

Scoping Data Access and Integration Needs to Facilitate Better Management of Research Innovation

P.S. Sriraj, Ph.D, Paul Metaxatos, Ph.D, and Scott Figler

Urban Transportation Center
University of Illinois at Chicago

WisDOT ID no. 0092-07-16

October 2011



RESEARCH & LIBRARY UNIT



MIDWEST REGIONAL UNIVERSITY
TRANSPORTATION CENTER
UNIVERSITY OF WISCONSIN-MADISON

WISCONSIN DOT
PUTTING RESEARCH TO WORK

DISCLAIMER

This research was funded by the Midwest Regional University Transportation Center, the Wisconsin Department of Transportation and the Federal Highway Administration under Project #00xx-0x-xx. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation,

University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof. The contents do not necessarily reflect the official views of the Midwest Regional University Transportation Center, the University of Wisconsin, the Wisconsin Department of Transportation, or the Federal Highway Administration at the time of publication.

The United States Government assumes no liability for its contents or use thereof. This report does not constitute a standard, specification, or regulation.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No. CFDA 20.701	
4. Title and Subtitle Scoping Data Access and Integration Needs to Facilitate Better Management of Research Innovation		5. Report Date August 9, 2011	
		6. Performing Organization Code	
7. Author/s P. S. Sriraj, Paul Metaxatos, and Scott Figler		8. Performing Organization Report No. MRUTC 08-10	
9. Performing Organization Name and Address Midwest Regional University Transportation Center University of Wisconsin-Madison 1415 Engineering Drive, Madison, WI 53706		10. Work Unit No. (TRAIS)	
		11. Contract or Grant No. DTRS 99-G-0005	
12. Sponsoring Organization Name and Address Wisconsin Department of Transportation 4802 Sheboygan Avenue Madison, WI 53707		13. Type of Report and Period Covered Final Report (6/1/2007 – 10/31/2011)	
		14. Sponsoring Agency Code 0092-07-16	
15. Supplementary Notes Project completed for the Wisconsin Department of Transportation by the Midwest Regional University Transportation Center.			
16. Abstract <p>Public agencies are charged with the difficult task of balancing the interests on each side of the data sharing equation. On one side are the users: private consultants, researchers, and other public entities. Proper planning and evaluation of public assets by these groups requires the best information available. The objective of this research is to understand the data access experience of one group of users, i.e., the researchers who have had a history of data requests made to the Wisconsin Department of Transportation (WisDOT) and provide recommendations to WisDOT for improving their data sharing practice and protocol.</p> <p>The report is divided into a literature Review documenting the theory and practice of data sharing in public agencies followed by the results of a Survey of researchers. The results of the survey are then used to identify a group of respondents who form the core for Case Studies involving their individual experiences in requesting data from WisDOT. The report concludes with a set of recommendations to the Wisconsin DOT; and some broad conclusions.</p> <p>The survey reflected that identifying the appropriate contact person was a major challenge, while the data processing time was the single biggest factor once the data was acquired. This also seemed to have an impact on the research findings with poor data quality leading to a compromised research effort. The respondents also seemed to agree that the WisDOT needed to have an integrated database or approach towards data dissemination or data sharing.</p>			
17. Key Words Data Sharing, Researchers, Data Acquisition, Data Quality, Data Preparation, Wisconsin, Transportation		18. Distribution Statement No restrictions. This report is available through the Transportation Research Information Services of the National Transportation Library.	
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. Of Pages 102	22. Price -0-

Final Report

Scoping Data Access and Integration Needs to Facilitate Better Management of Research Innovation

P.S. Sriraj, Ph.D.

Director, Metropolitan Transportation Support Initiative, and
Research Assistant Professor

Paul Metaxatos, Ph.D., Research Assistant Professor

And

Scott Figler

Graduate Research Assistant
Urban Transportation Center
University of Illinois at Chicago
412 South Peoria Street, Suite 340
Chicago, Illinois 60607
Phone: (312) 413-7568
Fax: (312) 413-0006
E-Mail: sriraj@uic.edu

June 30, 2011

TABLE OF CONTENTS

DISCLAIMER	2
Executive Summary	7
CHAPTER 1. Introduction AND LITERATURE REVIEW.....	9
Literature Review	9
I. Background of Data Sharing Policy	9
II. Issues in Data Sharing Policy.....	12
III. SCAN OF OTHER STATE PRACTICES (SEMI-STRUCTURED INTERVIEWS)	13
IV. Strategies for Facilitating Data Sharing	17
V. Experiences in other fields	21
VI. Key points.....	22
CHAPTER 2. Survey of Researchers.....	23
Survey Design	23
IRB Approval	23
Survey Administration	24
SURVEY RESULTS	24
1. Introduction.....	24
2. Exploratory Analysis.....	25
Critical Datasets	34
CHAPTER 3. Case Studies	38
Identification of Prominent Data Sets	38
Identification of Case Studies	39
Technical Details of Ranking Methodology	40
Interviews	41
Case studies	42

CASE STUDY I - Researcher A	42
CASE STUDY II - Researcher B	43
CASE STUDY III - Researcher D	44
CASE STUDY IV - Researcher H.....	45
Recommendations for WisDOT	46
Strategies for dealing with bureaucracy	46
Strategies for dealing with data limitations.....	47
CONCLUSIONS AND FUTURE WORK	48
References	49
Appendix A – SNAP SHOT OF ONLINE SURVEY QUESTIONNAIRE	51
APPENDIX B – CASE STUDY QUESTIONNAIRE	64

EXECUTIVE SUMMARY

Public agencies are charged with the difficult task of balancing the interests on each side of the data sharing equation. On one side are the users: private consultants, researchers, and other public entities and on the other side the agencies have to ensure that the sensitivity of the shared information is taken into account while making decisions that are likely to provide a worldview of the agency to outside interests. This challenging task of balancing expectations while at the same time ensuring that the information flow is streamlined and regulated sometimes results in skewed public perceptions of agencies and results in being attributed as a bottle-neck for fundamental and informed research. Proper planning and evaluation of public assets by these groups requires the best information available. The objective of this research is to understand the data access experience of one group of users, i.e., the researchers who have had a history of data requests made to the Wisconsin Department of Transportation (WisDOT) and provide recommendations to WisDOT for improving their data sharing practice and protocol.

The report is divided into three main sections: (1) A review of other state practices in the sphere of data sharing; (2) An online survey of researchers who had requested data from the DOT; and (3) case studies of a sub-set of researchers who participated in the online survey using a semi-structured interview format.

The literature review documenting the theory and practice of data sharing in public agencies was augmented with a structured interview of several state DOTs about their practice of data sharing. The literature highlighted the philosophies associated with data sharing and this allowed for a better perspective about some seemingly contradictory viewpoints. While some suggest that data sharing is a public service funded by taxpayer money, and therefore should be inherently as open as possible, there are others who are of the view that data sharing can be used as a business proposition. This view is further elaborated by the thinking that public agencies should use their information to leverage fair competition among private contractors. These divergent views shed light about the difficulties associated with data sharing from the perspective of public officials and researchers. The research team at UIC, based on the literature, identified public agency officials from various states to interview about their data sharing experiences.

This set of interviews illustrated some pertinent and relevant findings. Data sharing approaches at the state level have not been consistent across the board. Regardless, the interviews illustrated that for the most part the data sharing practices were dictated by three broad categories. They are (1) the goals of the agency for data sharing, (2) the economic condition of the agency, and (3) awareness about data on the part of the data requestor. The goals of the agency lead to either outsourcing of data management (due to security concerns) or trying to cope with the numerous requests for data with limited human capital in a fiscally constrained economy. A decade of economic downturn has left many state agencies with reduced payroll and this is cited as a constraining factor to an effective data sharing program at the state level. At the same time, public officials are also of the view data seekers are often unaware of how

to navigate the system and get what they need. Improving awareness amongst end-users of data about data availability is seen as a significant challenge toward a successful and seamless data sharing practice. Data providers all reported spending a considerable amount of time helping people determine what data they actually need.

The scan of other state practices was followed by the survey of researchers, administered online. The survey was administered to a group of researchers who had dealt with the DOT in the past with data requests. The survey was aimed at understanding each individual researcher's experience about data sharing and the context toward those experiences. The questions and the results were binned into four categories. They were (1) Data Acquisition – pertaining to time or effort taken to obtain the data, and identifying the appropriate person in the agency to make contact with; (2) Data Processing – pertaining to processing time for data entry, whether this was expected and planned for, and its impact on the research; (3) Data Quality – that included questions about the format of the data, quality of the data, and its impact on research findings; (4) Data Sharing.

The results of the survey indicated that there were specific issues that researchers felt had a significant impact on successful data sharing. These were: (1) identifying the appropriate contact person was a major challenge, (2) while the data processing time was the single biggest factor once the data was acquired. This also seemed to have an impact on the research findings with poor data quality leading to a compromised research effort. The respondents also seemed to agree that the WisDOT needed to have an integrated database or approach towards data dissemination or data sharing.

The final phase of the research involved selecting a few case studies from out of the list of researchers who responded to the online survey for an in-depth understanding of relevant and critical issues that can be used by the DOT to better inform their data sharing principles. A scoring method based on principal factor analysis was used to assign scores for each respondent to the survey and from this list, the individual researchers for participation in the case studies was identified. The case studies were focused on the individual experiences and were case specific. However, there were some generalizable themes that emerged from these discussions. Two areas that transcended the individual experiences were the bureaucracy and the data limitations.

The research has certain limitations. The sample for the survey was not randomly selected and the survey was impacted by low response rates. The survey also was geared toward obtaining feedback about specific experiences of data access from WisDOT and therefore restrictive to some extent. In order to improve on the findings from this research, it is recommended to expand the scope of research to multiple agencies and states and then pick a random sample of researchers and stakeholders to survey. This will ensure that the methodological limitations will be addressed and result in a generalizable set of findings.

CHAPTER 1. INTRODUCTION AND LITERATURE REVIEW

Federal, state, and municipal transportation agencies are the owners of a wealth of data related to the planning, structure, and performance of infrastructure. This data may be used for several purposes including policy analysis, commercial or academic research, advocacy, or education. Policies for sharing and accessing these data have received heightened attention in recent years as state agencies have had to respond to issues related to privacy, liability, and national security.

Public agencies are charged with the difficult task of balancing the interests on each side of the data sharing equation. On one side are the users: private consultants, researchers, and other public entities. Proper planning and evaluation of public assets by these groups requires the best information available. However, this often comes in conflict with the responsibilities of data providers who are legally obligated to preserve the confidentiality of individual records (microdata) as well as any classified information. To comply with requirements to preserve this confidentiality, agencies may need to severely limit the amount of detail included in shared data sets. This may limit the quality and scope of the analyses.

Better data access policies should be considered by DOTs for several reasons. As state governments continue to outsource work, private contractors are increasingly charged with the tasks previously done by these agencies. More streamlined procedures for sharing data are then necessary to ensure a quality of work equal to or greater than what the agency could provide for the public. Standardized procedures can also reduce costly duplication of effort by analysts, promote cross-disciplinary work, and provide a platform to build add-on applications. It is in this context that this research assumes significance.

The report is divided into five sections: (1) Literature Review documenting the theory and practice of data sharing in public agencies; (2) Survey of researchers; (3) Case Studies of select number of researchers' experiences; (4) Recommendations to the Wisconsin DOT; and (5) Conclusions.

LITERATURE REVIEW

I. BACKGROUND OF DATA SHARING POLICY

As government policy in the 1960s increasingly placed more power in the hands of the individual, it eventually became necessary to make government information available to the public. Understanding that some information must remain classified in the name of private interests, Congress passed the Freedom of Information Act (FOIA) of 1966 that established open access to information as the "default," with a list of exemptions to protect sensitive information and privacy of individuals. The Privacy Act of 1974 built on FOIA by establishing a code of practice that governs the collection, maintenance, use, and dissemination of personally identifiable information, or microdata. This law

required that agencies give the public notice of their systems of records by publication in the Federal Register. It also specified specific conditions for information disclosure, such as requiring written consent of the individual for whom information is requested.

With new protocols for data collection in place, a need became apparent for an oversight body to make sure this data was used effectively and in a transparent manner. In 1975, the United States Office of Management and Budget established the Federal Committee on Statistical Methodology (FCSM). The FCSM was an organization of statisticians working for the federal government who sought to maximize the value and integrity of statistical information. Their three listed objectives are:

- To protect the interests of data subjects through procedures that ensure privacy and confidentiality
- To enhance public confidence in integrity of statistical and research data
- To facilitate responsible dissemination of data to users

One subgroup within the FCSM, the Panel on Confidentiality and Data Access (PCDA)¹ (later to become the Confidentiality and Data Access Committee), was charged with reconciling issues between users who need information and providers who may be obligated to maintain confidentiality. The Panel has released several publications since its inception in 1995. These include the “Checklist on Disclosure Potential of Proposed Datasets” and “Restricted Access Procedures,” two documents that help state and local agencies determine what information should be treated more selectively; and “Confidentiality and Data Access Issues among Federal Agencies,” which explains how federal agencies have dealt with these issues and how other agencies can learn from these experiences. These documents are all available on the PCDA website for the convenience of state and local transportation officials. One of the earliest and most comprehensive works in the data access literature is the Panel’s 1993 publication, *Private Lives and Public Policies*, which summarizes the existing literature, gives detailed accounts of data access successes and failures in several fields, and recommends different types of agreements for dealing with data sharing issues. Many of the guidelines listed in *Private Lives and Public Policies* are taken from “Procedures for Restricted Data Access” (1993) by Thomas Jabine, a member of the Panel. The agreements include the following:

- *Free, unrestricted access* – universal access is given to any interested party with nothing expected in return.
- *Free access with cost recovery* – User is given free access but must pay for its own hardware/software, contribute value added material free of charge to agency, reimburse agency for costs of providing data, or make a general “in-kind” contribution to the agency.
- *Acknowledgment of source* – universal access is granted, but interested party may be punished if they do not acknowledge the original source.
- *Remote access* – data users are given access to exclusive public records, for example through the use of a file transfer server.

¹ The other four committees within the FCSM are the Committee on National Statistics, the Commission on Behavioral and Social Sciences and Education, the National Research Council, and the Social Science Research Council.

- *Visitor access* – select users are able to access data only on the premise of the public agency.
- *Non-disclosure/confidentiality agreements* – data users must sign an agreement stating that they accept the legal consequences of sharing the data to another party
- *Technical specifications* – Private firm is given control of data but public agency specifies the types of analyses and data masking procedures that must be used.
- *Exclusivity agreements* – One firm is charged with handling, analyzing, and disseminating important public data to ensure that its dissemination does not compromise public interest. If confidential information is disseminated, it is easier for the DOT to identify possible culprits if access has been limited to a small number of individuals.

According to this article, restrictions on data access should be based on the perceived risk to the agency if the data were to be disclosed. This article establishes a conceptual framework for restricting access to a dataset based on the following characteristics :

- *Type of data set* – Aggregate vs. microdata; microdata sets with individual identifiers should be masked.
- *Initial source of data* – Collected by statistical agency directly from respondents or obtained from another agency; if obtained from another agency, that source agency's restrictions normally must be complied with.
- *Number and type of variables* – The greater the detail in the data set, the greater the risk, and the more restrictive the conditions associated with its release should be.
- *Sensitivity of Information* – The law generally gives the agency discretion in determining what sensitive information is since this can be very context sensitive.
- *Statutory Restrictions on Release* – Agencies with no confidentiality provisions in their organic statutes must comply with provisions of the Privacy Act or Freedom of Information Act.
- *Pledges to Respondents* – If confidentiality pledges have been made to persons or organizations that supply information, these absolutely must be kept; however agencies should be careful when making these pledges, since they may greatly restrict data sharing.

