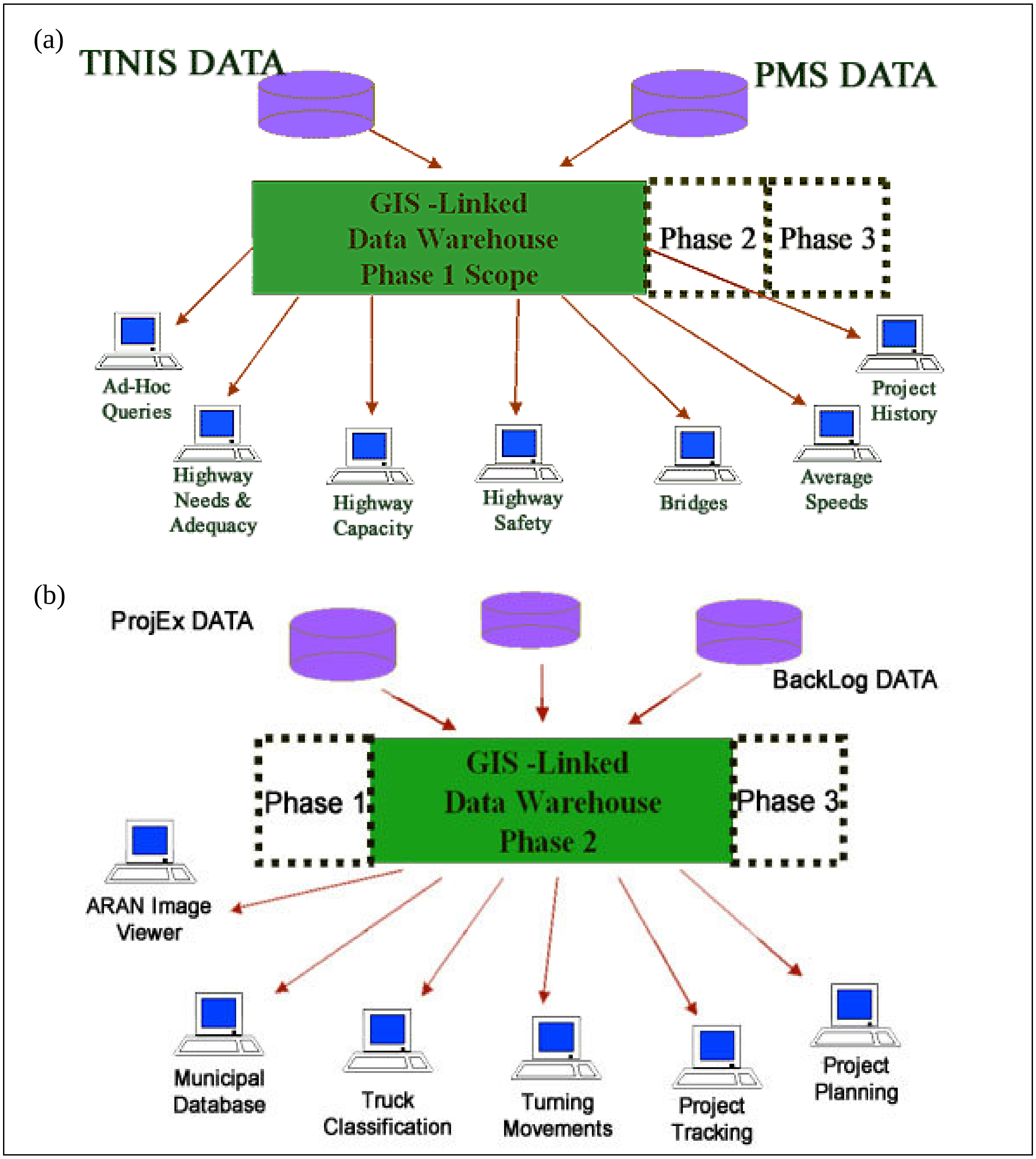


Figure 13 Phased implementation of MeDOT's TIDE data warehouse,

2.4. Summary

In summary, this chapter reviews past data integration efforts by identifying the motivation for these projects, the main issues that are to be resolved and the data integration tools and strategies that are adopted. Some of the common themes across all the projects presented above are summarized as follows.

- There has been an extensive application of GIS platform in all the projects to represent all the data in a compatible linear referencing system.

- Also, an Oracle based platform for querying or database management seems to be one of the popular tool for data integration. Another tool popular among states is using a java application with XML for web client application.
- Of all the data related to transportation engineering, there is a greater emphasis on including pavement and bridge data and hence these broadly come under the data integration requirement for asset management. Chapter 1 established that the transportation planning data requirements encompass a huge variety of data when compared to asset management. However, most projects in literature deal with data integration for asset management purposes only. This is a visible gap in current literature.
- Organizational issues play a key role in these projects where there could be a constant resistance to change from the staff. The fact that there would be varied knowledge of IT among DOT's staff could lead to inefficient process.
- Benefits of integration will lag its implementation and are not very easily quantifiable in short run.
- Preparing a data dictionary or a catalogue to maintain a good quality of metadata in the database has been given a key emphasis during data integration projects.

3. CURRENT DATA PRACTICES AND CHALLENGES

The first step to assessing the needs and best approaches for data integration is to establish a good understanding of the current data practices and challenges. As mentioned in Chapter 1, the research team conducted a survey and a series of interviews with WisDOT personnel for the purpose of establishing such an understanding pertaining to transportation planning in WisDOT. This chapter details the processes and findings of these efforts. Specifically, Section 3.1 describes the process and instrument involved in the survey to staff members of the Bureau of Planning and Economic Development (BPED) at WisDOT. Section 3.2 summarizes the responses of the survey respondents. Section 3.3 presents the major findings from the series of interviews with various WisDOT personnel, ranging from data users to data custodians and information systems managers. In Section 3.4, findings from the survey and interviews are summarized and their validity discussed.

3.1. Survey Process and Instrument

In the spring of 2008, the research team conducted an on-line survey to planners at the BPED of WisDOT to uncover the current data practices and future needs. The focus is to identify both technical and organizational barriers to using data sources internal and external to WisDOT for the purpose of statewide transportation planning. Invitation to the online survey was sent out in the format of an email to all BPED staff members. The email contained a link to the survey website, as well as a copy of the questionnaire in MS Word format to provide an off-line alternative to filling out the survey. The participants were asked to respond online or return the completed survey in MS Word format via email.

The questionnaire was divided into five sections as shown in Table 1. Part A collected basic information about the respondents, including frequently used software tools and job functionality. Parts B and C were both concerned with the following four aspects of frequently used data: accessibility, quality, format and timeliness. While part B focused on these aspects as they relate to data sources internal to WisDOT, part C focused on these aspects in the context of data shared between WisDOT and other organizations. In part D, the respondents were asked to report any unmet data needs that they currently faced. In part E, each respondent is asked to provide − to the best of their knowledge − the datasets that they use frequently to support their statewide planning activities. For each dataset, information about key characteristics such as data format, metadata availability and data custodian were also collected. The reader is referred to Appendix A for a complete copy of the survey questionnaire.

Table 1 Questionnaire structure

Survey Section	Theme
Part A	Basic information
Part B	Data accessibility, quality, format, and timeliness of data within WisDOT
Part C	Data accessibility, quality, format, and timeliness of data external to WisDOT
Part D	Unmet needs
Part E	Data inventory

3.2. Survey Responses

A total of nine BPED staff completed the survey (all online), yielding a response rate of about 30%. Responses were received from all three sections of the BPED (Urban Planning, Economic Development, and Travel Forecasting). The findings from the survey responses are summarized below by section. For the ease of discussion on the data issues that the respondents have experienced with specific data sets, we first present the commonly used data items as collected in Part E. This is then followed by our discussion of survey findings from parts A, B, C and D, respectively.

It should be noted that, given the response rate of 30%, care must be taken when interpreting our survey findings. While the experiences and views shared by these respondents provide valuable insights for this investigation, they do not necessarily apply to the remaining 70% of the BPED staff. The reader should also keep in mind that, as with any qualitative survey, the findings reported below do not necessarily reflect the 'true' data-related practices in WisDOT. Rather, they reflect the respondents' experiences and perceptions that may (or may not) vary from individual to individual. For the purpose of this project, information about inaccurate perceptions is just as useful and important as information about the true practice. While both type of information are reported in this section, our subsequent discussions (Section 3.3) will attempt to differentiate these two types of information from each other based on our more in-depth interviews.

3.2.1. Part E: Data inventory

The main purpose of this section of the questionnaire was twofold. First, it provided the research team an inventory of the frequently accessed datasets and insight into the flow of data within WisDOT. Second, this data inventory served as the starting point for the data catalog that is developed as a part of this project (see Chapter 4).

Identified by the survey respondents as frequently used datasets are:

- Wisconsin Local Roads system (WISLR) – a stand alone application developed mainly for the administration of local roads by regional agencies; contains a broad range of physical and administrative attributes of roadways (44).
- State Trunk Network (STN) – a GIS database of centerline files, shape files and tables for roadways in the STH system, containing attribute data about State,

Interstate, and National Highways that support the national roadway infrastructure within the State of Wisconsin.

- Financial Integrated Improvement Programming System (FIIPS) – a database of information relating to transportation improvement projects.
- Meta-Manager – short for Meta Management System, is currently used to integrate various data sets pertaining safety, mobility, pavement, and bridges; supports the Six-Year STN Highway Improvement Program (45).
- Photolog – includes images of Wisconsin highways and the adjacent environment (48).
- TRADAS – short for the TRAffic DAta System, is a database that contains traffic volume data (*e.g.*, roadway volume, speed, vehicle classification, length classification, and weight); supplied to WisDOT from a private firm Chapparal Systems inc. (47).
- U.S Census Population Data – centennial population demographic data provided by the U.S. Census Bureau; contains socio demographics of residents classified by blocks, block groups, tracts, counties, and other geographies.
- DOA Population projections – Wisconsin population projections developed by the Demographic Services Center; produced annually.
- *Reference*USA – an Internet-based reference service from the Library Division of *info*USA which provides data that relates to residential socio-demographics and business data.

3.2.2. Part A: Basic Information

All nine respondents reported that they were analysts (or advanced analysts) dealing with various WisDOT planning activities. They were all familiar with the concepts of GIS and were comfortable using the ArcGIS software.

3.2.3. Part B: Data sharing within WisDOT

In this section, respondents were asked about their experience relating to data accessibility, data quality, data format and data timeliness in the context of the agency's internal data. With regard to **data accessibility**, half of the survey participants disagreed with the statement – "I always have good access to all the data needed for my job function." Eight of the nine participants reported technical factors hindering their data access. For example, it was reported that there is no clear information about the specific data elements stored in the SDE[1], a platform employed by WisDOT to facilitate data sharing across the agency. Although a dataset may be available, not knowing its exact location details would hinder the data accessibility and cause delays in the planning activities. Organizational factor is also among the barriers to data access for certain items. For example, it was reported that "*Really only one person understands the traffic data*

[1] SDE is an oracle-based ArcSDE server which also stores the layers and tables in a geodatabase. ArcSDE is a server software developed by ESRI that spatially enables a Relational Database Management System (RDBMS). It integrates geographic information query, mapping, spatial analysis, and editing within a multi-user enterprise DBMS environment.

within the TRADAS system. If that person is not available, you might not have access to the data."

With regard to **data quality**, most respondents expressed satisfaction with responses such as "*Once I find the contact for the spatial data I need, it's good quality"* and *"Very rarely do I need to question those that I work with on the quality of the data. Most people in this type of work that are collecting data and reflecting data are very detailed and do their job very well."* However, the survey also revealed three main issues about data quality. First, two respondents raised concerns about the accuracy of STN and pointed out the differences between the STN and WISLR data. It was reported that the network representation in the WISLR system is more accurate than STN road network. Second, there is concern about the inconsistency among multiple sources of air and port/transportation data: *"There has been an issue where air freight data collected by the Bureau of Aeronautics (provided by individual airports) doesn't exactly match data collected by the FAA. Perhaps not a huge issue. But it makes you wonder whose data is "correct". We've had similar issues with freight data for Wisconsin ports. We've had some data provided specifically by each port. Other data has come from the Army Corps of Engineers. Whose data do we trust?"* Third, a few minor discrepancies have been reported about traffic counts data: *"While traffic data are inherently variable, is seems like few quality checks actually get done on the data. Finding traffic counts that inexplicably double or half from year to year is not uncommon."*

The responses were mixed regarding **data format** and interoperability of datasets across different platforms. Almost all the respondents reported that they *often* need to perform minor manipulation and *sometimes* need major manipulation to use the datasets. For example: "*One of the problems is that 'planned highway routes', for example, are not part of our data systems; so, these have to be created and re-created often."* "*Minor manipulations, such as aggregating communities by county for county-wide totals are completely understandable. Often not having organized records of the termini of planned or already constructed projects requires major manipulation."* In addition, not having consistent variable names or lack of organization between different versions of files would require some manipulation in the data as cited in the following response: "*Data formats and field (variable) names often change across different databases and even data tables (especially in TRADAS)."*

Regarding the **timeliness** of the data, most planners expressed dissatisfaction at the recentness of data that is available to them. Specifically, it was reported that the traffic accident or counts data seem to be lagging. For example, *"Much of the data lags by several years i.e. - traffic accident data."* and *"The monthly traffic data are usually timely, but there is often a several month delay in getting the annual data each year."* are two responses that point to the lack of timely updates in traffic counts/accidents data. As expressed by a respondent: *"Planning data is not up-to-date most of the time and there is no set time period (for example, once a year) when it is reviewed and updated."* An interesting dimension to this issue pointed out by a survey participant deals with the organizational issues in updating data: *"Some of the spatial data I acquire from the geobase (central location for GIS data) is old. For example, this may be data from the DNR. Whether this is the most recent data available or whether the DOT partly* *responsible for acquiring the data, has not had time to contact the DNR for new data, I'm not sure."*

3.2.4. Part C: Data sharing with agencies outside of WisDOT

Part C is concerned with the issues related to data sharing practices between WisDOT and other organizations. A host of federal, state and local agencies were reported to share data with WisDOT, including MPO's, RPC's, US Census, US Federal Railroad Administration (FRA), Department of natural Resources (DNR), Department of Commerce, Department of Tourism and Department of Administration (DOA). The two major issues with inter-agency data sharing are found to be (1) timeliness of data and (2) data interoperability.

