

## Improving Management of Transportation Information

### DETAILS

---

138 pages | 8.5 x 11 | PAPERBACK

ISBN 978-0-309-28339-7 | DOI 10.17226/22504

BUY THIS BOOK

### AUTHORS

---

Cambridge Systematics, Inc.

FIND RELATED TITLES

### Visit the National Academies Press at [NAP.edu](http://NAP.edu) and login or register to get:

---

- Access to free PDF downloads of thousands of scientific reports
- 10% off the price of print titles
- Email or social media notifications of new titles related to your interests
- Special offers and discounts



Distribution, posting, or copying of this PDF is strictly prohibited without written permission of the National Academies Press. (Request Permission) Unless otherwise indicated, all materials in this PDF are copyrighted by the National Academy of Sciences.

**NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM**

---

---

**NCHRP REPORT 754**

---

---

**Improving Management  
of Transportation Information**

CAMBRIDGE SYSTEMATICS, INC.  
Tallahassee, FL

*Subscriber Categories*

Administration and Management • Data and Information Technology • Policy

---

Research sponsored by the American Association of State Highway and Transportation Officials  
in cooperation with the Federal Highway Administration

---

**TRANSPORTATION RESEARCH BOARD**

WASHINGTON, D.C.  
2013  
[www.TRB.org](http://www.TRB.org)

## **NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM**

Systematic, well-designed research provides the most effective approach to the solution of many problems facing highway administrators and engineers. Often, highway problems are of local interest and can best be studied by highway departments individually or in cooperation with their state universities and others. However, the accelerating growth of highway transportation develops increasingly complex problems of wide interest to highway authorities. These problems are best studied through a coordinated program of cooperative research.

In recognition of these needs, the highway administrators of the American Association of State Highway and Transportation Officials initiated in 1962 an objective national highway research program employing modern scientific techniques. This program is supported on a continuing basis by funds from participating member states of the Association and it receives the full cooperation and support of the Federal Highway Administration, United States Department of Transportation.

The Transportation Research Board of the National Academies was requested by the Association to administer the research program because of the Board's recognized objectivity and understanding of modern research practices. The Board is uniquely suited for this purpose as it maintains an extensive committee structure from which authorities on any highway transportation subject may be drawn; it possesses avenues of communications and cooperation with federal, state and local governmental agencies, universities, and industry; its relationship to the National Research Council is an insurance of objectivity; it maintains a full-time research correlation staff of specialists in highway transportation matters to bring the findings of research directly to those who are in a position to use them.

The program is developed on the basis of research needs identified by chief administrators of the highway and transportation departments and by committees of AASHTO. Each year, specific areas of research needs to be included in the program are proposed to the National Research Council and the Board by the American Association of State Highway and Transportation Officials. Research projects to fulfill these needs are defined by the Board, and qualified research agencies are selected from those that have submitted proposals. Administration and surveillance of research contracts are the responsibilities of the National Research Council and the Transportation Research Board.

The needs for highway research are many, and the National Cooperative Highway Research Program can make significant contributions to the solution of highway transportation problems of mutual concern to many responsible groups. The program, however, is intended to complement rather than to substitute for or duplicate other highway research programs.

## **NCHRP REPORT 754**

Project 20-90  
ISSN 0077-5614  
ISBN 978-0-309-28339-7  
Library of Congress Control Number 2013943545

© 2013 National Academy of Sciences. All rights reserved.

### **COPYRIGHT INFORMATION**

Authors herein are responsible for the authenticity of their materials and for obtaining written permissions from publishers or persons who own the copyright to any previously published or copyrighted material used herein.

Cooperative Research Programs (CRP) grants permission to reproduce material in this publication for classroom and not-for-profit purposes. Permission is given with the understanding that none of the material will be used to imply TRB, AASHTO, FAA, FHWA, FMCSA, FTA, or Transit Development Corporation endorsement of a particular product, method, or practice. It is expected that those reproducing the material in this document for educational and not-for-profit uses will give appropriate acknowledgment of the source of any reprinted or reproduced material. For other uses of the material, request permission from CRP.

### **NOTICE**

The project that is the subject of this report was a part of the National Cooperative Highway Research Program, conducted by the Transportation Research Board with the approval of the Governing Board of the National Research Council.

The members of the technical panel selected to monitor this project and to review this report were chosen for their special competencies and with regard for appropriate balance. The report was reviewed by the technical panel and accepted for publication according to procedures established and overseen by the Transportation Research Board and approved by the Governing Board of the National Research Council.

The opinions and conclusions expressed or implied in this report are those of the researchers who performed the research and are not necessarily those of the Transportation Research Board, the National Research Council, or the program sponsors.

The Transportation Research Board of the National Academies, the National Research Council, and the sponsors of the National Cooperative Highway Research Program do not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the object of the report.

*Published reports of the*

### **NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM**

*are available from:*

Transportation Research Board  
Business Office  
500 Fifth Street, NW  
Washington, DC 20001

*and can be ordered through the Internet at:*

<http://www.national-academies.org/trb/bookstore>

Printed in the United States of America

# THE NATIONAL ACADEMIES

*Advisers to the Nation on Science, Engineering, and Medicine*

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. On the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, on its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

The **Transportation Research Board** is one of six major divisions of the National Research Council. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal. The Board's varied activities annually engage about 7,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation. **[www.TRB.org](http://www.TRB.org)**

**[www.national-academies.org](http://www.national-academies.org)**

# COOPERATIVE RESEARCH PROGRAMS

## **CRP STAFF FOR NCHRP REPORT 754**

**Christopher W. Jenks**, *Director, Cooperative Research Programs*  
**Crawford F. Jencks**, *Deputy Director, Cooperative Research Programs*  
**Andrew C. Lemer**, *Senior Program Officer*  
**Sheila A. Moore**, *Program Associate*  
**Eileen P. Delaney**, *Director of Publications*  
**Hilary Freer**, *Senior Editor*

## **NCHRP PROJECT 20-90 PANEL**

### **Field of Special Projects**

**John H. Thomas**, *Utah DOT, Salt Lake City, UT (Chair)*  
**Ellen M. “Leni” Oman**, *Washington State DOT, Olympia, WA*  
**Imad S. Aleithawe**, *Mississippi DOT, Jackson, MS*  
**Socorro “Coco” Briseno**, *California DOT, Sacramento, CA*  
**Mary “Rita” Evans**, *University of California at Berkeley - Institute of Transportation Studies, Berkeley, CA*  
**William G. Johnson**, *Colorado DOT, Denver, CO*  
**Jane West Minotti**, *New York State DOT, Albany, NY*  
**Karen Nicole Perrin**, *Illinois DOT, Springfield, IL*  
**Dawn Vanlandingham**, *FHWA Liaison*  
**Robert Cullen**, *AASHTO Liaison*  
**Amanda J. Wilson**, *Research and Innovative Technology Administration Liaison*  
**Lisa Loyo**, *TRB Liaison*  
**Jennifer L. Weeks**, *TRB Liaison*

  
FOREWORD

By Andrew C. Lemer

Staff Officer

Transportation Research Board

*NCHRP Report 754: Improving Management of Transportation Information* presents (1) a selective review of current practices of state departments of transportation (DOTs) and other agencies that collect, store, and use transportation data and information and (2) guidance on strategies and actions a DOT can implement to improve information capture, preservation, search, retrieval, and governance. The guidance is intended to be sensitive to the diversity of state DOTs; the range of transportation information that DOTs use (related, for example, to project delivery, environmental review, network configuration and design detail, and operations performance); and the variety of formats for transportation information (such as text reports, photographs, plans and drawings, geo-coded databases, and financial analyses).

---

DOT officials and other transportation professionals use information to make decisions influencing the performance of the nation's transportation infrastructure. DOTs and other agencies expend substantial amounts of time and money to ensure that their information is current, relevant, accurate, reliable, and available to users when it is needed, in forms that facilitate effective decision making.

The term "transportation information" has many embodiments: reports, manuals, maps, plans, and photographs printed on paper; electronic files and databases stored on computer networks or in machine-readable media; DOT business practices; and records and observations of transportation system configuration, operations, facilities design, materials, and construction. Such information is captured, stored, and made available by functional units at DOTs, libraries, organized bibliographic databases (such as the Transport Research International Documentation or TRID service), catalogs, and increasingly the open Internet. Transportation information may be available in forms suited to immediate or on-demand use by the public—for example, current traffic conditions or motor-vehicle registrations—or stored inaccessibly in agency archives.

Effective management of transportation information entails capturing the information from its various sources; organizing it in ways that facilitate its use by analysts, decisionmakers, and researchers; preserving it; setting policies regarding data security and who may have access; and making information findable and accessible by users. The explosive growth in capabilities to capture, store, and work with information has given rise to diverse terminologies and technical approaches to information management and challenged DOTs' capacity for change. There are no commonly accepted strategies or guidelines for a DOT's information governance and management or the capabilities required for agencies and staff to effectively manage the agency's transportation information across all information sources. Common problems within DOTs include uneven information access among operating

units, incomplete and inconsistent archiving of system information collected by multiple departments and private contractors during project development, and inadvertently lost ability to find information (for example, when text files such as construction specifications are stored only as images, rather than searchable files).

The objective of this research was to prepare guidance describing (1) practices DOTs can use for capture, preservation, search, retrieval, and governance of transportation data and information; and (2) strategies and actions a DOT can follow to implement current best practices for information management. A research team led by Cambridge Systematics, Inc., Tallahassee, FL, conducted the research.

The research team reviewed current DOT policies and practices and documented the practices of several agencies that have made particular efforts to improve their management of transportation information. A set of case studies and examples was developed to present principal findings from this review. The team also investigated current thinking in information management, enterprise content management, and related fields to identify issues of particular importance for transportation information management and technology trends likely to influence future DOT information management activities. The team gave particular attention to the current state of practice regarding standardization of terminology, categorization schemes for transportation information and enhancements in detail or scope of information that should be included in such schemes, and opportunities for developing a common terminology and categorization scheme that could be made available for use by all state DOTs. Drawing on this research, the team assembled a set of suggested policies, tools, and practices DOTs can use to improve their effectiveness in managing transportation information.

This document is written as a guide that DOT staff can use to supplement their understanding of information management principles and to consider practices they might adapt to improve their own agency's management of transportation information.



# CONTENTS

<b>1</b>	<b>Introduction</b>
2	Objectives
2	Use
3	Organization
<b>4</b>	<b>Part 1 Terminology and Categorization Standardization</b>
4	Types of Transportation Information That Need to be Managed
4	Maximize Information Use and Value
5	The Relationship Between Data and Content Management
6	Content
6	Structure
6	Processing
7	Linking Source Data to Published Analysis in Documents
7	Lifecycle, Workflow, Archiving
7	Metadata Standards
7	ISO 15836 (Dublin Core)
8	ISO 19115 and the Federal Geographic Data Committee (FGDC)
9	Open Data and Digital Government Initiatives
9	Digital Government Strategy
10	Open Data
10	Transparency, Application Programming Interfaces (APIs), Third-Party Applications
11	The Semantic Web
11	RDF Schema (RDFS) Vocabulary Description Language
11	Namespaces for XML Schema Elements and Attributes
12	Namespaces for Category Values for Named Entities
13	Review of Terminology and Categorization Schemes
13	Ways to Categorize Content Collections
14	File Directory Methods
14	Metadata Description Methods
15	Relevant Transportation and Related Terminology Resources
15	Transportation-Related Glossaries
15	Transportation-Related Thesauri
17	Library of Congress Subject Headings
17	FGDC Topic Categories
18	Developing a Common Categorization Scheme for Transportation
18	Requirements
21	Reengineering the TRT
26	Community-Based Vocabulary Management Model



28	Practices from Other Fields
28	Enterprise Content Management (ECM)
29	Records Management
31	Document Management
32	Information Science
36	Taxonomy Management Tools
38	Conclusions
<b>39</b>	<b>Part 2 Studies of Leading Practices</b>
39	Survey
40	Follow-Up Interviews with Survey Participants
40	Assessment of DOT Practices
44	Case Studies and Other Examples
48	Case Study: California DOT (Caltrans)
57	Case Study: Illinois DOT (IDOT)
61	Case Study: Minnesota DOT (MnDOT)
67	Case Study: Mississippi DOT (MDOT)
74	Case Study: Virginia DOT (VDOT)
81	Case Study: Washington State DOT (WSDOT)
92	Example: Alaska Department of Transportation and Public Facilities (ADOT&PF)
95	Example: Arizona DOT (AZDOT)
97	Example: Colorado DOT (CDOT)
98	Example: Georgia DOT (GDOT)
99	Example: Iowa DOT
100	Example: Kansas DOT (KDOT)
101	Example: Louisiana Department of Transportation and Development (LaDOTD)
102	Example: Maine DOT
103	Example: Nevada DOT (NDOT)
103	Example: New York State DOT (NYSDOT)
105	Example: North Carolina DOT (NCDOT)
106	Example: Ohio DOT (ODOT)
107	Example: Utah DOT (UDOT)
110	Example: Wisconsin DOT (WisDOT)
111	Example: FHWA
112	Example: National Oceanic and Atmospheric Administration (NOAA)
112	Example: Oak Ridge National Laboratory (ORNL)
113	Use of the Cloud
116	Example: Delaware
116	Example: Michigan
117	Example: Utah
118	Conclusions
<b>119</b>	<b>Part 3 Making Improvements in DOT Transportation- Information Management Practices</b>
119	Improving Information Management in a Transportation Agency
120	Types of Information
121	Information Management Processes at DOTs
123	Stakeholder Roles in the Information Management Process

123	General Approaches to Improving Information Management
125	Approaches Tied to Functions of the Information Management Process
125	Capture
127	Administer
131	Retrieve

**132 Glossary, Terms, Abbreviations**

---

Note: Many of the photographs, figures, and tables in this report have been converted from color to grayscale for printing. The electronic version of the report (posted on the Web at [www.trb.org](http://www.trb.org)) retains the color versions.

# Improving Management of Transportation Information

State departments of transportation (DOTs) need relevant and reliable information to support effective decision making. DOTs need this information to be readily available and accessible to users at a reasonable cost. Because significant amounts of time and money are expended to collect data, generate and store information, and make information available to users, DOTs need guidelines and good practices in managing transportation information

The information produced, managed, and used by transportation professionals has moved from static narrative documentation to dynamic databases that are used to produce reports and visualizations. Geographic information systems (GIS) have had a significant effect on transportation information because transportation infrastructure is location-based and geospatial visualization appeals to transportation information users. Narrative reports and other types of documents are still important, but most are now produced digitally. This shift to digital information has created challenges for information professionals in creating, managing, classifying, and storing transportation information so such information can be found and used by transportation professionals for their daily tasks.

Data and information have long been regarded as strategic assets of DOTs; now there is a surge of “big data.” The term refers not only to the burgeoning quantity and scope of data being collected and stored, but to the increased opportunities for extracting, from massive data, useful information that can help people gain understanding and make better decisions. Corporate giant IBM, for example, explains: “Every day, we create 2.5 quintillion bytes of data—so much that 90% of the data in the world today has been created in the last two years alone. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few. This data is big data.” (<http://www-01.ibm.com/software/data/bigdata/>; accessed 6/21/2013) The technologies enabling data collection and information management are advancing at a furious pace, often much faster than DOTs’ information-management capabilities.

It costs money to produce, use, and store information. Although DOT information is expensive to create, its value increases each time it is used. If information is not managed well and cannot be found when needed, then it has less value. The more efficient a DOT is in managing and using information, the greater the value of that information.

Measuring the value of information is difficult. Traditional measures of return on investment (ROI) focus on cost analysis and savings; however, exponential advances in information technology and data analysis have led to speculation that these measures are inadequate and that the economic benefits of information management are better captured by measures of customer satisfaction and quality of product offered. If a DOT can capture data on its transportation assets more efficiently and make that information available to users for various applications, then information management has increased value. This is especially true when

## 2 Improving Management of Transportation Information

considering the potential losses associated with redo loops for work recently performed, but for which information about the work is not accessible. Given the scarce resources at most DOTs, inability to find information (both recently generated and historical) is wasteful.

The rate, type, and amount of information created pose a challenge to effective governance and management of transportation information. Rapidly changing information technology is also a key factor. The varieties of terms used to describe the information, formats in which data and information are stored, and users' expectations for information continue to change rapidly. There are no commonly accepted guidelines for a DOT's information governance and management or standard capabilities required for agencies and their staff to manage the agency's transportation information across all information sources effectively.

This report presents the results of research conducted for NCHRP Project 20-90, "Improving Management of Transportation Information." This report will help meet the need for guidelines and good practices in managing transportation information at DOTs.

### Objectives

The objectives of NCHRP Project 20-90 were to provide guidance on (1) practices DOTs can use for capture, preservation, search, retrieval, and governance of transportation data and information; and (2) strategies and actions a DOT can follow to implement these practices. The guidance was intended to be sensitive to the diversity of state DOTs; the range of transportation information that DOTs use (related, for example, to project delivery, environmental review, and operations performance, as well as network configuration and design detail); and the variety of formats for transportation information (e.g., text reports, photographs, plans and drawings, geo-coded databases, and financial analyses). No single approach to managing transportation information is likely to meet the needs of all DOTs; guidance on management practices must provide flexibility for an agency to adapt recommended practices.

To meet these objectives, the guidance presented in this report includes three principal elements: (1) background explanations of key information management concepts, (2) summary case studies of management practices adopted by several DOTs that agency managers believe have improved agency performance and quality of transportation system service, and (3) suggested approaches for DOT personnel to take to improve management of transportation information. The primary audience for this report will likely be DOT information and data managers, but the contents may also be useful to any DOT leaders undertaking to communicate the value of information management to various audiences, promote establishment of strategic information management guidelines, and encourage collaboration among information technology (IT) professionals, data owners (e.g., project delivery and environmental review departments), and information managers (e.g., librarians and GIS managers).