These recommendations from the PCDA are intended to help DOTs find an optimal balance between protecting confidential information and providing transparent access to their data. While these guidelines are broad and generic in nature, the interpretation of these guidelines and the adoption of policies at the agency level and across the country need to be examined to comprehend the impact of data sharing on the end users of the data. In light of this, the research team at UIC developed a methodology that involved a mixture of gleaning information from two sources: (1) published sources such as journal articles, technical reports, government websites, and other digital sources, (2) telephone interviews of state and regional agencies to provide contextual references to published material and to understand the inherent institutional barriers. The agencies contacted for this purpose were the departments of transportation from Alaska, Florida, New York, Ohio, and Virginia along with the Dallas Metropolitan Planning Organization.

II. ISSUES IN DATA SHARING POLICY

The literature on data sharing is quite prolific and wide ranging. The scan of this literature led to the emergence of certain broad themes that impact the data access policies and procedures of agencies. The first is the philosophical approach of the agency to data sharing. Zimmerman et al. (2002) explores best practices of data sharing by surveying 34 DOTs. Based on this research, the author identifies three non-mutually exclusive data sharing philosophies that differ depending on the agency and on the dataset. The first view is that data sharing is a public service funded by taxpayer money, and therefore should be inherently as open as possible. Most agencies across the country adopt this view and provide as much data without charge and without restriction as possible. The second is that data sharing can be used as a business proposition. This is consistent with the belief that the government must wisely manage its resources and efficiently deliver services to the public. The revenue extracted would then be on behalf of taxpayers and would be used to make the agency more efficient in some way. An agency, therefore, might adopt an exclusivity agreement if they want to avoid dealing with too many parties or simply do not have resources in their budget for information dissemination. The final view is that public agencies should use their information to leverage fair competition among private contractors. This view is consistent with the second in that it believes in engaging the private sector, yet it discourages arrangements that give monopoly power to one firm, such as in the case of an exclusivity agreement. Rather, this view favors franchising agreements where firms compete for specific aspects of a data dissemination process and also engage each other in partnering on the project.

The survey results indicated that “acknowledgement of source” is the most common arrangement, with 25 of 34 agencies surveyed citing this as their most common policy. 18 reported using “restrictions on use,” 14 used “technical specifications,” and 4 used “other conditions.” According to these agencies, the most common type of traveler information provided was “traffic and road conditions,” “incident information,” “planned construction information,” and “special events information.” The most common types of data shared with both private organizations and public organizations, according to the survey, were “highway/real-time” and “highway-static” datasets. 18 out of the 34 agencies said that they had a formal data sharing policy in place, while 8 had no formal policy and no plans to issue a policy, 6 have no formal policy but plan to issue a policy, and 2 have no formal policy and their future plans were unknown. With regards to cost recovery mechanisms, 28 agencies reported that they made users pay for their own hardware, software, and other communications costs; 14 required users to make their value added information available to the agency; 8 required that users reimburse the agency for its costs in providing data; 7 mandated that the user share a portion of their revenue generated; and 7 specified that the user must make an “in-kind” contribution to the agency.

Once the philosophy to data sharing is firmly entrenched, the next step is to navigate the set of factors that influence the process of data sharing. They are: (1) Cultural and behavioral, (2) Technological, (3) Institutional and managerial, (4) Financial and budgetary, and (5) Legal and policy (Arzberger et.al, 2004). Different agencies have different procedures for data sharing and they are typically a reflection of these five themes. The regional and organizational culture as well as perceptions

about what the agency's role is in the data sharing experience to a great degree affects the perception of the stakeholders requesting the data from the agency. Hallenbeck et al. (1998) explore issues associated with data emerging from Intelligent Transportation Systems (ITS) and Advanced Traveler Information Systems (ATIS). The technological nature of these data streams lend themselves to being disseminated with the actual model for data sharing dependent on cultural, institutional, legal, and financial factors. However, the breadth of these arrangements can be captured within two types of models: the model in which the public sector generates, fuses, and disseminates the data and maintains a high level of control; on the other extreme is the franchising model, where the public agency collects the information but a private sector partner manipulates and diffuses it. In the latter case, the private sector would acquire the rights to market the data in exchange for improved access and use of it. As the literature and the interviews suggest, there are variations to these two extremes leading to many hybrid approaches to data sharing.

III. SCAN OF OTHER STATE PRACTICES (SEMI-STRUCTURED INTERVIEWS)

Interviews of state agencies to understand their data sharing policies and practices showed that the philosophy toward data sharing favoring unrestricted access as the default policy for public data requests. This is especially the case in Texas and Alaska, where data sharing is seen more as a free public service, and in the case of non-specific or aggregated data such as traffic counts or road centerline files. For requests made by university researchers and consultants in most states, a non-disclosure or confidentiality agreement is often used to clarify legal requirements and obligations. In New York, these are used as standard procedure to establish a minimum level of security as well as to "weed out" requests. NDAs and confidentiality agreements may have standard language or may be drafted on a case-by-case basis. The Ohio DOT typically issues an exclusivity agreement for researchers using employment and railway build data, as they reveal specific latitude and longitude coordinates of survey subjects. The Florida DOT only shares confidential data with agencies it is familiar with. All agencies reported masking all travel survey data to obscure individual identifiers, and all reported charging for data if the research intends to profit from a final product. Disaggregated freight data almost always requires an NDA because it involves specific origin and destination coordinates. While the philosophical stance within one agency is a guiding factor and in some instances looked at as being restrictive toward data sharing, it can potentially be a cause for concern when agencies are either mandated by law or are working on a project of regional or local significance warranting a pooling of resources.

RECONCILING DIFFERENT POLICIES IN PARTNERSHIPS

Issues may often arise in such partnerships or collaborative endeavors where multiple entities integrate their systems and procedures. A case study of the TransGuide Metropolitan Model Deployment Initiative (MMDI) project between San Antonio, Seattle, Phoenix, and New York by Booz-Allen (May 2000) highlights the difficulties caused by differences in data sharing philosophies among the participating agencies. The state of Arizona, for example, has an open data sharing policy with law

enforcement, though in the case of video monitoring they conceded this policy to conform to the policies of partnering agencies. Of the states that were studied for this research, Texas likely has the most transparent data sharing policy. In accord with the Texas Open Records act, minimal restrictions are placed on data and dissemination is achieved swiftly and comprehensively. New York, meanwhile, is much more active in distinguishing those who should and should not have access to data. These differences are reconciled by employing written agreements that specify what information would be allowed on changeable message signs, how to use closed circuit television (CCTV) cameras, assigning responsibility for maintenance and replacement of equipment, and the distribution of automated vehicle identification tags. The Booz-Allen study states that such agreements are needed to keep procedures consistent and to guard against liability, though in their opinion these policies are cumbersome to implement.

An interview with a NY DOT representative revealed more about the state's relatively stricter sharing policies. DOT officials are required to meet with legal council before entering projects with federal agencies and in the case of a request for data made under the Freedom of Information Act (FOIA). These policies have played a role in stifling a data sharing project between Port Authorities of New York, New Jersey, Pennsylvania, and Connecticut, as each state has its own concerns about disseminating waybill samples and other data.

DEALING WITH DATA LIMITATIONS

DOTs face a host of data sharing issues that are simply a result of inherent deficiencies in the data set. Those requesting and using transportation data are often not those collecting and organizing it. This may lead to differences in preferences for the organization and content of a dataset.

Dueker et al. (2000) investigate this problem in the context of Geographic Information Systems data and identify several issues. Different agencies have varying levels of GIS aptitude and capacity - some may be on the cutting edge of GIS while others are still growing accustomed to technologies that have been out for a few years. The authors report a general lack of agreement among transportation organizations on how to define transportation objects, as well as issues pertaining to the level of spatial accuracy required. Forest roads and private roads, for example, may not be accounted for by a DOT because they are not maintained with public money. This leads to difficulties conflating and integrating multiple views of the same or adjacent linear object. Another issue is that some state agencies may have trouble keeping data current enough for stakeholders who need the most up-to-date data. This is particularly pressing for E-911 computer-aided dispatch services where updates must happen in real-time. Among the recommendations made by the authors to alleviate these issues are relinquishing data maintenance duties to local agencies and private companies, using File Transfer Protocol (FTP) servers where data can be updated and improved by outside parties, and employing enterprise data warehouses such as Oracle and ESRI.

One of the important factors that dictate the sharing of data is the organizational needs of the agency in-charge of collecting and maintaining the data. The needs of the organization as well as the financial and budgetary limitations will drive the outlook toward data acquisition, storage, and sharing. Interviews with DOT representatives highlighted some of these issues. An agency might choose not to collect information because of financial and budgetary reasons. This was evidenced in the case of Alaska, where due to the cost associated with collecting GPS data in the most remote regions many of the state's transportation assets are unaccounted for. Some states may sacrifice the precision and accuracy associated with the data such as road files on the pretext that their work doesn't require the precision that a local agency would need. Meanwhile, the local agency may not have the resources to construct highly detailed GIS shapefiles. The Federal Highway Association (FHWA) has issued recommendations for states to use common shapefile formats but state representatives have expressed skepticism to this since it would require significant effort to adopt new federal procedures and formats that, as one DOT representative indicated, "could change again five years from now." This is further illustrated by the differences in agency practices when dealing with even publicly available data sets such as the decennial census.

The level of aggregation, particularly in census data, has often been a problematic one for researchers. Federal data may be collected as high up as the zip or county level when it is often needed at the block level or census tract. Some states (New York and Virginia) have responded by doing their own census add-on, which can be very costly. This is especially problematic with respect to freight data. Origins and destinations typically cannot be disclosed between agencies since they reveal flows between private sector carriers; instead freight flows are often aggregated to the county level. Yet this does the researcher little good in tracking specific flows. Furthermore, depending on who collects the data (the federal or state government, a private consultant, or a freight company reporting on their own shipments), categorizations by commodity type, mode type, and quantity may be inconsistent. This highlights the need for a uniform classification system to improve freight data quality.

DEALING WITH BUREAUCRACY

Many of the complications facing state DOTs stem from the fact that they are obligated to serve (or at least entertain) any data request while private firms are much more selective in who they deal with, allowing them to be much more specific in their scope and in who they respond to. This often leads to inefficiencies in public agencies, which slows the flow of information. Zimmerman et al. (2002) report that cooperation among agencies is "not what it needs to be" based on their survey and interviews with private and public sector representatives. One state employee in their survey cited a lack of leadership at the state level in enforcing a streamlined data sharing process and understanding the needs of data users. Another private sector partner claimed that public agencies are not as cooperative as they need to be and referred specifically to a dispute between a state DOT and an MPO regarding the dissemination of data that they share.

Many of the sentiments highlighted in Zimmerman et al. (2002) are echoed in the DOT interviews conducted as part of this research by UIC. According to an employee at the Virginia DOT who was part of the UIC interview, the state government in Virginia has outsourced its information technology systems to a private firm. The result of this has been more emphasis on security at the exchange of productivity and fluid exchange of information, evidenced by the frequent closing of the file sharing server. These issues lead to a thought amongst public sector employees that it might be better to exchange documents in person than to rely on the server.

A respondent from the NY DOT reported a great degree of difficulty in managing requests for data. It was believed that the problem was mainly due to state budget crises that necessitated the downsizing of state agencies. This typically leads to the loss of institutional knowledge as a result of frequent staff turn-over. Consequently, only a fraction of the people available decades ago is on hand now to field requests from data users. The data sharing experience then suffers as users must wait longer for requests and providers are overwhelmed by requests.

Another issue related to bureaucracy is that data seekers are often unaware of how to navigate it and get what they need. Several respondents to the semi-structured telephone interviews reported consistently having to field requests for data that they did not have. An employee at the Alaska DOT claimed that he frequently must turn down vehicle crash data because the DOT passes this information on to the Department of Traffic & Safety and does not maintain it. Employees at FDOT, ODOT, NYDOT, VDOT, and NCTCOG all mentioned that users often request the wrong data because they simply do not know what data is best for their project. Data providers all reported spending a considerable amount of time helping people determine what data they actually need.

OTHER ISSUES

DOTs must be mindful that the dissemination of confidential or private data can lead to liability issues with stakeholders. The Ohio DOT has adopted a strict confidentiality policy with stakeholders to protect the intellectual property of their research. If this information is revealed, it could compromise the competitive position of a private consultant or research institution and even result in a lawsuit. The Virginia DOT demands that anyone using their data sign a form stating that they accept responsibility of the product of that data. This is in response to the agency facing criticism for their data when researchers used it to do forecasts that turned out to be inaccurate.

A representative from the Florida DOT reported having observed issues in screening data users. This is a potentially critical problem with respect to security, since some datasets can reveal important information about an individual or a particular transportation facility. Though strict data sharing restrictions are enforced on key datasets to address this, it is not impossible to circumvent these channels. Virginia and New York also cited this as a common concern, and they follow strict screening policies for all researchers.

IV. STRATEGIES FOR FACILITATING DATA SHARING

DATA GOVERNANCE

A study for the Alaska DOT, by Cambridge Systematics (2009), to identify ways of streamlining their data sharing practices led to the establishment of a formal data governance initiative headed by a data governance board and characterized by clear roles and responsibilities for internal and external parties of interest. 5 core data areas – highway safety, traffic, road weather management, traveler information, and GIS services – were each assigned their own procedures to accommodate different stakeholders that would need each type of data. Finally, a single point of contact was established for each core data area: this minimized the risk of disclosing confidential data and created a more organized environment where stakeholders can be better served.

The report by Cambridge Systematics also included case studies from the Oregon DOT, Virginia DOT, and Kansas Department of Education. At ODOT and VDOT, a data council has responsibility to govern the development and implementation of an electronic data dictionary, templates for workflow processes, and data modeling repositories. The objective in these instances was for data to be collected once and to be used and updated many times by interested parties. A 7-step plan developed for these agencies includes the following steps:

- 1) *Assign Roles* – Business owners, data custodians, data stewards, data coordinators, data architects, etc.
- 2) *Produce Initial Systems Operations Data Catalog* – create templates and examples for data catalog elements; assemble information about each data product; data catalog is a living document that is built over time with the most critical information incorporated initially.
- 3) *Develop Systems Operations-Wide Data and Business Process Models* – In the case of Alaska, a business process model is developed for each data type (highway safety, traffic, road weather management, traveler information, and GIS).
- 4) *Develop Systems Operations-Wide Business Requirements for Data Access* – these requirements include defining the needs for reporting and analysis; how often the data must be updated; who it should be sent to on a daily, weekly, monthly basis, etc.; analyses that must be done for each data set.
- 5) *Estimate Data Acquisition and Maintenance Costs* – The development of a standard methodology for the collection and maintenance of data; this should include assessing the value of data programs in terms of “cost” factors.
- 6) *Initiate Annual Business Data Review Process* – This involves two steps:
 - Data Product Review – Review the status of each data product, improvement needs, and estimated costs of improvements; this should facilitate the annual IT budgeting needs to support these data systems.
 - Data Acquisition Review – Review opportunities for improved efficiencies and reduced costs.
- 7) *Establish Communication Protocols* – This crucial step involves establishing clear lines of communication between stakeholders and those responsible for maintaining the various

data programs and establishing channels of dissemination (web, internet, intranet, email, FTP, etc.).

While the two DOTs have only recently adopted the changes, the Kansas DOE has been pleased with the outcomes. In the latter case, the report cites more consistent treatment of requests, more efficient prioritization of requests, and the ability to automate the tracking of requests. Apart from these efforts by the states, there have been efforts at the national level to advance the understanding and improve the practice of data sharing.