With regard to **timeliness**, one of the respondents said: "*Many times the data we access comes from the MPO long range plans, which are static documents and a snapshot in time, so the timeliness is the biggest issue."* Organizational factors also impede frequent updating of data. For example, "*The issue about timeliness relates in that one of our functions is to analyze data and the other is to report it. Depending on if FHWA has 'approved' the data, we wait to report it. It's almost like having two accounting systems."*

With regard to **interoperability**, two responses cited inconsistent data format: *"FRA data models are not consistent with Wisconsin DOT models. It makes data entry from federal database cumbersome."* This response referred to the inconsistent field (variable) names between FRA models and WisDOT models. Another respondent pointed to the lack of compatibility between data from Army Corps of Engineers and individual ports. "*Army Corps of Engineers waterway freight data is organized terribly. The data for individual ports is available via PDF documents, but they are not organized alphabetically. They seem to be organized roughly by region, but you have to dig through the documents. Larger ports like Chicago are presented in a variety of ways. Some summaries include certain components of the inland waterway system, but not others. It's unclear which data a user should actually us."* Also, it was reported that getting population data from DOA requires major manipulation as a lot of effort have been into mapping that data to a GIS format. A survey respondent pointed out that finding the correct contact person at DNR is sometimes an issue and it was unclear about what data items WisDOT and DNR host.

3.2.5. Part D: Unmet needs

The respondents were asked to identify any unmet or future data needs that they might have. Apart from the issues and needs mentioned above, there were quite a few interesting suggestions on how to make planning more efficient:

- *I think the biggest area of concern I have is having a one-stop shop for planning data within WisDOT and having the resources to be able to maintain that data. It should be for all modes. It should be standardized. It needs to be detailed enough for regional planning to get some use out of it.*
- *A concern that comes to mind is the data source ReferenceUSA (provided by InfoUSA). There is a lot of weeding and manipulation associated with this data.*
- *I would like have access to the annual traffic data I need to estimate annual VMT in a more timely manner.*

- *WisDOT needs to develop a set of freight data that we trust and can get on a regular and timely basis. This will take time and will require dedication on WisDOT's part.*
- *Environmental impact analysis - obtaining data from DNR can be challenging; locating available data can be challenging - don't know what is available internally vs. what needs to be asked for externally.*

3.2.6. Respondents' recommendations

In each of the questionnaire components discussed above, the respondents were also asked to recommend ways for dealing with the issues they mentioned. Several recommendations were related to the current data sharing practice in WisDOT. Survey respondents felt that the existing SDE should be beefed up with more data. While the biggest challenge currently for the WisDOT staff lies in locating the most relevant data, respondents expressed the need for a central clearinghouse for all the data items. A few respondents suggested that, as a short term solution to this problem, a data catalog/repository should be developed to provide the information about data custodians and specific data items that can be found in various databases. The specific recommendations are quoted below:

- *For GIS, the data is unorganized and unwieldy. There are so many folders with random names filled with awkwardly titled files that are inaccessible to someone who doesn't have the contents memorized. If we could just get our data organized, I think we would have more success.*
- *Having more than one person responsible for TRADAS data might help improve the annual data timeliness issues*
- *Update planning data once a year during the STN/WISLR update time period. Work on incorporating all modes - not just highways in this update*
- *Merge the STN and WISLR systems to get better linework alignments and correct some of the issues mentioned about. Figure out a more 'usable' methodology to incorporate meta and STN data analysis along with others in the department.*
- *Implement better traffic data collection guidelines -- such as don't count when traffic is affected by construction or detours -- and quality control checks of the data collected.*
- *Develop a central repository for all data, or at least a contact list. Develop a database/list of when data updates can be expected. Identify one person to be responsible for ensuring data is updated regularly (based on schedule). Communicate with all staff that data sharing is a priority.*

3.3. Interviews

Following from the survey to BPED staff, the research team conducted a series of interviews with WisDOT personnel for the following purposes:

- To follow up with survey respondents and ask clarifying, more in-depth questions regarding the data challenges they reported and datasets referenced;

- To obtain additional inputs from BPED administrator and middle managers regarding data needs and challenges;
- To collect information from data custodians – as identified by survey respondents and other subsequent interviewees – regarding the data content, access method, update cycle, original source/alternate custodian (if apply), and other metadata needed for the development of a data catalog; and
- To gain first hand information from database application managers and agency-wide IT support staff regarding existing data management systems and any on-going or anticipated data access improvement efforts.

The interview process revealed the following key pieces of information that suggest miscommunication between business units and possibly inaccurate perception:

- The survey respondent's caution against counting traffic where there is construction or detours is most likely an inaccurate perception as such practice is not expected within the guidelines of traffic data collection.
- On-going effort is led by the Data Management Section to unify the STN and WISLR network alignment
- Most of the highway data, particularly those that describe the historic or current state of the highway operations and infrastructure, which are used for planning activities are frequently updated and well integrated under Meta-Manager.
- Geobase is a dump place for informal data exchange and was not the recommended way for systematic data sharing; Geobase is on the way out; rather, SDE is the agency's official platform. Because the system is not centrally managed, confusion arises as to data versions, quality, etc. This is the main misperception
- Due to their disparate sources, data about non-highway modes are the problematic ones that present access and interoperability problems
- The Data Management Section at BSHP was in the process of developing a catalog of highway-related datasets at the time of the interview.
- There is a work group consisting of WisDOT upper management and BITS staff looking at agency-level strategic directions for data integration

3.4. Summary

This chapter has presented the process and findings from a survey conducted with several WisDOT staff in the Bureau of Planning and Economic Development. The focus of this survey was to sample the experiences, needs and concerns with accessing and using planning-related data. It should be noted that, while our survey respondents provided valuable insights into the current practices and perceived barriers to effective data use for planning purposes, their view represent about 30% and not all of the staff in the BPED. Caution has to be exercised when interpreting these individual responses. Nevertheless, a few general themes did emerge from the survey regarding the challenges and needs for planning-related data housed within WisDOT. These general themes warrant special attention in our subsequent development of data integration recommendations:

3.A. Data access appears to be hindered by the perceived lack of up-to-date information regarding what data are available and where. This perception is a result of having no clear documentation of data or users not aware of data documentation that exists.

3.B. Planning staff generally found the data quality to be satisfactory. An exception is when multiple sources appear to be available for a data item. The question arises as to which source is the most accurate.

3.C. Most respondents reported that they often need to perform minor manipulation and sometimes need major manipulation to use planning data. This is because the wide range of planning-related data often comes in differing formats and level of spatial scale.

3.D. Concerns were raised about the timeliness of data. The differing, and sometimes irregular, cycles by which datasets are updated further hinders users' ability to locate the latest version of a dataset.

3.E. Two major issues were identified regarding data shared between WisDOT and other agencies. First, as data from external agencies are sometimes in hard copy and represent a snapshot in time, the timeliness of such data is of concern. Second, the data models used by WisDOT and other agencies (such as *FRA and Army Corps of Engineers*) are often different, making data interoperability an issue.

It is not surprising that our survey revealed data challenges similar to those discussed in our literature survey of past data integration efforts. The solutions recommended by our survey respondents also coincide with those having been implemented in other state DOTs. These recommendations include the development of a catalog of the update schedule and custodians of planning-related datasets, as well as a "one-stop shop" for these datasets. These ideas will be revisited in Chapter 5 of this report.

4. DATA CATALOG AND FLOW MODELS

Through the survey and interviews, the research collected metadata and custodian information regarding existing planning-related datasets, as well as information pertaining the business process that drive the flow of planning data. Based on this information, we undertook a data cataloging and data flow modeling exercise. The resulting catalog and data flow model serve at least two purposes. First, they allow us to verify the accuracy of the survey response (*i.e.* to distinguish the issues attributable to biased perceptions from those issues that truly exist across-the-board). Second, the information contained in the catalog and flow models will help break down some of the perceived data barrier and form part of the data integration solution. Our data cataloging and flow modeling efforts are described in the remainder of this chapter.

4.1. Data Catalog

Our data cataloging effort built on the catalog ready developed by the Data Management Section of BSHP (hereafter referred to as the Highway Data Catalog)[2]. As such, we adopted the structure of the existing Highway Data Catalog (as of March 16, 2009) so that our catalog can be viewed as an extension of and be subsequently incorporated into the Highway Data Catalog. Also, we focus our cataloging effort on those planning datasets (as identified by our survey respondents and interviewees) that contain non-highway data. These include multimodal data as well as demographic and economic data that support various planning activities. Below, we first describe the structure of the Highway Data Catalog, followed by the proposed catalog as an extension to the existing catalog.

4.1.1. Catalog structure

The Highway Data Catalog is a database in MS Access format. It consists of three inter-related tables (see Figure 14): Systems, Attributes and Descriptions. The Systems table describes the content of and contact information for seven datasets, each associated with a unique System ID. The Attributes table provides the definition of the unique data attributes contained in the datasets that are listed in the Systems table. Each entry in the Attributes table is also assigned a unique Attribute ID. The Descriptions table relates the items in the Systems table to the corresponding items in the Attributes table. That is, each record in the Descriptions table suggest that the attributed identified by Attrib_ID is present in the database system identified by System_ID. This data catalog schema allows the user to know if an attribute in one or more systems.

[2] The Highway Data Catalog is being disseminated through the WisDOT Library Services.

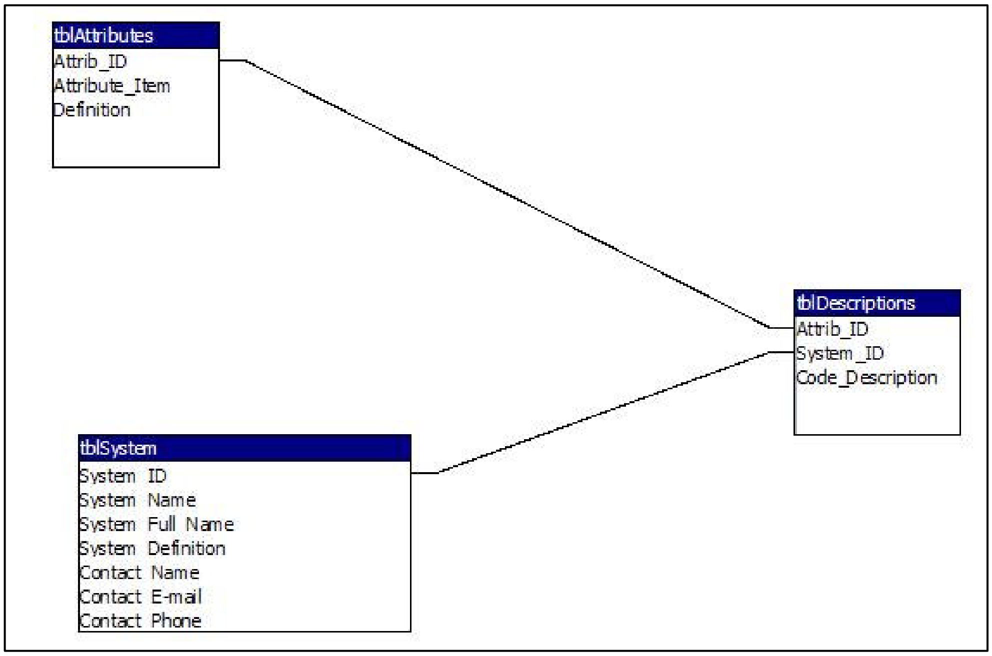


Figure 14 Database schema describing the structure of WisDOT's existing Highway Data Catalog

4.1.2. Proposed catalog extension

Our catalog of non-highway data follows the structure discussed above and also contains three tables, as shown in Figure 15, Figure 16, and Figure 17. Our Systems table lists a total of 13 new systems, with ID numbered from 1001 to 1013[3]. Similarly, out Attributes table lists a total of 50 new attributes, with ID numbered from 1001 to 1050[4].

[3,4] The IDs start with 1001 in both the Systems and Attributes tables to allow room for easy expansion of the highway component of the combined catalog.