### Use

To get the most benefit from the material presented in this report the following approach is recommended:

1. **Assess the DOT's existing practices** for managing transportation information. This includes, but is not limited to, identifying needs and gaps; categorizing needs as to where they fit in the information management process; and prioritizing the most critical (i.e., top three) needs.
2. **Identify good practices** to be adapted for use at the DOT.
3. **Develop an information management plan** to guide the improvement of existing information management programs or the development of new information management programs.

4. **Evaluate and document risks** associated with implementation and support of the information management plan.
5. **Develop an action plan** to implement the information management plan. The action plan identifies champions in business areas of the agency (information management professionals at the DOT) to help support implementation of the information management plan.
6. **Evaluate, at least annually**, the information management plan to improve the processes and tools and therefore the management of transportation information.

## Organization

The report has three main parts:

- **Part 1, Terminology and Categorization Standardization**, summarizes the current transportation information management landscape and discusses the following topics, which provide context for Parts 2 and 3:
  - Developing a common categorization scheme for transportation information management and identifying enhancements in detail or scope of information that should be included in such schemes;
  - Strategies for developing a common terminology and categorization scheme that could be made available for use by DOTs; and
  - Practices from other fields that may be adapted for improving DOT management of transportation information (e.g., file formats, naming conventions, and information preservation strategies).
- **Part 2, Survey, Case Studies and Examples, and Use of the Cloud** (1) describes the survey approach and summarizes the major research findings from the survey; (2) provides case studies and examples; and (3) summarizes cloud computing and states' activities in adopting cloud computing policies and installations. Part 2 presents the research efforts and specific findings.
- **Part 3, Making Improvements in DOT Transportation-Information Management Practices**, defines the common elements of data and information within DOTs, proposes a model of the processes used for information management, discusses stakeholders' roles in information management, and presents information DOTs can use to capture, administer, and retrieve information, as well as govern transportation data and information activities. Part 3 presents ideas that DOTs can use to improve their management of transportation information.

In addition, a list of terms, acronyms, and abbreviations is provided.



## PART 1

# Terminology and Categorization Standardization

The information produced, managed, and used by transportation professionals has undergone a transformation from primarily static narrative documentation to dynamic databases used to produce reports and visualizations. GIS have heavily affected transportation information—transportation infrastructure and use is location-based, so geospatial visualization appeals to transportation data users. Narrative reports and other types of documents are still important, but most of these are produced digitally, in serial versions that need to be managed and may need to be synchronized with the related data sources. These factors have created new challenges for information professionals to identify, describe and manage transportation information items so that they can be found and used by transportation professionals to accomplish their daily tasks.

Part 1 provides an overview of the current transportation information management landscape, and makes recommendations related to

- Developing a common categorization scheme for transportation information management and identifying enhancements in detail or scope of information that should be included in such schemes.
- Strategies for developing a common terminology and categorization scheme that could be made available for use by state DOT.

Part 1 also discusses practices from other fields that may be adapted for improving DOT management of transportation information (e.g., guidance for file formats, naming conventions, and information preservation strategies).

### **Types of Transportation Information That Need to be Managed**

DOTs are responsible for numerous reports and considerable information and data, including project information, systems conditions/performance data and reports, research reports, administrative information and inventories. A common way to assess the types of information an organization needs to manage is to consider the key business functions or activities of the organization and the types of information produced or consumed in accomplishing each function. Exhibit 1-1 illustrates how the scope of transportation information created as part of common DOT business functions includes data or data-generated information formats such as computer-aided design (CAD), GIS and other computer-generated graphics. Often these data-based visualizations are generated for use as part of a document or presentation as a document.

### **Maximize Information Use and Value**

DOT information is created to support a business function or activity, although the intended use of the information may not always be readily apparent. It is even more difficult to anticipate and envision future secondary and potential tertiary uses of that information. For example,

**Exhibit 1-1. Types of transportation information by function.**

Functions	Information Types	Formats				
		Documents	Data Tables	CAD	GIS	Graphics
Project Information	Engineering (e.g., Drawings)					
	Specification					
	Performance Test					
	Planning Study					
	Technical Reports (e.g., Materials Research)					
	Environmental Report					
System Condition/ Performance	Traffic Data					
	Safety Data					
	Performance Report					
Research	Research Report					
Administration	Financial					
	Contact Information					
Inventory	GIS Data					
	Asset Inventory Database					

information may be generated as part of an immediate operational activity such as accessioning assets,<sup>1</sup> which is part of the DOT asset management function. Later that same information may be analyzed to produce an asset maintenance plan. If DOT information is to realize its full value as a resource, it must be created and maintained in a form that will support such primary, secondary, and tertiary activities. To maximize the value of DOT information, consider how information can be structured to maximize its potential uses. Also consider when it is appropriate to archive and/or purge it from an active collection.

In order to maximize the usefulness of DOT information, it is necessary to consider the key information management activities in the transportation domain. Consider the primary purpose of each application and how it acquires, organizes, retrieves, secures, and maintains information to support specific objectives.

## The Relationship Between Data and Content Management

The specific types of resources that transportation data and information applications have handled have changed over time. Prior to the mid-1980s, most information was in databases and was primarily numerical in its basic format. Although word processing was available prior to the adoption of the personal computer (PC), this was primarily a centralized operation organized along the lines of a typing pool. As the PC emerged, word processing became ubiquitous and the production of text files and presentations led to an explosion of documents. The emergence of information networks and widespread adoption of email and the web accelerated the trend; today social media continues to drive the growth. These trends

<sup>1</sup>According to AASHTO, “Transportation Asset Management is a strategic and systematic process of operating, maintaining, upgrading, and expanding physical assets effectively throughout their lifecycle.” Accessioning assets is the process of adding a new asset such as streets, signs, curbs, gutters, rights-of-way, etc. into an asset management system. See FHA Asset Management Overview, December 2007 ([http://www.fhwa.dot.gov/asset/if08008/assetmgmt\\_overview.pdf](http://www.fhwa.dot.gov/asset/if08008/assetmgmt_overview.pdf)).

## 6 Improving Management of Transportation Information

have motivated an evolution in thinking about transportation information, from data management to content management.

**Data management** comprises processes and technologies for collecting and managing data so that it can be used to accomplish tasks effectively. For transportation information management, we are primarily concerned with (1) ensuring the quality of authoritative data, (2) identifying and providing appropriate access to authoritative data to accomplish business functions, (3) generating visualizations or analyses based on processing data, and (4) synchronizing visualizations and narratives with source data. A key aspect of data quality is to ensure that data values exist and that they are consistent. Other aspects of data quality include accuracy, timeliness, validity, and completeness. When data quality needs to be maintained across data sources, extra work is required to obtain a set of consistent values (e.g., for the names of organizations such as government agencies or contractors). Data values that have been assembled and mapped from across multiple data sources are called reference data. Alternately, a common set of data values (e.g., ISO 3166) may be used to identify countries.

**Content management** encompasses processes and techniques for collecting, managing, and publishing information so that it can be found and used to accomplish tasks effectively. For transportation information management, we are primarily concerned with (1) identifying and providing appropriate access to authoritative versions of content items, (2) providing adequate descriptions of content items so that they can be found and used, (3) linking to data sources for visualizations and analyses included in narratives, and (4) linking to related content items. Effective content management requires the generation and use of complete and consistent metadata values. Although system-generated metadata (e.g., unique identifiers and last-modified dates) are readily available, descriptive metadata (e.g., topic) frequently does not exist because it is not required by organizational business processes. With full-text search readily available, the value of author-generated metadata is not always considered worth the effort.

### **Content Structure**

Although the definitions of data and content management are aligned, the types and formats of information being managed are different. Given that most information forms today include structures amenable to automated processing, it is more useful to describe content items using a continuum of more to less structured.

Every type of content item has some data or metadata associated with it. However, data management applications are typically structured to answer types of questions such as “What is the current balance in a program fund account?” It is much more difficult to ask “What was the balance in a program fund account a year ago?” Similarly, computer file servers can reveal the last-modified date of a piece of content, but it is usually not possible for them to reveal the effective date of a piece of content—unless a human editor has added that information. For example, the effective date of a regulation is normally different from the date of the legislation, or the date of an announcement in the *Code of Federal Regulations* (CFR).

### **Processing**

Another trend is the processing of content to identify meaningful patterns (e.g., identifying patterns among the words and phrases in the text and extracting named entities, such as locations, organizations, or people mentioned in the text) and presenting these patterns using



some form of visualization such as GIS maps. Analytics is the processing of content into a data representation, where all types and forms of content can be reduced to a set of data values.

Managing content across an enterprise encompasses both data and content management. Sometimes this is referred to as enterprise content management and sometimes simply as data management.

### **Linking Source Data to Published Analysis in Documents**

An important challenge is managing heterogeneous content (i.e., narrative content), which may be based on structured data sets and include visualizations of that data. Providing dynamic methods to directly link narrative content to such source data is becoming necessary. It is no longer sufficient to manage such narrative content simply as a static content item. For example, a research report on highway safety that includes tables of data, charts, and maps needs to be linked explicitly back to the data sources so that further analysis of the same data set can readily be replicated, or new analyses performed.

### **Lifecycle, Workflow, Archiving**

A content set will typically evolve through drafts and versions and will often have associated annotations and commentary. Today's information manager must manage and synchronize multiple versions of overlapping sets of heterogeneous sources. For example, a PowerPoint report on material properties of highway surfaces will typically be developed through many drafts and versions for different audiences (e.g., engineers and budget analysts). The manager must keep track of the multiple versions and determine which is the most current or which is the official document of record. Although this is a difficult task, it can be addressed with versioning software available in document management systems.

### **Metadata Standards**

Metadata standards provide basic guidelines to ensure common description of content so that it can be found and used within and across applications, repositories, and organizations. Metadata should be associated with all types of content items, including documents, data sets, and visualizations. Metadata may be embedded in the content item or in a separate metadata database with identifiers to link the content item and the metadata database record. Metadata will be generated when the content item is created as well as each time the content item is used throughout its lifecycle. Ideally, metadata should provide a longitudinal record covering the life of the content item.

Two ISO metadata standards are particularly relevant to transportation information: ISO 15836, *Dublin Core*; and ISO 19115, *Geographic Information—Metadata*. Another potentially relevant ISO standard is ISO 11179, *Information Technology—Metadata Registries* (MDR).

### **ISO 15836 (Dublin Core)**

Referred to as the Dublin Core (which refers to Dublin, Ohio, the site of the meeting where this standard originated), ISO 15836<sup>2</sup> is the standard for describing content published on the web. This ISO standard defines 15 properties for use in resource description as shown in Exhibit 1-2.

---

<sup>2</sup><http://www.dublincore.org/documents/dces/>.

**Exhibit 1-2. Dublin Core properties.**

Core Property	Metadata Grouping	Purpose
Subject	Subject	Answers the questions What, Where, and Why
Type		
Coverage		
Date	Use	Answers the questions When and How
Language		
Rights		
Identifier	Asset	Answers the question Who
Creator		
Title		
Description		
Publisher		
Format		
Contributor		
Source		
Relation		

Dublin Core properties can be expressed as HTML meta tags or as RDFa, an HTML extension useful for marking and publishing metadata as linked data (i.e., publishing structured data so that it can be interlinked with data items from different data sources). Dublin Core has been widely adopted in government and business as the basic properties for describing content items. OMB Circular A-130 recommends that federal public websites use encoded Dublin Core metadata in the headers of their HTML pages.<sup>3</sup> These properties are available in commercial software such as in Microsoft SharePoint (MS SharePoint) where it is represented as metadata columns.

### ISO 19115 and the Federal Geographic Data Committee (FGDC)

ISO 19115 is the standard for describing geographic information and services. The FGDC endorsed this geographic information metadata standard in 2010. This common metadata standard for the description of GIS data sets has enabled the building of the National Spatial Data Infrastructure (NSDI) Clearinghouse Network. Geographic data, imagery, applications, documents, websites, and other resources have been cataloged for the NSDI Clearinghouse Network. ISO 19115 can be searched to find geographic data, maps, and online services. A description of the NSDI Clearinghouse Network from the FGDC Data and Services webpage follows:

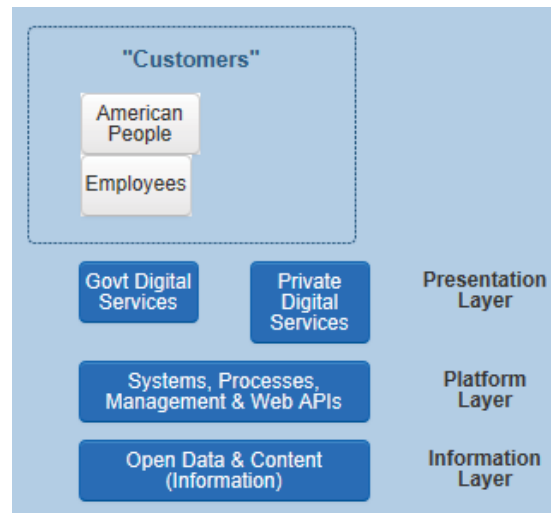
The Clearinghouse Network is a community of distributed data providers who publish collections of metadata that describe their map and data resources within their areas of responsibility, documenting data quality, characteristics, and accessibility. Each metadata collection is hosted by an organization to advertise their holdings within the NSDI. The metadata in each registered collection is harvested by the [geo.data.gov](http://geo.data.gov) catalog to provide quick assessment of the extent and properties of available geographic resources.<sup>4</sup>

DOTs can participate in the NSDI Clearinghouse Network to help meet their GIS needs and use their GIS resources. This is one way to use the adoption and use of the FGDC metadata standard.

<sup>3</sup><http://www.howto.gov/web-content/manage/categorize/meta-data>.

<sup>4</sup><http://www.fgdc.gov/dataandservices>.

**Exhibit 1-3. Digital government services model.<sup>1</sup>**



<sup>1</sup>Conceptual Model. In: Office of Management and Budget. *Digital Government Strategy* (<http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government.html>).

## Open Data and Digital Government Initiatives

Under the past two administrations, the White House has promoted information management good practices in U.S. government agencies that take advantage of the current and emerging networked information ecosystems. These initiatives have sought to improve customer service, efficiency, effectiveness, accountability, and transparency.

### Digital Government Strategy

Early websites consisted of static HTML pages usually authored by hand. Web content management (WCM) applications provide interfaces for authoring web pages and include features such as being able to preview what the HTML coding will look like before the page is published to a website. WCM applications also provide templates for creating different types of content and for publishing those pages with a particular design. Such templates can query a database or content repository to find content items to populate presentation templates, thus enabling an early version of dynamic web pages. This model of separating content creation from content publishing has become a key part of ECM strategies.

The Digital Government Strategy is a conceptual model with three service layers: (1) the information layer, (2) the platform layer, and (3) the presentation layer.<sup>5</sup> The information layer contains structured, digital information (e.g., traffic and highway safety data) as well as unstructured information (content) (e.g., fact sheets, press releases, and compliance guidance). The platform layer includes the systems and processes used to manage this information. The presentation layer defines how information is organized for presentation to users via websites, mobile applications, or other modes of delivery. These three layers, as illustrated in Exhibit 1-3, separate information creation from information presentation, thus allowing content and data to be created once and then used in different ways.

<sup>5</sup><http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government.html>.

There are similarities between the Digital Government Strategy and the Transportation Knowledge Network (TKN) as described in *NCHRP Report 643*.<sup>6</sup> The TKN vision is one of a portal enabled by information standards as well as communities of practice (COPs). The intention is very much the same as the Digital Government Strategy but the architecture has been updated. Whether implemented as a portal, wiki, or service-oriented architecture, the intention is the same—to enable wider access to transportation information at the local, regional, and national levels.

## Open Data

For many years, publishing government data (both datasets and bibliographic metadata) was done primarily by commercial publishers, and database publishing continues to this day. With the advent of the web, public government data began to be more widely and freely published, but on a voluntary basis.

The President's 2010 Memorandum on Transparency and Open Government, OMB Memorandum M-10-06, the Open Government Directive,<sup>7</sup> made it a priority for the public to be able to easily find, download, and use datasets generated and held by the federal government. Data.gov was created as a catalog to provide descriptions of these datasets. The federal open data policy has greatly accelerated the trend to make datasets from all levels of government publicly available. Under the data.gov model, DOTs can develop applications based on the datasets they have for traffic, safety, and other areas of public interest or simply publish the datasets and let third parties develop those applications. Exhibit 1-4 is a dynamically generated traffic map made available by a third party based on the California DOT (Caltrans) data

## Transparency, Application Programming Interfaces (APIs), Third-Party Applications

The Digital Government Strategy is intended to facilitate the development of services that use the information layer through the federal government open data policy and thus provide transparency. These services may be developed by the government agencies that produce the data, or they may be developed and deployed by other government agencies or other parties outside government. There are several architectures for doing this: (1) download the dataset and implement it as a stand-alone service on your own server, (2) dynamically query the information layer, or (3) link to the information layer. Dynamically querying the information layer can be done using web services description language (WSDL) or, if not supported, via an API or custom application. Linking to the information layer requires the agency to publish their information as a linked data service (i.e., publishing structured data so that it can be interlinked with data items from different data sources). (Linked data is the basis for the semantic web, which is discussed in the next section.)