The National Cooperative Highway Research Program's Report #525, Volume 5 (2005) is a synthesis of best practices and recommendations for transportation agencies in managing sensitive information. Though it was drafted in response to terrorism threats after September 11, many lessons can be applied with respect to other liability and safety issues. Because state and local agencies cannot always rely on methods used by federal officials to control sensitive information, this report is meant to provide a robust framework for evaluating key datasets and creating guidelines for dissemination. There is significant overlap between the Cambridge Systematics recommendations and those listed in the NCHRP report and these are:

- 1) Create a data oversight committee with representation from a wide variety of stakeholders.
- 2) Identify sensitive information. These could be reports that give detailed information about the vulnerability of infrastructure, risk assessments or infrastructure, countermeasure procedures, emergency response plans, or structural plans for buildings.
- 3) Datasets that are not available from other sources and may require special handling procedures should be earmarked as sensitive.
- 4) A clear channel of information flow should then be established with a single point of contact (i.e. data governance).
- 5) Once these steps are taken, the agency should identify individuals who have a legitimate need to know certain information.
- 6) All documents deemed sensitive should be marked with clear warnings and this information should be stored responsibly by removing from shared computer networks, making it password protected, etc.
- 7) Finally, all employees should be informed of the agency's data sharing policies to make sure procedures are followed throughout.

DEAL WITH REQUEST BASED ON DATA TYPE/STAKEHOLDER TYPE

With the exception of Alaska and Virginia, none of the agencies surveyed had a formal data sharing governance policy. This could be because data sharing practices are too firmly entrenched in the organization's culture, not enough staff is on-hand to make it worth adopting, or data sharing may simply not be an agency priority. In most cases data requests are handled on a case-by-case basis subject to the best judgment of those overseeing the data, though common practices are seen for some datasets.

All agencies surveyed reported masking unique identifiers and using nondisclosure agreements for household travel survey data and employment/wage data. Virginia and New York reported using remote access agreements for some data sets. Ohio DOT and NCTCOG report charging the researcher if they intend to profit on the dataset or if the data preparation effort is especially extensive. The NYDOT reported having to consult a FOIA department for all Freedom of Information Act data requests and the agency legal department before entering contracts with the federal government.

In handling data requests, NCHRP (2005) recommends utilizing practices consistent with state “sunshine laws” that encourage free access to as much information as possible. These laws are often based on the federal Freedom of Information Act. However, there may often be cases where the state looks for exemptions in the legislation as a legal basis to withholding sensitive data, and NCHRP Report 525, Vol 5, provides these. FOIA has two key exemptions that are also likely to apply at the state level:

- **Exemption 2** – This exempts records “related solely to the internal personnel rules and practices of an agency.” Courts interpret this as including two types of information: routine internal administrative matters (or “low 2” info) and more substantial internal matters, the disclosure of which would risk circumvention of a statute or agency regulation, (“high 2”). The underlying concept is that FOIA disclosure should not facilitate the violation or circumvention of the law.
- **Exemption 5** – This protects “inter-agency or intra-agency memorandums or letters which would not be available by law to a party...in litigation with the agency.” It includes “pre-decisional” documents whose disclosure would inhibit open dialogue among officials regarding agency policy. Documents protected under this exemption include security analyses and recommendations as well as related draft letters and memorandums.

The following is a list of key exemptions to state “sunshine laws:”

- *New York* – release of information can be restricted if it “endangers the life or safety of any person.”
- *Illinois* – information can be withheld if it endangers “the life or physical safety of law enforcement personnel or any other person.”

Some DOTs have changed sunshine law language, adding specific exemptions to safeguard critical infrastructure data. This allows them to legally deny requests for these documents:

- *Florida* – can exempt “building plans, blueprints, schematic drawings, diagrams, drafts, preliminaries, final formats, which depict the internal layout and structural elements of a building, arena, stadium, water treatment facility, or other structure...
- *Maryland* – can exempt “response procedures or plans prepared to prevent or respond to emergency situations, the disclosure of which would reveal vulnerability assessments, specific tactics, specific emergency procedures, or specific security procedures

- *Missouri* – can exempt information about “existing or proposed security systems and structural plans of real property owned or leased by a public governmental body, the public disclosure of which would threaten public safety.”
- *Texas* – can exempt information if “the information is collected, assembled, or maintained by or for a governmental entity for purpose of preventing, detecting, responding to, or investigating an act of terrorism or related criminal activity
- *Virginia* – can exempt “plans and information to prevent or respond to terrorist activity, disclosure of which would jeopardize safety of any person, including 1) critical infrastructure or structural components 2) vulnerability assessments, operational, procedural, transportation, and tactical planning or training manuals, staff meetings minutes or other records, and 3) engineering or architectural records
- *Washington* – can exempt records that have been maintained to respond to criminal terrorist acts as well as “specific and unique vulnerability assessments or emergency response plans intended to prevent or mitigate criminal terrorist acts.”

OTHER STRATEGIES

A few pieces of literature touch on specific strategies for facilitating data sharing between public agencies. DeBlasio (1999) explores data sharing agreements between Washington, Texas, Arizona, and New York DOTs in the MMDI ITS pilot project. The author cites the importance of explicit written agreements in determining how to distribute ownership and transfer of data and assets. Written policies were also negotiated to find an optimal policy for sharing data with other state agencies and researchers. Liu et al. (2002) scans best practices in ITS data management and archiving. This article provides a useful synopsis of DOT data collecting methods, key data elements useful to specific practitioners, and case studies of pioneering technologies. One of these, the Performance Measurement System (PEMS) developed by CalTrans, obtains 30-second loop detector data in real time and presents this information in various forms to assist managers, traffic engineers, planners, freeway users, researchers, and value-added resellers or travel information service providers.

Watson (2008) presented ideas for improving data sharing between public and private partners at a GIS conference in 2008. The author here recommends separating core managed data (data formatted for optimal data maintenance tasks and restricted access) from derived published data (data optimized for easy access, simplified for a wider audience, and generally unrestricted access internally). Some of this core data could then be transformed to derived products that can then be loaded into a Director Resource Site (DRS). Referenced DRS metadata is updated regularly and replicated among partner sites. In this way, core data from partner agencies can become local data benefitting the functions of the “mother” agency. Updates will then occur regularly and all data and applications can be kept up-to-date more easily. Other technologies recommended are FTP servers, Web Map Services (WMS), Web Feature Service (WFS), Keyhole Markup Language, File Replication Service (FRS), cloud servers, and geoprocessing.

Holman (2007) suggested recommendations for the North Carolina Geographic Information Coordinating Council, a data governance council responsible for the warehousing of all geographic data in the state. With respect to geographic data, the author made the following recommendations:

- *Avoid formal agreements* – they are often unnecessarily restrictive, although exceptions should be made for records that are deemed confidential or that could pose a public safety or security risk.
- *Web access* – Making data widely available on the web will allow; use secure access for more sensitive information.
- *Single point of contact* – This is consistent with the data governance procedures mentioned above.
- *Regional solutions* – Local agencies should collaborate with others if they lack the necessary technical capacity or resources.
- *Official outlets* – Data can only be accessed from the original source. Secondary sources should obtain permission to redistribute and should include original source in the metadata. This will make it easier for primary agencies to keep track of who is modifying their data.
- *Archive/long-term access* – Data producers should evaluate and publish long-term access, retention, and archival strategies for historic data.
- *Keep data free* – this is consistent with most literature on best practices in data sharing; allowing free access to non-sensitive data provides the best chance for the information to be maximized. The legal basis for the principle of free access to data is established by OMB Circular No. A-130, which encourages these practices on the basis that data is collected with taxpayer money and should therefore, be used to improve public welfare.

V. EXPERIENCES IN OTHER FIELDS

Data sharing issues are common to all types of organizations in all fields. As mentioned above, the Kansas Department of Education implemented a formal data governance policy to improve its data sharing and research. The National Center for Education Statistics, meanwhile, has formal internal review groups that evaluate all proposed microdata releases. The National Agricultural Statistics Service has a formal procedure for obtaining written permission statements from survey respondents. This is necessary because these respondents often work in rural areas and their activities account for a relatively large proportion of area activity.

Perhaps the best example of public benefits to open data dissemination is the National Weather Service (NWS), discussed in Weiss (1997). Good weather data are valuable for improving travel safety and reducing the costs of travel for business and individuals. The NWS therefore discourages giving exclusive rights to data to private organizations for a fee because it believes in keeping the data more accessible and stimulating competition in the short- and long-run. However, the NWS makes a few conditions clear to anyone using their data: users assume all risk and liability related to use of the data, users must state that NWS data is not subject to copyright protection, and users cannot present modified data as official government material. These stipulations clear the agency of any legal threat.

The author gives three arguments in favor of free access to data. First, the exchange of data between government agencies introduces a shell game where money is moving but not benefitting the overall economy. Second, there is generally a high elasticity of demand for government data. Most

companies and organizations wouldn't buy it even if the price were very low. Finally, the authors cite the problem of *cream skimming* of markets. This means that if the cost of purchasing data is equal to or more than gathering it independently, private companies will simply collect it themselves and they will start with the most lucrative markets. Government agencies will then be left gathering the most expensive or unprofitable but necessary data. To illustrate this point, the author uses the example of the government of England, which at one point charged for maps of most cities. Private mapmakers then began making better maps of the same geographies, and eventually the only maps that the government was responsible for making were of the most remote regions where collecting geographic information was extremely costly.

VI. KEY POINTS

A few common themes reoccur in the literature. They are:

- Sharing information freely should be standard procedure. Interviews with DOT and MPO representatives implied that this is indeed the case. Still, much can be done to improve the availability of some data.
- Agencies should refer to the website for the Panel on Confidentiality and Data Access (PDCA) to learn more about how and when to apply restrictions on confidential data sets.
- Several sources recommend a formal data governance plan. This could be useful in helping agencies identify and protect sensitive data, organize their data dissemination processes, improve service to data seekers, and make it easier to know who has used what data set.
- Clear written agreements are best when entities with different data sharing policies engage in partnerships.
- Agencies should utilize File Transfer Protocol servers and File Replication Servers to facilitate a more dynamic interplay with stakeholders. Local and private sector counterparts can help update and maintain data through these servers, which can greatly relieve the burden of the state agency.
- Despite the pressing need to improve security in IT, agencies should make sure this does not come at the expense of productivity and flow of information.

CHAPTER 2. SURVEY OF RESEARCHERS

SURVEY DESIGN

The survey was designed in conjunction with the project advisory panel and was aimed at understanding the stakeholder experiences pertaining to data requests from the DOT. The research team obtained a list of researchers from the Wisconsin Department of Transportation and from MRUTC. The literature review and the scan of other state experiences helped the research team in identifying critical issues in the realm of data access vis-à-vis public agencies and research communities while navigating relevant and tangential issues such as privacy, academic freedom, etc. The list of questions included questions aimed at understanding the context of each respondent about their experience in dealing with Wisconsin DOT, followed by a section about the projects that required accessing data from the DOT. In this regard, the research team made use of the information provided by MRUTC about research projects funded by MRUTC or by Wisconsin DOT and managed by University of Wisconsin. This provided a list of projects for each researcher dating back to the year 2000. If the researcher indicated that they had requested data from the DOT for the projects listed, then they can proceed to the survey. The data sets were self-reported by the respondent along with a reflection of their experience with that data request. The survey itself was divided into three sections, one each for the best, worst, and most important data sets that they had requested from the DOT. The same set of questions was repeated for each of their experiences. This allowed the research team to compare different data request experiences ranging from worst to best to most important for the same sets of criteria.

The next step in the survey design was the vehicle for survey administration. The research team developed a customized web page for the survey. The survey was customized for each respondent by including the projects that they had worked on. Each respondent received an email from the research team explaining the objectives of the project along with a link to a web page. This information was in the database that UIC received from University of Wisconsin/Wisconsin DOT. If a researcher had worked on three projects, say, then the clickable link will lead the researcher to their survey questionnaire customized to reflect their input on projects that they had worked on.

IRB APPROVAL

The University of Illinois at Chicago monitors and regulates research involving human subjects in order to ensure that the subjects' rights are not violated. The monitoring is done through the Institutional Review Board (IRB) and the research protocol including the survey questionnaire was screened by the IRB before subjects could be contacted. Every respondent to the survey was required to be made aware of their rights pertaining to the participation in the survey. The survey instrument was designed with the informed consent form being a part of the survey. If any subject indicated that they were not voluntarily willing to be part of the survey, they were removed from the survey pool. In this research, one researcher indicated that they did not want to proceed with the survey. The other aspect of the IRB governs the privacy of the research subjects. Respondents had the option of informing the research

team if they did not want their names to be included in the report in order to maintain their anonymity/privacy.

SURVEY ADMINISTRATION

The survey was administered online using a customized link for each recipient. The population was contacted via email describing the goals of the research, with a link to a web-page (see appendix). The link took the recipient to the survey page which described the project objective and provided an informed consent form. If the recipient agreed to be a part of the survey, they proceeded to answer the questions discussed in the previous section.

SURVEY RESULTS

1. INTRODUCTION

Addressing the issues associated with transportation infrastructure requires increasingly sophisticated data. The formulation of effective public policy and efficient public spending is intimately correlated with the policymaker's understanding of behavioral patterns not only on a macro level but especially on the disaggregated level. Unfortunately, the dissemination of such data between agencies is often subject to friction in the form of privacy and confidentiality regulation, issues with the format or neatness of the data, and issues associated with procurement. Public agencies especially have difficulty implementing optimal data access standards because they are held directly accountable for both sides of the equation: transparency on one hand and security and protection on the other. These hurdles are made more complicated when multiple institutions, with their own unique cultures, procedures, and preferences, are working together to achieve a common objective. Having considered these issues, it is easy to see the need for these institutions to streamline the process of accessing and sharing data.

Researchers at the University of Illinois – Chicago's Urban Transportation Center, in conjunction with the Midwest Regional Urban Transportation Center at the University of Wisconsin – Madison, are working toward improving the flow of data between the Wisconsin Department of Transportation (WisDOT) and their private and public partners. The objective of the project is to identify problematic datasets and data procurement practices and amend them using existing literature on best practices as well as feedback from stakeholders. An online survey was employed to distinguish the stakeholders that had negative experiences obtaining data from WisDOT as well as the datasets they had problems with.

Candidates for the survey were identified with the help of WisDOT, and the link was sent to 87 researchers that have worked with them in the past. 11 of them did not have valid email addresses thus shrinking the universe of cases to 76. Of the remaining 76, 26 people consented to taking the survey. This represents a 34% response rate which is acceptable for a web-based survey. Out of the 76 valid email addresses, one refused to provide consent to taking the survey. Thus, 27 out of the 76 people that received the email link to the survey, clicked the link to open the survey page. Of these, 23 individuals representing 13 research institutions submitted responses. Each individual is currently working on one or more projects for WisDOT or had worked on projects in the past, and reported procuring at least one dataset from the agency. The relationship is not necessarily one to one: one

researcher reported working on 7 WisDOT-funded projects but was only using 3 datasets while another was working on only one project but using two datasets. In all, respondents reported 39 WisDOT-funded projects and - coincidentally - 23 datasets. The results of the survey are stratified by dataset rather than project or respondent name since we are primarily interested in finding datasets that presented the most inconvenience in access. Datasets are considered to be problematic if the respondent(s) who used it gave multiple “problematic” responses in the survey with respect to their experience with that data². A list of all survey questions can be found in Appendix A.

2. EXPLORATORY ANALYSIS

23 respondents representing 13 organizations participated in the survey. The respondents were mostly from academic institutions with some respondents from private firms as well as a two state agencies and one federal agency. 24 of these projects are funded under the Wisconsin Highway Research Program (WHRP), 9 by the Council on Research (COR), and 6 by the Midwest Regional Urban Transportation Center (MRUTC) in Madison (Table 1). The projects that the researchers had worked on were provided by the DOT along with their names and contact information. This dataset also contained the project categories which revealed that 18 projects fell in the category of pavement (asphalt and or concrete), with policy and planning and technology transfer projects combining to 9 out of the 39.