System ID	System_Name	System_Full_Name	System_Definition	Contact_Name	Contact_E-mail	Contact_Phone
1001	Rail	Railroads data	Railroads data consist of all the railroad network attributes including services, facilities, system condition, passenger frequency and railroad related projects information.	Maria Hart	maria.hart@dot.state.wi.us	(608) 266-8968
1002	Transit	Transit data	Transit data consist of all the public transit network attributes including services, facilities, system condition, passenger frequency and transit related projects information.	John Alley	john.alley@dot.state.wi.us	(608) 266-0189
1003	Water	Ports and harbors data	Ports and harbors data consist of all the locations of ports and inland waterways and various other attributes such as shipping services, facilities, system condition, freight services.	N/A	N/A	N/A
1004	Air	Airports data	Airports data consists of all the locations of airports and various other attributes such as airlines information, fee and fare structure, facilities, runways and thier condition and air freight services.	N/A	N/A	N/A
1005	Bikeped	Bicycle and Pedestrian data	Bicycle and pedestrian data consist mainly the current existing bicycle paths, sidewalks and other project related data including ongoing and proposed construction projects.	Thomas Huber	thomas.huber@dot.state.wi.us	(608) 267-7757
1006	Economic	US Macro economic model	The US macro economic model features around 2000 economic, financial and business concepts including aggregate statistics related to domestic spending, inflation, tax policy, domestic income and other financial data.	John Cherney	john.cherney@dot.state.wi.us	(608) 264-8142
1007	WORKNET	WORKNET	Wisconsin's workforce and labor market information system	N/A	N/A	N/A
1008	Reference USA	Reference USA	Reference USA databases contain product data about a large number of businesses and further includes many databases pertaining to residential units.	Liat Lichtman	liat.lichtman@dot.state.wi.us	(608) 267-3614
1009	Population	Population demographics	The population demographics obtained from Census contain socio demographics of residents classified by blocks, block groups, tracts, counties, and other geographies.	Joleen Nelson	joleen.nelson@dot.state.wi.us	(608) 266-2571
1010	CTPP	Census Transportation Planning Package	CTPP is a set of special tabulations from decennial census demographic surveys designed for transportation planners. CTPP 2000 is divided into three parts: Part 1 contains residence end data summarizing worker and household characteristics, Part 2 contains place of work data summarizing worker characteristics, Part 3 contains contains journey-to-work flow data.	Joleen Nelson	joleen.nelson@dot.state.wi.us	(608) 266-2571
1011	Pop_projection	Population projections	Wisconsin population projections are developed by the Demographic Services Center in accordance with Wisconsin Statute 16.96. These projections are based on past and current population trends, and are intended as a baseline guide for the users. Users are urged to examine any other available forecasts that incorporate additional information such as land usage, zoning regulations, and planned or proposed developments. Users may also compare the projections with the population estimates that are produced annually.	N/A	N/A	N/A
1012	Rail_safety	Railroads safety data	Visitors to this website have access to railroad safety information including accidents and incidents, inspections and highway-rail crossing data. From this site users can run dynamic queries, download a variety of safety database files, publications and forms, and view current statistical information on railroad safety.	N/A	N/A	N/A
1013	EIA	Environmental Impact Analysis information	Environment impact analysis data available from Wisconsin DNR typically consist data about air monitoring, boat and developed shore fishing access sites, chronic wasting disease, inventory of natural heritage, hazardous substances and wetland indicators.	Christy Abing	christy.abing@dot.state.wi.us	(920) 492-5713

Figure 15 Proposed extension to the Systems table of WisDOT's data catalog

Attrib_ID	Attribute Item	Definition
1001	Railroads network attributes	Network data of both privately and publicly held railroads (linework and attributes)
1002	Railroads Capacity	Railroads capacity and current utilization of all the principal routes
1003	Railroads access	Data describing intermodal access opportunities, constraints etc.
1004	Cities serviced	Major cities serviced by the railroad network
1005	Passenger service frequency	Passenger service frequency on all the principal railroad routes in Wisconsin
1006	Passenger and Freight cars	Number of passenger and freight cars
1007	Passenger track miles	Number of passenger track miles traveled during one month on the railroads in Wisconsin
1008	Railroads Inventory	Inventory of facilities at each stop, rail road crossings and other infrastructure
1009	Age of railroad facilities	Age of various railroads elements such as cars, tunnels, bridges, railroad crossing equipment, tracks.
1010	Transit network attributes	Network data of public transit routes in Wisconsin (linework and attributes)
1011	Transit Capacity	Transit capacity and current utilization of all the principal routes
1012	Ports and waterways	Ports and inland waterway segments representation in GIS
1013	Locks	Location and capacity of Locks in Wisconsin
1014	Ship services	Ship lines/Ferry Service lines serving each port
1015	Sailing frequencies	Sailing frequencies by destination
1016	Barge lines	(F) Barge lines serving each port
1017	Multimodal facilities	opportunities for multimodal connections at ports
1018	Dredging schedule	Dredging schedules of ports
1019	Age of port facilities	Age, service records, maintenance schedules (docks, berths, navigation aids, locks etc.)
1020	Airport runways	Airport runways (location, number and lengths)
1021	airports data	Number and location of airports, service providers, cities served and airport facilities
1022	Air Freight	(F) Freight service frequency
1023	Intermodal access	Intermodal access and connections
1024	Airports fee structure	Fare or fee structure (range of prices, prices per passenger mile)
1025	Passenger transfer facilities	Passenger transfer facilities (bus stops, train stations, parking)
1026	Airport facilities	Inventory of airport facilities (gates, walkways, etc.)
1027	Airport State projects	List of all State projects (proposed and recommended in the long range plan)
1028	Airport projects history	Project history data (past capacity expansion and maintenance, project information such as dates, type of construction, etc.)
1029	Sidewalks	GIS representation of Sidewalks, age and condition
1030	Bikepaths	GIS representation of Bike paths, age and condition
1031	Bicycle project data	List of all State and MPO projects related to bicycle infrastructure proposed for next 3 years (minimum)
1032	Income	Income data by household and region -- historical, current and projected
1033	Employment	Employment data by SIC code and region -- historical, current and projected
1034	Vehicle ownership	Vehicle ownership data by household and region
1035	Industrial operations	(F) Industrial operations (Location, SIC code and employment)
1036	Wholesalers and distributors	(F) Wholesalers and distributors (Location, SIC code and employment)
1037	Commodity production	(F) Commodity production data by SIC and geographic detail -- historical, current and projected
1038	Commodity consumption data	(F) Commodity consumption data by SIC and geographic detail -- historical, current and projected
1039	Export/Import data	(F) Export/import data by point of exit/entry (seaports, airports and highway and rail border points)
1040	Projections for Commodity production and consumption	(F) Proxy data for projecting commodity production and consumption data (projections of employment, income, etc., by geographic area)
1041	Population and Labor force	Population and labor force data (e.g., population size, density, geographic distribution) -- historical, current and projected
1042	Household characteristics	Household characteristics (e.g., household size, number of children, number of licensed members) -- historical, current and projected
1043	Commodity flow	Commodity flow data by O-D
1044	Modal split	Modal split on commodity flow data by O-D
1045	Mode choice data	Mode choice data (e.g., air, rail, highway, port, transit fare matrices, parking costs, mode availability variables such as vehicle ownership and percent of houses and jobs within walking distance to transit, etc.) -- historical, current and projected
1046	User preferences	User preference data (e.g., willingness to pay, rider preferences, carpooling, ridesharing, etc.) -- historical, current and projected
1047	Traffic incidents	Incident data (e.g., number, type, location, and duration of traffic incidents, etc.)
1048	Traffic crashes	Accident data (e.g., number of accidents, deaths, injuries by mode)
1049	Security data	Security data (number and type of security incidents by mode and service populations, etc.)
1050	Transit performance data	Transit performance data (e.g., average system speed, on-time performance, vehicle hours per trip, etc.) -- historical, current and projected

Figure 16 Proposed extension to the Attributes table of WisDOT's data catalog

Attrib ID	System ID	Code Description
1001	1001	Network data of both privately and publicly held railroads in Wisconsin (linework and attributes)
1002	1001	Railroads capacity and current utilization of all the principal routes
1003	1001	Data describing intermodal access opportunities, constraints etc.
1004	1001	Major cities serviced by the railroad network
1005	1001	Passenger service frequency on all the principal railroad routes in Wisconsin
1006	1001	Number of passenger and freight cars
1007	1001	Number of passenger track miles traveled during one month on the railroads in Wisconsin
1008	1001	Inventory of facilities at each stop, rail road crossings and other infrastructure
1009	1001	Age of various railroads elements such as cars, tunnels, bridges, railroad crossing equipment, tracks.
1010	1002	Network data of public transit routes in Wisconsin (linework and attributes)
1011	1002	Transit capacity and current utilization of all the principal routes
1050	1002	Transit performance data (e.g., average system speed, on-time performance, vehicle hours per trip, etc.) -- historical, current and projected
1012	1003	Ports and inland waterway segments representation in GIS
1013	1003	Location and capacity of Locks in Wisconsin
1014	1003	Ship lines/Ferry Service lines serving each port
1015	1003	Sailing frequencies by destination
1016	1003	(F) Barge lines serving each port
1017	1003	opportunities for multimodal connections at ports
1018	1003	Dredging schedules of ports
1019	1003	Age, service records, maintenance schedules (docks, berths, navigation aids, locks etc.)
1020	1004	Airport runways (location, number and lengths)
1021	1004	Number and location of airports, service providers, cities served and airport facilities
1022	1004	(F) Freight service frequency
1023	1004	Intermodal access and connections
1024	1004	Fare or fee structure (range of prices, prices per passenger mile)
1025	1004	Passenger transfer facilities (bus stops, train stations, parking)
1026	1004	Inventory of airport facilities (gates, walkways, etc.)
1027	1004	List of all State projects (proposed and recommended in the long range plan)
1028	1004	Project history data (past capacity expansion and maintenance, project information such as dates, type of construction, etc.)
1039	1004	(F) Export/import data by point of exit/entry (seaports, airports and highway and rail border points)
1029	1005	GIS representation of Sidewalks, age and condition
1030	1005	GIS representation of Bike paths, age and condition
1031	1005	List of all State and MPO projects related to bicycle infrastructure proposed for next 3 years (minimum)
1040	1006	(F) Proxy data for projecting commodity production and consumption data (projections of employment, income, etc., by geographic area)
1032	1007	Income data by household and region -- historical, current and projected
1033	1007	Employment data by SIC code and region -- historical, current and projected
1032	1008	Income data by household and region -- historical, current and projected
1035	1008	(F) Industrial operations (Location, SIC code and employment)
1036	1008	(F) Wholesalers and distributors (Location, SIC code and employment)
1037	1008	(F) Commodity production data by SIC and geographic detail -- historical, current and projected
1038	1008	(F) Commodity consumption data by SIC and geographic detail -- historical, current and projected
1039	1008	(F) Export/import data by point of exit/entry (seaports, airports and highway and rail border points)
1032	1009	Income data by household and region -- historical, current and projected
1033	1009	Employment data by SIC code and region -- historical, current and projected
1034	1009	Vehicle ownership data by household and region
1035	1009	(F) Industrial operations (Location, SIC code and employment)
1036	1009	(F) Wholesalers and distributors (Location, SIC code and employment)
1042	1009	Household characteristics (e.g., household size, number of children, number of licensed members) -- historical, current and projected
1032	1010	Income data by household and region -- historical, current and projected
1033	1010	Employment data by SIC code and region -- historical, current and projected
1034	1010	Vehicle ownership data by household and region
1043	1010	Commodity flow data by O-D
1044	1010	Modal split on commodity flow data by O-D
1045	1010	Mode choice data (e.g., air, rail, highway, port, transit fare matrices, parking costs, mode availability variables such as vehicle ownership and percent of houses and jobs within walking distance to transit, etc.) -- historical, current and projected
1046	1010	User preference data (e.g., willingness to pay, rider preferences, carpooling, ridesharing, etc.) -- historical, current and projected
1041	1011	Population and labor force data (e.g., population size, density, geographic distribution) -- historical, current and projected
1047	1012	Incident data (e.g., number, type, location, and duration of traffic incidents, etc.)
1048	1012	Accident data (e.g., number of accidents, deaths, injuries by mode)
1049	1012	Security data (number and type of security incidents by mode and service populations, etc.)
1050	1002	Transit performance data (e.g., average system speed, on-time performance, vehicle hours per trip, etc.) -- historical, current and projected

Figure 17 Proposed extension to the Descriptions table of WisDOT’s data catalog

4.2. Data Flow Modeling

Based on the survey responses and follow-up interviews, the research team documented the flow of frequently used planning data in the form of a data flow diagram, as shown in Figure 18. It should be noted that the data model presented herein does *not* document the complete data flow process. This is because many of the datasets of interest are originally collected by entities outside of WisDOT and have to undergo multiple stages of manipulation by more entities (internal or external) before they are in a format accessible to the planners at WisDOT. Therefore,

The data flow diagram serves to:

- Help identify the organizational units and technology involved in the collection, maintenance and updating of each dataset.
- Illustrate possible data redundancy and lack of centralized access.
- Help formulate recommendations to increase the efficiency of the current data flow mechanism.

For the ease of discussion and readability, we break the above data flow diagram into multiple sub-diagrams that focus on a few selected data items at a time. These sub-diagrams are discussed in turn below.