Making APIs available for web developers has been popular. The Google maps API is well known and widely used to represent data that has geospatial information on a Google map. All mapping applications have APIs to represent geospatial data using one or more of their map layers. Exhibit 1-4 is an example of a mashup<sup>8</sup> which combines traffic data with a GIS map representation. Several information services make APIs available so that developers can access and reuse data. For example, the *New York Times* makes APIs available for article search, Congressional information, location concepts, and many others.<sup>9</sup>

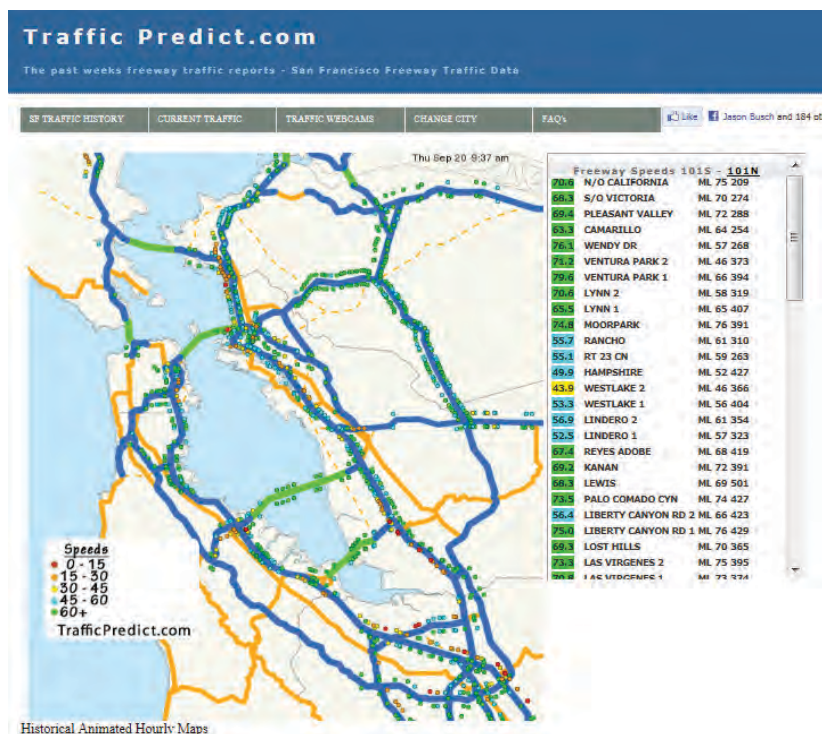
<sup>6</sup>[http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp\\_rpt\\_643.pdf](http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_643.pdf).

<sup>7</sup>[http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda\\_2010/m10-06.pdf](http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-06.pdf).

<sup>8</sup>A mashup is a web application that combines data from two or more sources and presents it in a single webpage.

<sup>9</sup><http://developer.nytimes.com/docs>.

**Exhibit 1-4. Traffic data map generated using web API (www.trafficpredict.com).**



## The Semantic Web

The term “semantic web” refers broadly to an extension of the World Wide Web that enables people to share content beyond the boundaries of applications and websites, the principles and practices underlying that extension, and the loose organization of people and institutions engaged in developing and institutionalizing those principles and practices. Underlying the semantic web is the idea that meaningful content—such as names of people or organizations, locations of events or resources, and dates, for example on web pages—can be tagged or presented in standard formats to facilitate the finding and presentation of the content. This semantic coding enables, for example, web browsers to recognize semantic content so that unstructured documents can be used like structured data and linked dynamically to data sources and visualizations. The semantic web is giving rise to various tools and methods such as the Standard Generalized Markup Language (SGML), Resource Description Framework (RDF), and Extensible Markup Language (XML); some of these most likely to be relevant to transportation information management are described below.

### RDF Schema (RDFS) Vocabulary Description Language

The Resource Description Framework (RDF) is a general-purpose language for representing information in XML on the Web. RDFS is an RDF extension used to describe groups of related resources and the relationships between these resources. A vocabulary description language is a way to represent the components of a schema and the relationships between these components. For example, Dublin Core or ISO 19115 can be represented in RDFS.

### Namespaces for XML Schema Elements and Attributes

A namespace is a specification for unique identifiers for labels. A namespace disambiguates labels that are otherwise the same (e.g., homographs), with unique and referenceable identifiers.

**Exhibit 1-5. Dublin Core Metadata Initiative® elements namespaces.**

Term Name	URI
contributor	<a href="http://purl.org/dc/elements/1.1/contributor">http://purl.org/dc/elements/1.1/contributor</a>
coverage	<a href="http://purl.org/dc/elements/1.1/coverage">http://purl.org/dc/elements/1.1/coverage</a>
creator	<a href="http://purl.org/dc/elements/1.1/creator">http://purl.org/dc/elements/1.1/creator</a>
date	<a href="http://purl.org/dc/elements/1.1/date">http://purl.org/dc/elements/1.1/date</a>
description	<a href="http://purl.org/dc/elements/1.1/description">http://purl.org/dc/elements/1.1/description</a>
format	<a href="http://purl.org/dc/elements/1.1/format">http://purl.org/dc/elements/1.1/format</a>
identifier	<a href="http://purl.org/dc/elements/1.1/identifier">http://purl.org/dc/elements/1.1/identifier</a>
language	<a href="http://purl.org/dc/elements/1.1/language">http://purl.org/dc/elements/1.1/language</a>
publisher	<a href="http://purl.org/dc/elements/1.1/publisher">http://purl.org/dc/elements/1.1/publisher</a>
relation	<a href="http://purl.org/dc/elements/1.1/relation">http://purl.org/dc/elements/1.1/relation</a>
rights	<a href="http://purl.org/dc/elements/1.1/rights">http://purl.org/dc/elements/1.1/rights</a>
source	<a href="http://purl.org/dc/elements/1.1/source">http://purl.org/dc/elements/1.1/source</a>
subject	<a href="http://purl.org/dc/elements/1.1/subject">http://purl.org/dc/elements/1.1/subject</a>
title	<a href="http://purl.org/dc/elements/1.1/title">http://purl.org/dc/elements/1.1/title</a>
type	<a href="http://purl.org/dc/elements/1.1/type">http://purl.org/dc/elements/1.1/type</a>

For example, an application could query a schema via a web service when it is necessary or useful to interact with an application that uses that schema. Exhibit 1-5 shows the Dublin Core namespace, which is used to discover the semantics and syntax of the Dublin Core elements. In this way, the semantics of XML schema elements and attributes are a type of vocabulary, i.e., a controlled list of values with a specific meaning and purpose.

### Namespaces for Category Values for Named Entities

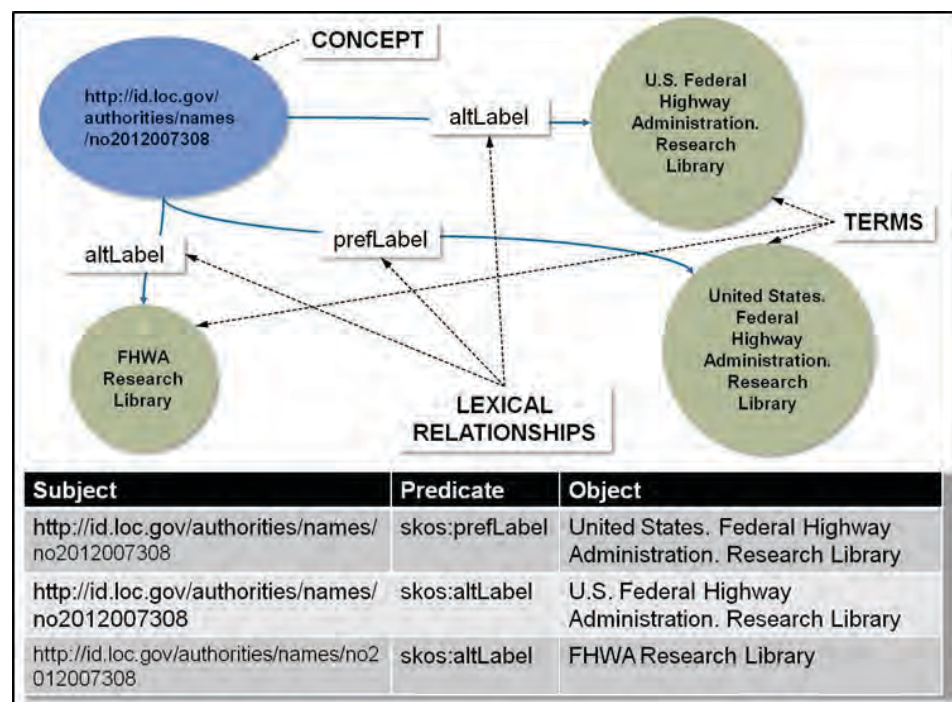
Other types of vocabularies that are important for the semantic web are “named entity” vocabularies. Named entities are the names of people, organizations, locations, events, topics, and other things with proper names or other specific, controlled names. These types of resources are named entity vocabularies. The RDFS vocabulary description language can also be used to describe the values in named entity vocabularies, the relationships among those values, and related values within a particular vocabulary or within another namespace.

SKOS (Simple Knowledge Organization System) is the World Wide Web Consortium (W3C) specification for representing knowledge organization systems using RDF. SKOS make an important distinction that the discrete components of a categorization scheme (often called nodes in a categorization scheme) are “concepts” and that various labels and other types of information can be associated with a concept.

Exhibit 1-6 illustrates an example of a concept—the name for the U.S. FHWA Research Library. The actual concept is an identifier, in this case the Library of Congress Name Authority File identifier for the U.S. FHWA Research Library (<http://id.loc.gov/authorities/names/no2012007308>). Each lexical relationship—in this case, equivalent relationships—can be represented as a subject-object-predicate triple shown in the table at the bottom of the diagram. An SKOS representation can easily be extended to add information about a concept by adding another row in the triple table referenced to the concept’s identifier as the Subject.



**Exhibit 1-6. Some semantic relationships for the U.S. FHWA research library.**



An emerging trend in library authority files and other types of authoritative lists of named entities (e.g., people, organizations, locations, events, and topics) is to publish them on the web using universal resource identifiers (URI). The most common URIs are uniform resource locators (URLs) or webpage addresses. URIs are unique identifiers on the web, so assigning URIs to authority records allows them to be referenced persistently on the web. The idea is to enable organizations to publish and reference named entities on the web. For example, FHWA could publish and maintain the authoritative list of FHWA agency names, programs, projects, and so forth. By doing this, the authoritative list of FHWA names would be available to DOTs as well as application developers—just as traffic and safety data is available for mashups.

## Review of Terminology and Categorization Schemes

This section provides a brief overview of controlled vocabularies, which can be used to describe and categorize transportation-related content to make such content easier to find and use.

### Ways to Categorize Content Collections

Working with digital content, users can either (1) browse for content using a file manager or (2) search for content using a local search engine. Browsing for content relies on how file directories and files themselves are organized and named. This method is often ad hoc—even when files are kept on a shared file store. Searching relies on how the internal search engine has been configured, including considerations such as what content has been indexed, whether indexing is full-text or metadata-driven, how search results are presented, and whether or not search refinements can be made to the query or the results.

## File Directory Methods

The most common way to organize content is to put it in a physical directory. This is similar to the single access method used with paper files where content items are filed in folders in filing cabinets. Everyone who has a PC has to decide how to set up file manager directories and folders. When there are shared network drives, a standard method of naming directories and folders is usually established.

Folders are organized differently depending on the business activity. Examples are as follows:

- **Records management** is based on a record retention schedule. These schedules are typically set up by business functions (e.g., accounting, administration, environment, finance, and human resources), then by content type (e.g., annual report, best practice, correspondence, datasheet, handbook, and form), and then by date.
- **Project management** is based on work breakdown structure (WBS), usually by technical discipline, subdivided by task, and then chronologically.
- **Administrative** files are often organized chronologically by date in alphabetical order (i.e., A-Z).

## Metadata Description Methods

In addition to putting content in a file directory, it is helpful to associate descriptive metadata with the content item. Metadata supports content retrieval for authors and content managers so they can

- Reliably find whether a content item exists,
- Determine ownership of a content item and whether it can be re-utilized or not,
- Enable alerts to new content or subscription to a pre-defined query, and
- Keep content items current, accurate, and in compliance with regulations.

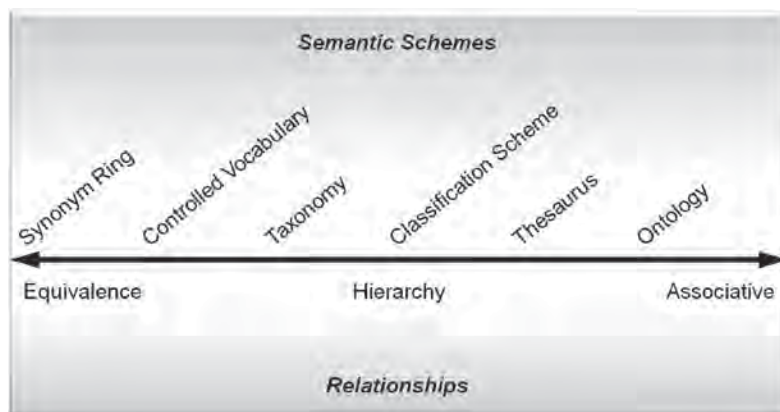
Metadata supports content publishing and general use such as the following:

- Faceted search based on metadata properties
- Search optimization
- Dynamic content delivery based on standard categorization
- Content reuse in multiple distribution channels (e.g., web, mobile, and really simple server [RSS] alerts)
- Content reuse in FAQs (Frequently Asked Questions) on specific topics and other categories
- Orienting those searching on public websites (even when they land on a page 15 layers deep)
- Ensuring consistent values for analytics across channels.

Exhibit 1-7 presents various types of semantic schemes along a continuum based on the types of relationships that characterize the scheme. Generally, the schemes are arrayed from simple to complex relationships in terms of the difficulty in making these types of relationships:

- **Synonym Ring.** A synonym ring is a set of words or phrases that can be used interchangeably for searching (e.g., fringe parking and park and ride).
- **Controlled Vocabulary.** A controlled vocabulary is a list of preferred terms (e.g., a pick-list in a data entry form).
- **Taxonomy.** A taxonomy is a system for identifying and naming things and classifying them according to a set of rules (e.g., a biological taxonomy or even most shopping websites which aim to arrange products according to a set of rules).
- **Classification Scheme.** A classification system is an arrangement of knowledge usually enumerated, but that does not follow taxonomy rules (e.g., the Dewey Decimal System).



**Exhibit 1-7. Semantic schemes: simple to complex.<sup>2</sup>**

<sup>2</sup>After: Amy Warner. *Metadata and Taxonomies for a More Flexible Information Architecture*.

- **Thesaurus.** A thesaurus is a tool that controls synonyms and identifies the semantic relationships among terms (e.g., Transportation Research Thesaurus (TRT)).
- **Ontology.** An ontology is a faceted taxonomy, but uses richer semantic relationships among terms and attributes, and strict specification rules.

## Relevant Transportation and Related Terminology Resources

This section discusses authoritative transportation-related terminology resources. The resources were found primarily by organic web searching. No libraries or specialized resources were consulted. Thus the resources listed are not exhaustive and are intended to be representative resources. The research team considers federal, state and some industry associations to be authoritative and has not included resources whose provenance could not be verified.

### Transportation-Related Glossaries

A glossary is an alphabetical list of terms in a specific subject area. Dictionaries usually have a broader scope. In the context of this project, glossary and dictionary mean the same thing. Glossaries typically include a definition for each term entry. Glossaries are often created to support the function of an organizational unit, a project, a policy initiative, compliance with a legislative mandate, or some other specific purpose. As such, glossaries are useful resources when building a terminology scheme for a subject area. They represent important concepts in the domain. Because glossaries have definitions, they are helpful in understanding the nuances that may be associated with certain terms and phrases in a discipline. Glossaries can also be a useful source of synonyms and quasi-synonyms; abbreviations, initialisms, and acronyms; and other term variants. Exhibit 1-8 lists authoritative vocabularies related to transportation in general and air, ground, rail, and water transportation more specifically. Exhibit 1-9 lists thesauri that are not specific to transportation.