Table 1 WisDOT Funded Research Programs

WisDOT funded Programs	Total
WHRP	24
COR	9
MRUTC	6
(blank)	
Total number of projects	39

The start date for the list of projects provided by WisDOT extended back to 2003 and the end date was as recent as 2009. The survey responses were classified into five categories, mainly based on factors identified in the literature. These are (1) Data Acquisition – pertaining to time or effort taken to obtain the data, and identifying the appropriate person in the agency to make contact with; (2) Data Processing – pertaining to processing time for data entry, whether this was expected and planned for, and its impact on the research; (3) Data Quality – that included questions about the format of the data, quality of the data, and its impact on research findings; (4) Data Sharing - issues connected with data sharing/ have you published the findings/ were there restrictions placed on the use of data/ were these restrictions limiting the research/ how did you reconcile the data restrictions/ if so how?; and (5) their overall experience in accessing data from the DOT.

² A problematic response is defined as the most negative or undesirable response to a survey question.

DATA ACQUISITION (Time or effort taken to obtain the data/how long did it take to obtain the data/was the request granted/appropriate contact information)

On the subject of time and effort taken to obtain the data, only 17% indicated that they spent “great” time and effort in obtaining the data, with another 22% indicating they had to expend “moderate” time and effort for the same (Figure 1). The process of obtaining the data is closely related to identifying the appropriate contact person at the DOT who will be charged with handling the request for the data and delivering the same to the researcher. This is where many of the respondents indicated facing a difficult time. 41% of the respondents indicated that they had at least a cumbersome experience in getting through to the right person within the department (Figure 2). To the DOT’s credit, 59% indicated that identifying a contact person was “easy”. This may be tied to the fact that the DOT does a good job of assigning individuals to the project advisory panel who in turn become the de-facto contact person for procuring the data for the researcher.

The researchers when asked about the difficulties in obtaining the right contact information, procuring the data, etc, impacting the quality of their research, replied in the negative. Only two out of the twenty-two indicated that the quality of their findings was compromised because of the difficulty with data access.

Data sets associated with respondents indicating a great degree of difficulty were the As-built Plan, Pavement Information Files, Pavement Performance Data, and Structural Data. A majority of the respondents indicated that they were able to get the data set they wanted within two weeks’ time. Waits of over three months in procurement were reported for the As-built Plan, Pavement Performance, Permit Vehicle Data, and Structural Data (Figure 3). With exception of one request, the rest of the respondents indicated that their request for data was granted by the DOT.

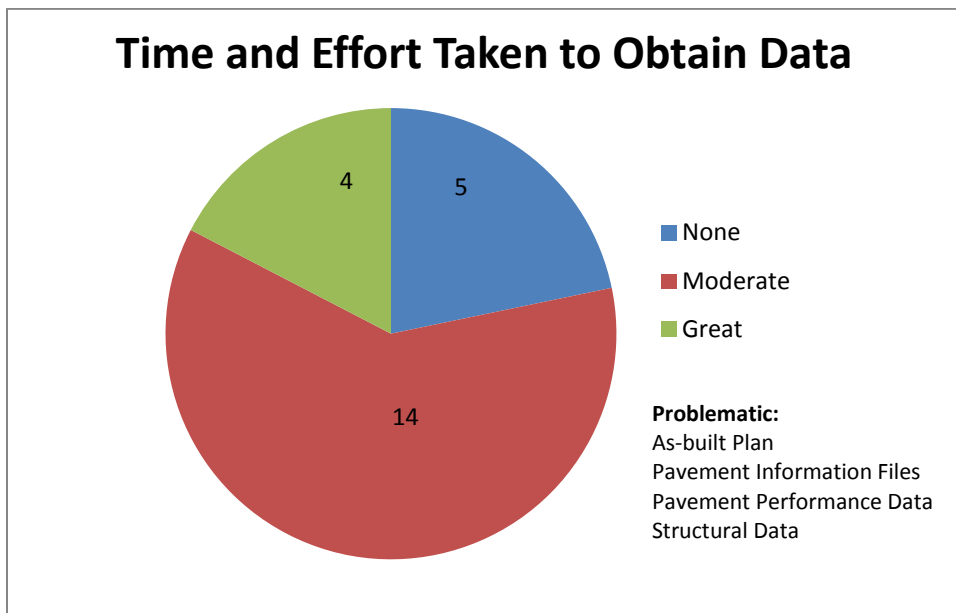


Figure 1 Time and Effort Taken to Obtain Data

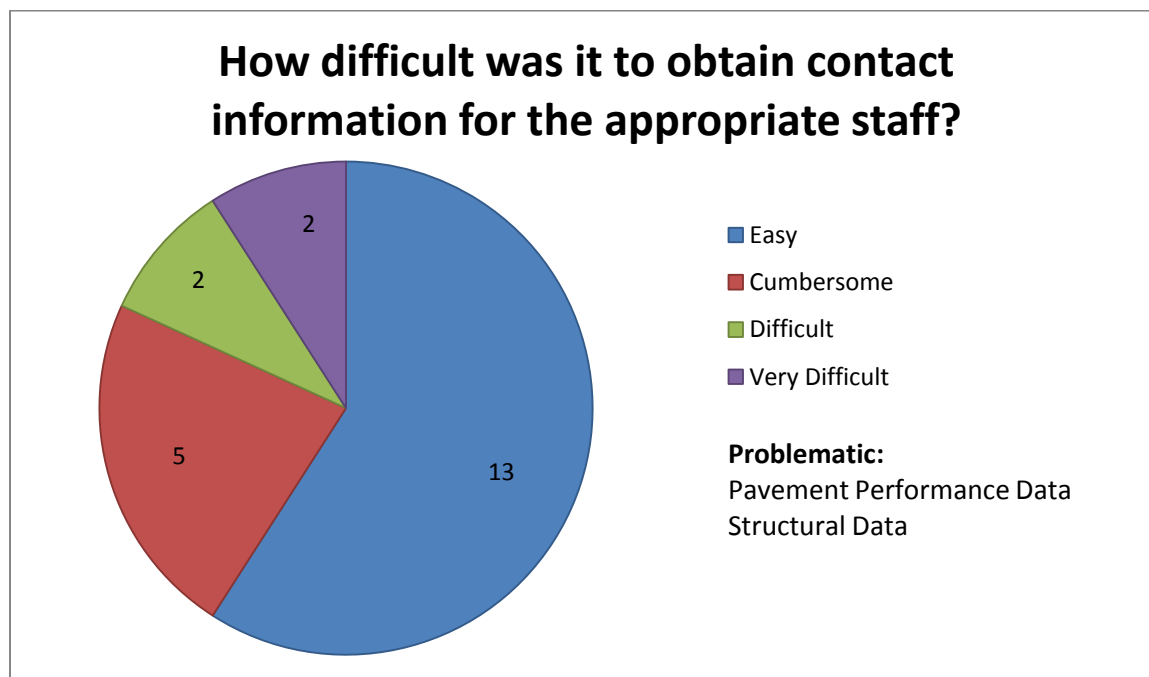


Figure 2 Difficulty in Obtaining Contact Information

How long did it take you to obtain the data you requested?

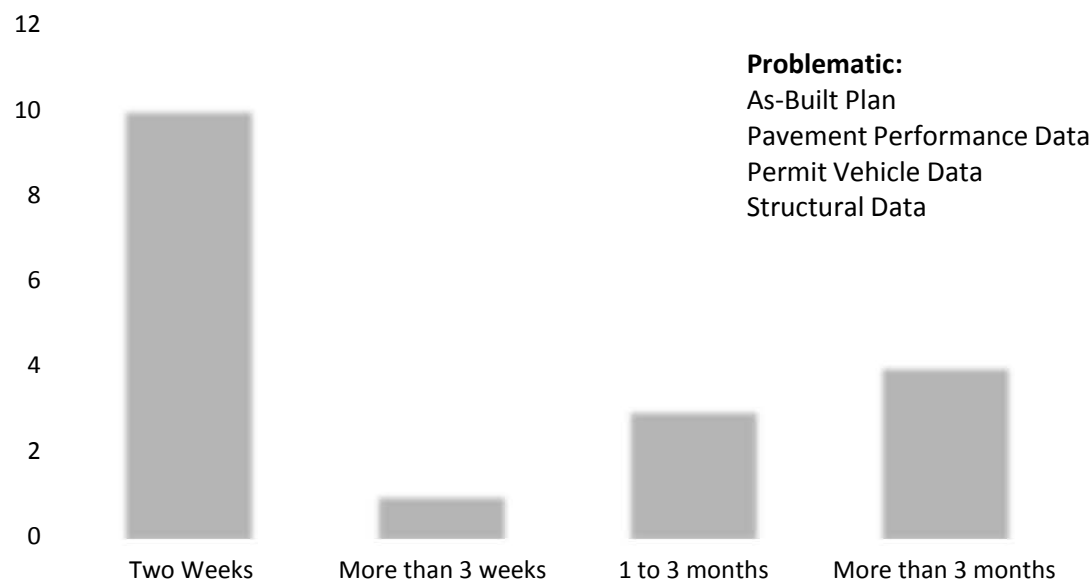


Figure 3 Time to get the Data

DATA PROCESSING (processing time for data entry/time spent preparing the data/was this expected/was it planned for ahead of time/was research delayed due to preparation of data)

The same datasets that were reported as taking more than three months' waiting time were also cited as presenting a "great" degree of difficulty in processing time. The results from Figure 4 show that the delay was not anticipated for by these researchers. Each of these - along with Real Estate Maintenance Procedures, Ridership Data, and Socioeconomic Data – required over a month of preparation once they were obtained. Pavement Performance, Structural Data, Permit Vehicle Data, and List of Environmental Commitments were all reported as having caused delays in the research process resulting from excessive time in data preparation. These along with the As-built Plan, MetaManager, and Pavement Information Files were cited in the survey as those presenting difficulty associated with formatting. In multiple cases there were issues with datasets being maintained in Microsoft Access while the researcher had a preference for SAS. Researchers indicated that they had not planned for the delay stemming from data processing ahead of time resulting in unanticipated delays for four of the data sets (Figure 5). The delay in processing the data played a part in the overall research timetable being extended. Three of the researches indicated that their research was delayed due to the unanticipated delays associated with data processing.

Was the time you waited expected?

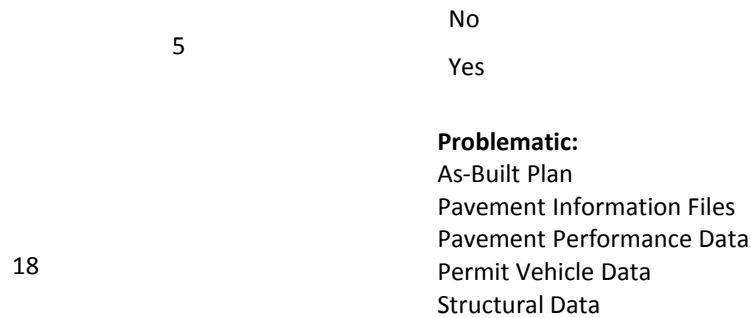


Figure 4 Expectation for Wait Time

Processing Time (due to incorrect format, data entry issues, etc.)

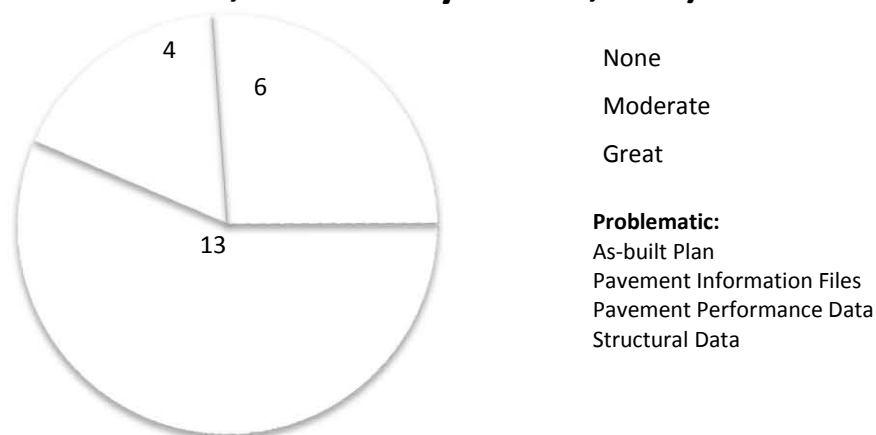


Figure 5 Processing Time

DATA QUALITY (was data delivered in the form desired/was quality of data compromised/were the findings compromised)

This viewpoint is supported when looking at the question pertaining to the format of the data set(s) requested. Almost all of the researchers indicated that they received the data set in the desired format, except for five respondents who did not receive the data in the format desired (Figure 6). This is also reflected in their responses pertaining to data processing and data preparation (Figure 7). A

majority of the respondents indicated that it took them longer than three weeks to prepare the data once they received it from the DOT. Of these, four indicated that their research efforts were delayed due to the data preparation time (Figure 8). As for whether the research quality was compromised due to the data quality, there were four experiences that replied in the affirmative (Figure 9). Of these four, there were two experiences indicating that their overall findings were compromised due to the difficulties associated with data.

Were the data you requested delivered in the format required?

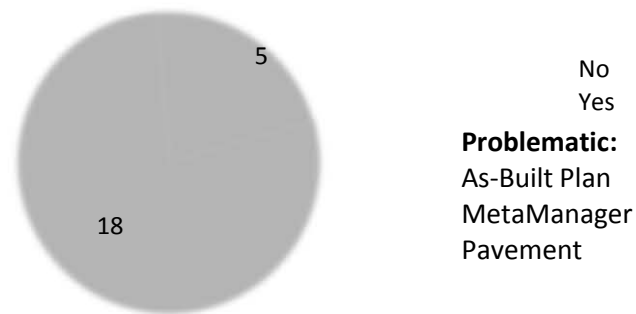


Figure 6 Data Format

How long did it take you to prepare your data?

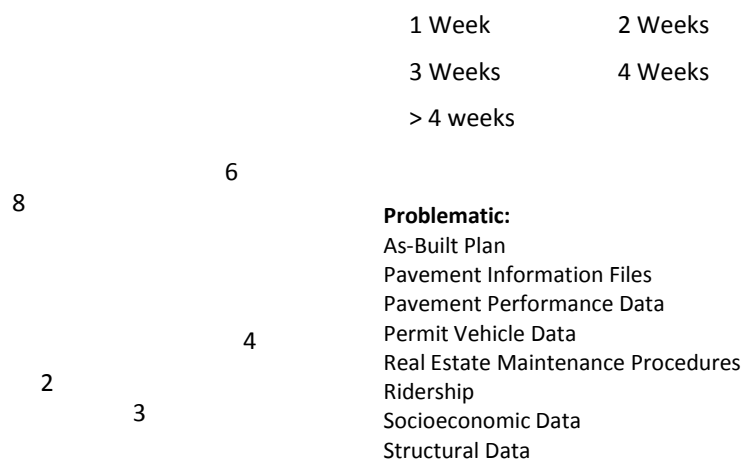


Figure 7 Data Preparation

Was your research delayed by preparation?

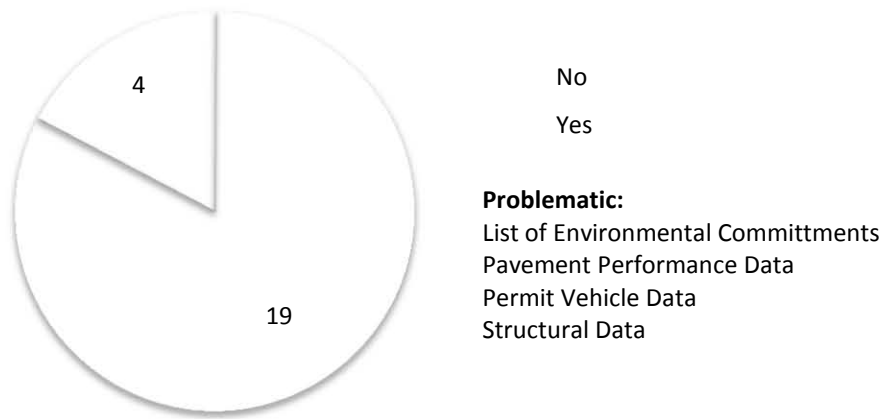


Figure 8 Delay in Research Due to Preparation

Was the value of your findings compromised by the data?