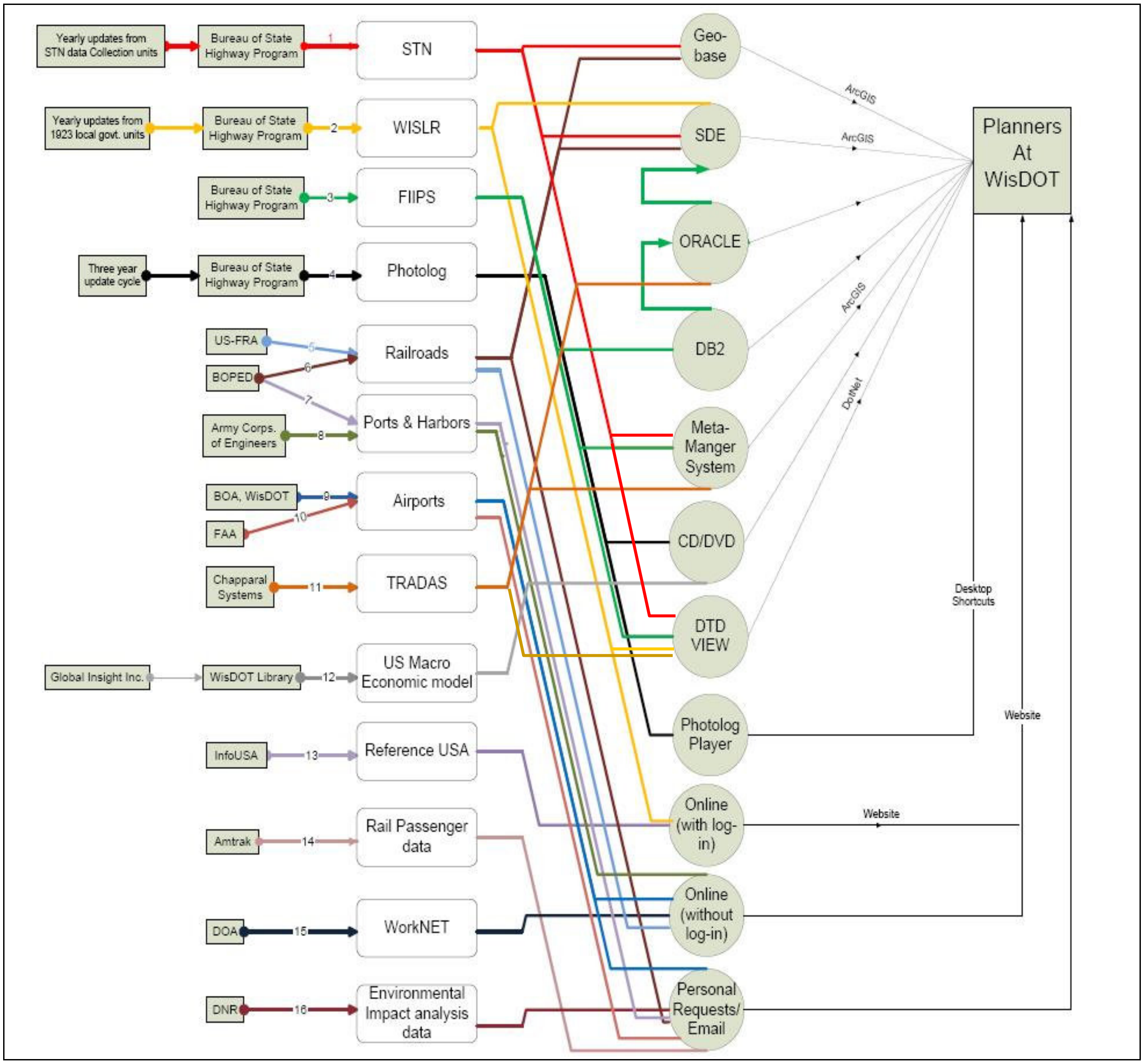


Figure 18 Data flow diagram depicting the flow of key data used to support planning activities

4.2.1. STN, WISLR, and FIIPS

Figure 19 depicts the flow of three datasets: STN, WISLR, and FIPPS.

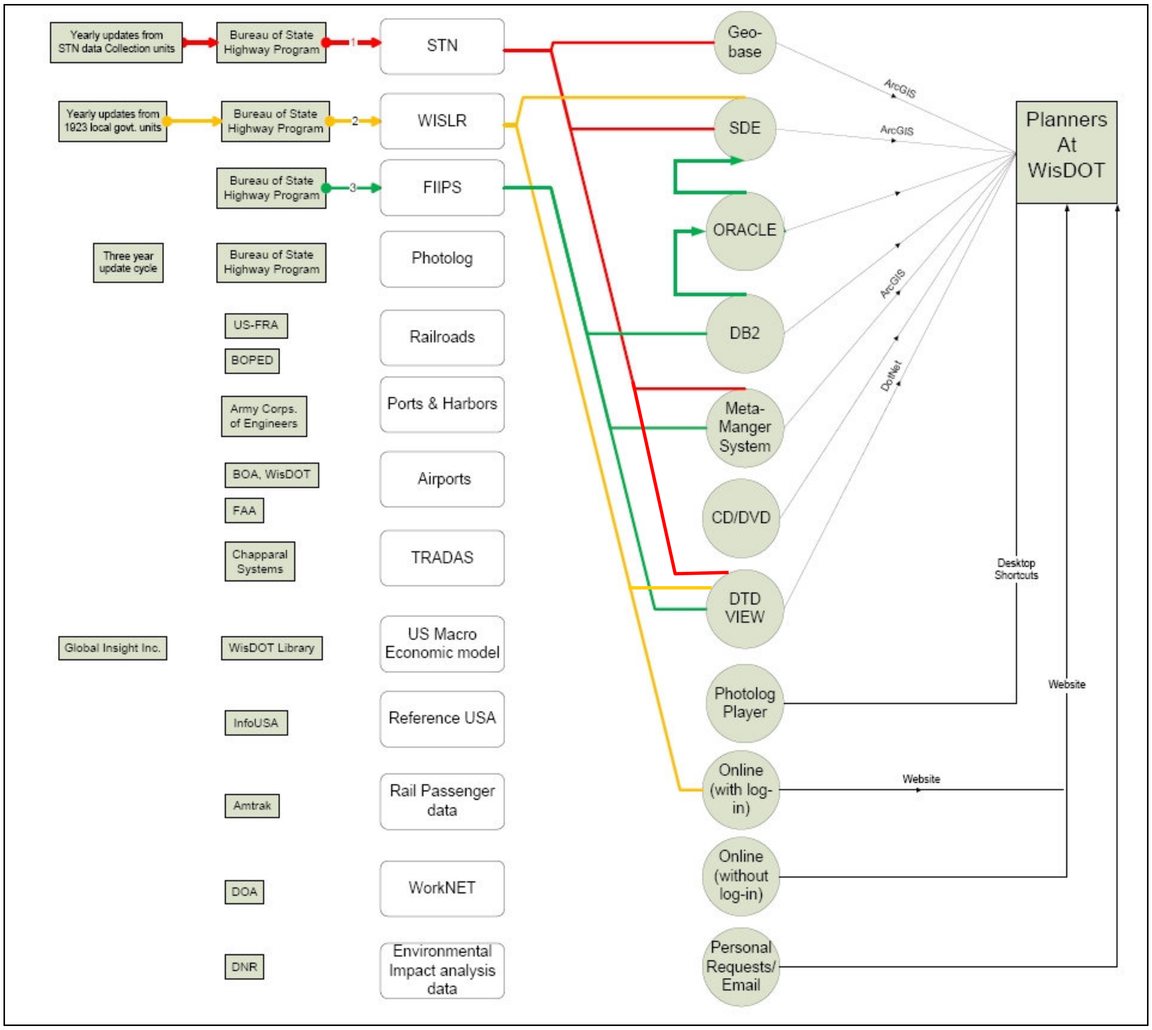


Figure 19 STN, WISLR, FIIIPS

The STN data is collected by individual data collection units assigned by the BSHP at WISDOT. BSHP is responsible for the maintenance and updating of STN data. Planners at WisDOT can access the linework of STN through enterprise GIS data sharing in two ways: Geo-base and SDE. The Meta-manager also contains the most recent STN data.

The WISLR database is collected and directly updated by 1923 local government units. However, the actual changes in the linework are updated only through the BSHP after receiving updates on the changed geometry. Generally, these updates are done in a yearly cycle. Currently, WISLR can be accessed through the SDE or directly online. Only authorized representatives of local governments are eligible to obtain WISLR access. Other parties are able to request view access through the state portal.

The FIIPS database is maintained by the BSHP. It currently exists as a DB2 database. A part of the "live" data is converted to Oracle and is spatially enabled for planners to access through SDE. FIIPS data is also integrated to meta-manager system. In addition, FIIPS can also be accessed through DTD view, which is a web based tool for viewing enterprise data. The planners at WisDOT can access it via internet browser using the intranet of WisDOT.

4.2.2. Photolog, railroad, and ports/harbors data

Figure 20 depicts the flow of the Photolog, railroad data, and ports/harbors data.

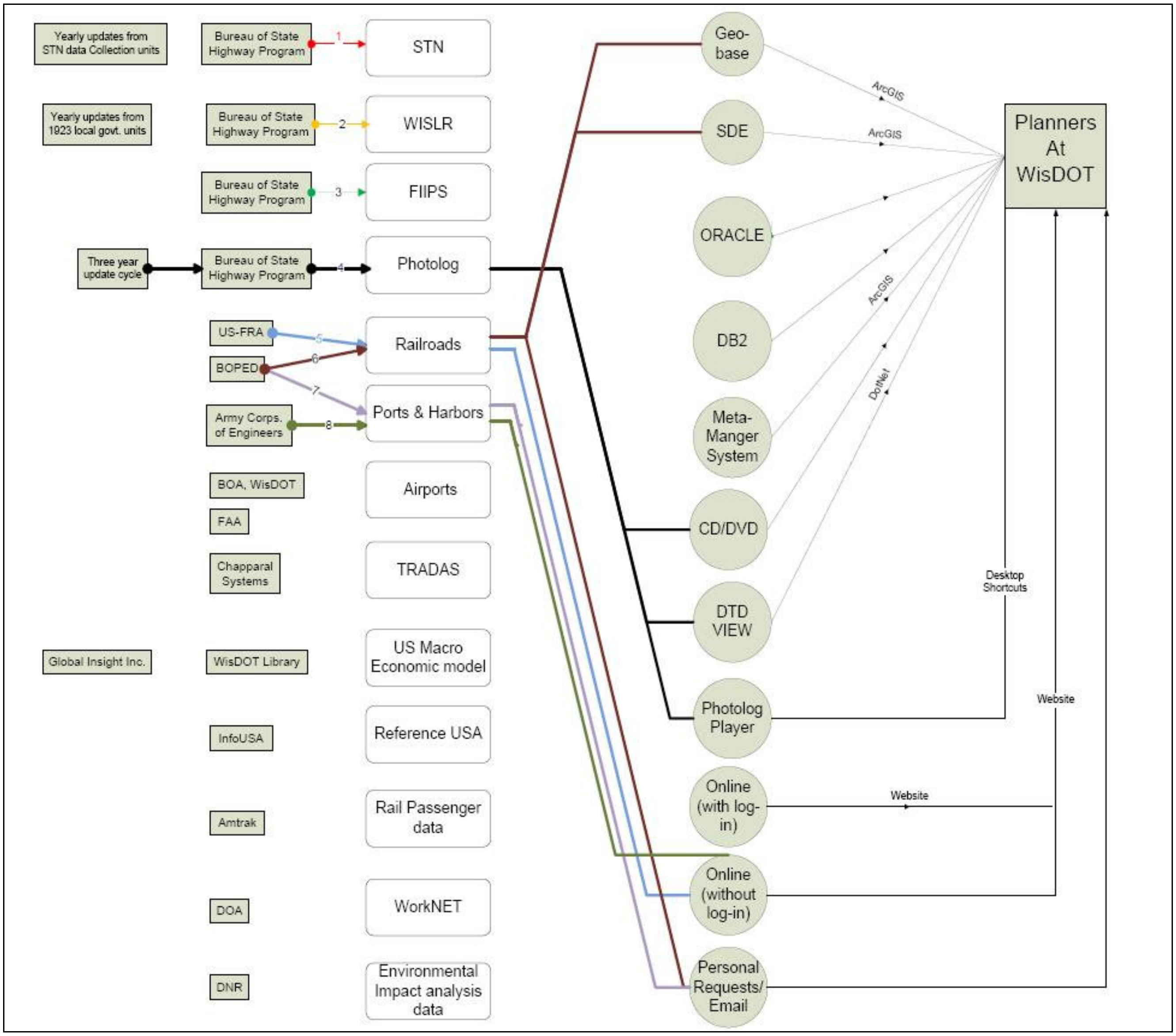


Figure 20 Photolog, railroads, ports and harbors

Photolog data are collected on all State Trunk, U.S., and Interstate highways in both directions on a yearly basis within a three-year cycle. BSHP is responsible for the maintenance of the Photolog data. The collected digital images are compressed and associated with differentially-corrected GPS data and DMI data. Images include a 120-degree composite view of the roadway, electronic capture of location, geometric data, right-of-way of the road, and visual pavement distress. Photolog digital image files and associated data are uploaded to several network server hard disks and can be accessed via the DTD view or the Photolog Player.

The railroad data is maintained by staff at the BPED. Also, US-FRA hosts the railroad inventory on the internet. Planners can access the data directly from US-FRA or from the data custodian at BPED. SDE and Geo-base are two enterprise GIS platforms that host the inventory and linework of the railroads data. However, the most up-to-date railroads data is not available to the whole organization and needs to be obtained through personal requests to the custodian.