### Transportation-Related Thesauri

A thesaurus is a controlled vocabulary. A thesaurus that conforms to the Z39.19 standard includes equivalent, hierarchical, and associative relationships among terms and is intended

**Exhibit 1-8. Transportation-related glossaries.**

Category	Title	Source	URL
General	Bureau of Transportation Statistics Dictionary	USDOT. RITA. Bureau of Transportation Statistics.	<a href="http://www.bts.gov/dictionary/">http://www.bts.gov/dictionary/</a>
	Glossary of Terms	USDOT. RITA. Intelligent Transportation Systems.	<a href="http://www.standards.its.dot.gov/terms.asp">http://www.standards.its.dot.gov/terms.asp</a>
	MnDOT Glossary	Mn DOT.	<a href="http://www.dot.state.mn.us/information/glossary.html">http://www.dot.state.mn.us/information/glossary.html</a>
	Transportation Glossary of Terms and Acronyms	Florida DOT. Office of Policy Planning.	<a href="http://www.dot.state.fl.us/planning/glossary/glossary.pdf">http://www.dot.state.fl.us/planning/glossary/glossary.pdf</a>
Air	Air Cargo Glossary	Cargo Airlines	<a href="http://www.cal.co.il/glossary">http://www.cal.co.il/glossary</a>
	Air Traffic Management Glossary of Terms	USDOT. FAA	<a href="http://www.fly.faa.gov/Products/Glossary_of_Terms/glossary_of_terms.html">http://www.fly.faa.gov/Products/Glossary_of_Terms/glossary_of_terms.html</a>
	Glossary of Airport Acronyms	USDOT. FAA	<a href="http://www.faa.gov/airports/resources/acronyms/">http://www.faa.gov/airports/resources/acronyms/</a>
	Glossary of Civil Aviation and Air Travel Terminology	airodyssey.net (blog)	<a href="http://airodyssey.net/reference/glossary">http://airodyssey.net/reference/glossary</a>
Freight	Glossary of the American Trucking Industry	Wikipedia	<a href="http://en.wikipedia.org/wiki/Glossary_of_trucking_industry_terms_in_the_United_States">http://en.wikipedia.org/wiki/Glossary_of_trucking_industry_terms_in_the_United_States</a>
	Planning Glossary	USDOT. FHWA. Office of Planning, Environment, & Realty	<a href="http://www.fhwa.dot.gov/planning/glossary/">http://www.fhwa.dot.gov/planning/glossary/</a>
	Truck and Bus Glossary	University of Michigan Transportation Research Institute	<a href="http://www.umtri.umich.edu/divisionPage.php?pageID=201">http://www.umtri.umich.edu/divisionPage.php?pageID=201</a>
	TWNA Glossary - Trucking terms	TWNA.org	<a href="http://www.twna.org/trucking_terms.htm">http://www.twna.org/trucking_terms.htm</a>
Rail	Glossary of North American Railway Terms	Wikipedia	<a href="http://en.wikipedia.org/wiki/Glossary_of_North_American_railroad_terminology">http://en.wikipedia.org/wiki/Glossary_of_North_American_railroad_terminology</a>
	Intermodal Glossary	Union Pacific	<a href="http://www.uprr.com/customers/intermodal/integlos.shtml">http://www.uprr.com/customers/intermodal/integlos.shtml</a>
	Railroading Glossary	Kalmbach Publishing Co	<a href="http://trn.trains.com/glossary">http://trn.trains.com/glossary</a>
Maritime	Glossary of Maritime Terms	American Association of Port Authorities	<a href="http://www.aapa-ports.org/industry/content.cfm?itemnumber=1077&amp;navitemnumber=545">http://www.aapa-ports.org/industry/content.cfm?itemnumber=1077&amp;navitemnumber=545</a>
	Glossary of Shipping Terms	USDOT. Maritime Administration	<a href="http://www.marad.dot.gov/documents/Glossary_final.pdf">http://www.marad.dot.gov/documents/Glossary_final.pdf</a>

to support the indexing and retrieval of documents. Several controlled vocabularies are currently in use for transportation information. The TRT is used for indexing the Transportation Research Information Services (TRIS) Database. The TRT originated in research sponsored by the NCHRP in the 1990s and is designed to cover all modes and aspects of transportation. It is managed by TRB with input and content development done by the TRT Subcommittee. Since 2010, TRB has assumed responsibility for the content of and further development of the TRT. (Readers can refer to *Principles for the Organization of the TRT* and *TRT Development and Maintenance Procedures*, for more information about the TRT<sup>10</sup>.)

<sup>10</sup>Available at <http://ntl.bts.gov/tools/trt/>.

**Exhibit 1-9. Non-transportation-specific thesauri.**

Title	Source	URL
Gale Transportation Thesaurus	Cengage	Not publicly available.
Intellisophic Urban Transportation Systems	Intellisophic	Not publicly available.
NASA Thesaurus	NASA	<a href="http://www.sti.nasa.gov/sti-tools/">http://www.sti.nasa.gov/sti-tools/</a>
Thesaurus for the Australian Transport Index	ARRB Group	<a href="http://www.arrb.com.au/admin/file/content2/c7/Aust%20Transpt%20Index%202010.pdf">http://www.arrb.com.au/admin/file/content2/c7/Aust%20Transpt%20Index%202010.pdf</a>

**Library of Congress Subject Headings**

The Library of Congress Subject Headings (LCSH) is the standard subject vocabulary used by university and research libraries in the United States to categorize materials in collections. The LCSH was designed as a pre-coordinated list of subject headings for subject indexing in library catalogs. Although originally intended for content published in books and for library card catalogs, the LCSH has evolved and been adapted to more contemporary models of subject indexing and information search and retrieval. The basis for determining entries is still literary warrant (i.e., based on the emergence of concepts and labels in publications). LCSH headings consist of topics or proper names (e.g., locations, events, and personal or organization names) which can be subdivided by genre, location, and/or time period. The entries include references to preferred labels, some cross-references showing relationships, and some parent and child thesaurus relationships have been added when appropriate. The LCSH includes terminology directly and indirectly relevant to transportation.

Recently the Library of Congress (LC) implemented a linked data service for authorities and vocabularies.<sup>11</sup> The LC linked data service includes the LCSH, as well as the Name Authority File (i.e., entries for the names of people and organizations) and other miscellaneous vocabularies. The linked data service is intended to enable people and machines to access LC authority data.

**FGDC Topic Categories**

FGDC (ISO 19115) Topic Categories, a set of 19 high-level subject categories, provides a standardized way to quickly sort and access thematic information. The FGDC Topic Categories are as follows:

- Farming
- Biota
- Boundaries
- Climatology/meteorology/atmosphere
- Economy
- Elevation
- Environment
- Geoscientific information
- Health
- Imagery/base maps/earth cover
- Intelligence/military

<sup>11</sup><http://id.loc.gov/>.

- Inland waters
- Location
- Oceans
- Planning/cadastre
- Society
- Structure
- Transportation
- Utilities/communication

## **Developing a Common Categorization Scheme for Transportation**

While having a single, generally accepted scheme for categorizing and organizing transportation information might facilitate information management, no such scheme is likely to emerge spontaneously from the past work summarized here. A focused effort would likely be required, either led by an authoritative body or self-organized from the information-user community. The following paragraphs discuss some of the principal issues that such an effort would have to address.

### **Requirements**

The TRT was originally developed to support bibliographic information retrieval, but the transportation information management landscape has changed since the 1990s. Today most transportation information is “born digital.” There is an enormous volume of data about transportation assets, particularly related to their location and use over time. Although asset-based information has not yet been fully integrated, we can expect further integration over time so that eventually there will be an up-to-date digital representation of every asset throughout its life, as well as a longitudinal record that can present a series of asset information snapshots over time. We should expect that reports making use of data and visualizations will be linked directly to those data sources. The two big future trends are (1) more data and (2) more integration. The key requirement to support these trends is metadata that will enable integration. That metadata will require common topical categories, as well as common sets of proper names for organizations, persons, computer programs, places, and other relevant named entities.

### *Flexibility*

The transportation information categorization scheme must be applicable to all types of content in all presentation formats. Transportation information takes many forms and exists in many formats. Specifically, the categorization scheme must be readily applicable to database schemas, dataset metadata, document properties, static and time-based visualizations, as well as heterogeneous content forms that include one or more of these content types or need to be synchronized across types (the dataset associated with a visualization in a document).

The scheme needs to be applicable to a collection, an item, and a component, and these should be able to inherit properties from the more general to the more specific level.

It should not matter where or how the container holding the values associated with a particular item is implemented—it may be embedded within the object or stored as an external database record referenced to the object using an identifier.

### *Semantics*

The transportation information categorization scheme should be based on standards and good practices for representing semantics between and among categories including (1) ANSI/NISO Z39.19-2005 Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies<sup>12</sup>, as well as (2) SKOS, the W3C specification on how to represent knowledge organization systems using RDF.

The scheme needs to support hierarchical, equivalent, and associative relationships. Hierarchical means broader and narrower concepts. The scheme needs to handle multiple parent (broader) concepts called polyhierarchy. Equivalent means synonyms and quasi-synonyms, near synonyms, abbreviations, acronyms, and other alternate labels. The scheme needs to support identifying regional variations where one label might be preferred in one region while another label would be preferred in another region (e.g., Park and Ride vs. Fringe Parking). Associative relationships means related concepts, called Related Terms (RTs) in TRT. OWL, the Web Ontology Language, may also be used to specialize associative relationships.<sup>13</sup>

### *Currency and Governance*

A transportation information categorization scheme must be frequently updated to reflect the terminology needs of DOTs and remain current in a rapidly evolving area that overlaps many related disciplines and locales. However, it is also important that the scheme is stable and is not changed just in reaction to immediate events. A governance process is needed that (1) defines roles and responsibilities, (2) identifies appropriate policies and procedures, and (3) provides a communication plan to promulgate the scheme and governance processes and communicates changes to all parties.

### *Localization*

A transportation information categorization scheme must support localization so that a DOT can choose which of several alternative labels available for a concept to use, but still be able to align the label with the core concept. The relationship among alternative labels must be defined so as to enable retrieval of federated search results from across different agencies, which may use different alternative labels. For example, although “Fringe parking” is the preferred LCSH (and the preferred TRT category), “Park and ride” or some variation of this label is frequently used instead. Exhibit 1-10 shows alternate labels for “Fringe parking” based on the LCSH linked data service.<sup>14</sup> The TRT also has the category “Fringe parking” which is represented by the TRT unique notational code identifier “Brddf”. Localization can also take the form of subsets of categories to support specific user communities; however, those communities may be defined (e.g., locale, function, expertise, project, and any other subdivision).

### *Ease of Use*

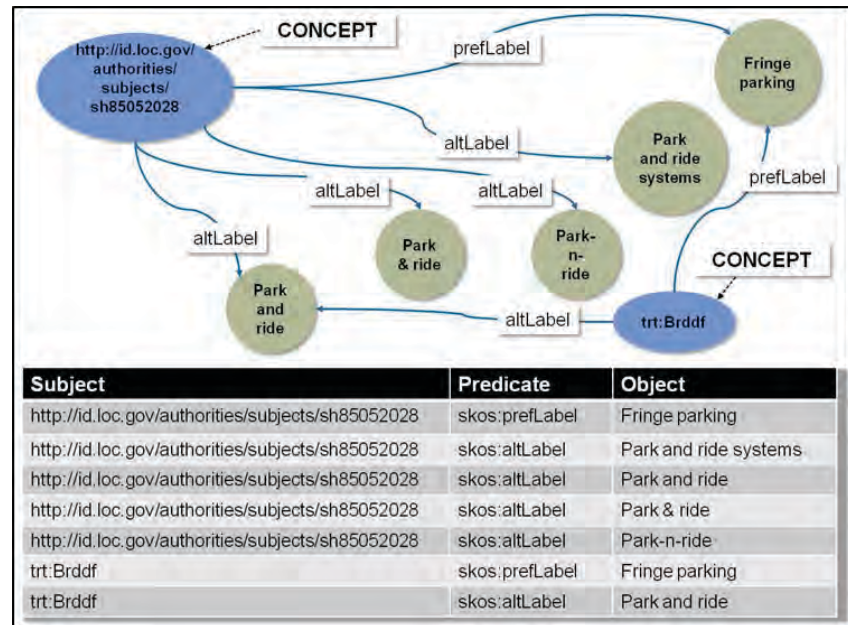
A transportation information categorization scheme must be easy to use for all stakeholders. The report classifies DOT information managers as (1) information professionals, (2) IT professionals, or (3) data managers. Additional stakeholders include DOT GIS offices and all levels of management, transportation professionals in local and regional transportation agencies, federal transportation-related agencies, transportation research organizations, and academic institutions. Those doing business with DOTs and other transportation-related agencies,

---

<sup>12</sup>[http://www.niso.org/apps/group\\_public/download.php/6487/Guidelines%20for%20the%20Construction,%20Format,%20and%20Management%20of%20Monolingual](http://www.niso.org/apps/group_public/download.php/6487/Guidelines%20for%20the%20Construction,%20Format,%20and%20Management%20of%20Monolingual).

<sup>13</sup><http://www.w3.org/TR/owl-ref/>.

<sup>14</sup><http://id.loc.gov/authorities/subjects/>.

**Exhibit 1-10. Alternate labels for the concept *Fringe parking*.**

NGOs, and citizens who have an interest in or have a need or want to know about transportation information should be added to this list of stakeholders.

Although DOT information managers may be the primary audience, all stakeholders need to be able to understand and use the categorization scheme to some extent. The expectation on the web is that categorization schemes need to be understandable without any training—they need to be as easy to use as Google. This does not mean that training, experience, and subject matter expertise are not important and valuable for obtaining more effective use of a categorization scheme, but it does mean that, on the surface, the scheme needs to be usable “out of the box.” The usability of a categorization scheme is usually measured by (1) discreteness of broad categories, (2) consistency in indexing information, and (3) consistency in finding information.

**Card Sorting.** The discreteness of categories can be measured by closed card sorting. In this test, commonly used terms selected from query logs and analytics are sorted into broad categories by representative users. The sorting results are compared to a baseline to measure consistency. This provides an independent assessment of how distinct the scheme’s broad categories are perceived to be. Seventy to eighty percent consistency is considered a high usability validation for a categorization scheme. Card sorting is usually done using online tools with iterative sets of 15 to 20 participants sorting up to 50 terms. There is some debate in the usability community about how many participants are required to provide meaningful results and at what point adding participants does not add meaningfully to the results.<sup>15</sup> The research team’s opinion is that 15 to 20 participants provide meaningful results.

**Indexing Consistency.** Inter-indexer consistency can be measured by having representative users index a set of representative types of information—data, visualizations, and documents. The values used to index each information item are compared to the baseline to measure

<sup>15</sup>See for example: Jakob Nielsen. “How Many Test Users in a Usability Study?” *Alertbox* (June 4, 2012) <http://www.useit.com/alertbox/number-of-test-users.html>.



completeness and consistency. Alternative values are identified, assessed, and applied to the scoring scheme as appropriate. This provides an independent assessment of the categorization scheme's usability for indexing. Seventy to eighty percent consistency is considered a high validation. Indexing is usually done as a paper exercise with iterative sets of 15 to 20 participants indexing 5 to 10 information items.

### *Findability*

Findability can be measured by having representative users look for, or describe how they would look for, a set of representative content items. The categories and subcategories used to find content items are compared to a baseline to measure consistency. Alternative category paths are identified, assessed, and applied to the scoring scheme as appropriate. This provides an independent assessment of the categorization scheme's usability for finding information. Seventy to eighty percent consistency is considered a high validation. Finding is done as a computer- or paper-based exercise with iterative sets of 15 to 20 participants searching for 5 to 10 information items.

## **Reengineering the TRT**

The previous section outlined the key requirements for a transportation information categorization scheme. Exhibit 1-11 shows scoring by the research team of the strengths and weaknesses of the TRT based on these requirements. The scale used is 1 to 5, with 1 being a major weakness and 5 being a major strength. "0" denotes non-applicable. The major strengths of the TRT are in the areas of Semantics and Currency. The major weaknesses of the TRT are in the areas of Flexibility, Localization and Ease of Use. Overall, the TRT is a well-designed thesaurus, but may not be suitable for digital information in the future.

### *TRT Strengths*

The major strengths of the TRT are in the areas of Semantics and Currency. The research team considers Semantics a strength primarily because the TRT follows the Z39.19 NISO standard for thesaurus construction. However, the TRT does not follow, use, or support SKOS or OWL.

**Strength 1: Semantics.** The TRT uses RT both to associate siblings (e.g., Private Trucking and Tank Trucking, Trucking and Bus Transportation, and so forth) and to associate terms in different facets. Although the Z39.19 standard discusses several RT relationships between terms in the same hierarchy—overlapping sibling terms, mutually exclusive sibling terms, and derivational relationships—these associations are problematic to model in other standards. For example, in SKOS, one cannot relate concept A to B if concept A is already broader than B. Also, in the interest of usability, it is often the editorial practice not to make associative relationships to terms that are siblings, or even to terms that are nearby in the same tree (e.g., child to grandparent). Overuse of generic associative relationships diminishes their usefulness and value. This adds complexity to the syndetic<sup>16</sup> structure without necessarily adding any value.

**Strength 2: Currency.** The research team considers Currency a TRT strength because of the well-defined and documented governance processes. Areas for improvement include expanding participation in the governance process by subject matter experts from the wider group of information management stakeholders, particularly to include data and GIS managers who can represent the needs and requirements of DOTs. The modes and formats of communication need to use the latest methods for receiving and routing term requests and for subscribing to

---

<sup>16</sup>The cross-references and relationships between terms.

**Exhibit 1-11. TRT scored against requirements.**

Requirement	TRT	Category
<b>Flexibility</b>		<b>2.14</b>
Applicable to all types of content		
Database schemas	1	
Data set metadata	2	
Document properties	4	
Static visualizations	3	
Time-based visualizations	2	
Heterogeneous	3	
Synchronize across types	0	
<b>Semantics</b>		<b>3.00</b>
Support standards		
Z39.19	4	
SKOS	0	
Relationships		
Hierarchical		
Parent/Child (BT/NT)	5	
Polyhierarchy	5	
Equivalent		
Synonyms (UF)	5	
Alternates	0	
Associative		
RTs	5	
OWL support	0	
<b>Currency and Governance</b>		<b>4.33</b>
Roles and responsibilities	4	
Processes	5	
Communications	4	
<b>Localization</b>		<b>0.00</b>
Regional variations	0	
Other subsets	0	
<b>Ease of use</b>		<b>2.33</b>
Broad category discreteness	3	
Indexing information	2	
Finding information	2	
<b>OVERALL SCORE</b>		<b>2.65</b>

and being sent notifications of all term request status changes. This (together with the recommendation for handling subsets and syndication of other authoritative sources discussed in the next section) will help improve the currency of the TRT.