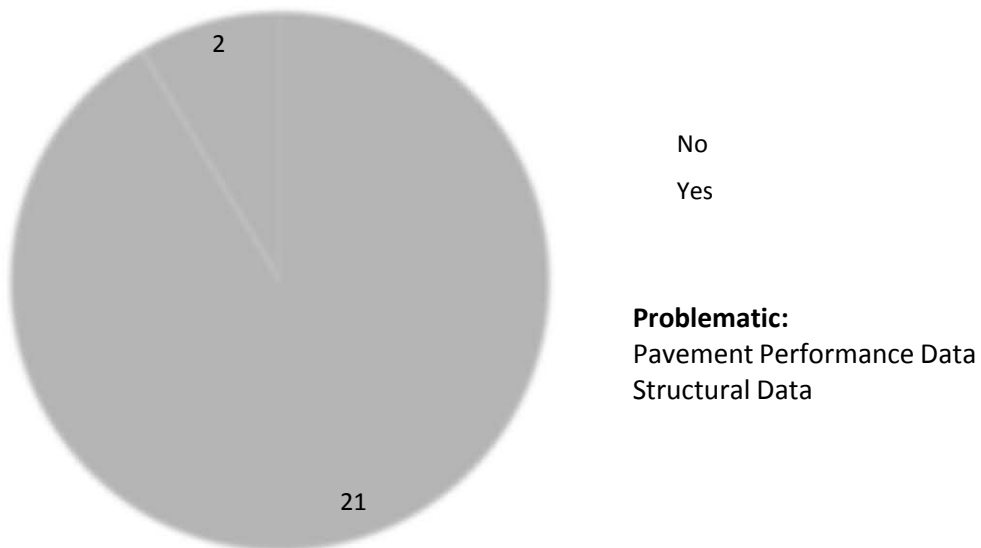


Figure 9 Value of Findings

DATA SHARING (issues connected with data sharing/ have you published the findings/ were there restrictions placed on the use of data/ were these restrictions limiting the research/ how did you reconcile the data restrictions/ if so how)

While the time and effort taken to obtain the data and processing it was perceived as somewhat problematic in certain instances, the feedback pertaining to data sharing and publication of the results of the analysis did not register as a major issue according to the survey respondents. Only 2 out of the 23 indicated that they had difficulty when it came to publishing the results of their research (Figure 10). This was attributed mainly due to the difficulty associated with acquiring the data sets and not necessarily due to any restrictions placed by the DOT regarding data sharing.

Issues related to data sharing and publication

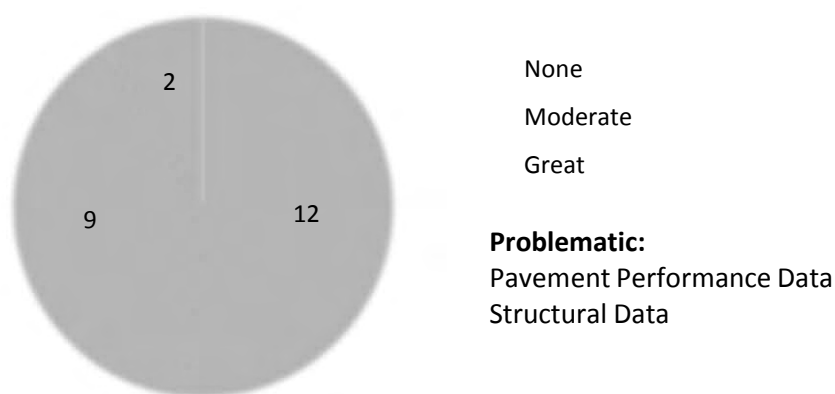


Figure 10 Data Sharing and Publication

The respondent who used Pavement Performance Data and Structural Data reported a “great” degree of difficulty in sharing and publishing data, a “very difficult” experience in obtaining contact information for the appropriate staff, and highly compromised value of research findings.

OVERALL EXPERIENCE

The overall experience about data access was captured in two different questions – one about the best and worst experiences the respondent had had in accessing data from the DOT and the other in a open-ended question that allowed them to elaborate on their responses to the entire survey. The table below provides the responses about best and worst data sets (Table 2). As can be seen from the table, there is no dataset with more than one experience in either the best or worst categories. The only exception was the Pavement Information Files (PIF) which was reported as the best experience in two responses.

Table 2 Datasets with Best and Worst Experience

Best Experience Datasets	Worst Experience Datasets
Approved repair materials list	
Contacts in region	As-built
culvert survey	DB 100
Introduction to internal WisDOT procedures for maintenance and real estate related to assessing potential to effect cultural properties	Existing bridge plans
PIF	List of environmental Commitments
Ridership	MetaManager
Soil boring information	Pavement Performance Data
Statewide Weigh in Motion Data	PCC Mix Design Data
Structural Data	Traffic Data

The following table illustrates the responses received from the respondents regarding their overall experience in accessing data from the DOT. The experiences ranged from those that highlighted the many missing details in the data requested, to those that expressed complete satisfaction with the process of obtaining the data. However, the survey did highlight the fact that accessing data was still a tedious process especially when it involved collecting information from multiple sources and data owners.

Table 3 Respondents' Overall Experience

S. No.	Overall Experience
1	The data missed the details needed but that is very usual around the country
2	There were many challenges to pulling together the information from a variety of formats and data owners.
3	The mix design data and related compressive strengths did not correlate with the actual mixtures used for construction
4	The statewide data was extremely useful in developing reliability-based traffic factors for use in pavement design/rehab
5	Pavement thickness was obtained from design plans
6	The information received was complete and within a reasonable time period
7	Obtaining the data was easy and was necessary for initial planning purposes
8	It was characterized as the worst because the Historic Structures Database is cumbersome to use
9	Can not do anything without. It is in a format of MS Access. Relatively user-friendly.
10	Everything has to be done manually

The respondents were also asked an open-ended question to provide feedback about the data accessing practices and these responses are listed below in the table. This question received 8 responses (35% response rate). It is interesting to note that almost all of the responses gathered for this question echoed the same sentiment - that of an integrated database operation within the DOT.

Table 4 Respondents' Feedback

S. No.	Comments
1	I believe that WisDOT can still benefit from some type of integrated database on operation similar to other state such as Michigan DOT, where that data are integrated into PTMS.
2	TIME/EFFORT: No centralized source for the information, pieced together through many different systems in the regions.
3	There are many data sets that are difficult to access and take staff time to prepare. Providing access (whether on site or with data feeds) would be valuable to the research community.
4	Integration of all pavement traffic/design/materials/performance databases would be extremely useful for research applications
5	The foundation of the project was based on getting construction, materials, and pavement data and it took more than two years to get that information
6	Keep up the excellent work
7	Sharing data benefits WisDOT
8	The data should be integrated

Finally, the responses to all the questions in the survey were aggregated according to problematic (negative) responses by question and then by the data set they were responding about. The results of this analysis are represented in figure 11 and Figure 12. A look at the negative responses by each question (Figure 11) reveals that a little more than a third of the respondents (34.8% or 8 out of 23) had indicated that they had problems with the time associated with data preparation, followed by those that indicated problems associated with the quality of the data. Relatively less number of negative responses was attached to questions about data acquisition or data sharing.

When looking at the problematic responses by data sets (Figure 12), it is evident that a majority of the data sets which formed basis for the responses from the survey participants reflected a positive experience with zero or one problematic response per data set. At the same time, Pavement Performance Data and Structural Data were associated with negative responses in 13 out of 23 (56%) respondents followed by As-built Plan, and Pavement Information Files. The responses to the individual questions discussed earlier also support this finding. This analysis reveals that Wisconsin DOT has done a great job of meeting stakeholder expectations as pertaining to data access from the DOT. This is illustrated by the fact that 17 out of the 23 data sets (74%) discussed in the survey had less than 2 (1 or 0) negative responses about the data access process.

CRITICAL DATASETS

In considering the results of the survey, a few points should be taken into consideration. Some of the survey questions are likely to be more critical than others depending on the project or the client involved. Some stakeholders may value freedom in the sharing and usage of data while others might be

more concerned with obtaining the data in a timely manner. Furthermore, some data might be inherently more difficult to obtain because of privacy concerns, resolution, or a number of other factors. Results should be interpreted with this in mind.

That said, one of the objectives of this report is to identify datasets associated with the best experiences as well as those with the worst experiences. Figure 12 shows that many data sets 9 out of 23 (39%) were not reported to have resulted in any negative experiences for the stakeholders. In all, with which WisDOT stakeholders had either a positive or negative experience, and the results of the survey do implicate a few candidates. Structural data and pavement performance data seem to be negatively implicated throughout the survey. Both took a considerable amount of time to obtain³, both presented issues in processing, formatting, and sharing, and both compromised the research process. This respondent clearly had the worst experience with WisDOT data and should be contacted.

Other data sets that are worth looking into are the As-Built Plan, Pavement Information Files, Real Estate Maintenance Procedures, and Permit Vehicle Data. These showed mostly positive feedback, though a few problems were reported for each. The As-built Plan received criticism mostly for its format, which is not integrated and requires excessive manual manipulation. Permit Vehicle Data and Pavement Information Files received complaints on procurement and processing time. Pavement information Files were the only data used by multiple survey respondents, so it would be recommendable to contact both. Finally, the main issue with Real Estate Maintenance Procedures was that the records contained information relating to the location of properties in culturally or archeologically sensitive areas, resulting in certain restrictions on use.

These datasets represent important issues in data access and their users could provide interesting insights and suggestions. The recommendation of this report is for the team working on the WisDOT data access project to contact the survey respondents who used the datasets highlighted in this section.

³ The respondent who used Pavement Performance Data reports waiting over two years to obtain his data.

Problematic Responses by Survey Question

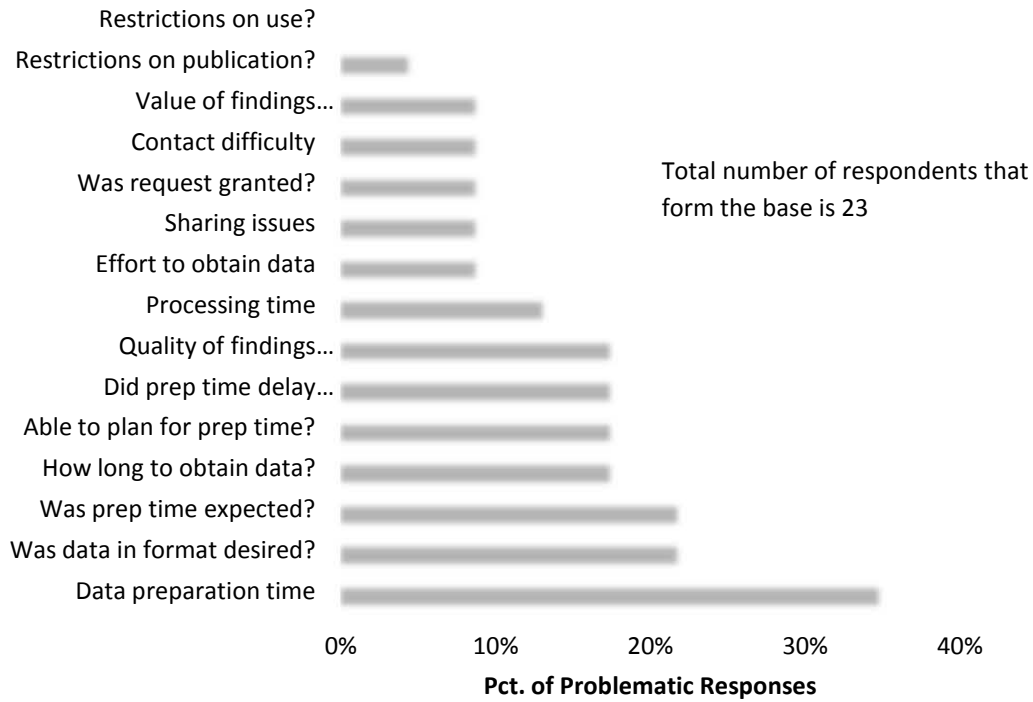


Figure 11 Problematic Responses by Question

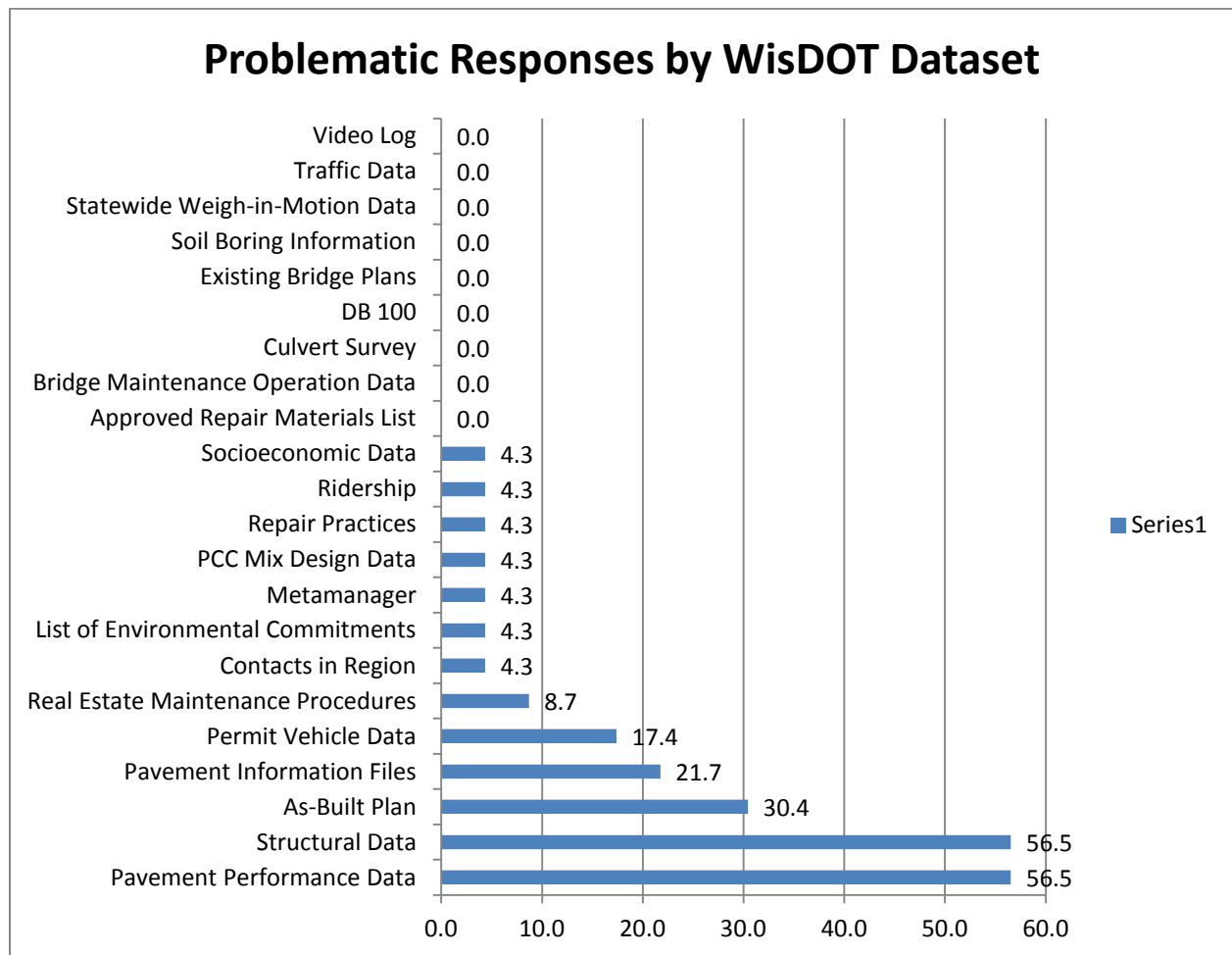


Figure 12 Problematic Responses by Dataset

CHAPTER 3. CASE STUDIES

IDENTIFICATION OF PROMINENT DATA SETS

A scoring procedure to identify data sets characterized as ‘best’, ‘worst’, and ‘most important’ by the survey respondents was developed for this research. Pre-coded survey responses were re-coded per the scoring key in Table 5. Note that all items are scored so that a high score defines a more favorable experience with the particular data set. Each data set is scored on a 0 to 100 range so that 0 represents the ‘most onerous experience’ and 100 the ‘most painless experience’ in each of the dimensions measured.