The ports/harbors data is currently maintained by and accessed through the BPED. The data can also be accessed directly through the Army Corps. of Engineers' website.

4.2.3. Airports, TRADAS, and economic data

Figure 21 depicts the flow of the airports, TRADAS, and economic forecast data.

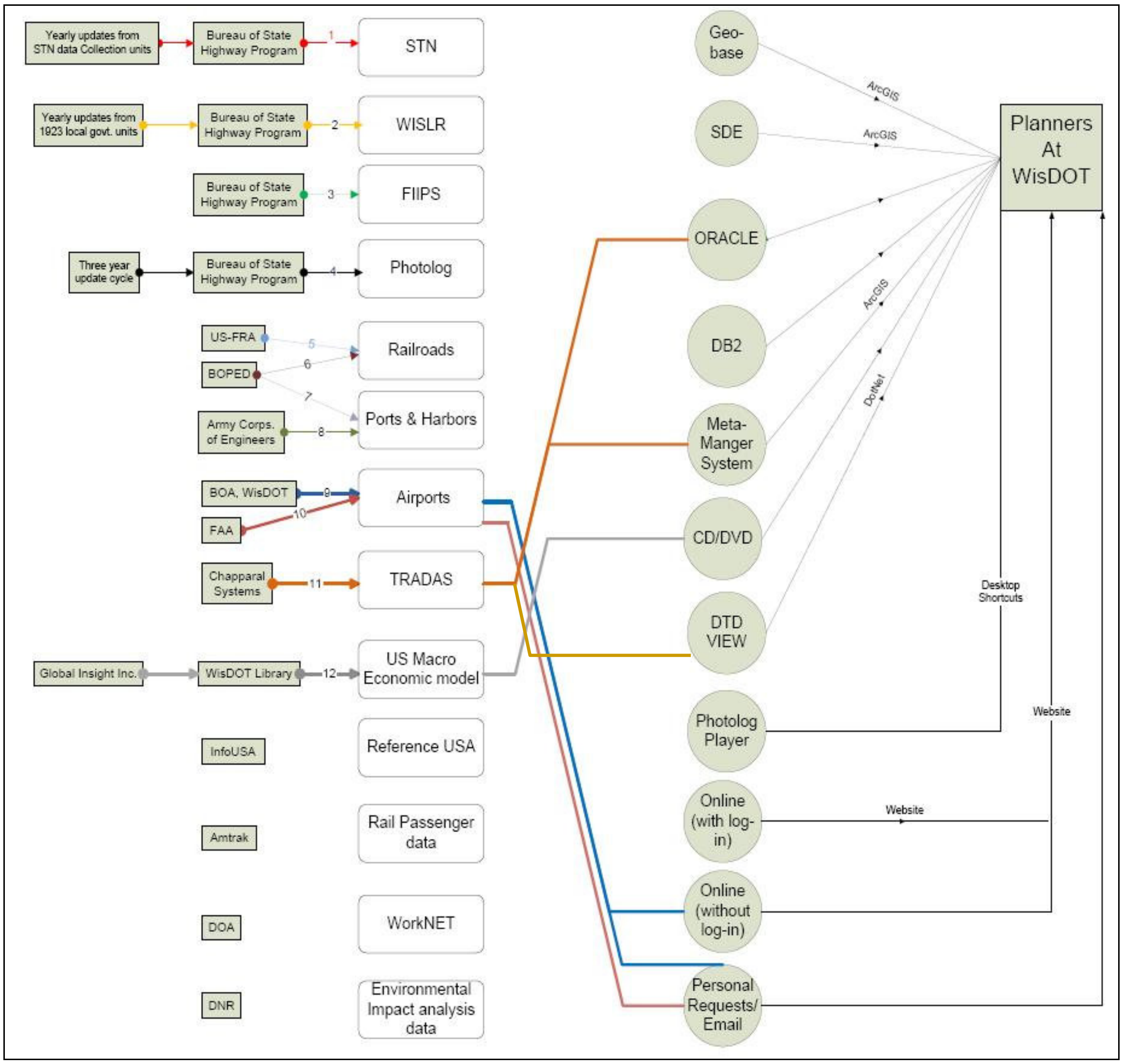


Figure 21 Airports, TRADAS, and economic data

The original source of airports data is the FAA. Within WisDOT, it is the Bureau of Aeronautics (BOA) that maintains the airports data. The users can gain access through personal requests to BOA or directly through FAA's website.

TRADAS is provided to WisDOT by Chapparal Systems Inc. Users at WisDOT can access this data set through either the Oracle server or the Meta-manager.

The US macro economic model and forecast data are originally provided by an external vendor Global Insight Inc. The data is received in the form of DVD and can be checked out through the WisDOT Library.

4.2.4. Reference USA, Rail passenger, WorkNET, and environmental data

Figure 22 depicts the flow of the airports, TRADAS, and economic forecast data.

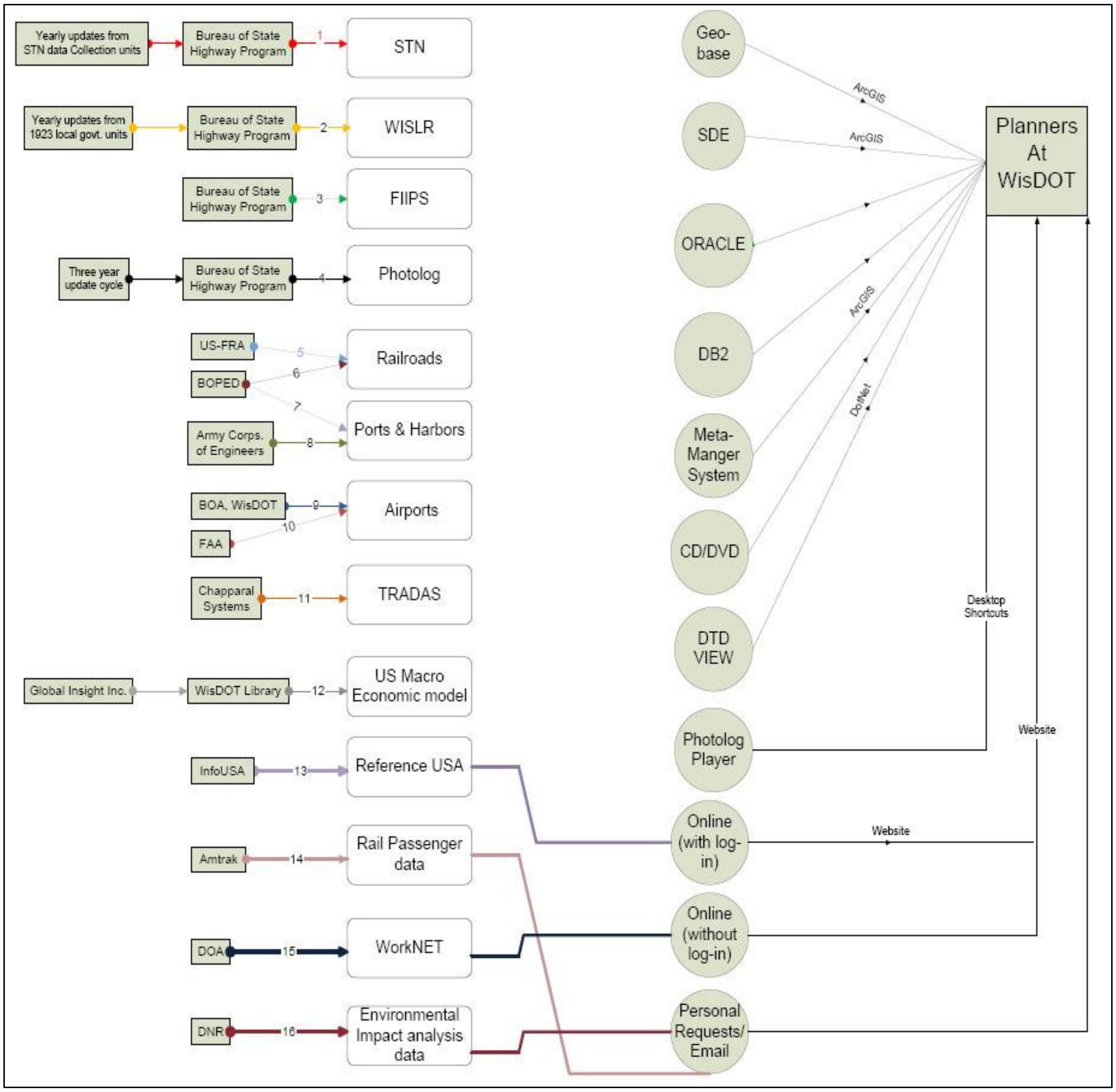


Figure 22 Reference USA, Rail passenger, WorkNET, and environmental data

The ReferenceUSA data is purchased by WisDOT from InfoUSA. The license allows designated WisDOT users to access through secured log-in.

Rail passenger data is obtained via Amtrak through email every month. This email would then be used for further analysis and can be provided by the designated planner at BPED.

The WorkNET data provides the employment data and is maintained by the Department of Administration (DOA). It is typically accessed by users through the web portal of the same name.

The Environment Impact Analysis (EIA) data and reports are obtained from the DNR either through online or personal requests based on a project-by-project basis.

4.3. Summary

The data cataloging effort and the data flow modeling process allow us to better understanding the overall process of the data sharing mechanism within WisDOT. Our primary findings are:

4.A. There are well established database systems and integrated platforms for highway related data – this is consistent with the responses from the survey. Much of the data sharing mechanism is placed on collecting, maintaining and reporting the highway related data. Many of the non-highway data are not available as an enterprise resource that is easily accessible and personal data requests have to go to designated individuals. For example, the staff at BPED maintains and updates the railroad data and there is no set mechanism or time period to make the updated data available to planners among other sections and bureaus. Hence, most up-to-date rail data is not available to all planners unless personal requests are made to the staff concerned. Second, there is no streamlined process to update the railroads information. For example, any change or construction updates for railroad crossings is not informed immediately to the data custodians at WisDOT and, hence, there is often delay in updating the railroads data accordingly. The primary reason for such delays stems from the fact that WisDOT needs to contact a number of outside organizations (such as private railroad owners) to obtain updates on their assets. Since there is no integrated platform for these private stakeholders to update their data, a lot of staff time is spent in contacting each individual party.

4.B. The seemingly emphasis on highways in the current data management systems is largely driven by the business processes. For example, it is mandatory for all state DOTs to report the HPMS data to federal agency once every year. This has led to the implementation of effective highway data management strategies in Wisconsin and other states. However, the statewide transportation planning takes place in a multimodal context and requires a much broader range of data beyond highways. Currently, a streamlined process for maintaining multimodal data that parallels the existing process for managing highway data remains lacking. As revealed in our literature review, other states in the country are also facing such a challenge as the states putting increased emphasis on the multimodal context of statewide transportation planning.

4.C. The data flow diagram shows that more than one access channels exist for some datasets. This does not necessarily indicate data duplication. For example, photolog can be accessed through DTD view, CD/DVD or desktop shortcut to photolog player. However, the photolog data itself is not duplicated and the multiple access methods offer great convenience to the users. In this case, all that is needed is for the end users to be aware that the data obtained through the alternative access channels is the same and most up-to-date. In other cases, however, duplicate copies of data or variants of the same data do exist. For example, Geo-base has been used as an informal platform for storing and sharing GIS data (relating to airports, railroads, etc.) among BPED staff. Since there is no formal mechanism set up to manage these data, different users could access the same file, process, and save the file again using similar file names. This would result in variants of the same source data and cause confusion for the next user looking up the data on Geo-base.

4.D. The disparity across the planning-related data in their sources and flow paths – as reflected in the data flow diagram – further supports the idea that a well structured and managed central clearinghouse for planning data would be a valuable asset to BPED.

4.E. The disparate format, quality, and update cycle of data sources from other state and federal agencies present a data challenge that is beyond WisDOT's administrative boundary and requires federal leadership to achieve standardization across planning data providers and users.

5. RECOMMENDATIONS AND CONCLUSIONS

In Sections 3.4 and 4.3, we highlighted a number of data challenges and opportunities uncovered through the various stages of this research project. In this chapter, we present a set of recommendations to address these data challenges and opportunities. As summarized in Table 2, the recommendations presented here are based on lessons learnt through the literature scan, information collected from the survey and interviews, and the research team's diagnosis during the data cataloging and flow modeling process. Each recommendation is mapped to at least one of the four types of strategies discussed in Chapter 2. The primary issue(s) that each recommendation targets at are also identified. Moreover, under the last three columns of Table 2, we provide ballpark estimates of the relative cost requirement, the implementation time requirement, and the anticipated scope of impact on planning practices associated with each recommendation. Our recommendations include not only strategies that are called for by the immediate business needs, but also strategies that are envisaged to help enhance data sharing internal and external to WisDOT in the long run. These recommendations are discussed in turn below.

5.1. Information Dissemination

As the transportation planning process places increasing emphasis on coordinating among all modes of travel, it is essential for planners to have easy access to up-to-date data for planning investments in modes including highway, railroads, transit, ports, pedestrian and bicycle. One of the first steps that the agency can take is to raise awareness of where data are and provide users with key information regarding data quality, format, custodians, etc.