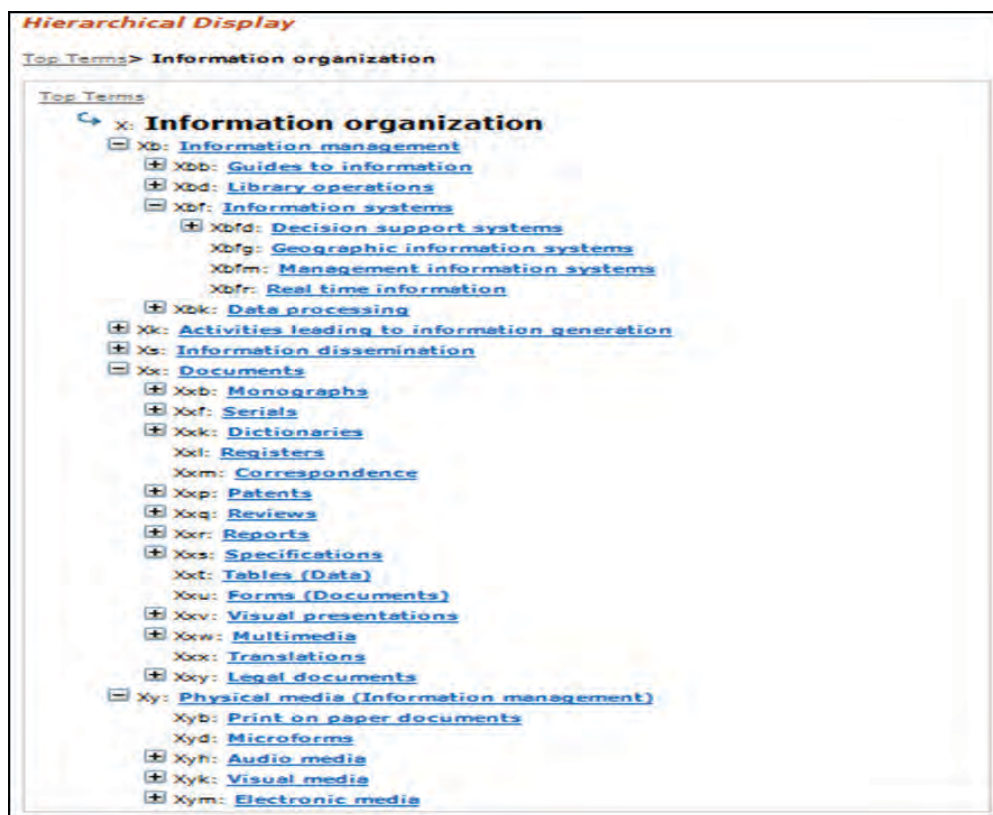
### *TRT Weaknesses*

The major weaknesses of the TRT are in the areas of Flexibility, Localization, and Ease of Use.

**Weakness 1: Flexibility.** The TRT has been designed to be used primarily on document-based content, especially technical reports. For example, as illustrated in Exhibit 1-12, the *Information organization* facet of the TRT includes a set of terms for types of *Documents*, but not for datasets or data-based visualizations. GIS is a category under *Information management > Information systems*. In this example, the TRT mixes information organization functions with



**Exhibit 1-12. TRT information organization facet showing selected subcategories.**



information forms. The current best practice is to separate categories that are forms from categories that are functions into separate broad divisions, facets, or, even into separate schemes.

An important information management requirement is to synchronize data with its representation in documents. While synchronization itself requires a system of identifiers that will be understandable to people and also be interpretable by systems, a common categorization scheme to describe both datasets and documents is crucial.

**Weakness 2: Localization.** The TRT offers no way to identify alternate terms and use them instead of the TRT preferred form. The Z39.19 NISO standard for thesaurus construction provides guidance on how and when to choose one form of a term or another as the preferred form based on criteria related to usage, spelling, abbreviations, jargon, trade names, popular names, loan words, and proper names. The reality is that these criteria are somewhat subjective and variations may be preferred for similar subjective reasons. Therefore, as many variations as possible must be collected in a categorization scheme, and it must be possible for users to apply their own criteria to choose which variation will be the preferred term in their local implementation. As discussed above, SKOS solves this problem with the representation of a concept as a unique identifier and terms being associated with that identifier via a type of relationship. Localization is a key functional requirement for the categorization scheme.

**Weakness 3: Ease of Use.** The TRT has been designed for “indexers, content managers and librarians in the transportation community,”<sup>17</sup> and not the broader group of transportation

<sup>17</sup>TRT “About” webpage (<http://ntl.bts.gov/tools/trt/>)

stakeholders. The TRT is intended to be comprehensive and exhaustive. It is very large, but it does not include any proper names for geographic locations, DOTs, or federal transportation programs or provide links to authoritative sources for such relevant named entities. The current expectation for ease of use is that a categorization scheme must be intuitive and “natural” to use for tagging and finding content. Practically, the TRT needs to provide methods to create more usable subsets to support specific applications, while maintaining a master authoritative resource.

### *Recommendation to Address Weaknesses: Microthesaurus*

The Z39.19 standard for thesaurus construction briefly discusses the concept of a microthesaurus as a subset of a broader thesaurus created to be used in “specific indexing products.”<sup>18</sup> According to the standard, microthesaurus requirements include the following:

- There should be a defined scope for a specialization of the broader thesaurus.
- Terms and relationships extracted from the broader thesaurus should have integrity (i.e., there should be no orphan terms).
- Additional specific descriptors can be added, but should be mapped to the structure of the broader thesaurus.

Providing a way to enable DOTs and other TRT users to generate and maintain microthesaurus subsets would address the flexibility, localization, and usability weaknesses discussed.

There are two broad models for implementing a microthesaurus—centralized and decentralized. In the centralized model, a new term property and set of controlled values is defined to be used to code any thesaurus term so that it can be included in a microthesaurus. In this model, the so-called broader thesaurus is, in effect, a microthesaurus so each term must be identified. To define a microthesaurus, preferred and variant terms would be individually selected. In this model, term relationships generally remain unchanged from the broader thesaurus. Exceptions are automatically handled so that (1) broken hierarchical relationships will be collapsed, (2) orphaned equivalent and associative term relationships will be ignored, and (3) relationships will be added for any new more specific terms. As shown in Exhibit 1-13, a thesaurus management tool that can provide distributed access is usually required to implement such a centralized service. Solid lines indicate tight integration; dotted lines indicate loose integration.

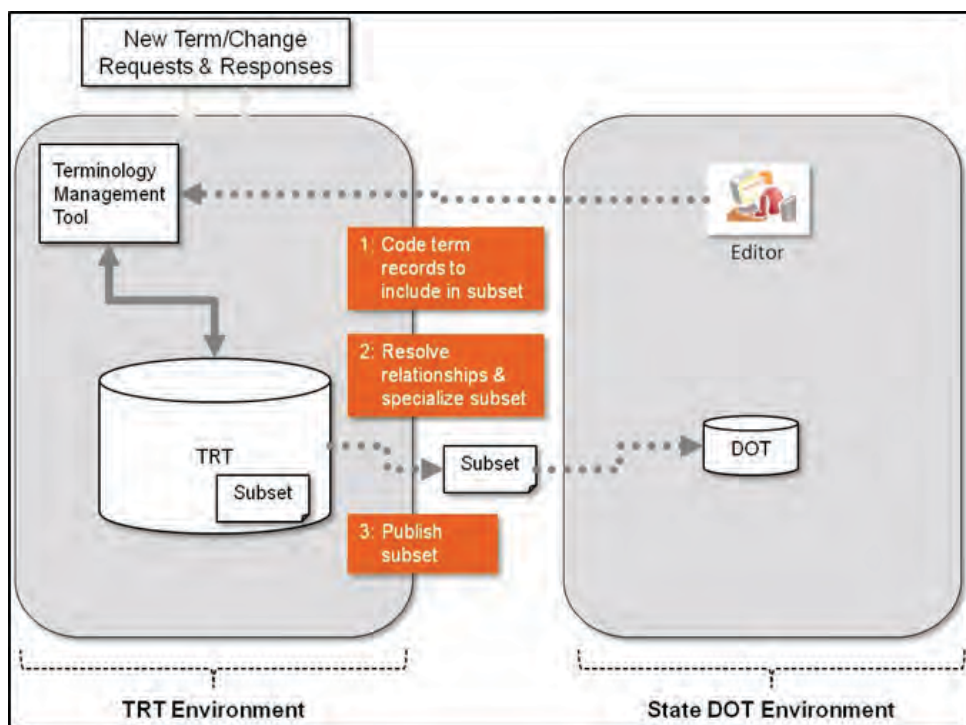
Exhibit 1-14 illustrates a decentralized service that would be decoupled (or only loosely coupled) to the broader thesaurus infrastructure. In this model, all or selected TRT terms would be downloaded to a microthesaurus site. That site would have its own thesaurus management system—tools and processes. No coding would be required to identify that a TRT term is to be included in the microthesaurus. Exceptions to relationships would be handled by editors so that (1) broken hierarchical relationships would be rationalized, (2) orphaned equivalent and associative relationships would be resolved, and (3) relationships would be added for any new more specific terms. If more specific terms are outside the scope of the TRT, then they would not be submitted as candidate terms. Instead, TRT would maintain a catalog or registry of microthesauri.

It would also be valuable for TRT to maintain a catalog of external terminologies for proper names from authoritative sources such as the Library of Congress.

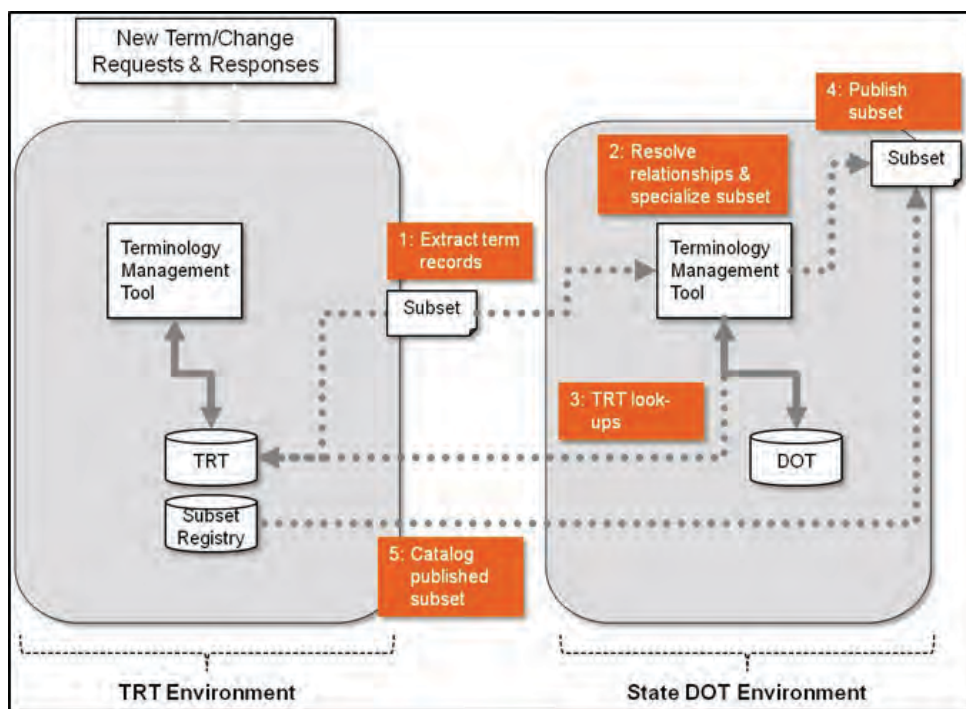
The decentralization is not only applicable to microthesaurus building, but also to distributing the work of identifying, building, and maintaining terminology subsets as a general vocabulary management model.

<sup>18</sup>Z39.19-2003, p. 33.

**Exhibit 1-13. Centralized microthesaurus/subset model.**



**Exhibit 1-14. Decentralized microthesaurus/subset model.**



## Community-Based Vocabulary Management Model

This section describes how a community-based transportation information terminology management model could work.

### *What Does a Community-Based Model Look Like?*

A community-based model for transportation information terminology management delegates responsibility for creating and maintaining terms and groups of terms to distributed responsible organizations such as DOTs. The central office functions as the overall terminology editor and coordinates the distributed effort by

- Managing the terminology management environment;
- Identifying who will be responsible for what term subsets;
- Routing candidate terms requests to distributed editors based on expertise, volume, or other criteria;
- Communicating editorial policies and answering editorial questions;
- Providing training to distributed editors;
- Communicating additions and changes to subscribing users; and
- Coordinating communication with the community at large.

### *Governance Model*

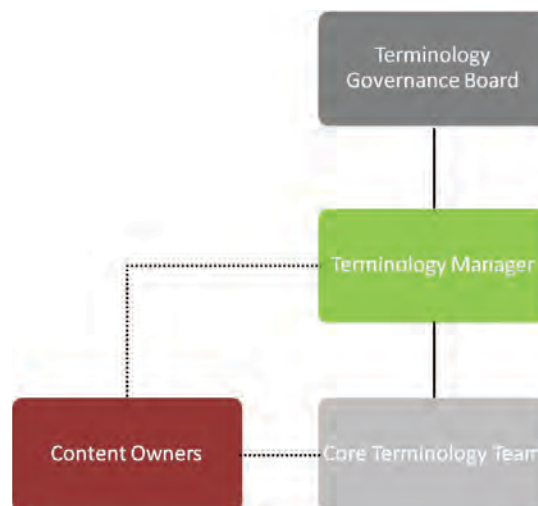
Terminology governance models need to specify the (1) roles and responsibilities, (2) policies and procedures, and (3) plan for communicating with the community about the program as well as providing notifications about terminology additions and changes.

**Roles and Responsibilities.** Exhibit 1-15 provides an overview of the terminology governance roles and the relationships among them.

**Governance Board.** The governance board provides executive sponsorship, arbitrates disputes and disagreements, is responsible for long-term decision making, and endorses policies and procedures as needed.

The governance board provides leadership and final decision-making authority for the core team of terminology editors on matters related to the transportation information terminology

**Exhibit 1-15. Terminology governance roles.**



(hereafter referred to as TransIT). The role of the governance board is to define the vision of improving transportation information retrieval using XML tagging, metadata, and controlled vocabularies. The board should be limited to four or five members, representing the major stakeholder groups, to foster agile decision making. This governance board would arbitrate disputes and disagreements between stakeholders regarding usage of TransIT. Members would meet quarterly with the TransIT core team to review execution of strategic plans and to ratify recommended new policies and procedures.

**Terminology Manager.** The terminology manager provides day-to-day operational support for TransIT including

- Routing change requests to distributed editors or directly handling them,
- Responding to requests regarding content classification,
- Enforcing the terminology governance policies and procedures,
- Reviewing requests for changes that may affect the overall terminology structure, and
- Ensuring that procedures for tagging content are effective.

The terminology manager will ensure execution of the daily activities needed to support TransIT governance and strategy. The manager will be a primary owner of the terminology management tool to document, edit, and publish the TransIT. The manager also needs to understand the context of key terms, term relationships, and vocabularies necessary for users to retrieve information from transportation information management systems. The manager will advise on how to consistently and appropriately use TransIT to populate metadata in transportation information management systems.

**Core Team.** The core team represents various stakeholders. The team will

- Be responsible for creating, updating and cataloging terminology subsets that have been assigned for their stewardship;
- Keep the terminology dynamic, updated and relevant; and
- Understand terminology editorial principles.

These individuals will be the primary point of contact for the terminology manager on terminology issues and should have a good understanding of taxonomy and metadata principles as well as the overall vision of the governance board.

**Content Owners.** These individuals determine appropriate content as well as access control, content provisioning, and compliance with regulations. Content owners advise the core team and terminology manager on initiatives to improve accuracy and usability of metadata and vocabularies when applied to various types of content.

### *Policies and Procedures*

The TransIT terminology should be governed by a documented set of policies and procedures. The current TRT procedures could be used as a model and updated to reflect a community-based vocabulary management model. At a high level, these principles can be summarized as follows:

- All primary actions related to editing the terminology should be driven by policies and procedures;
- To ensure sufficient review and communication prior to adding or editing a concept, a change request should be submitted;
- All change requests should be logged and acknowledged by the terminology manager; and
- An assessment of the benefits and effect of change requests should be completed and communicated quickly.



The foundation for any terminology subset service is to implement a persistent method to identify TRT terms. The current system of unique notational codes is not an adequate term identifier system. The TRT should adopt a Uniform Resource Identifier scheme.

### *Communication Plan*

As the TransIT terminology continues to evolve, communications and outreach to more constituents will be imperative for long-term success. It will be important for the terminology governance board to ensure consistent communications are released that

- Explain the process to request a change;
- Explain the actions of the terminology governance board and associated roles, including overall goals, terminology and metadata strategy, and decision-making process; and
- Present the value of the terminology in a meaningful and concise manner.

As a first step, the TransIT terminology governance strategy should be made publicly available to all current and future stakeholders. This will enable all interested parties to understand the immediate scope, efforts, and decision-making constraints of the taxonomy. In short, by making clearly visible the processes governing the TransIT, the team will be encouraging a greater understanding of its efforts for all transportation information management constituents. Some of the greatest benefits of this approach should be an increase in valuable feedback, more complete requests and up-to-date terminology.

## **Practices from Other Fields**

This section discusses good practices from other disciplines and applications that can improve the management of transportation information. Each section provides a short overview of the discipline or application, followed by some specific practices applicable to transportation information management.

### **Enterprise Content Management (ECM)**

ECM includes processes and applications to create and manage all types of content across an organization. Data and content management are converging. The practical implications are that data management policies and processes are being applied to more information types. A best practice relevant to transportation information management is the emergence of faceted taxonomy for categorizing and retrieving information with ECM applications.

A taxonomy is a system for identifying and naming things and classifying them according to a set of rules. A faceted taxonomy is a set of mutually exclusive taxonomy divisions that can be used to classify a content item. Faceted taxonomies are frequently used in online shopping websites to help customers find products to purchase. For example, shopping for shoes online would typically include facets for gender, brand, style, material, and color. The same principles used in designing a shopping website are being applied to classifying content in ECM applications. For transportation applications, the taxonomy facets might be content type, location, asset type, asset material, and asset age.