Table 5. Recoding of Responses

Dimensions Measured Based on Questions Asked	Possible Answers (Recoded Responses)
Time or effort taken to obtain the data	Great/moderate/none (0/50/100)
Processing time in case of data in the incorrect format, data entry issues, for example entering information from paper records	Great/moderate/none (0/50/100)
Issues connected with data sharing and publication	Great/moderate/none (0/50/100)
Was your data requested from WisDOT granted?	Yes/no (100/0)
How long did it take to obtain the data requested?	Two weeks/1 month/1 to 3 months/more than 3 months (100/67/33/0)
How difficult was it to obtain contact information for the appropriate staff to request the data?	Easy/cumbersome/difficult/very difficult/never succeeded (100/75/50/25/0)
Were the data you requested delivered in the form you desired?	Yes/no (100/0)
How long did you spend on preparing the data?	1 week/2 weeks/3 weeks/4 weeks/longer (100/75/50/25/0)

The recoded numeric responses to the eight questions in Table 5 were used to conduct a principal components factor analysis. All eight variables were loaded on a single factor and the resulting factor scores were assigned to each data set. Table 6 shows the data sets with highest and lowest scores in each of the ‘best’, ‘worst’, and ‘most important’ categories. The scores can be best thought of as standardized variables with a mean of 0 and a variance of 1.

Table 6 reveals that certain data sets were characterized as either ‘best’ and ‘most important’, or ‘worst’ and ‘most important’. Given the recoding protocol above, the highest scores represent the best-case scenario and the lowest scores the worst-case scenario.

Table 6. Data Sets with Highest and Lowest Scores

Category	Score	Data Set	Value
Best	High	Contacts in Region	1.18
	Low	Structural Data	-2.47
Worst	High	PCC Mix Design Data	0.91
	Low	As-built plan	-1.39
Most Important	High	Bridge Maintenance Operation Data	0.36
	Low	Permit vehicle data for both annual permits and single-trip permits	-0.24
Best and Most Important	High	Soil Boring Information	0.95
	Low	PIF	-1.25
Worst and Most Important	High	List of Environmental Commitments	0.31
	Low	Pavement Performance Data	-2.47

IDENTIFICATION OF CASE STUDIES

Given the incomprehensive nature of surveys, the stories behind the responses are a fundamental supplement that will shed light on how Wisconsin DOT could have better facilitated these projects. A scoring procedure was used to identify case studies that could provide the most insight. As shown in Table 5, pre-coded survey responses were recoded numerically. Each data set is scored on a 0 to 100 range so that 0 represents the “most onerous experience” and 100 the “most painless experience” in each of the dimensions measured. Respondents classified certain datasets as their “best dataset,” “worst dataset,” and “most important dataset.”

The recoded numeric responses to the eight questions in Table 5 were used to conduct a principal components factor analysis. All eight variables were loaded on a single factor and the resulting factor scores were assigned to each data set. Table 6 shows the data sets with highest and lowest scores in each of the “best,” “worst,” and “most important” categories. The scores can be best thought of as standardized variables with a mean of 0 and a variance of 1. Given the recoding protocol above, the highest scores represent the best-case scenario and the lowest scores the worst-case scenario.

TECHNICAL DETAILS OF RANKING METHODOLOGY

A *common* factor is an unobservable, hypothetical variable that contributes to the variance of at least two of the observed variables. The unqualified term "factor" often refers to a common factor. A *unique* factor is an unobservable, hypothetical variable that contributes to the variance of only one of the observed variables. The model for common factor analysis posits one unique factor for each observed variable. In this regard, the model producing the factor scores can be written as:

$$y_{ij} = x_{i1}b_{1j} + x_{i2}b_{2j} + \cdots + x_{iq}b_{qj} + e_{ij}$$

where, y_{ij} is the value of the i th observation on the j th variable, x_{ik} is the value of the i th observation on the k th factor, b_{kj} is the regression coefficient of the k th factor for predicting the j th variable, and e_{ij} is the value of the i th observation on the j th unique factor. The assumptions are: (a) the unique factors e_{ij} (playing the role of residuals) are uncorrelated with each other; (b) the unique factors are uncorrelated with the common factors x_{ik} ; (c) the common factors x_{ik} are uncorrelated with each other and have unit variance.

These assumptions imply that the common factors are, in general, not linear combinations of the observed variables. In fact, even if the data contain measurements on the entire population of observations, one cannot compute the scores of the observations on the common factors. Although the common factor scores cannot be computed directly, they can be estimated in a variety of ways. To illustrate the scoring of our data set, assume that the raw data in one of the observations – data sets to be scored – is:

50 50 50 100 67 100 100 0

The contribution of this observation to the value of the common factor is:

$$\begin{aligned} & \frac{50 - 52.1}{31.9} \times 0.17 + \frac{50 - 54.3}{33.4} \times 0.15 + \frac{50 - 71.7}{33.1} \times 0.17 + \frac{100 - 91.3}{28.8} \times 0.15 + \frac{67 - 65.2}{38.2} \times 0.16 \\ & + \frac{100 - 81.5}{25.2} \times 0.13 + \frac{100 - 78.2}{42.1} \times 0.16 + \frac{0 - 47.8}{41.9} \times 0.14 = -0.07 \end{aligned}$$

This computation involves multiplying the standardized raw scores by the standardized scoring coefficients for each variable in Table 6 and adding them up. The same is done for each observation in the original data. The end result is the computation of the contributions of each observation to the unique factor (in this case), which is the estimated score for the particular data set (observation). The highest and lowest scores produced in this manner are shown in Table 6. These scores correspond to 7 researchers, who we will name A through G to maintain confidentiality (Table 7).

Table 7. Datasets with highest and lowest scores

Researcher	Category	Score	Data Set	Value
Researcher A	Best	High	Contacts in Region	1.18
Researcher B		Low	Structural Data	-2.47
Researcher C	Worst	High	PCC Mix Design Data	0.91
Researcher D		Low	As-built plan	-1.39
Researcher E	Most Important	High	Bridge Maintenance Operation Data	0.36
Researcher F		Low	Permit vehicle data for both annual permits and single-trip permits	-0.24
Researcher G	Best and Most Important	High	Soil Boring Information	0.95
Researcher D		Low	PIF	-1.25
Researcher A	Worst and Most Important	High	List of Environmental Commitments	0.31
Researcher B		Low	Pavement Performance Data	-2.47

INTERVIEWS

This group of 7 researchers includes a mix of academic, private sector, and public sector employees as well as good, bad, and average experiences with WisDOT data. Though all were contacted for follow-up interviews, only 3 agreed to do so. An extra interviewee who did not respond to the preliminary survey was eventually added to provide additional information.

The research team then compiled interview questions that touched on data sharing issues seen in the literature and issues seen in the survey responses. Some of these include the effect that data restrictions may have had on the project, delays in receiving data, currency and overall reliability of the data, and comparisons between the Wisconsin DOT and other agencies they may have worked with in the past. For the interviewee who did not take the initial survey, interview questions echoed many questions from the survey so that this person's case study could be somewhat comparable to the others.

CASE STUDIES

CASE STUDY I - RESEARCHER A

The individual with the most experience working with the Wisconsin DOT is Researcher A. He has worked at a university and has made data requests to WisDOT frequently. Researcher A felt that the project he responded about was important because of the multi-agency involvement of the data and the need subsequently for a centralized repository of the same. According to Researcher A, with a centralized database, different agencies can understand their objectives and track their progress more effectively over time. The end goal according to him was for these agencies to avoid situations where they invest money in maintenance, repairs, or upgrades and – in doing so – they undermine investments that other agencies are making.

Compiling the necessary data for this project was a challenge according to the researcher because nothing of the sort previously existed at WisDOT. He established several points of contact at the agency and received relevant information electronically via email. 1 to 3 months had passed before he had sufficient information to begin compiling the database, yet this wait was not unexpected and they did budget for this time. He recognized that it was not an easy task for the Wisconsin DOT; it was somewhat obscure information that they were not accustomed to collecting and DOT employees were stretched thin as it was. To their credit, he indicated that WisDOT personnel did keep him updated on the status of the data requested. No data sharing agreement was required by the DOT when the data was shared with him.

When asked about another past experience that was not as pleasant, Researcher A recalled a project that was difficult for several reasons. According to him, there were significant issues with freight data because it is compiled in different years in different states, thus making it difficult to compare across jurisdictions. Second, a data sharing agreement (though not very restrictive) was required in which he must notify the data owner of any publication and strip out personal identifiers such as individual companies. Finally, Researcher A claimed that the DOT indicated that they would provide the necessary information on shipments and where they were broken down into smaller vehicles. However, the data they provided did not have nearly the level of detail that was needed. He attributed this to a communication error during the scoping phase in which the two sides failed to clearly state their responsibilities and assets with respect to the project.

Researcher A offered several points of constructive criticism for the Wisconsin DOT. First of all, he emphasized the need for a centralized data point of contact. His familiarity with DOT personnel had allowed him to circumvent traditional channels of communication and seek out the data librarian, who he said, often helped him find pertinent information. He acknowledged that this familiarity had been a huge advantage for him, and recommended that the same data librarian officially be made the point of contact for all data inquiries so that other researchers can benefit from the same type of experience that he has had. Researcher A also stressed the need for tighter performance management standards in data

access, as these are currently rather weak internally. As he put it, “If you don’t regularly report on data, then it just gets buried. But once you have performance management systems in place that need data to support them, then it becomes somebody’s responsibility and it becomes well cared for.”

CASE STUDY II - RESEARCHER B

Researcher B, a professor at a university had the most negative survey responses about his experience in accessing data from the DOT. He was awarded two contracts from the Wisconsin Highway Research Program. One of them was a paper study with no data needs; the other required several data elements from the DOT. This case study draws on his experiences in the latter case.

The purpose of the project was to investigate the composition of pavements for better performance. The scope called for studying a select few roadway sections and observing several specific points about each segment. The data needs fell into three categories: 1) photographs of construction practices, 2) post-construction roadway testing, and 3) structural assessment of the pavement (i.e. structural data). A point of contact was established within WisDOT to facilitate the communication for the project.

According to Researcher B, the photographs of construction practices were time-sensitive and thus represented a crucial data element. As in most Midwestern states, construction season occurs within a small window in Wisconsin, necessitating the research team to send photographers out during the construction phase to ensure the timely completion of data collection for the project. Researcher B requested a month’s notice before construction season started so that he could alert his photographer to the task. Due to various unforeseen circumstances involving staff turnover, information about the construction schedule slipped through the cracks, and Researcher B never received the critical photographs that he needed to evaluate the road construction. The Wisconsin DOT did however, supply him with historical information such as hard copies of old photographs of roadway sections and sectional drawings of roadways. This was better than nothing, but was no replacement for the up-to-date photographs that he needed. He also obtained pavement photographs from another state and attempted to fill the data gap in that manner. The final report was done using the other state data, though many limiting assumptions had to be made.

By the time the final report was issued, the project had taken approximately significantly longer to complete - well over a year more than what was anticipated. The lack of data had significantly compromised the project to the point where the results were of little value to the Wisconsin DOT. With the cost overruns and time delays, the project was completed almost a year late. A less quantifiable but significantly more important factor is that the DOT never got the answers to the questions that they originally sought.

In discussing what went wrong in this process, Researcher B recalled that there was no physical meeting between the points of contact. However, he acknowledged that WisDOT employees are required to wear many hats and are stretched thin. He also conceded that this project was a lot to ask

of his point of contact because it went above and beyond typically data sharing duties; it required coordination between several teams that was likely to be very time consuming. Still, he concluded that there is a need for clearer communication. He referred specifically to the frequent reports and updates he sent that were met with complacent approval void of any critical feedback. He also faulted the DOT for not being honest in the amount of responsibility they were actually willing to take on. “Going in,” he stated, “if the DOT had told me ‘you will have to step up and dig up the data for yourself; that is part of the project. You will have to reach out to these people, we don’t have the time.’ Then it is OK...I think that the level of cognizance of this problem could have been budgeted in the RFP. I would have budgeted a graduate student to take photographs...that lack of understanding of their own limitations hurt the project.”

CASE STUDY III - RESEARCHER D

Researcher D works for a private engineering firm that frequently collaborates with the DOT. He has worked extensively with WisDOT, having obtained well over 100 datasets and design plans from them. Researcher D responded to the UIC survey based on his experience in accessing two datasets: soil boring information and existing bridge plans. Soil boring data is a necessary precursor for developing bridge plans as it contains logs of detailed soil features, subsurface elevation, and drilling suitability. The bridge plans provide engineering designs as well as information about materials and machinery used.

These data were found on the DOT’s Historical Structure Information (HSI) database, and the process was facilitated by an informal remote access agreement. No restrictions were made on usage or sharing; the only requirement was a simple registration process. Researcher D asserted that the HSI system is a tremendous upgrade over the previous system, where stakeholders had to find a DOT employee who could find the information and relay it to them weeks or months after the request was made. Still, many critical issues need to be addressed.

To find information about a particular bridge plan, the user must search by that bridge’s unique identifier. However, if this is unknown then finding the relevant bridge is a laborious process – Researcher D recalled having gone through hundreds of pages of data manually to find the bridge of concern, and he added that the DOT was not helpful in this process. To improve this experience, he suggested that the agency provide additional information in the HSI on which the data could be sorted and searched such as bridge classification, specific bridge features, and features of the surrounding landscape. Users are able to search by county, but with hundreds of bridges in some counties, this is often too coarse. He also recommended using a Gridpoint Statistical Interpolation (GSI) system that would allow the user to zoom in on a map interface to select a particular bridge.

Researcher D also raised concerns about the reliability of older soil boring logs. Because many of these are from the 1950s and 1960s, their condition has often deteriorated over the years; some documents even have essential information missing because of a rip. In some situations the user can use his or her discretion and make an assessment based on other available data and images. Yet in

other cases where the obscured information is absolutely fundamental, the researcher has had to go to the physical site and collect his own data. This can lead to significant delays and cost overruns. Researcher D mentioned one case when the data was so compromised that they had to completely discard a boring site and start a new one; consequently, the project was delayed by two years.

Based on this person's experiences, he made a few recommendations to the Wisconsin DOT. He proposed that they attach notes to original documents, pictures, and data where information has been obscured or compromised. He also suggested that the DOT can get more meaningful data by including geotechnical reports with more comprehensive boring and lab data for subsurface exploration reports. Consulting companies would do these tests and give them to the DOT without any additional contractual obligations other than standard payment. This information could then be uploaded electronically to the HSI site. Finally, he contended that date and temperature information would be extremely useful if included in boring data because of the effect they have on soil conditions. Researcher D concluded with a discussion of best practices for file sharing websites. He identified Michigan, California, and Ohio DOTs as especially adept in compiling and sharing information on electronic file sharing servers.

CASE STUDY IV - RESEARCHER H

The next researcher did not participate in the online survey – he was contacted in the later stages to add supplemental information. This person is affiliated with a university. Though he has limited experience working with the Wisconsin DOT, he is currently involved in three projects, each requiring different types of data and different data access procedures. One of the projects sought to improve traffic flow in work zones; the next project was about freeway closures and it looked at how traffic was diverted in the vicinity of a major freeway when it was closed for a weekend; the last project was a Corridor Rehabilitation project in Wisconsin investigated traffic diversion during a repaving project. His experiences seemed to have been overwhelmingly positive.