BSHP's catalog of highway-related data and the data catalog developed in this project (see Section 4.1) together represent the first part of this solution. The two catalogs combined provide key contact and content information regarding most of the datasets needed to support planning activities. The catalogs are set up so that a user can find out where to locate, or who to contact for, a data element of interest (e.g. traffic counts). The existing catalogs, however, do not contain detailed metadata regarding each data attributes in a given system. This type of information is undoubtedly important for someone looking for a specific data item, but is excluded from the catalogs by design because of the potentially dynamic nature of this type of information, That is, if and when a data element is updated (say, from a 2-digit land use code to a 4-digit land use code), the change would not be automatically propagated into the data catalog. Instead, it would require the catalog administrator being informed of this change and update the catalog accordingly. Maintaining the catalog content at this level of detail would be time-consuming and laborious if such data updates are to occur often (this is quite possible, especially for data that require frequent manual cleaning and processing at irregular cycles).

Table 2 Summary of recommendations

	Recommendation	Type of Strategy *G: GIS-based solution* *A: Data architecture* *D: Documentation* *H: Human resources*	Issue Addressed 3.A ~ 3.E (Sec 3.4) 4.A ~ 4.E (Sec 4.3)	Cost • *low* •• *medium* ••• *high*	Time • *short* •• *medium* ••• *long*	Impact • *low* •• *medium* ••• *high*
1	Information dissemination	D	3.A/B	●	●	●●
2	Centralized data platform	A	3.B 4.A/C	●	●	●●
3	Designated data coordinator	H	3.D 4.A	●●	●●	●●
4	Data access tool for long range planning	G, A	3.C/D 4.B/C	●●●	●●	●●●
5	Data standardization	D	3.E 4.E	●●●	●●●	●●●

To supplement the data catalogs and to provide more detailed information about data contents, we suggest making the metadata about existing datasets more complete and easily accessible. This could be accomplished through building an online catalog of SDE data so that the GIS data in the server can be explored (but not directly accessed) through a web browser in a manner similar to how it can be viewed in Arc Catalog, as illustrated in Figure 23. The implementation would require building a web application using a client-side, JavaScript framework (49). Such an application would provide the following features:

- All the geospatial data layers in the current SDE can be explored in a user friendly browser environment. Tables can be viewed and a preview of maps can be obtained without physically accessing the data.
- Provided that detailed Metadata are maintained by the respective data custodians and managers through the SDE, detailed and timely information regarding each data layer is readily accessible to all users within the agency (and the general public if so desired).
- The fact that all the metadata is available online makes it easy add additional search functionality (using the standard search engines) to provide more detailed key word search than what the existing data catalogs support.

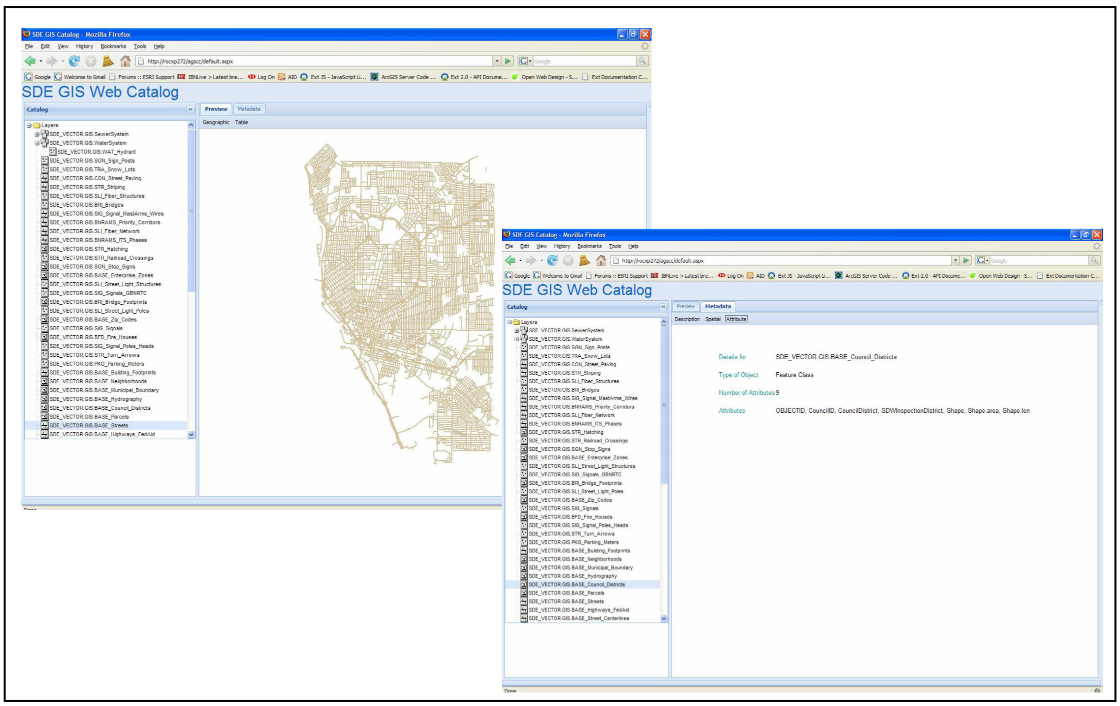


Figure 23 A sample SDE web-based catalog (source: 49)

Another approach to improving information dissemination across the BPED is the ECM system described in Section 2.2.3. The ECM could be considered as an extension of the current data catalogs and would be best pursued as an agency-wide data cataloging strategy for capturing all relevant transportation data. Due to its agency-wide impact and the leadership needed from higher management, the ECM idea is beyond the immediate scope of this project and was not included in our assessment of the implementation cost and time in Table 2.

5.2. Centralized Data Platform

One of the main objectives of this research effort is to identify any data redundancy arising from either storing the same data at various locations or allowing more than one way to access the same data. As revealed in the survey (Chapter 3) and reflected in the data flow diagram (Chapter 4), two platforms have been used as spatial database engines for accessing similar enterprise spatial data: Geo-base and SDE. Geo-base is a file-based spatial data server which was originally set up as a place to store and access ArcInfo coverages and ArcView shapefiles. SDE is an oracle based ArcSDE server which also stores the layers and tables in a geodatabase. Table 3 below summarizes the similarities and differences between Geo-base and SDE.

Table 3 Geo-base Vs. SDE

Characteristics	**Geo-base**	**SDE**
Access methodology	Connect through ArcView	Connect to a SDE server
Data	Mostly contains the linework of highway, railroads at various resolutions.	Tables and layers of highways, railroads, bridges, STN inventory etc.
Documentation	No formal documentation.	Documentation of the table name and unabbreviated name available through WisDOT's intranet (dotnet).
Data standards	Informal data sharing in the enterprise.	Formal data sharing that needs to meet the data standards set by Bureau of Information Technology Services at WisDOT.

Geo-base is an informal data sharing platform which contains data about airports, highways, railroads etc. Similar data items can be accessed using SDE, although it is a more formal data sharing mechanism adhering to the standards set by the Bureau of Information Technology Services in WisDOT. Such multiple sources can potentially create confusion among planners since they may not know which data is most recent and accurate for their purposes. Based on our interviews with BITS staff, the use of Geo-base is in fact not recommended and is meant to be for temporary sharing of data only (i.e. data should be removed after the recipient has downloaded the data). Therefore, we propose that BPED identify the data items currently residing on the Geo-base platform that are worth keeping and move these data items to the SDE under the management of BITS. This approach would largely address the concerns of data redundancy with regard to spatial data at WisDOT. The main cost of implementing this recommendation lies in the staff resources that would be needed to upload and maintain data in SDE. As noted earlier, SDE data need to adhere to WisDOT's IT standards and could prove to be a resource intensive task. This approach also requires the buy-in from other Bureaus that are providers of some of the non-highway data. The benefits of such an effort can be realized in a short to medium term since such a centralized platform would encourage and allow all the planners to use the same and consistent set of enterprise GIS data.

5.3. Designated Data Coordinator

Designating data coordinators is a common practice within organizations dealing with large amounts of data across multiple platforms. As discussed in sections 3.4 and 4.3, while highway-related databases are centrally coordinated through the Data Management Section of BSHP, the up-keeping of non-highway data typically falls on individuals in the BPED who are frequent users of respective data. This limits data maintenance to specific individuals in an uncoordinated fashion and knowledge could be easily lost when an individual is on leave or moves onto a different position or organization.

The alternative to the current approach is to centralize the data management duties by designating a data coordinator, whose job function would be to streamline the data business process for BPED and provide the needed data services and products to support data applications, planning processes, and business objectives. The data services and products would range from data modeling to quality assurance, storage, updating, security, and catalog of non-highway data. The data coordinator would structure these services, coordinate among organizational units that provide the original data, coordinate with the data management units in other Bureaus (e.g. BSHP and BITS), and ensure the services are maintained and running. Specific tasks for the data coordinator may include:

- to ensure that the data business plan is implemented
- to partner with IT to develop and advance data programs and services
- to coordinate development of data architecture to support business
- to act as a point of contact for data quality issues and ensure that data has accurate definitions, context, coding standards and usage notation
- to implement data changes following the updated management process
- to coordinate and maintain master data catalog and data dictionary
- to document all activities for contingency planning

This recommendation could be resource intensive and would require reassigning job functions among existing BPED staff. Alternatively, it would require setting up a new appointment that would be more costly. The return is a much more streamlined process to managing the many non-highway datasets that are increasingly in volume and complexity.

5.4. Data Access Tool for Long Range Planning

This recommendation builds on the previous two ideas regarding data coordinator and centralized data platform. The focus here is to facilitate the business processes of long range planning, for which much staff time and efforts are devoted to consolidating, analyzing, and mapping multi-modal data. For example, maps of thirty-seven priority corridors were generated for Wisconsin's latest long range plan, *Connection 2030*. Each corridor map includes a comprehensive list of existing multimodal facilities, WisDOT's priority project action areas, and WisDOT's priority support areas. Planners at the BPED have to obtain information about the many data item from different bureaus within WisDOT, including the bureau of aeronautics (airports data), bureau of state highway programs (information about highway construction projects) and bureau of transit, local roads, rails and harbors (waterways data). Additional data from MPO's, regional offices and other stakeholders are also needed and collected by planners within BPED to include in the corridor maps. Based on the current data acquisition and compilation processes, updating corridor maps entails planners to repeat this time- and resource-intensive exercise each time when these maps need to be updated. Once the necessary data are assembled, further efforts are also needed to publish maps in a presentable and consistent format with standardized legends.

To meet the strong business need for working with and integrating multimodal data, we propose a data access tool as depicted in Figure 24. There are three main components of the proposed data access framework:

- All the essential data items that are available within WisDOT across all bureaus should be made as shared enterprise resource through SDE. Currently, only highways and railroad data is available on this platform. This is the recommendation presented in Section 5.2.
- Once multimodal enterprise data are available, a data access tool can be developed based on the 'layer file' feature in ArcGIS. A layer file by definition stores the paths to a source datasets and stores other layer properties, including symbology, built-in queries, legend and other information used to generate maps. The primary purpose of a layer file is to act as a set of pointers to the original data sources. Hence, the creation of a layer file entails pointing the layer to the needed data items on SDE. This would be a one time effort to set up the multimodal data needed for each frequently performed mapping or analysis activity. In addition to pulling together multiple spatial datasets, layer files also provide the following features:
    - Symbology: The symbols for each layer can be set according to the standard legend that is used for defining frequently generated maps (e.g. corridor maps). This would greatly reduce the time taken to adjust the symbols every time the maps are updated.
    - Display settings: Settings such as transparency and border lines can be saved into the layer file and retrieved later.
    - Scale dependent display properties can also be customized for the different datasets combined in the layer file.
    - Joined table/related table information: The information in the tables can be joined or related to the data layers in the layer file. In the case of corridor maps, this feature can be particularly useful if we need to add Oracle tables to existing data layers.
    - Definition query: Queries can be defined to build a data layer in the layer file. For example, in WisDOT the project information on STN is stored in FIIPS database which is in a DB2 database format. To link any data from FIIPS to this layer file would require a query to be set up that links to the appropriate fields in the tables of DB2. Only the query would be stored in the layer file and hence, every time the FIIPS database is updated, the layer file would be automatically updated providing the planners access to the latest data.
    - Labels and annotation can also be saved into a layer file.