For ECM, some taxonomy facets emerging as common practice include

- Document type;
- Business function;
- Organizational unit;
- Location;
- Time period; and
- Topic.

**Exhibit 1-16. Common taxonomy facets used in ECM with transportation examples.**

Facet Name	Description	Transportation Examples
Type	The type of resource	Technical Report, Data Set, Whitepaper, Drawing
Function	The purpose of the content and any business activity or function that the content is about or related to	Procurement, Human Resources, Engineering, Policy Making
Organization	The organizational unit that the content is about or related to	FHWA; TRB; VDOT
Location	The geographic location or facility that the content is about or related to	Vancouver, WA; San Antonio Federal Complex; 800 E Leigh St, Richmond, VA
Coverage	The period when the content is effective or that it refers to	January 1, 2010–December 31, 2010; September 30, 2009, 5:00 pm; Twentieth Century
Subject	Other themes that the content is about or related to	Maintenance Practices, Domestic Transportation, Liability

Exhibit 1-16 describes facets generally applicable to all types of content and some transportation examples.

Although it seems as if there would be standard lists of values for types of resources, there are not. Dublin Core includes a type vocabulary as part of the Dublin Core® Metadata Initiative (DCMI) Terms<sup>19</sup>; however, the list of values (i.e., Collection, Dataset, Event, Image, Interactive Resource, Moving Image, Physical Object, Service, Software, Sound, Still Image, and Text) is not appropriate for ECM. A resource type vocabulary is almost always uniquely developed for each organization. The lists may be similar from one organization to another, but one agency’s “whitepaper” may be another agency’s “technical report.” Resource types may be a component of a records management schedule as discussed below.

A records retention schedule is the listing of the types of organizational records, with guidelines for their transfer or disposal, including the length of time records need to be retained. Most states have regulations about the retention and disposition of state and local agency records, including those from DOTs. Records retention schedules are usually organized by function and then by type.

## Records Management

Records management is the set of processes related to managing an organization’s information resources over their lifecycle from the time they are created until they are disposed of. While ECM includes creating, managing, and using the resources to support organizational activities, records management includes collecting, classifying, storing, and destroying or preserving the resources. DOTs need to manage their records to

- Comply with state regulations;
- Preserve an historical record; and
- Support the long-term maintenance of assets and applied research.

<sup>19</sup><http://dublincore.org/documents/2012/06/14/dcmi-terms/?v=terms>.

DOTs as public agencies need to comply with state requirements to maintain records according to an appropriate retention schedule and also need to be able to make their records available on request from the public as required by the Freedom of Information Act (FOIA). DOTs also have a broader responsibility to preserve records with historical value and need to apply appropriate criteria to select those records, which should be preserved even beyond the required retention period. Finally, DOTs have to maintain transportation assets that may have long lifecycles (e.g., bridges and tunnels). This includes maintenance, repair, and replacement of transportation assets, as well as gathering information over the whole lifecycle that can be used for applied research to inform transportation engineering decisions and good practices over time.

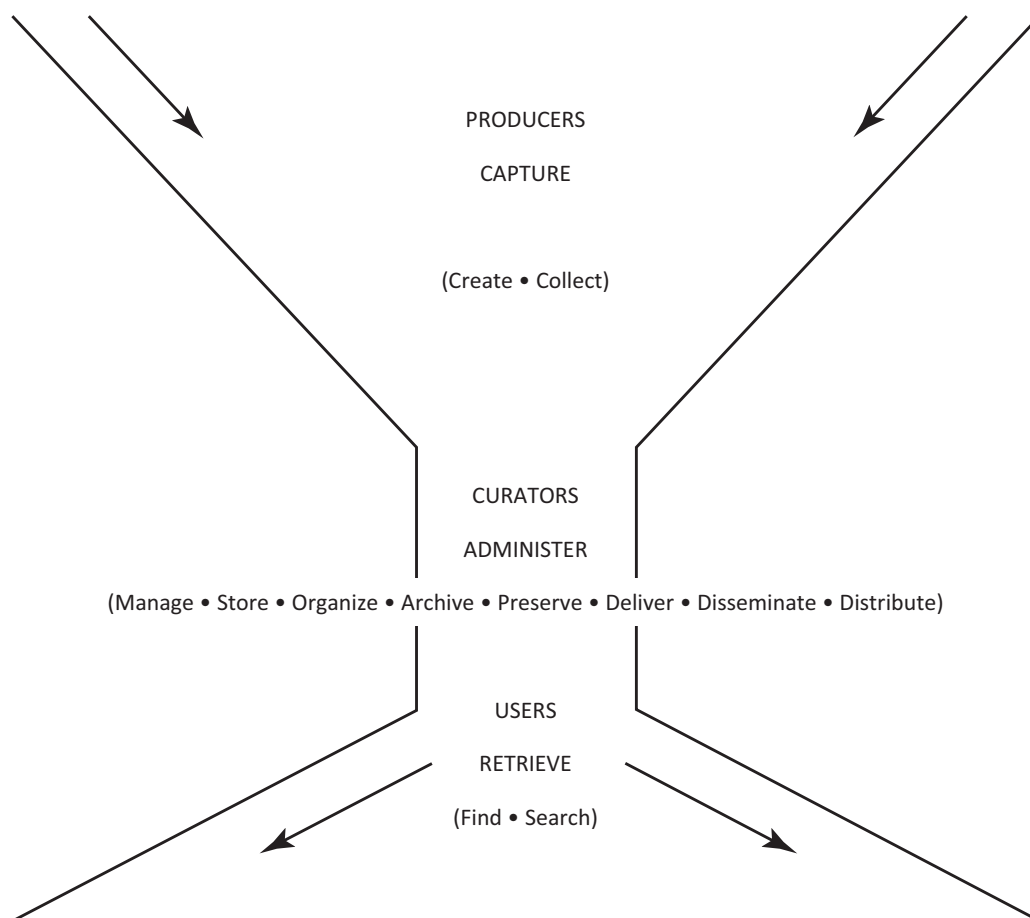
### *Content Framework*

An information management framework, referred to hereafter as the information management process, is provided in Exhibit 1-17.

### *Digital Preservation*

Most resources today are born digital—They are created, published, and used electronically. The electronic version of the content is the resource of record. This may be a PDF file for a technical report, a set of database layers for a GIS visualization, or a CAD file format for a blueprint. All of these are electronic files that require specialized software to be rendered human readable. Simply preserving the bits will not ensure that the files will be readable in the future. Although an electronic version of a technical report may be retrievable based on the words and phrases

**Exhibit 1-17. Information management framework.**





that occur in the file, GIS, CAD, and many other types of visual electronic content do not contain descriptions that can be used to retrieve them. Descriptive information, called metadata, needs to be created and linked to the file so information can be stored and retrieved.

Because most DOT resources are being produced electronically, digital preservation is becoming the primary focus of records management. Digital preservation, as with other forms of records management, requires processes to identify, collect, classify, store, and preserve digital resources.

### *Content Harvesting*

With most resources being produced electronically, there is the opportunity to systematically collect resources using automated methods such as spiders and robots that are the same as those used by web search engines.

## **Document Management**

Document management is a set of processes and applications for managing and storing electronic documents. Generally, these sorts of applications manage document-like content but not databases or visualizations generated from data (e.g., GIS or CAD). However, information output from a database such as a spreadsheet, table, or other form of columnar report no longer dynamically tied to the source database is a “document” for the purposes of document management. Similarly, a static map, diagram, or blueprint output from a GIS or CAD system is a “document.” For DOTs, many documents are generated as part of the administration, project management, performance, research, inventory, and other business functions.

### *File Directory Structures*

When documents are uploaded to a document management system, they need to be put in a shared file directory structure. Thus an important part of configuring such a system is to design a specification for how the directory structure will be set up and maintained. The goal of a file directory structure scheme is to organize the directory and name files so that it will be easy to find documents based on the directory structure and sort order of the file names. Thus file directory structures should provide an intuitive method that will enable (1) consistent filing of documents in the appropriate place in the folder structure and (2) a reliable method for finding documents in the system.

**Personal Working Documents.** According to Hicks, “the most common criteria for naming directories are the purpose or function (85%), the name of the project (55%) and the date (20%).” (p. 23:36) Working documents are commonly organized by project, then by context (such as function), then by year, and then by type. Although these components are commonly found, the order of these components vary depending on the function that owns the file store. Hicks studied personal electronic files which are likely to be found on a personal or shared file store. Engineers do not typically set up document management systems, so the particular scheme described by Hicks is not necessarily the one likely to be found in such a system. The best practice for personal working documents is a folder structure based on document type, chronology, subject or work flow.

**Organization Shared Files.** The best practice for shared file stores is to use the organizational structure as the overall directory structure. This practice dates from the era of paper files in filing cabinets and is still common practice.

The problem with using organizational structure is that when there is a re-organization, the directory structure needs to be changed. This was true with paper files in filing cabinets and is still the case with electronic files. However, it is much easier to change a directory name than it was to change the name of a department in paper files.

With electronic files, problems may arise when content items are linked to (hyperlinked or referenced) if the directory structure and file name are used as the identifier for the content item. So for content items stored on a shared file store, it is not a good idea to change the name of a directory, folder, or file name.

The preferred way to handle the effect of re-organization on a file structure based on organizational structure is to stop using the old folder and create a new folder for the re-named or new organizational entity. After all, content should be stored (and labeled) with the name of the organizational entity that existed when the content was created. That was a valid entity at one time.

Although this principle is useful, it is often better to create a directory structure based on a scheme that will be more stable than organizational structure. The best practice is to use business function categories. For example, even though the organizational entity called “Risk Evaluation, Analysis & Management” may be changed to “Risk Evaluation & Management,” the business function is still “Risk Management.”

**Website Content Management.** For a website (as opposed to a general-purpose file store), it is also useful to group content by type. This facilitates maintenance, administration, and approvals because different types of content have different lifecycles. For example, a “Policy” has a different approval process than “Home Page.”

Root folders should be labeled with the content types, for example

- . . . /Application/Model/
- . . . /Document/Policy/
- . . . /Image/Photo/
- . . . /Web Page/Home-Page/
- . . . /Component/Org-Chart/

### *File Naming Conventions*

Files should be named so that they are grouped usefully by alphabetical and chronological sorting. This should be done using a multi-part structure, usually with three segments such as purpose or function, short title, and date.

File names should start with a function code based on a controlled vocabulary of business functions (see Exhibit 1-18).

Short titles should be no more than 20 characters long. Titles could automatically be the first 20 characters of the Microsoft Office Document Properties Title field or a title defined by the document creator. Titles should describe the purpose and intent of the content. Articles, pronouns, and conjunctions should be avoided as much as possible. Dates should be related to the content (e.g., the report date explicitly shown in the document), not the system-generated last-modified date. Dates should be formatted in a standard way that sorts chronologically (e.g., YYYYMMDD). Exhibit 1-19 illustrates the basic file naming structure and provides examples.

## **Information Science**

Information Science is an interdisciplinary field at the intersection of library science, computer science, linguistics, philosophy, communications, and other disciplines. This section discusses information retrieval (IR) methods that can automate content classification. Automated categorization and text mining use various analytical methods to (1) identify words and phrases relevant to the meaningful use of content items and (2) generate statistical representations of content items based on an analysis of those words and phrases. Using IR methods can help

**Exhibit 1-18. Example of controlled vocabulary of business functions.**

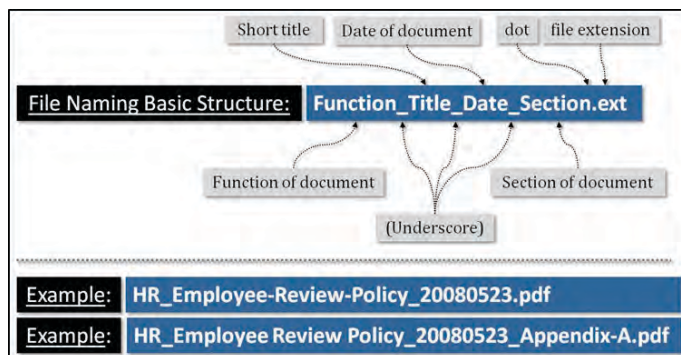
File Name Function Code	Top-Level Function Category
Adm	Administrative Resources
ETS	Engineering & Technical Services
Fin	Finance
Gov	Governance & Ethics
HES	Health, Environment & Safety
HR	Human Resources
IT	Information Technology
Lead	Leadership
PA	Public Affairs
Res	Research
Risk	Risk Management
SCM	Supply Chain Management

DOTs classify large volumes of content without hiring armies of librarians. This can help DOTs comply with records management regulations and open government policies and also participate in data publishing initiatives like data.gov.

### *Automated Categorization*

When people categorize content it is labor intensive, and the results may be incomplete and inconsistent. Once configured, automated methods are less expensive and very consistent, but it is difficult to consistently produce accurate results. There is always a tradeoff between accuracy and completeness, which is called “precision” and “recall” in information retrieval. The best scenario is when automated methods are used to suggest classifications and subject matter experts (SMEs) review and improve those classifications. Several ways exist for computers to tag content, ranging from very simple to quite complex. This section reviews several of these techniques.

**Keyword and Regular Expression Matching.** The most obvious method to automatically tag content is to associate several keywords and key phrases with each category. The text of each

**Exhibit 1-19. Basic file naming structure and examples.**

content item is searched. If there is a match, the item is assigned to the matching category. Many tools provide keyword matching capability as part of their capabilities. Another technique is to match regular expressions—patterns that include wildcards and other features—to offer broader or narrower matches than simply an exact match against a keyword or key phrase. Simple keyword and regular expression matching is likely to lead to false positives and false negatives. False positives are when an item is incorrectly part of a result set; false negatives are when an item is not retrieved but should be. The accuracy of this approach can be improved through more complex approaches to the rules and various ways of preprocessing the text before the rules are applied (e.g., to narrow the matching to words that are nouns, and phrases to noun phrases through part of speech (POS) analysis).

**Templates and Business Rules.** Another basic method is for metadata values to be hard-coded into the template used for adding content to the system. For example, the Procurement office of a DOT may have a template staff use for creating Requests for Quotation (RFQs). That template may have several metadata fields with pre-defined values (e.g., the Type, Organization, Function, Project, and Format). This method is cheap and pragmatic—it helps reduce the workload on the people entering the content, keeps tagging costs low, and reduces staff resistance to tagging. Its flexibility can be increased by adding small dropdown picklists based on the identity of the person filling out the form, the business process in which the form is being used, or other simple contextual interactions.

**Complex Pattern Categorizers.** There are many ways to improve keyword matching. One way is to introduce scoring, instead of a simple match/no match decision. For example, the number of matches for each category within a content item can be counted, and the item assigned to the most frequently occurring category. Another enhancement is to weight the matches by where they appear and what is being matched. Matches in the title are normally weighted more heavily than matches in the body. Phrase matches are normally weighted more heavily than single word matches. Negative matches, known as exclusions, can be used to reduce false positives. For example, the “financial institutions” category might match “bank,” but not “river bank” or “Georges Bank.” Clues from the surrounding context can be used to boost or reduce the score of a match. Each category needs its own rules to describe the things that match the category. Those rules start very simple, frequently just the category name and a synonym or two. They are iteratively improved by analyzing ever-increasing amounts of content, tagging the content with rules, then looking for the false positives and false negatives. This can be quite a task if the taxonomy is large and/or the content uses a diverse vocabulary.

**Entity Extraction.** Entity extractors are software routines that scan the text and find mentions of entities such as people, places, organizations, and products, as well as addresses, dates, currency amounts, job titles, and topics. Extractors can assign a suggested category to these terms and make these terms available for consumption by a target application. Recognizing entities does not, by itself, categorize the content. Categorization software can, however, score the occurrences of the entities and assign categories based on that score, similarly to the way text matches can be weighted and scored.

Recognized entities can be more robust than simple text matches for several reasons. First, entity recognition does a better job of distinguishing company and personal names than simple text matching does. For example, is “Charles Schwab” a mention of a person or a company? Entity recognizers can use clues such as POS tagging of the content and patterns in the neighborhood of the text to identify noun phrases that are more likely to be a company name than a personal name. Second, entity recognizers typically make use of large authority files—lists of the names of people, places, organizations, products, and so forth which also include variations

on those names. Authority files provide patterns to be recognized that are not obvious variations of the full name of a company. In addition to complex patterns, entity recognizers also make use of heuristics (i.e., guesses). For example, a number followed by a proper name that ends in St., Dr., or Blvd. is probably a street address. A noun phrase ending in a gerund, such as Minnesota Mining and Manufacturing, is probably an organization name.

**Trained Categorizers.** Machine learning techniques can be applied when a collection of items has already been categorized. A collection of pre-categorized items is known as the training set. Using the training set, the categorizer “learns” the words that occur, and co-occur, in each category. Although it is called learning or training, what is really happening is that the algorithm is building lists of words and computing statistics about them. Once the algorithm has built up information on the relative frequencies of various words in the different categories, new documents have their word information compared to the data stored for each category. The document is then tagged with the category that is best matched.

**The Vector Space Model.** The Vector Space Model was first used by the SMART Retrieval System developed by Gerald Salton at Cornell University in the 1960s. In this model, each document is represented by a term vector (i.e., a column of numbers generated by analyzing the document’s content). When a new document needs to be categorized, one compares its vector to all the others and sees which one or ones it is closest to. Bayesian Classifiers, Support Vector Machines, Hidden Markov Models, and other advanced categorization techniques are all based on this foundation.