Tube detectors, wavetronics, and Volume & Speed Occupancy (VSPOC) were supplied by the DOT to provide mainline traffic counts. For historical traffic counts dating back to 1996, the research team was granted visitor access to the DOT's internal Traffic Database System (TRATIS) at the Traffic Operations Center (TOC). For the freeway closure project, collaborating consultants provided access to a Traffic Responsive Signal System, an FTP site with speed and volume data on arterial corridors.

Researcher H recalled DOT personnel being extremely accommodating. Tube detector and wavetronic data, which required internal processing, were disseminated within two weeks of request. They showed him how to get data from the file-sharing server so that he could come in and download it instantaneously whenever he needed. The only holdup came about when the research team requested approval for a visitor access agreement to the Traffic Operations Center – this process lasted about month. He was required by the DOT to fill out paperwork, though it was only for security purposes and did not concern data sharing issues. This individual indicated that once he was "in the system," things

were very easy. DOT personnel even offered constant follow-up on the projects, providing input at every step in the process. Datasets were either downloaded on site or were emailed electronically in text or excel files.

He admitted that certain datasets required substantial manipulation. VSPOC traffic counts conflicted with data collected by Automatic Traffic Recorders, however the DOT supplied software that helped “balance” the two. He also reported that other technical flaws exist with the VSPOC data, but added that the DOT was aware of them and were attempting to fix those. Overall, he believed that the Wisconsin DOT’s data sharing protocol is extremely effective.

RECOMMENDATIONS FOR WISDOT

The case studies shed more light on the experiences of individual researchers as it relates to data access from the DOT and allowed the research team to get a better sense of the underlying issues that were perceived of as hurdles to seamless data sharing. In this section, the issues identified from the case studies are tied to the findings from the literature review to provide an outline for broad solutions.

The case studies do not provide any indication that there are significant issues with respect to overly burdensome data sharing agreements. In this respect, WisDOT appears to favor the open dissemination of public data over data security. The case studies do suggest, however, that the agency struggles with two of the issues seen in the literature: bureaucracy and data limitations. This subsection will discuss some potential solutions based on the pertinent literature.

STRATEGIES FOR DEALING WITH BUREAUCRACY

As Zimmerman et al. (2002) point out and interviews with DOT practitioners confirm, bureaucracy is a common impediment to effective data sharing protocols all over the country. To better navigate different departments and offices, several sources recommend developing a comprehensive and holistic data governance strategy (Cambridge Systematics 2009), (NCHRP 2005). The agency could divide data responsibilities into subgroups (i.e. highway, transit, GIS, bridge, etc.) as the Alaska DOT did, each with its own channels of communication and protocols. An overall data manager would then oversee all internal and external data dissemination and ensure that each subgroup is adequately facilitating the flow of good, reliable data.

Whoever this data manager would be, Holman (2007) suggests that they not only be well integrated into WisDOT’s data processing channels, but also those of other local agencies (see discussion of “other strategies” in Section V of literature review). This is necessary because the scope of the DOT’s projects will often overlap with other departments and stakeholders will benefit more if agencies or departments are less stratified. Additionally, better integration among these entities will result in less task duplication and more synergic opportunities.

Aside from streamlined communication, proper data governance should entail regular reporting on the agency's data inventory. This would be achieved through a "data process review" and a "data acquisition review." A process review periodically assesses the status of data products, improvement needs, and estimated costs of improvements. An acquisition review examines opportunities for improved efficiencies and cost reductions. Such strategies would guarantee better data archiving and updating protocols. Furthermore, it would put Wisconsin DOT in a better position to evaluate its assets at the onset of a project and consequently avoid confusion with stakeholders about what each party will contribute.

Each researcher interviewed would have benefited directly from a data governance protocol. For Researcher A, environmental commitments could have been obtained from one person instead of several. Furthermore, when this database is compiled, a data governance system will ensure that it is continually updated and used. Researcher B would have been better able to assess WisDOT's ability to provide the necessary construction photographs if the lines of communication were simplified. Had they not been able to, the DOT should have assigned a co-researcher either from UW-Madison or a private firm. As it happened, the DOT didn't meet its obligations on the project because they failed to objectively evaluate their ability to contribute in the initial stages of the project. Finally, Researcher D would have benefitted from data governance because it would call for routine updating and improvement of the HSI database, as well as other databanks. Old, damaged photographs and boring logs that this individual needed would have been better maintained.

STRATEGIES FOR DEALING WITH DATA LIMITATIONS

Relative to other state DOTs such as Virginia and New York, WisDOT seems to place less emphasis on liability and security and more emphasis on the stakeholder's experience. This is an excellent strategy that is widely advocated in the literature. However, given the current fiscal environment, the agency could take steps to move in the direction of partial privatization of some services. As Zimmerman et al. (2002) point out such a policy would be consistent with the idea that government must wisely manage its resources and efficiently deliver services to the public.

In cases where an agency is simply overstretched and unable to provide information needed by stakeholders, franchising these tasks with *technical specifications* agreements (see discussion of PCDA in Section II of literature review) could be an effective solution. In this arrangement, certain services would be outsourced to private firms on a competitive basis. These firms would then be free to pursue the lowest cost and highest value-added arrangement, though the agency retains authority in specifying the types of analyses and procedures that would be performed. To gain a better understanding of these arrangements, WisDOT should talk to officials at the Virginia DOT where these practices have become common. However, lessons should be interpreted to context as VDOT's focus on data security, according to one interviewee, at times has hampered stakeholders' data sharing experience.

Finally, Dueker et al. (2000) recommend engaging stakeholders in a more interactive FTP environment where data can be updated and improved upon by outside parties. This strategy, otherwise known as *crowdsourcing* has become a powerful and cost effective technology in all sectors, harnessing the value of decentralized knowledge sharing. Researcher D recommended this as an effective solution to maintaining an updated HSI database.

CONCLUSIONS AND FUTURE WORK

This research aimed at understanding data access from public agencies using three different approaches. The first was a scan of other state data sharing practices. The second involved an online, web-based survey of researchers' experience in obtaining data from Wisconsin DOT, and the third was a series of semi-structured interviews with researchers. Each of these approaches helped project a different dimension to data sharing, the sum of which gives rise to a synergistic feedback to the Wisconsin DOT.

The scan of other state practices along with the literature review revealed that there is a wide range of data sharing practices across the various states. At the same time, there were some common threads to data sharing which are summarized here. There was consensus on the fact that data sharing should be standard and free (to overcome institutional bureaucracy) while adhering to the recommendations in the report by the Panel on Confidentiality and Data Access (to address data limitations). The presence or formulation of a data governance plan seems to have an impact in regulating the data sharing from the agencies' perspective. The literature advocates for the infusion of technology to enhance the data sharing process, albeit not at the expense of data flow and productivity.

The online survey of researchers who had dealt with WisDOT provided an in-depth look at the ground realities of data access. The survey reflected that identifying the appropriate contact person at the agency was a major challenge in the data acquisition phase, while the processing time for data preparation was the single biggest factor once the data was acquired. This also seemed to have an impact on the research findings with poor data quality leading to a compromised research effort. The respondents also seemed to agree that the WisDOT needed to have an integrated database or approach towards data dissemination or data sharing. There was also agreement on the fact that the DOT was doing a satisfactory job of data sharing but can always improve on its current performance by addressing the technological issues (data integration) and quality issues (data format, etc).

The case studies which were a subset of the survey respondents provided a broader canvas for the respondents to expand on their responses to the survey and at the same time talk about their experience in the context of their project. The case study participants were from the public sector (academic) and private sector (one consultant). The feedback was that the WisDOT does an adequate job of data sharing but can improve in certain areas associated with data sharing. Almost all of the participants in the case study reinforced the findings from the survey about the need for an integrated

database. At least in a couple of instances, there was a comparison of WisDOT to its peers in the mid-west who have “better” data integration and data sharing. This is an area of study to pursue and understand the best practices from other state agencies and incorporate them at WisDOT.

The research has certain limitations. The sample for the survey was not randomly selected and the survey was impacted by low response rates. The survey also was geared toward obtaining feedback about specific experiences of data access from WisDOT and therefore restrictive to some extent. In the future, it would make sense to design a survey to understand and compare data access experience from different agencies, and not just restrict it to WisDOT. This will allow for a better bench-marking exercise for the DOT against other similar public agencies. The other area to include in future research is to conduct a separate research on identifying best practices of data sharing in the nation.

REFERENCES

- Zimmerman, Carol A. et al. (2002). “Sharing Data for Traveler Information: Practices and Policies of Public Agencies.” Prepared for U.S. DOT ITS Joint Program Office. January 2002.
- Cambridge Systematics (2009). “Data Governance, Standards, and Knowledge Management: Alaska Data business Plan.” Prepared for Alaska Department of Transportation and Public Facilities.
- DeBlasio, Allan (1999). *Successful Approaches to Deploying a Metropolitan Intelligent Transportation System*. Chapter 9: Develop Written Policies. Prepared by U.S. DOT Research and Special Programs Administration. March 1999.
- Liu, Henry et al. (2002). “A Literature and Best Practices Scan: ITS Data Management and Archiving.” University of Wisconsin at Madison. Submitted to the Wisconsin Department of Transportation, May 2002.
- Holman, Bill (2007). Recommendations for Geospatial Data sharing. North Carolina Geographic Information Coordinating Council. A Report by the Local/State/Regional/Federal Data Sharing ad hoc Committee.
- Clinger, S. and D.R. Abedon (2001). “Summary of Findings for CCTV Privacy Policy Survey and Examples of Existing Privacy Policy.” U.S. Department of Transportation, Federal Highway Administration. Eastern Resource Center, Baltimore, MD.
- Weiss, P.N. (1997). “International Information Policy in Conflict: Open and Unrestricted Access versus Government Commercialization.” *Borders in Cyberspace*, B. Kahin and C. Nesson. Cambridge: MIT Press.
- Watson, Hal (2008). Distributed Data sharing Among State Agencies and Partners. Presentation at the 2008 GIS/LIS Conference, Rochester, MN.
- NCHRP Report 525, Volume 5. (2005): *Guidance for Transportation Agencies on Managing Sensitive Information*. Prepared by TransTech Management, Inc. Transportation Research Board of the National Academies.
- “Checklist on Disclosure Potential of Proposed Datasets.” Panel on Confidentiality and Data Access. 1999.
- “Restricted Access Procedures.” Confidentiality and Data Access Committee and Federal Committee on Statistical Methodology. November 16, 2001.
- “Confidentiality and Data Access Issues among Federal Agencies.” Confidentiality and Data Access Committee and Federal Committee on Statistical Methodology. November 16, 2001.

- Jabine, T. B. (1993). *Procedures for Restricted Access*. Journal of Official Statistics, Vol. 9, No. 2, ppg 537-589.
- Duncan, George T., Thomas B. Jabine, and Virginia A. de Wolf (1993). *Private Lives and Public Policies*. Panel on Confidentiality and Data Access. National Research Council.
- Hallenbeck, Mark E. (1998). "Choosing the Route to Traveler Information Systems Deployment: Decision Factors for Creating Public/Private Business Plans." U.S. Department of Transportation and Federal highway Administration. Washington D.C.
- Booz Allen and Hamilton (2000a). "Enhancing ITS Public/Private Partnerships in Wisconsin, Case Studies: San Antonio TransGuide MDI." Report prepared for the Wisconsin Department of Transportation.
- Booz Allen and Hamilton (2000b). "Enhancing ITS Public/Private Partnerships in Wisconsin, Case Studies: Seattle SmartTrek MDI." Report prepared for the Wisconsin Department of Transportation.
- Dueker, K.J. and Butler, J.A. (2000). "A Geographic Information Systems Framework for Transportation Data Sharing." Elsevier Science, Incorporated.

APPENDIX A – SNAP SHOT OF ONLINE SURVEY QUESTIONNAIRE

MUTRC Data Dissemination - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.utc.uic.edu/wisdot/sv_aa.php?hash=80570

MUTRC Data Dissemination

MRUTC Data Dissemination

Dear Sriraj :

The Wisconsin Department of Transportation (WisDOT) would like to improve its data dissemination practices.

We are working on a research project funded by the Midwest Regional University Transportation Center and the Wisconsin Department of Transportation. The objective is to recommend innovative ways of data dissemination by the DOT to its stakeholders in need of data. It is in this context that we are reaching out to you. Based on our records, you have worked on the following projects either funded or associated with WisDOT.

You have participated in these WisDOT projects:

- Sample Project**

If this is a complete list of your substantially completed projects with WisDOT, and you **made some use of data provided by WisDOT**, we would like your feedback. Please click the green button below.

If you **have not obtained data** from WisDOT for any project, then there is no need to complete this survey. Please click the "No Data" button below.

If you **have used WisDOT data but this list is inaccurate or incomplete**, please click the red button and we will be in contact shortly to correct it.

Thanks.

Intro	Consent	List Datasets	Select	First Dataset	Second Dataset	Third Dataset	Final Question	Thanks
-------	---------	---------------	--------	---------------	----------------	---------------	----------------	--------

Welcome screen

MRUTC Data Dissemination

University of Illinois at Chicago
CONSENT FORM

Project Title: Scoping Data Access and Integration Needs to Facilitate Better Management of Research Innovation

You are being asked to participate in the research conducted by Research Assistant Professor P. S. Sriraj, or by staff under the supervision of Dr. Sriraj.

Purpose of the Study: The primary objective of this study is to gain better understanding of the issues associated with data access and dissemination of Wisconsin DOT by the stakeholders. This will allow the research team from UIC in developing plans and policies to assist Wisconsin DOT and its data dissemination practices. As part of the study, the researchers from UIC are conducting a survey of the stakeholders from both the public and private sectors to collect information about topics such as the best, worst, and most important data request/access experience of stakeholders.

If you agree to participate in the research, you will be asked to participate in a survey aimed at understanding your experience in obtaining and using data from Wisconsin DOT. The survey should take approximately 30 minutes.

You understand that your participation in this study is entirely voluntary and that you can withdraw from the study at any time without penalty. The research team will exclude your name from any reports and likewise will do so with regards to maintaining your privacy. You understand that your responses will be used in aggregation and individual responses will not be published with your name associated with it.

You understand that your participation in this research will not pose any physical risks to you personally and that you can skip any questions you are not comfortable answering.

You understand that you will not directly benefit from participating in the research, but that the research may be of benefit to the future of data access from Wisconsin DOT.

If you have any questions about this study, feel free to ask them now or anytime throughout the study by contacting: Dr. P. S. Sriraj, Senior Associate & Research Assistant Professor Urban Transportation Center University of Illinois at Chicago Phone: (312) 413-7568 e-mail: sriraj@uic.edu If you have any questions about your rights as a research subject, you may write or call OPRS at the following address:

Office for the Protection of Research Subjects (OPRS)
1737, W. Polk Street, M/C 672
203 Administrative Office Building
Chicago, Illinois - 60612.
Phone: (312) 996 1711 or toll free: 866-789-6215
Email: uicirb@uic.edu

Agreement to Participate in Research:

MUTRC Data Dissemination - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.utc.uic.edu/wisdot/sv_ba.php?hash=80570&step=Proceed+to+Survey

MUTRC Data Dissemination

dissemination of Wisconsin DOT by the stakeholders. This will allow the research team from UIC in developing plans and policies to assist Wisconsin DOT and its data dissemination practices. As part of the study, the researchers from UIC are conducting a survey of the stakeholders from both the public and private sectors to collect information about topics such as the best, worst, and most important data request/access experience of stakeholders.

If you agree to participate in the research, you will be asked to participate in a survey aimed at understanding your experience in obtaining and using data from Wisconsin DOT. The survey should take approximately 30 minutes.