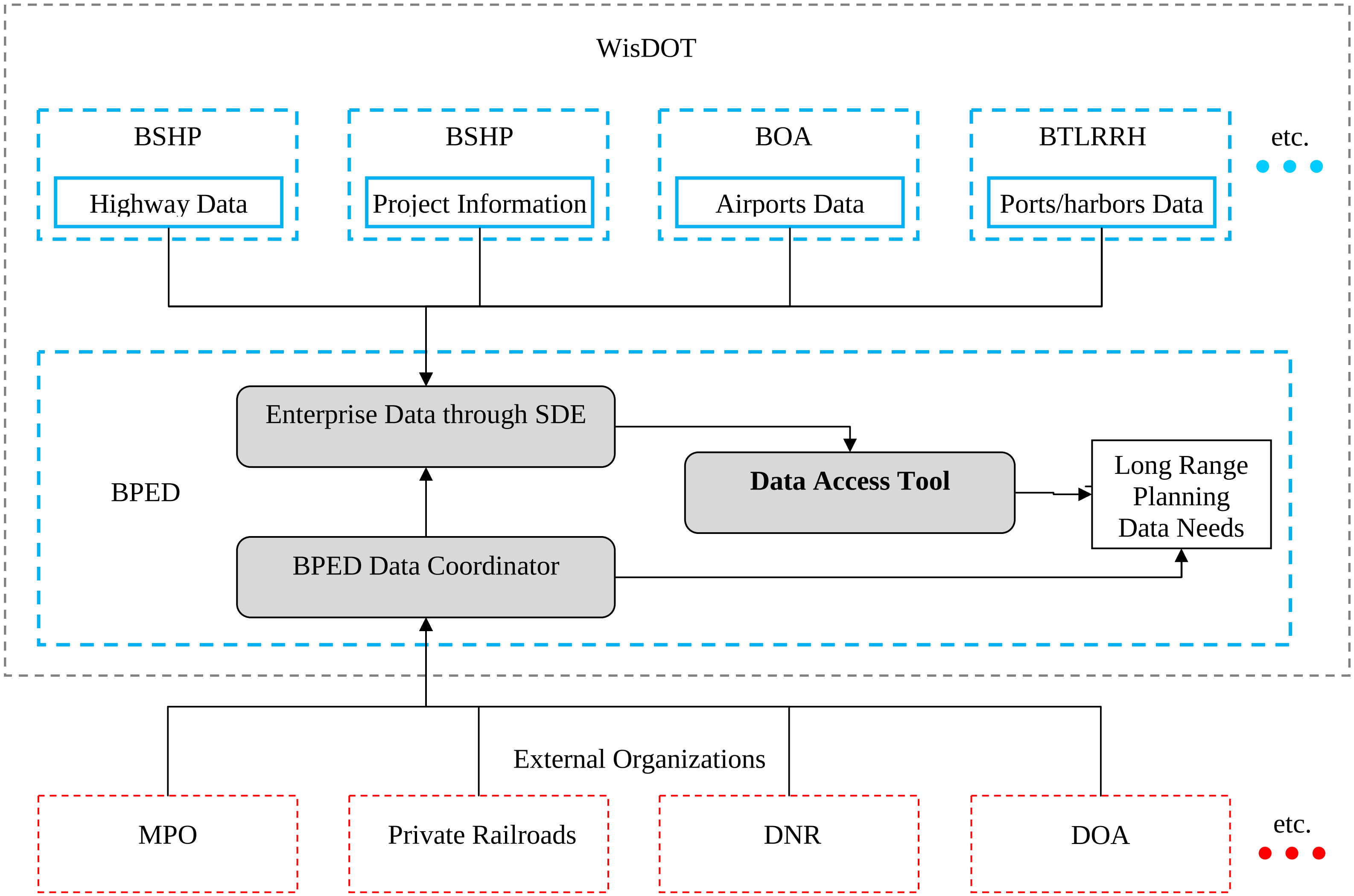


Figure 24 Proposed data access tool to support long range planning activities

- For data that are provided by organizations external to WisDOT, the data coordinator (as suggested in Section 5.3) would be responsible for gathering and consolidating the data into predefined formats and store the final data in a designated location (if differing from the SDE) that is linked to the proposed data access tool.

An alternative to implementing the data access tool using layer files is to build a graphical user interface (GUI) application on SDE. Such a user interface could provide planners a chance to interactively select the data layers required in the map using dropdown menus or clicking appropriate buttons. Also, advanced users can write additional queries to refine their results and obtain appropriate data. The end product could allow data users to perform a broad range of frequent tasks with just a few clicks. However, this development of such a GUI would be more expensive and involved than that of the layer-file based tool.

5.5. Data Standardization

Sharing information between agencies requires the management of all agencies to consider both their own needs as well as the needs of the other agencies. Often a balance needs to be struck between competing needs and resources. Regarding resources, it is important to recognize the multiple benefits that can be obtained from data integration. Recognition of the benefits may help justify the allocation of the necessary resources. For example, to ensure the local MPO's co-operation with WisDOT in maintaining corridor maps, resources must be allocated by both agencies to improve data exchange practices. Facilitating web based access, developing local and regional visions for spatial data archival and exchange, improving software and hardware capabilities at MPO's are some efforts that are likely to yield mutual benefit.

A solution that would have the most long term impact on enhancing data exchange and interoperability across multiple agencies is to establish data standards. There has been ongoing research effort in developing standards for transportation data. For example, the TransXML effort was launched as a research project under the auspices of the National Cooperative Highway Research Program (NCHRP). The objectives of this project were to develop broadly accepted public domain XML schemas for exchange of transportation data, and a framework for development, validation, dissemination, and extension of current and future schemas. The initial focus of this project has been on four key business areas: 1) Survey/Roadway Design, 2) Transportation Construction/Materials, 3) Highway Bridge Structures, and 4) Transportation Safety. Ultimately TransXML will encompass a broader set of schemas for all crucial transportation business areas (including planning, for example). The XML provides a foundation for data exchange among transportation applications and a common vocabulary and information structure for transportation agency activities and assets. WisDOT can adopt and expand these established standards to incorporate planning data.

5.6. Summary

As summarized in Table 2, the implementation of our five recommendations would entail differing levels of costs and time requirements. Information dissemination and centralized data platform are the low-hanging fruit that would address several data challenges with a relatively short timeframe. The remaining three recommendations require more financial and time investment and also stronger agency commitment to changing the current business practices. However, they are expected to yield high benefits in the long term. The recommendations are also interrelated. For example, the development of the data access tool would be best built on the centralization of data and designation of data coordinator. Such a data access tool is expected, however, to have high payoff, as having been experienced by other agencies such as FDOT. The interrelationship among our recommendations suggests that an incremental approach to implementation would be appropriate. Finally, the establishment and adoption of data standards represents a long-term investment that has begun to gain support at the federal level and may revolutionize how we manage multimodal planning data in the future. This is not an approach that is ready for adoption, but a direction for future practice that warrants planning agencies' attention.

6. REFERENCES

1. Fact sheets on highway provisions: Statewide planning. FHWA, 2007. Acccessed September 1, 2007. http://www.fhwa.dot.gov/safetealu/factsheets/statewide.htm
2. Data Partnerships: Making Connections for Effective Transportation Planning. Transportation Research Circular E-C061, November 2003.
3. T. Kirk, A. Y. Levy, Y. Sagiv, and D. Srivastava. The Information Manifold. In Proceedings of the AAAI 1995 Spring Symp. on Information Gathering from Heterogeneous, Distributed Enviroments, pages 85–91, 1995.
4. M. Lenzerini. Data integration: a theoretical perspective, Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, Madison, Wisconsin, 2002.
5. Data Integration Glossary: Federal Highway Administration Office of Asset Management, U.S. Department of Transportation, August 2001.
6. Preliminary report of TRB peer exchange, 2001. Last accessed October 24, 2008. http://webservices.camsys.com/trbcomm/docs/peerex03-22-2001.htm
7. Data Integration Primer, Report no. FHWA-IF-01-016, U.S. Department of Transportation, August 2001.
8. Integrating Highway Information Preliminary Report of the TRB Peer Exchange http://webservices.camsys.com/trbcomm/docs/peerex03-22-2001.htm , March 22-23, 2001. Last accessed October 24, 2008.
9. Operations data for Planning Applications: Identifying needs, Opportunities and best practices. Transportation Research Circular E-C095, April 2006.
10. GIS cap Grant Project, Pend Oreille county, September 2006, Progress report. Last accessed October 24, 2008. http://www.fgdc.gov/grants/2006CAP/Reports/096-06-5-WA-InterimReport.pdf
11. Essen V. Ian and Ernst Shawna. Integrating Distinct Transportation Data into a Single Geo Database Framework. A presentation at Pend Orielle County, Washington, January 2006. Accessed September 30, 2008. http://www.wsdot.wa.gov/mapsdata/transframework/project_presentations/WASHDOT%20Integrating%20Distinct%20Transportation%20Data%20into%20a%20Single%20Geodatabase.pdf
12. Wiggins Lyna, Deuker Kenneth, Ferreira Joseph, Merry Carolyn, Peng Zhong-ren and Spear Bruce. Application challenges for Geographic Information Science: Implications for Research, Education and Policy for transportation management and policy, URISA journal, No.12-2, spring 2002.
13. Review of Data Integration Practices and their applications to transportation Asset Management, Report no. FHWA-IF-03-023, U.S. Department of Transportation, July 2003.
14. Geographic Information System Strategic Plan, Prepared for Kansas DOT. Prepared by Geo Decisions in association with Triple I Corporation, March 2000.
15. Geographic Information Systems Standards. A report prepared for the Wyoming Department of Transportation. Prepared by geographic information systems oversight committee, September 10, 2001.

16. Data Integration and Automation in Polk County, Florida: Transportation Applications and lessons learnt. Last accessed October 24, 2008. http://gis.esri.com/library/userconf/proc96/TO300/PAP296/P296.HTM
17. Data integration – what is it anyway. Last accessed October 24, 2008 http://www.dmreview.com/dmdirect/20040305/8232-1.html
18. Sutton C. John. Role of Geographic Information systems in Regional Transportation Planning , Transportation Research Record 1518, Journal of the Transportation Research Board, pp no. 25-31, April 1995.
19. ARCGIS Server. Last accessed September 30 2008. http://www.esri.com/software/arcgis/arcgisserver/index.html
20. Traffic Road Book Quality Control using ARCGIS server. Prepared for the County of San Bernardino, California, Last accessed September 30 2008. http://www.esri.com/library/fliers/pdfs/cs-sanbernardino.pdf
21. Pendyala, Ram. Data integration procedures in support of statewide transportation modeling and planning processes, Summary of final report BC-353-20, Research, University of South Florida, June 2003.
22. Ross, Dan. Building a Stable Transportation Model for Minnesota, A presentation. Office of Information Technology, MNDOT. Last accessed September 30, 2008. http://www.metrogis.org/teams/cc/meetings/03_0618ross.pdf
23. Markow J Michael and Racosky Joe. Asset Management Implementation Plan And Tiered System Process. Colorado Department of Transportation Research, Report No. CDOT-DTD-R-2001-13, September 2001.
24. LRS background – Literature review, ITS and evolving Spatial Technologies, Location Referencing system requirements analysis, RoadsOne Project, Vic Roads, November 2005.
25. Adams, Teressa; Koncz and Vonderhoe, Alan. Guidelines for the Implementation of Multimodal Transportation Location Referencing Systems. NCHRP Report 460 [Project 20-27(3)], 2001.
26. Ries T, Integrating traffic management data via an enterprise LRS, GeoAnalytics presentation at North American Travel monitoring exhibition and conference, June 2000.
27. Ilir Bejleri. A New database framework for Florida's Transportation Planning: Integrating work program, multimodal transportation networks, planning and environmental Databases. Prepared for Florida DOT. Prepared by Department of Urban and Regional Planning at the University of Florida, Final Report No. BD545-11, February 2006.
28. GPS to LRM: Integration of Spatial Point features With Linear Referencing Methods, the Iowa DOT CTRE Management Project 00-68, October 2001.
29. Transportation Metadata: Role of data and information technology Section, Final Metadata Working Group Report, July 2005. Last accessed September 2008 http://www.trb.org/committees/datasection/DataSection-Metadata.pdf
30. Geospatial Metadata – Federal Geographic Data Committee, Website of the Federal Geographic Data Committee. Last accessed September 30 2008. http://www.fgdc.gov/metadata
31. How is CDOT Getting There? – The Colorado Experience. FHWA Asset management project case studies. Last accessed September 30, 2008. http://www.fhwa.dot.gov/infrastructure/asstmgmt/dico05.cfm

32. Wigan, Marcus; Grieco, Margaret and Hine, Julian. Enabling and Managing Greater Access to Transport Data Through Metadata Transportation Research Record 1804, Journal of the Transportation Research Board, Paper no. 02-3064, 2002.
33. Data warehouse concepts, Oracle. Last accessed September 30 2008. http://download.oracle.com/docs/cd/B19306_01/server.102/b14220/img/cncpt170.gif
34. Hall, James; Robinson, Rob; Paulis, Mary Ann. Enterprise wide Spatial data integration of legacy systems for Asset Management: the case of the Illinois Department of Transportation, Transportation Research record 1917, Journal of the Transportation Research Board, pp 11-17, 2005.
35. King County GIS center: Spatial Data Catalog. Last accessed on September 30,2008 http://www.metrokc.gov/gis/sdc/index.htm
36. New Hampshire DOT GIS section data catalog. last accessed on September 30,2008 http://www.nh.gov/dot/bureaus/planning/GISDataCatalog.htm
37. Andy Everett. Transportation Research Thesaurus: WSDOT Use Cases. Transportation Librarians Roundtable, February 14, 2008.
38. Where Data Groups Are Organizationally Located Within State DOT's. Requested by AASHTO Standing Committee on Planning, NCHRP Project 08-36, Task 46, National Cooperative Highway Research Program, Transportation Research Board, Prepared by ICF consulting, Virginia, December 2005.
39. Data Integration: The Michigan Experience, August 2005. Last accessed on September 30, 2008. http://www.fhwa.dot.gov/infrastructure/asstmgmt/dimi05.cfm ,.
40. Paul O'Packi; Dubois, Rick; Armentrout, Nancy and Steve Bower. Maine's Approach to Data Warehousing for State Departments of Transportation, Transportation Research record 1719, Paper no. 00-1037**,** Journal of the Transportation Research Board, 2000.
41. Zhou, Yushuang; Cambridge Systematics. Building an Enterprise Geodatabase for Transportation Planning and Modeling Puget Sound Regional Council, GIS-T Symposium, March 2004.
42. Data Management Techniques. A presentation. Integrating Databases at the Puget Sound Regional Council, AMPO Travel Modeling Work Group, September 28, 2004.
43. Puget Sound Regional Council's Data integration efforts. Last accessed on September 30, 2008. www.pagnet.org/Documents/LandUse/LUModel/PSRCGeodata07-2005.ppt
44. WISLR brochure from WisDOT website. Accessed July, 2008. http://www.dot.wisconsin.gov/localgov/docs/wislr-brochure.pdf
45. Meta-Manager, Management System Database, Division of Transportation Investment management, Bureau of State Highway Programs, Program Development and Analysis Section, April 2008.
46. Bob St.Clair , WisDOT's Meta-Manager, presented at the *4th National Transportation Asset Management Workshop*, September 23 through 25, 2001, Madison, WI.
47. Chappral Systems Inc. Accessed July, 2008. http://www.chapsys.com/products/tradas.php
48. Wisconsin department of Transportation website link describing photologging. Accessed July, 2008. http://ww.dot.wisconsin.gov/business/econdev/maps-data.htm
49. Vijay, L.S. SDE Web Catalog, 3rd prize winner in ArcGIS server code challenge 2008, Buffalo, New York, USA.