### *Text Mining*

While automated classification focuses on classifying content, text mining focuses on turning text into analytics that can be processed to identify patterns and trends. Text mining tasks may include categorization, entity extraction, summarization, taxonomy generation, and sentiment analysis. However, text mining focuses more on collection, than on the individual content item. This section presents some important uses of the output of automated classification in the context of text mining.

**Identifying Candidate Categories and Synonyms.** A simple form of text mining is keyword extraction, where a string of text is parsed by POS to identify nouns and noun phrases. A de-duplicated list of nouns and noun phrases, with the exception of a list of excluded words and phrases, is what is usually provided as a raw keyword extraction result. As described above, keywords are the raw material on which entity extractors work. After keyword extraction results are processed against an authority file, anything left that has not been matched can be considered candidate categories and/or synonyms to be reviewed by the authority file editor.

**Collection Analysis to Identify Trends and Anomalies.** The categorizations of content collections—whether by content type, function, or topic—tend to exhibit a Zipf distribution,<sup>20</sup> where roughly 80% of the content falls into 20% of the categories, and where there is a steep drop-off, called “the long tail,” which contains roughly 80% of the categories and 20% of the content. Any content collection can be analyzed to identify (1) which categories fall within which part of the Zipf distribution and (2) any divergences from the expected Zipf distribution. For example, the distribution of types of content should not be even across a collection—they should

---

<sup>20</sup>Zipf’s law states that the frequency of any word is inversely proportional to its frequency ranking. Thus the most frequent word will occur approximately twice as often as the second most frequent word, three times as often as the third most frequent word, etc.

follow a Zipf distribution. Some types of content will dominate, while other types will occur much less frequently, and categorization schemes should be designed and maintained to preserve this occurrence behavior.

From a classification management perspective, all types of classification schemes when applied to a collection should exhibit a Zipf distribution. If they do not exhibit the Zipf distribution, then the categorization scheme needs to be revised to split categories occurring too frequently and to merge categories that are too sparse in the collection. From a collection management perspective, the distribution of content in the collection should exhibit a Zipf distribution. If not, then content in over-represented categories should be reviewed for deletion, and content in sparsely represented categories should be targeted for acquisition or development of more content.

### Taxonomy Management Tools

Taxonomy management systems (TMS) enable an organization's users to view, apply, and modify a common set of controlled vocabulary lists to classify enterprise content. TMS tools are commonly associated with functions such as metadata management, indexing, and search.

TRB uses an in-house system developed specifically to maintain and publish the TRT. Reengineering the TRT requires more microthesaurus functionality, including mapping and publishing capabilities that support more complex and localized versions of the Thesaurus.

#### Core Functionality

Exhibit 1-20 provides a typology of taxonomy tool functions, which are further described in this section. A TMS generally provides the following basic activities around taxonomies and individual categories within them:

- Adding: The ability to easily add taxonomies and categories, including batch adding.
- Editing: The ability to easily edit taxonomies and categories, including batch editing.

**Exhibit 1-20. Typology of taxonomy tool functions.**

Functional Area	Functions
Taxonomy Development	Create a taxonomy User roles and permissions
Taxonomy Maintenance	Add, edit, move, delete items Assign or modify privileges to one or a group of items Activity logging
Taxonomy Governance	Approval workflow for additions and changes
Metadata Controlled Vocabulary	Assign attributes to a category Associate controlled vocabulary with metadata field Thesaurus capabilities
User Interface	Search and browse Drag and drop Multiple windows
Reporting	Alphabetical, hierarchical and other views Visualizations Importing and exporting taxonomies
Application Integration	APIs (WSDL, scripts, Java, etc.) Application integration (CMS, DMS, search engine, etc.)



- **Deleting:** The ability to easily delete taxonomies and categories, including batch deleting.
- **Mapping.** The ability to easily and automatically map taxonomies and categories from source taxonomies, such as topics from the TRT and a DOTs applications.
- **Importing/Exporting.** The ability to easily import and export taxonomies from and to other source taxonomies, including batch importing and exporting, while maintaining semantic and structural integrity. Common file formats such as database APIs and XML must also be supported.

Additional capabilities must include the ability to easily configure the software to work with common data structures (e.g., hierarchical and polyhierarchical taxonomies, controlled vocabularies, and bibliographic catalogs). A TMS tool should enable users to

- Store multiple taxonomies in a standard format, accessible through a common interface.
- Browse a taxonomy to explore categories and view information about classification terms, including definitions, relationships, explanations, and examples.
- Search the taxonomy and related vocabulary lists, for example, to identify the authorized term synonymous with the search term.
- Edit the properties of elements within the taxonomy, such as relationship types, facets, terms, and notes.
- Represent multidisciplinary through polyhierarchies with multiple “parent” relationships.

Other potentially valuable features in a TMS are

- Support for multiple simultaneous users.
- Role-based security (e.g., varying access control for taxonomy owners, authors, and general users). Roles must be well-defined and enforced within the TMS. Roles should include managers (approvers), curators (editors), and coordinators and support who help manage and stage workflow. End users must be able to submit change requests. Varying degrees of permissions for end users should also be possible.
- Audit trail and tracking of taxonomy editing and revision activities, as well as capabilities for analysts to resolve and reconcile multiple footnotes, allowing them to provide descriptions of data exceptions. This functionality must also include the ability to search time series and audit trails to find justifications for footnoting. There is also a requirement to represent taxonomy items as they existed in specific time periods and to be able to track, represent, and report their changes over time.
- Editorial support tools such as task prioritization because numerous change requests are anticipated.
- Integration with other workflow and collaboration software.
- Online collaboration for taxonomy creation and change (e.g., voting).

### ***TMS Benefits to DOTs***

An examination of the typical features of a TMS shows how such a system can benefit DOTs:

- **Central repository for the TRT or a successor global transportation taxonomy.** A TMS would constitute a single resource where all DOT data managers, analysts, librarians, and other users could find any transportation taxonomy and view that taxonomy, its history, and its relationship to other related taxonomies.
- **Version control and role-based permissions.** Any changes to a given taxonomy would be made only by an authorized taxonomy “custodian” assigned responsibility for maintaining the authoritative version of that taxonomy. When the custodian makes a change, the details of each change and the reasoning behind the change could be noted for future reference. Taxonomy custodians (who could be survey managers, publication managers, or dataset managers) would have local control over their own taxonomies, but all appropriate staff could access the taxonomies and use them in their own work products.



- Taxonomy harmonization and standards of practice. The TMS would require users to follow specific procedures to make modifications to a given taxonomy and could prompt taxonomy owners to make changes when necessary (e.g., when a taxonomy is deemed to be outdated). This would encourage greater rigor in taxonomy development and enhance the quality and consistency of survey taxonomies. The TMS can also link local taxonomies and taxonomy practices to the systems architecture and information management policies of the entire enterprise. For example, FHWA could implement a TMS to ensure compliance with overall information and data policies.
- Taxonomy crosswalks and comparisons. The TMS would store crosswalks between taxonomies, which would be especially useful to DOT users who need to integrate multiple datasets (for analysis and data integration). Robust TMS tools issue each taxonomy category a unique identifier and preserve that unique identifier to allow for comparisons between taxonomies (e.g., tracking terms with identical labels but from different taxonomy sources).
- Voting and collaborative editing. In cases where changes to a taxonomy would affect multiple DOT stakeholders, the TMS could provide a platform for facilitating a collective decision on those changes.

### *Taxonomy Tools Vendors*

Taxonomy management systems constitute a very small market within the enterprise software environment and are not covered by any technical analysts such as Forrester, Gartner, or IDC. A few TMS vendors compete for a very small market share (estimated to be between \$16.5 and \$66M in 2012).

## **Conclusions**

Part 1 has summarized the current transportation information management landscape in the context of broader information management trends and developments. Part 1 provides many examples of practices that may be adapted in order to improve DOT management of transportation information (e.g., guidance for file formats, naming conventions, and information preservation strategies). We have discussed the convergence of data and content in this landscape, manifested most visibly in the combination of data and content from multiple sources on the web. Although data.gov is leading to wider publication and access to raw datasets, a rich categorization scheme such as the TRT that gathers and maps the semantics of transportation information should facilitate the interchange, combination, and use of heterogeneous sources. The semantic strengths of the current TRT should be used and the weaknesses addressed to make the TRT more flexible and easy to use for DOTs. The primary recommendation is to develop a way to enable DOTs to identify and generate TRT subsets or microthesauri easily. At the same time, DOTs need incentives to be more engaged in the TRT governance.

# Studies of Leading Practices

Part 2 describes the survey approach and summarizes the major research findings from the survey; (2) provides case studies and examples and discusses good practices related to policy, practices, tools, and outreach; and (3) summarizes cloud computing and states' activities in adopting cloud computing policies and installations.

The research entailed a review of current transportation information management practices. The review included a survey and personal interviews. The research team gave particular attention to cloud-based storage because that is an emerging trend that the team believes will have important consequences for practitioners.

## Survey

The objective of the project's information management survey was to identify and understand the technical, institutional, and governance challenges faced by a sample of DOT information managers—including information management and the data business planning functions—in managing transportation information. The primary tools for identifying and understanding these challenges were an online survey, supplemented by in-person and/or telephone-based follow-up interviews as appropriate.

The survey was designed to do the following:

- Evaluate the current state of practice among DOTs and their information management professionals.
- Assess the practices each DOT uses to store, organize, and retrieve transportation information.
- Identify factors that enhance or impede the user acceptance of the current tools and methods for accessing and using information.
- Collect information to permit classification of the information governance structure of the DOTs.
- Examine the technical and institutional challenges and the extent of coordination between DOT information management and data business planning functions.
- Investigate funding, institutional, governance, and other factors that drive overall effectiveness in information management.

The survey was distributed electronically to 197 individuals representing DOTs, commonwealths/territories, national agencies, academia, local agencies, and the private sector. Responses were received from 144 individuals (73% response rate). Of these, 117 individuals completed the survey fully, while 27 completed the survey partially. Fifty-three individuals opted out or did not respond.

## Follow-Up Interviews with Survey Participants

Follow-up interviews were conducted with select survey participants in March 2012. The list of participants for the follow-up survey was divided into three categories: (1) those who responded to the survey and indicated that they would like to be contacted for further information (42 respondents); (2) those who responded to the survey but did not indicate they were willing to be contacted (15 respondents); and (3) those who did not answer the survey (6 respondents). Emails were sent to all of them thanking them for their participation, describing the purpose of the study, and requesting information on what their state is doing relative to information/data management in the areas of (a) governance, (b) capture, (c) preservation, (d) search, (e) retrieval, and (f) making the information usable. Participants were also asked how guidance in these areas would help them.

The combination of survey results with telephone interviews resulted in a thorough understanding of the overall status of information management at DOTs, detailed examples of good practices in many areas, and insight regarding desired guidance. Analysis of the survey and interview results is provided in the following sections.

## Assessment of DOT Practices

Based on the results of the online survey and follow-up interviews with DOT professionals responsible for managing transportation information, an assessment was conducted to identify the current state of DOT policies and practices regarding the following:

- How governance and technology support sharing of transportation information, including some of the technical, institutional, and governance challenges faced by a representative sample of DOT information managers responsible for managing transportation information;
- Methods/tools used for capture, preservation, search, and retrieval of such information; and
- Making transportation information findable and accessible to information users, including identifying factors that drive overall effectiveness in information management.

The following research was conducted to complete the assessment of DOT practices:

1. In-depth analysis of the survey
2. Use of results of interviews with panel members
3. Follow-up interviews conducted with survey respondents

### *Governance and Technology*

**State of the Practice.** Only 26% of DOT respondents reported that their agency has a formal governance structure that clearly defines responsibility for information management and sharing, while 40% reported they do not. (Q16c)

Less than half of respondents from DOTs (40%) believe departments within their agency have clearly defined roles for ownership, quality assurance, and maintenance of data and information, while 37% disagreed. Respondents whose primary role is in GIS and information technology/systems reported that departments do not have clearly defined roles for data governance (50% and 60%, respectively). (Q12a)

Most respondents from DOTs believe their organization is effective at information management (61%) and information sharing (56%). (Q6, Q8)

Institutional issues were cited by respondents as the biggest challenge to information management, particularly in the areas of governance (31%) and resources (29%). Both technical

and institutional issues were cited as the biggest challenge to information sharing, particularly in the areas of standards (36%) and institutional resources (26%). (Q7, Q9)

Departmental responsibility for information sharing appears to be less clearly defined across DOTs than that for information management. (Q12b, Q12c)

Financial support for information management and sharing appears to be a significant issue for many DOTs. Respondents whose primary role is in information technology/systems reported the greatest confidence in the financial support for their department, while respondents in library services reported the least. (Q12d)

Respondents whose primary role is in information technology/systems reported having the most use to bring about improvements in information management and sharing, while respondents in library services reported the least. (Q12e)

Only 36% of DOT respondents believed their library, data management, and GIS groups work effectively together, while 48% believe they do not. Respondents whose primary role is in information technology/systems reported the greatest confidence in these groups working effectively together (60% agreed and 40% disagreed), while respondents whose primary role is in library services reported the least confidence (8% agreed and 84% disagreed). (Q12f)

**Assessment.** Although some DOTs have implemented formal data governance programs, many are still struggling in this area. According to participants, executive management often lacks understanding of the importance of information management, not only the collection of information (e.g., research results, technical reports, and publications), but (1) the management and stewardship of the information, (i.e., making it findable and usable—arranging, indexing, cataloging—storing and retaining), and (2) a recognition of the value of historical research, versions, and editions of materials.

Many DOTs continue to operate legacy systems that are not interoperable with other systems, essentially creating data silos that require redundancy in data storage and data entry. Middleware and modernization are providing some relief from this issue, but without a fully interoperable system, data is not being governed as effectively as it could be. For some, different sections of the organization are creating/maintaining their own information resources but not necessarily coordinating or sharing the information with others.

Organizational changes are presenting additional challenges to governance. When an organization is constantly changing and being reorganized, the organization charts are never current or complete (i.e., they do not cover all units, bureaus, and people); hence, it can be difficult for staff to know who is responsible for what.

Data business planning is another area for improvement. This process helps DOTs develop a clear plan for information management by defining current business processes, information products, and user needs related to information management and sharing.

Many agencies are having challenges with their IT department being an obstacle to progress and innovation. In the online survey, only 29% of DOT respondents believed their IT department could be perceived as a positive partner in promoting information management and sharing, while 43% responded that they are not. Further investigation uncovered a sentiment that the IT department, whose existence relates to supporting the business units, instead is perceived as requiring groups to adopt business work flows which may be contrary to GIS good practices. One respondent noted that, too often, information *technology* management is being confused with information *content* management. Or, if this is understood, the end user or creator of information is tasked with guiding information content management, but they, too, often lack real understanding of how to effectively find management information for long-term retrievability. Until there is some fundamental understanding of the need for better

information practices and acknowledgment that competencies are needed to do this well, DOTs will lack resources and have limited improvements.

### *Methods/Tools for Capture, Preservation, Search, and Retrieval*

**State of the Practice.** Types of information managed and shared most effectively at DOTs are traffic data (90% of respondents), GIS data (84% of respondents), and research reports (83% of respondents). Types of information managed and shared least effectively are asset inventory databases (23% of respondents) and performance reports (13% of respondents). (Q10)

Nineteen percent of DOT respondents reported having a structured data management program in place, and 19% have one in development. Most existing programs tend to be formal bodies that establish policies and procedures, while most programs in development are committees that coordinate data collection and dissemination practices. (Q17, Q18)

Only 2% of DOT respondents reported having a formal policy on digital preservation based on the Open Archival Information System (OAIS), while 17% have a policy not based on OAIS. Most respondents indicated that they have limited resources (technical, manpower, financial) to fully implement this policy. (Q25, Q26)

The three most common technical standards in place at DOTs appear to be: (1) a document management system (88%); (2) an image management system (49%); and (3) effective publication standards and templates (43%). Clear metadata requirements and technical standards for keywords used on Internet pages are not as common, but are still used by roughly one-third of all respondents. (Q28)

Most DOT respondents (52%) reported their agency effectively inventories and controls access to information available only in print form, while 21% reported they do not. Most respondents (77%) reported that additional funding is critically needed for digital conversion and preservation. (Q29)

**Assessment.** Success in data capture appears to vary across DOTs. States using enterprise class technology reported having good success in capturing data and reducing the number of silo databases/applications within separate offices. For other states (e.g., California), it is often a challenge to obtain and collect information from other offices. The Caltrans Library and History Center is working with the Deputy Director for Administration to develop a communication plan for the library. They are trying to come up with a directive from the Chief Deputy Director that says that the various offices shall provide information to the library, via email, if a particular office publishes a document. At a minimum, that office should send an email or information about a web link to published documents, so that this information can be made available at the library. Other DOTs struggle with data not being updated in a timely fashion and with bad data being loaded. These issues prevent DOTs from providing executive-level reporting. Another challenge is locating and keeping track of where information is housed, particularly in terms of historical information that has been updated or otherwise revised over time.

Many states appear to have effective methods/tools for preserving data/information. Oracle databases, Microsoft SharePoint, and Microstation Project Wise appear to be most commonly used as document management tools for storing and preserving electronic documents. These tools provide for change control on documents and engineering diagrams, security controls, and attaching metadata to documents so that they can be queried and easily searched. For Minnesota, the use of Project Wise was one of the first real successes in the area of data governance, in that it helped to implement CAD standards in which users were required to put data in a specified CAD file layer. Some states (e.g., Mississippi) noted that, although SharePoint has been an effective document management system, there were problems associated with implementation because the entire system was implemented at one time. It would have

been preferable to prioritize implementation of the system, by implementing higher priority components first and other components later.