You understand that your participation in this study is entirely voluntary and that you can withdraw from the study at any time without penalty. The research team will exclude your name from any reports and likewise will do so with regards to maintaining your privacy. You understand that your responses will be used in aggregation and individual responses will not be published with your name associated with it.

You understand that your participation in this research will not pose any physical risks to you personally and that you can skip any questions you are not comfortable answering.

You understand that you will not directly benefit from participating in the research, but that the research may be of benefit to the future of data access from Wisconsin DOT.

If you have any questions about this study, feel free to ask them now or anytime throughout the study by contacting: Dr. P. S. Sriraj, Senior Associate & Research Assistant Professor Urban Transportation Center University of Illinois at Chicago Phone: (312) 413-7568 e-mail: sriraj@uic.edu If you have any questions about your rights as a research subject, you may write or call OPRS at the following address:

Office for the Protection of Research Subjects (OPRS)
1737, W. Polk Street, M/C 672
203 Administrative Office Building
Chicago, Illinois - 60612.
Phone: (312) 996 1711 or toll free: 866-789-6215
Email: uicirb@uic.edu

Agreement to Participate in Research:
I understand that in signing this consent form, I am agreeing to participate in the research and give Professor Sriraj, and his associates, permission to present this work in written and oral form, without further permission from me.

Hit the "I Consent" button if you agree.

MUTRC Data Dissemination - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.utc.uic.edu/wisdot/sv_da.php?&hash=80570

MUTRC Data Dissemination

MRUTC Data Dissemination

For the study listed earlier, please identify the data you requested from WisDOT. If there were more than five, please pick the five most important or the five most likely to give us insight into WisDOT's practices.

1. Sample Project

a.

b.

c.

d.

e.

[Next Page](#)

Intro	Consent	List Datasets	Select	First Dataset	Second Dataset	Third Dataset	Final Question	Thanks
-----------------------	-------------------------	-------------------------------	------------------------	-------------------------------	--------------------------------	-------------------------------	--------------------------------	------------------------

MUTRC Data Dissemination - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.utc.uic.edu/wisdot/sv_ea.php?hash=80570

MUTRC Data Dissemination

MRUTC Data Dissemination

Pick the dataset with which you had the **best** experience, the one with which you had the **worst** experience, and the one which was **most important** to your research.

On the following pages, one for each identified dataset, you will be asked to share your experiences as they relate to each of the identified datasets.

Name of Dataset	Best	Worst	Most Important
Small Data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Medium Data	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Large Data	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

Next Page

Intro	Consent	List Datasets	Select	First Dataset	Second Dataset	Third Dataset	Final Question	Thanks
-------	---------	---------------	--------	---------------	----------------	---------------	----------------	--------

MUTRC Data Dissemination - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.utc.uic.edu/wisdot/sv_fa.php?hash=80570&attention=1

MUTRC Data Dissemination

MRUTC Data Dissemination

With respect to the dataset you called
Small Data

In each of the general areas listed below, please describe any difficulties you may have had with data, and rate the level of difficulty, if any, your research encountered when dealing with data and with the sponsor.

Time or effort taken to obtain the data.	Level of Difficulty
	<input type="radio"/> Great <input type="radio"/> Moderate <input type="radio"/> None
Processing time in case of data in the incorrect format, data entry issues, for example entering information from paper records.	Level of Difficulty
	<input type="radio"/> Great <input type="radio"/> Moderate <input type="radio"/> None
Issues connected with data sharing and publication.	Level of Difficulty
	<input type="radio"/> Great <input type="radio"/> Moderate <input type="radio"/> None

Was your data request from WisDOT granted?
☐ Yes ☐ No

How long did it take to obtain the data requested?
☐ Two weeks ☐ A month ☐ 1 to 3 months ☐ more than 3 months

How difficult was it to obtain contact information for the appropriate staff to request the data?
☐ Easy ☐ Cumbersome ☐ Difficult ☐ Very Difficult ☐ Never succeeded

Were the data you requested delivered in the form you desired
☐ Yes ☐ No

How long did you spend on preparing the data
☐ 1 week ☐ 2 weeks ☐ 3 weeks ☐ 4 weeks ☐ Longer

Was this time period expected and routine?
☐ Yes ☐ No

Were you able to plan for the preparation time?
☐ Yes ☐ No

Was your research delayed by preparation?
☐ Yes ☐ No

Detailed Information about First Dataset, part 1

MUTRC Data Dissemination - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.utc.uic.edu/wisdot/sv_fa.php?hash=80570&attention=1

MUTRC Data Dissemination

☐ 1 week
 ☒ 2 weeks
 ☐ 3 weeks
 ☐ 4 weeks
 ☐ Longer

Was this time period expected and routine?

☐ Yes ☐ No

Were you able to plan for the preparation time?

☐ Yes ☐ No

Was your research delayed by preparation?

☐ Yes ☐ No

Was the quality of the data you used compromised?

☐ Yes ☐ No

Was the value of your findings compromised?

☐ Yes ☐ No

Have you published the findings of your research?

☐ Yes ☐ No

Were there restrictions placed on the use of the data by WisDOT for publication purposes?

☐ Yes ☐ No

Were there restrictions placed which interfered with or were too limiting for your research?

☐ Yes ☐ No

How did you reconcile the data restrictions?

If your research output was compromised by data restrictions, please describe how

Finally, again, with respect to the dataset you called

Small Data

You characterized it as one of the best experiences you had with WisDOT data dissemination. Could you tell us, briefly, why this was the case?

Drawing 1: Detailed Information about First Dataset, part 2

MUTRC Data Dissemination - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.utc.uic.edu/wisdot/sv_fa.php?hash=80570&attention=1

MUTRC Data Dissemination

☐ Yes ☐ No

Were there restrictions placed which interfered with or were too limiting for your research?

☐ Yes ☐ No

How did you reconcile the data restrictions?

If your research output was compromised by data restrictions, please describe how

Finally, again, with respect to the dataset you called

Small Data

You characterized it as one of the best experiences you had with WisDOT data dissemination. Could you tell us, briefly, why this was the case?

[Next Page](#)

Intro	Consent	List Datasets	Select	First Dataset	Second Dataset	Third Dataset	Final Question	Thanks
-----------------------	-------------------------	-------------------------------	------------------------	-------------------------------	--------------------------------	-------------------------------	--------------------------------	------------------------

MUTRC Data Dissemination - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.utc.uic.edu/wisdot/sv_fa.php?hash=80570&attention=2

MUTRC Data Dissemination

MRUTC Data Dissemination

With respect to the dataset you called
Medium Data

In each of the general areas listed below, please describe any difficulties you may have had with data, and rate the level of difficulty, if any, your research encountered when dealing with data and with the sponsor.

Time or effort taken to obtain the data.	Level of Difficulty
<input type="text"/>	<input type="radio"/> Great <input type="radio"/> Moderate <input type="radio"/> None
Processing time in case of data in the incorrect format, data entry issues, for example entering information from paper records.	Level of Difficulty
<input type="text"/>	<input type="radio"/> Great <input type="radio"/> Moderate <input type="radio"/> None
Issues connected with data sharing and publication.	Level of Difficulty
<input type="text"/>	<input type="radio"/> Great <input type="radio"/> Moderate <input type="radio"/> None

Was your data request from WisDOT granted?
☐ Yes ☐ No

How long did it take to obtain the data requested?
☐ Two weeks ☐ A month ☐ 1 to 3 months ☐ more than 3 months

How difficult was it to obtain contact information for the appropriate staff to request the data?
☐ Easy ☐ Cumbersome ☐ Difficult ☐ Very Difficult ☐ Never succeeded

Were the data you requested delivered in the form you desired
☐ Yes ☐ No

How long did you spend on preparing the data
☐ 1 week ☐ 2 weeks ☐ 3 weeks ☐ 4 weeks ☐ Longer

Was this time period expected and routine?
☐ Yes ☐ No

Were you able to plan for the preparation time?
☐ Yes ☐ No

Was your research delayed by preparation?
☐ Yes ☐ No

Detailed Information. Second Dataset. partial (remainder same as first dataset)

MUTRC Data Dissemination - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.utc.uic.edu/wisdot/sv_fa.php?hash=80570&attention=3

MUTRC Data Dissemination

MRUTC Data Dissemination

With respect to the dataset you called
Large Data

In each of the general areas listed below, please describe any difficulties you may have had with data, and rate the level of difficulty, if any, your research encountered when dealing with data and with the sponsor.

Time or effort taken to obtain the data.	Level of Difficulty
<input type="text"/>	<input type="radio"/> Great <input type="radio"/> Moderate <input type="radio"/> None
Processing time in case of data in the incorrect format, data entry issues, for example entering information from paper records.	Level of Difficulty
<input type="text"/>	<input type="radio"/> Great <input type="radio"/> Moderate <input type="radio"/> None
Issues connected with data sharing and publication.	Level of Difficulty
<input type="text"/>	<input type="radio"/> Great <input type="radio"/> Moderate <input type="radio"/> None

Was your data request from WisDOT granted?
☐ Yes ☐ No

How long did it take to obtain the data requested?
☐ Two weeks ☐ A month ☐ 1 to 3 months ☐ more than 3 months

How difficult was it to obtain contact information for the appropriate staff to request the data?
☐ Easy ☐ Cumbersome ☐ Difficult ☐ Very Difficult ☐ Never succeeded

Were the data you requested delivered in the form you desired
☐ Yes ☐ No

How long did you spend on preparing the data
☐ 1 week ☐ 2 weeks ☐ 3 weeks ☐ 4 weeks ☐ Longer

Was this time period expected and routine?
☐ Yes ☐ No

Were you able to plan for the preparation time?
☐ Yes ☐ No

Was your research delayed by preparation?
☐ Yes ☐ No

Detailed Information, Third Dataset, partial (remainder same as first dataset)

MUTRC Data Dissemination - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.utc.uic.edu/wisdot/sv_ja.php?hash=80570

MUTRC Data Dissemination

Info page 7 of 7

MRUTC Data Dissemination

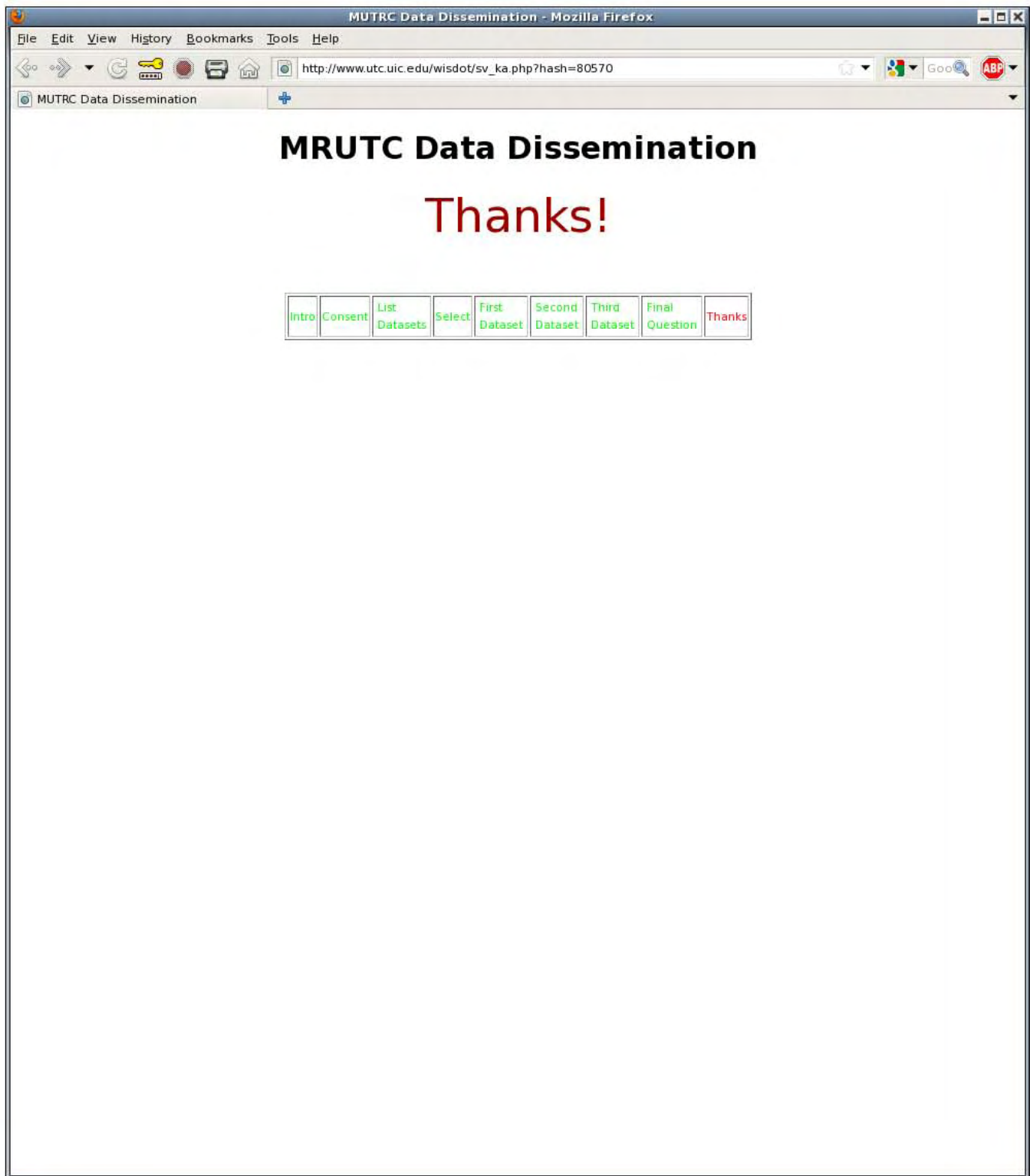
A final question:

What feedback would you give to WisDOT regarding data sharing and publishing?

Next Page

Intro	Consent	List Datasets	Select	First Dataset	Second Dataset	Third Dataset	Final Question	Thanks
-------	---------	---------------	--------	---------------	----------------	---------------	----------------	--------

Drawing 2: Final Feedback Solicitation



Thanks

APPENDIX B – CASE STUDY QUESTIONNAIRE

Interview questions for Case Studies

1. How did you make your data request (phone, email, formal letter, etc)?
2. Is this your first time obtaining data from WisDOT? If not, please describe how this experience compares with other instances.
3. When you request data from WisDOT do you typically know exactly what you are looking for, or do you expect someone at WisDOT to help you figure out what you are looking for?
4. Was the staff able to guide you in the data request to focus exactly on the data items required? Did you have to search around a bit to find the right staff to help you?
5. What avenues did you go through to obtain your data (WisDOT personnel, internet, etc.)? If internet, was your data easily accessed or was the website hard to navigate and find what you needed?
6. When the data became available were you able to find and access it (e.g., storage medium, web access)?
7. What type of agreement, if any at all, did you accept in order to obtain data from WisDOT? Were there any restrictions on using the supplied data?
8. Has having to sign any type of agreement ever discouraged you from seeking a dataset from WisDOT?
9. Please describe the content of the dataset you requested (i.e. most important attributes, scope, etc.)
10. Why do you think it took you so long to receive the PIF and As-Built plan? Did the dataset include information that WisDOT may have needed to mask or modify for liability or confidentiality purposes? Did someone at WisDOT explain to you why it took a long time to get this data?
11. Was the dataset you requested for a project specific to Wisconsin or for a broader scope? If broader, did you have any issue linking this dataset to the same data from a different state?
12. Did the data you received from WisDOT conflict with similar data from another source?
13. How current was the data you obtained from WisDOT?

14. Please describe some of the issues you experienced with respect to the format of the data and issues in sharing the data.
15. Do you have any suggestions or recommendations for helping WisDOT improve its data access policy?
16. *Please compare your experience in accessing data from another public agency, compared to best experience in accessing data from WisDOT. How does WisDOT stack up? How many data requests have you made with WisDOT?*