50. XML Schemas for the Exchange of Transportation Data, National Cooperative Highway Research Program (NCHRP) Project 20-64. Accessed January, 2009. http://www.transxml.net/

APPENDIX A. SURVEY QUESTIONNAIRE

The University of Wisconsin-Madison Transportation and Urban Systems Analysis Lab would like to welcome you, and thank you for taking the time to participate in this survey. Our team of researchers is working on a project to help WisDOT identify data access and interoperability issues pertaining to the statewide transportation planning process. We will produce an inventory of available data housed within and outside of WisDOT to support current and future planning-related analysis and business processes. We will also develop a set of recommendations regarding future courses of action to improve data integration and partnership for planning in Wisconsin.

With these objectives in mind, we have prepared this five-part survey to help us start this process. Your responses are very valuable to us in assessing current conditions and making recommendations. Part A of this survey asks you to provide some basic information about your role within WisDOT. In parts B and C, we ask you to reflect on your experiences as a data user along four different themes: data accessibility, data quality, data format, and data timeliness. Part D is regarding your unmet data needs and Part E asks you to provide information about the data items and/or the databases that you work with.

To participate in this survey, please fill out this Microsoft Word document and send it back to the survey administrator Sasanka Gandavarapu at gandavarapu@wisc.edu. Alternatively, an on-line version of this survey is also available and can been accessed at: http://tusal.cee.wisc.edu/limesurvey/index.php?sid=17825&newtest=Y.

We request you to complete the survey by March 31, 2008. Please be as detailed and complete as possible when filling out the survey. If any questions should arise, please don't hesitate to contact me at (608) 890-1064.

Thank you again for taking the time to complete this survey.

Sincerely,

Jessica Guo
Assistant Professor of Civil and Environmental Engineering
Director of the Transportation and Urban Systems Analysis (TUSA) Lab
University of Wisconsin-Madison

Data Integration for Statewide Transportation Planning

A Questionnaire for Planners in WisDOT

6.1. Part A: Basic Information

Name	
Section	
Bureau	
Job Title	
Job Function (how your work relates to statewide transportation planning process)	
Primary analysis tools used (Ex: GIS/DBMS/EXCEL/SAS etc)	

Before you proceed, please take a moment to read the following definitions of terminology used in this questionnaire.

Data accessibility: Data accessibility refers to the ease with which data can be reached. The factors that hinder data access can be generally classified into three categories:

- Technical (E.g., Complex computer applications to access data each time),
- Organizational (E.g., Staff coordination among participating agencies/sections) or
- Financial (E.g., Cost incurred in sharing data).

Data quality: Quality refers to the accuracy, precision and the level of detail in the data. For example, the quality attributes of spatial data would be accuracy of representing a physical location or resolution of the map etc.

Data format: Sometimes, even if a dataset is accessible and of good quality, the format in which it is made available could make it difficult to use the data. For example, the format might not be compatible to work with your typical choice of analysis tools. This may result in additional effort and time to manipulate data and delay planning activities.

Data timeliness: Timeliness of data refers to the degree of recentness of data. Often in transportation planning, having frequently updated data is critical for the quality of models. Hence, this issue deals with the access of recent data to transportation planners.

6.2. Part B: Data Experience within WisDOT

This part contains questions about the experience regarding transportation planning data sharing within WisDOT.

6.2.1. B.1 Experience with data access

1. To what extent do you agree/disagree with the following statement?

"I always have good access to all the data needed for my job function."

☐ Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree

2. To what extent have technical, organizational, and financial factors hindered your access to data in WisDOT? (Please choose one option for each row)

	Always	*Often*	*Sometimes*	*Never*
Technical Factors	☐	☐	☐	☐
Organizational Factors	☐	☐	☐	☐
Financial Factors	☐	☐	☐	☐

3. Please briefly describe specific occasions that you have had problems accessing the data you need within WisDOT?

4. Please provide your suggestions as to how data access may be improved across WisDOT.

B.2 Experience with data quality

1. To what extent do you agree/disagree with the following statement?

"I am pleased with the quality of data that I receive."

☐ Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree

2. Please explain your response to the question above. What specific issues have you encountered regarding data quality, if any?

3. Please provide your suggestions as to how data quality may be improved across WisDOT.

B.3 Experience with data format

1. Please respond to the following statements that relate to your experience with data formats. (Please choose one option for each row)

"The format in which I receive data makes it…"

	Always	*Often*	*Sometimes*	*Never*
1. *Usable immediately*	☐	☐	☐	☐
2. *Usable after minor manipulation*	☐	☐	☐	☐
3. *Usable after major manipulation*	☐	☐	☐	☐
4. *Not usable*	☐	☐	☐	☐

2. Please explain your response to the question above. If applicable, please provide examples of minor and major manipulations that you need to perform.

3. Please provide your suggestions to improve data format interoperability within WisDOT.

B.4 Timeliness of data

1. To what extent do you agree/disagree with the following statement?

"The data items that I acquire and use for planning are timely and up to date."

☐ Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree

2. Please explain your response to the question above and specify any data items that you would like to see updated more frequently to better support your planning activities.

3. Please provide your suggestions as to how data timeliness may be improved across WisDOT.

Part C: Data Experience with other Agencies

The questions below are concerned with your experience in obtaining and working with data from other agencies at the federal, state and local level.

1. Please list the primary agencies (federal/state/local) from which you frequently acquired data for your planning purposes. (e.g FHWA, Department of Natural Resources, Department of Administration , Department of Revenue, local RPC/MPO, etc.)

2. To what extent do you agree/disagree with the following statements as they relate to your data experience with these <u>external agencies</u>? (Please choose one option for each row)

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
"I always have good <u>access</u> to all the data needed for my job function."	☐	☐	☐	☐	☐
"I am pleased with the <u>quality</u> of data that I receive."	☐	☐	☐	☐	☐
"The <u>format</u> of data I receive is convenient and compatible to my planning applications."	☐	☐	☐	☐	☐
"The data items that I acquire and use for planning are <u>timely</u> and <u>up-to date</u>."	☐	☐	☐	☐	☐

3. What are the main issues relating to data sharing with these external organizations? Please describe specific problems with regard to data access, quality, format or timeliness.

4. Please provide any suggestions to improve the existing data sharing mechanism between WisDOT and other organizations.

Part D: Unmet Needs

Overall, what planning activities would you like to perform/ perform more efficiently, but you are currently unable to because of the various data issues identified in the earlier parts of this survey? Are there any other unmet or future data needs that you have in relation to your work? Please describe briefly below.

Part E: Data Inventory

In the table below, please provide information regarding the data items/databases that you frequently work with. Examples of data items include land use data, environmental data, economic data, Census data, network files (highways, streets, bike paths, etc.), WISLR, MetaManager, WisTransPortal, PhotoLog, TraCS, etc.

Name of data item/database	Type of data (geospatial, text, numerical, imagery etc)	Metadata Available (yes/no)	Software tool Used (ArcGIS, Excel, SAS, MS Word etc)	Custodian (name and contact info, if possible)	Comments (relating to data accessibility, format interoperability, quality, timeliness, etc.)
1.					
2.					
3.					

Name of data item/database	Type of data (geospatial, text, numerical, imagery etc)	Metadata Available (yes/no)	Software tool Used (ArcGIS, Excel, SAS, MS Word etc)	Custodian (name and contact info, if possible)	Comments (relating to data accessibility, format interoperability, quality, timeliness, etc.)
4.					
5.					
6.					
7.					

Name of data item/database	Type of data (geospatial, text, numerical, imagery etc)	Metadata Available (yes/no)	Software tool Used (ArcGIS, Excel, SAS, MS Word etc)	Custodian (name and contact info, if possible)	Comments (relating to data accessibility, format interoperability, quality, timeliness, etc.)
8.					
9.					
10.					
11.					

APPENDIX B. DISCUSSION

Comments from the Wisconsin Department of Transportation

The Wisconsin Department of Transportation (WisDOT) recognizes the longstanding positive and productive relationship with the University of Wisconsin-Madison in the areas of research and engineering and looks forward to continuing these relationships.

On project 0092-07-23, "Data Integration and Partnership for Statewide Transportation Planning," the UW researchers worked closely with WisDOT staff to identify and respond to feedback and comments on the conduct of the research and its findings. Most of the feedback and comments were accepted and incorporated by the researchers. The department recognizes and respects the purview of the researchers to include elements they consider appropriate and necessary to communicating the results of the research project.

However, certain issues remain unresolved and remain a concern for the department. These issues include the application of the literature review, the scope of the survey conducted in the research and the limited reflection of WisDOT's current data management efforts.

Specifically, WisDOT notes the following concerns regarding this project:

- A substantial portion of the research project is based on publications that are five years or older. Efforts to address and stay current with technology and data management needs are an ever evolving and changing forum. What was true one or more years ago are oftentimes no longer applicable. While the report's literature review offers insight into the data issues faced by DOTs nationwide, many of the issues identified have already been resolved or are in process to be resolved in Wisconsin.
- The survey used to assess the current value and data needs in WisDOT's planning business areas offered an unbalanced analysis perspective of needs and available resources. The report does not offer an analysis of whether the results of the survey are isolated to an individual experience or widespread throughout the organization. This paper does not verify the responses; thus, assumptions on WisDOT practices are not accurately reflected. As a result, work efforts underway or in the future pipeline are not acknowledged.

These comments reflect input from these WisDOT business areas:

- Bureau of Planning
- Bureau of State Highway Programs
- Research & Communication Services Section

These comments are submitted by Daniel Yeh, Chief, WisDOT Research & Communication Services Section on September 30, 2009.

Authors' Response

In response to the above comments provided by WisDOT on September 30, 2009, the research team offers the following clarifications.

- The literature review presented in the report reflects the literature scan conducted initially in early 2008 (the original timeline for the task) and later expanded and updated in late 2008. Several databases were used to search for published materials, including TRB publication catalogs, TRIS online, and Web of Knowledge. These are databases commonly used by transportation professionals. Web search was also conducted by the research team to locate web sites describing relevant data practices. Due to the broad scope of the topic, various keyword combinations were used and the hit lists carefully screened through to exclude many irrelevant studies. The research team had strived to locate the most recent and relevant information in published format. However, considering that information technology is constantly evolving, planning agencies' experiences in data integration may not be well documented or published in a timely fashion. This set a limit on what the research team could uncover through the literature review process. The research team did recognize and acknowledge this limitation. The literature review findings were therefore used primarily to provide the readers some insights into the spectrum of data-related challenges typically faced by transportation planning agencies and the range of solutions that have been considered in the past. The research team did not intend to suggest that the same challenges and solutions would be directly applicable to WisDOT. Rather, these past experiences and lessons learnt of other agencies served as background information and thinking points for our investigation of the practice in WisDOT. While some of the ideas drawn from the literature review did apply and feed into the research team's recommendations for WisDOT, our recommendations were developed based on several other research tasks that attempted to examine the local issues (as depicted in Figure 1).
- It was the intent for section 3.2 to truthfully report the information provided by the survey respondents. As emphasized on page 26 of the report, *"… as with any qualitative survey, the findings reported below do not necessarily reflect the 'true' data-related practices in WisDOT. Rather, they reflect the respondents' experiences and perceptions that may (or may not) vary from individual to individual.*" Also highlighted in the executive summary on page viii: "*[o]ur survey to BPED staff resulted in a response rate of about 30% (9 responses). Thus, caution needs to be exercised when interpreting these individual responses.*" The research team did devote significant effort verifying the user experience. In some cases, the research team was able to identify the underlying issues that may have led to user misperception of the actual WisDOT practice. These were reported on page 31 of the report. The research team recognizes and acknowledges that there is limitation to what the research team was able to assess. This limitation is largely attributed to the fact that data management is a dynamic process involving a multitude of business units across WisDOT. Given the limited scope of and the resources available to this project, not all business units and decision makers involved in the process were interviewed as part of this project. Thus, further work is warranted to better document all of WisDOT's efforts that are underway or in the future pipeline.

Wisconsin Highway Research Program
University of Wisconsin-Madison
1415 Engineering Drive
Madison, WI 53706
608/262-2013
www.whrp.org