The survey results indicate that many DOTs are seeking guidelines on organization of data and standardized data formats to facilitate search and retrieval. Several states (e.g., Virginia and Maryland) reported that while sharing data within their own group is effective, users in other groups have to know who is responsible for the data to find it. Some states (e.g., Minnesota) noted that they have challenges in providing ways for staff or others to search for data or other information. There is no single portal or method to do this now. They have information available through their website, but most users do not know how to access or get information from their data warehouse. Access requires assistance from MnDOT staff responsible for the various databases.

For states using SharePoint, Microsoft Search and Microsoft FAST search are often used for searching SharePoint archives. Searches are also done using ApplicationXtender. Searches on an agency's intranet site can also be done within SharePoint if that is where the intranet site resides.

### *Making Information Findable and Accessible*

**State of the Practice.** Most DOT respondents reported that their agencies work actively to ensure digital publications are available through the Transport Research International Documentation (TRID) Database (63%); work actively to ensure online information resources are findable via the TRID Database (53%); make information submitted through the TRID Database also available through the National Transportation Library (NTL) (61%); and make regular use of the TRID Database to conduct research and find reports (69%). (Q4)

More than one-half of DOT respondents reported that they have some degree of knowledge and expertise in the following: the TRT (52%); TKNs (58%); and the TRID Database (61%). Fewer respondents reported that they have knowledge and expertise in the National Information Exchange Model (NIEM) (19%) and Information Exchange Package Documents (IEPD) (15%). (Q22)

Only 5% of DOT respondents reported that their agencies routinely register map services, images, or geographic data sets with Geospatial One-Stop, while 35% reported that they do not. (Q27)

**Assessment.** One of the biggest challenges in managing information is making resources findable and accessible to users. This task is often assigned to the library services division of an agency. Although some states are effective at collecting and distributing information in a consistent, organized, and useful manner, many others reported that they lack the procedures, standards, and tools to do this. Other agencies have effective tools in place, but lack of user training on their information systems hinders discovery and simple data retrieval.

For many, a lack of resources (e.g., staff, time, funding) is the problem. Often, there is only one librarian or information manager who has to catalog, shelve, and provide reference for the entire agency. There are an overwhelming number of sources of information—particularly online—and it is impossible to read everything and decide which ones are the most relevant and valuable and should be shared with colleagues and customers. However, if agencies simply forward all information apparently related to a key individual's role, much of it will not be useful. Eventually that individual will not value what is sent. Some states (e.g., Utah) have resolved this by implementing a new template to differentiate between "general" versus "specific" information when pushing information to recipients.

Organization of information also appears to be a major challenge for many DOTs. Several agencies reported that they have a collection of information (e.g., books and reports) that is



not organized in any way, although they are cataloging this information to have better access. Many reported that they are uncertain about the new Resource Description and Access (RDA) cataloging rules. Based on the research, a taxonomy and recommendations for organization structure of information are needed. States also cited a need for a roadmap on how to better use library science resources to organize information.

Outward sharing of information is often made difficult by the findability of reports/publications on an agency's website. States cited a need for guidelines on how to better organize their websites and better key words for tagging information to ensure findability. However, responsibility for an agency's website often falls outside the library science realm, so librarians are limited in providing input or making significant improvements. There is also a lack of direction from executive management on how to treat information or manage it as the asset it is.

Outward sharing of information is also challenging when ineffective data management practices are in place. Agencies are seeking guidance in the form of a tool kit or checklist format to know where to post research documents regularly. One agency representative noted that, given the many silos of information created within his agency, it is challenging to ensure that information is captured and submitted to the TRID Database and NTL. The agency representative has to seek out the information, which is time consuming, and he receives the information through mechanisms such as newsletters and email distribution, which is inconsistent and not comprehensive.

## Case Studies and Other Examples

The case studies presented herein illustrate recommended good practices in managing transportation information at DOTs. Good practices are defined as those that incorporate some or all of the information management process (i.e., capture, administer, and retrieve) to manage transportation information.

The case studies reflect interviews with the states. The terminology used to describe the information management processes in the states is consistent with use of the terms in the individual states and may not be consistent with those used in other parts of this report. For quick reference, Exhibit 2-1 summarizes lessons learned from the case studies and examples.

**Exhibit 2-1. Case study agencies.**

Agency/ Contact Person/ Focus Area-Key Themes	Lessons Learned
ADOT & PF Jack Stickel DBP	<p>Use of a DBP helps an agency to</p> <ul style="list-style-type: none"> <li>• Establish goals for data programs</li> <li>• Assess agency data programs</li> <li>• Establish a data governance framework</li> <li>• Ensure proper use of technology/tools to support data management</li> </ul> <p>Once implemented, a DBP needs to be reviewed regularly (at least annually) and updated to reflect changes in policies, standards, business processes, technology/tools, or organizational realignments that may affect management of data programs and information at the DOT.</p> <p>Establishing Communities of Interest (COIs) (for transportation data programs) and continuing outreach to the COIs helps the DOT to make appropriate decisions on developing/implementing new hardware/software solutions to support data programs relied on by the user community.</p> <p>Use of a DBP helps everyone in the DOT to understand their roles in providing high-quality data and information in a timely and efficient manner for all stakeholders.</p>



## Exhibit 2-1. (Continued).

Agency/ Contact Person/ Focus Area-Key Themes	Lessons Learned
<b>AZDOT</b> Dale Steele Digital Collection and Arizona Records Retention schedule	<p>Market the type of information available from the source agency to the widest audience possible, using newsletters, fliers, meetings, and the Internet.</p> <p>Maintain a central location or web portal as a shared space to store multiple types of information provided from many sources. A single portal with easy access enables quick search and retrieval of information by the user community.</p> <p>Use WCM software to document statistics on the usage of the website (daily, weekly, monthly, annually). This enables the sponsoring agency to adjust the content displayed on the website as needed. In the case of transportation information, this is particularly useful for supporting emergency management operations by providing additional information on evacuation routes or road closures during times of emergencies.</p> <p>Policies regarding preservation of records in paper or electronic format should be instituted by law and/or supplemented with department policies at the DOT. Such laws or department regulations for the proper retention and/or disposal of records is critical in defending the agency against charges of spoliation or tampering with evidence in the case of litigation.</p> <p>Policies and laws regarding retention/disposal of records should be made available to all DOT employees who may have responsibility for managing records such as reports, documents, photos, digital images, and videos that may contain information that can be used in litigation or other types of inquiries.</p> <p>As federal or state laws or department policies governing such procedures change, the records retention/disposition procedures used at the DOT should be updated as needed and re-distributed to the appropriate department personnel.</p>
<b>Caltrans</b> Coco Briseno Janet Coles Harold Feinberg DBP; GIS	<p>Understand the business process and prepare systems based on data-driven decision making;</p> <p>Embrace IT processes and procedures and work in collaboration with IT; and</p> <p>Develop information management systems that are as user friendly as possible.</p>
<b>CDOT</b> William Johnson Data Governance	<p>Following a multi-step approach for implementing governance similar to the approach at CDOT ensures that assessments of existing programs are completed and needs and gaps are identified which can be addressed as part of a DBP using data governance.</p> <p>As part of the data governance framework, CDOT also identified good practices in the use of technology/tools used to display data and information pertaining to performance measures.</p> <p>Data governance at CDOT provided a way to coordinate with a large number of stakeholders through the use of stakeholder groups organized around specific data programs used to support performance measures.</p>
<b>GDOT</b> Jane Smith Teague Buchanan Data Governance and MS SharePoint	<p>Implementing data governance has helped GDOT establish policies, roles, and responsibilities, which can be clearly communicated to all staff responsible for managing data and information at the department.</p> <p>GDOT is benefiting from the partnering of business units with the IT office through the Data Resource Management Steering Group. This group is empowered to make decisions about changes in information systems at the DOT. This helps ensure that business needs are considered as part of any enhancements to application systems.</p> <p>Software development time has been drastically reduced because of the use of BI tools within a GIS environment.</p> <p>Use of BI tools also helps to ensure that critical documents are secured and properly stored in systems such as SharePoint and Project Wise, with metadata and tags to identify the purpose of the documents. This helps ensure that the correct information can be located to respond to requests.</p>
<b>IDOT</b> Karen Perrin	<p>Participate in knowledge sharing groups where resources and information are shared among many, and the overall collective intelligence of the group is higher than any individual on his/her own.</p> <p>Always have an elevator speech ready and find someone in upper-level management to understand the work and become a champion of information management. Convincing someone to work toward an information management governance policy is a huge step for the agency.</p> <p>The issue of funding is also important. Designation of some funds to address electronic file management issues is an important step toward meeting information management needs of the future.</p>

(continued on next page)

**Exhibit 2-1. (Continued).**

<b>Agency/ Contact Person/ Focus Area-Key Themes</b>	<b>Lessons Learned</b>
<b>Iowa DOT</b> Hank Zeletal data collection guide Image indexing and historical images - SharePoint	<p>Automated indexing software for cataloging digital photos, images, or videos, such as that used at Iowa DOT, can be implemented at other DOTs to eliminate time-consuming, error-prone manual processes associated with indexing.</p> <p>Enhancements can also be made to the indexing software as business needs dictate. These include the need to distinguish between images and objects which appear to be trucks or cars or pedestrians.</p> <p>Automated indexing software can also be used to provide easier access to a larger user base seeking photo and video information. The indexing information about photos and videos can be transmitted to other enterprise records management systems to facilitate retrieval of the photos and videos from their original repositories.</p>
<b>KDOT</b> KTOC – Facebook, Videostreaming, etc.	<p>DOTs can use social media venues such as Twitter, YouTube, Facebook, Flickr, and blogs to provide information to a wider audience interested in transportation-related information. This audience may include the general public or stakeholders who use the data and information in their daily business operations.</p>
<b>LaDOTD</b> Jim Mitchell Enterprise GIS system	<p>Using BI tools such as GIS applications can provide multiple types of integrated transportation data in a map-based visual format, to a wide audience (using a central database repository for geospatial data).</p> <p>Use of GIS applications requires that the user community have access to the technology tools (such as ArcView or ArcGIS Explorer) needed to query, retrieve, and analyze the data. For more experienced users, an enterprise GIS system can be used to integrate data layers from other areas in a DOT to support typical DOT operations, including asset management, emergency management, travel monitoring, and safety. Use of GIS also offers the capability to perform a visual QA/QC on data and correct any errors as needed.</p>
<b>Maine DOT</b> Nate Kane GIS is primary tool used to manage and access location info	<p>Gaining buy-in from external GIS data providers is essential for maintaining an up-to-date geospatial database at the DOT. Allowing access to the enterprise GIS to all providers of the data layers helps to ensure continued support of the GIS for all users.</p> <p>The persons/offices responsible for developing and maintaining GIS systems at the DOT need to continue to market the benefits of this technology to internal and external customers, including the legislature.</p> <p>Identifying champions at the Executive level and in business units to support the continued enhancement and use of GIS technology within the organization is critical for sustaining an agency's ability to integrate multiple types of data.</p>
<b>MnDOT</b> Cassandra Isackson	<p>Use governance with flexibility.</p> <p>Integrate governance with other initiatives. This helps the business offices/units realize how exercise of good data and information management practices can safeguard the agency against potential risks, not just those identified by a particular business area.</p> <p>Use a balanced approach for governance with a bottom-up and top-down approach to help convince business areas of governance needs in support of overall agency strategic needs.</p> <p>Business offices/units must lead; business must put resources toward good data management practices and they should also hold other business areas accountable; data management is not IT's responsibility.</p> <p>Focus on data for decisions at all levels; ensure that data is useful and usable; remember the purpose of data and for what it is used.</p> <p>Establish standards that are nimble.</p> <p>Establish policies that allow for transparency and innovation in information management.</p> <p>Data management is not the end, it is the means to other ends (e.g., providing information needed by the public, policy makers, and snowplow drivers).</p> <p>Establishing governance takes time, but, it is worth the effort.</p> <p>One does not always need more data, just better access, better tools, better understanding.</p>
<b>MDOT</b> Imad A. Aleithawe Denise Jones Ray Barksdale SharePoint	<p>Information management will be ongoing and may never be fully complete.</p> <p>Using a more agile project management method is more beneficial than the standard project management method. One must be able to adapt and produce solutions that best fit the agency.</p> <p>It is also important to work at a manageable pace when working with the agency's end users and business owners.</p>

## Exhibit 2-1. (Continued).

Agency/ Contact Person/ Focus Area-Key Themes	Lessons Learned
<b>NDOT</b> Sena Loyd EOS.Web Express	<p>Select an information management technology solution the right fit for the agency or organization. In the case of the Nevada DOT library management system, an enterprisewide solution was not needed. The DOT was able to implement a cost-effective solution using EOS.Web that provided the functionality needed for managing the library's collection of transportation materials.</p> <p>Use other repositories and archives (outside the sponsoring agency) as available for retention of historically significant or otherwise important documents, reports, maps, photos, digital images to supplement the collections maintained at the DOT's library and/or business units.</p>
<b>NYS DOT</b> Governance process specific to their IT department	<p>Establishing a well-defined governance structure for IT empowers an agency to make appropriate decisions regarding current and future technology investments in support of agency goals and objectives.</p> <p>Documenting the governance policies in a format such as a Strategic Plan (e.g., IT Strategic Plan) provides a platform to educate all persons/offices on the expectations and goals and objectives for use of governance within the agency.</p> <p>Establishing mission and vision statements for governance at the agency helps to set forth the direction for current and future policies, practices, and standards to be used in managing data programs to support business needs. Mission and vision statements <i>for data programs</i> are separate from agency mission and vision statements.</p>
<b>NCDOT</b> Dr. Mrinmay Biswas	<p>Use all sources available for adding to the collection of transportation-related materials (reports, documents, maps, photos, digital images, videos, etc.) maintained and managed by the DOT. This includes providing links to external sources of information or purchasing additional collections of materials as needed. The sources of additional information and materials may be provided and/or used by the DOT's divisions, offices, and user communities, such as MPOs, transit authorities, law enforcement, and other state and local transportation agencies.</p> <p>Enter into sharing and exchange of information agreements or arrangements with state and national organizations to obtain copies of transportation-related publications, which will enhance the collection of materials maintained at the DOT.</p> <p>DOTs should consider the use of newer available technologies for cataloging and indexing of transportation information maintained by the libraries at the DOTs.</p> <p>EDM systems such as SharePoint are an option that can be considered for establishing a central repository of information, including documents related to transportation information. Several DOTs, including North Carolina (future) and Virginia (now) are using or plan to use SharePoint to facilitate management of documents and information.</p> <p>Electronic Document Management Systems (EDMS) facilitate easy querying and access to information through categorization and taxonomy schemes within the EDMS.</p>
<b>ODOT</b> Dave Gardner IT Governance Council	<p>Implementing a Governance Council or Board helps the organization designate the persons/offices responsible for establishing policies, procedures, and standards for the collection and use of data and information within the organization.</p> <p>The Governance Council is a high-ranking authority with responsibility to oversee any initiatives to develop or enhance existing application systems, to meet the needs of the organization, with consideration of the overall goals and objectives of the agency.</p> <p>Establishing clear roles and responsibilities for data governance within the organization helps everyone to understand how their job contributes to the efficient management of data and information.</p> <p>Responsibilities for governance should be shared between the IT office and the business units. The business units should retain responsibility for management of specific data and information needed to support their business needs and the IT office should ensure that hardware and software infrastructure is in place to support the business application systems.</p>
<b>UDOT</b> John Thomas Cameron Kergaye UPlan Library functions	<p>Providing access to integrated datasets through a single portal shortens the time to evolve data into information and ultimately into knowledge for use in decision making.</p> <p>Integration of data in a similar manner illustrated in uPlan can result in higher benefits/ lower costs by eliminating redundant data collection processes and saving time in data analysis by using GIS tools for geospatial analysis.</p> <p>Use of integrated data helps facilitate cooperative working relationships among departments and improves understanding of issues and needs regarding data across the department.</p> <p>UPlan provides a mechanism for one office to demonstrate a particular need or issue to the other departments by displaying a particular type of data using the uPlan GIS tools.</p>

(continued on next page)

**Exhibit 2-1. (Continued).**

Agency/ Contact Person/ Focus Area-Key Themes	Lessons Learned
	Libraries in DOTs can use various methods to market their services to the internal and external user community and promote the types of collections/information available through the library. The methods used for marketing services can include newsletters, Internet advertisements, and social media options including YouTube, Twitter, Facebook, etc.
<b>VDOT</b> Ken Winters Maureen Hammer	Capture lessons learned from previous (project) experiences. Document information from previous projects and make it available across the agency (such as through the use of a KM system). Use consistent indexing terms for quickly locating and accessing the repository of lessons learned. Do not try to do everything at once. Focus on what services can be provided by each specific office responsible for information management whether it is the library at the DOT, the KM office, or perhaps the Administrative Services Office.
<b>WSDOT</b> Leni Oman Gordon Kennedy Kathy Szolomayer Andy Everett	Understand the organization's business needs for data and information. Have the messy conversations to bridge gaps in awareness and understanding amongst managers of information resource types and between business units and information management professionals. Understand that content is different from the technology system that it resides in. The information content may be used by business units other than the primary creators/users. The content may also be needed again decades later. Managing content so that it can address this business need requires attention. Library science has demonstrated successful strategies for managing content across diverse user groups and extended time periods. We all create and use information daily. We are all responsible for managing our portion of information responsibly. Having clear principles and intent and strategy to carry them out is critical to success. It is beneficial to have staff resources and knowledge that considers the big picture for information management. It is also valuable to have staff who can investigate experiences, guidance, and standards associated with information management to determine usefulness and usability. It will only get worse if we wait.
<b>WisDOT</b> John Cherney	Developing similar data management tools and allowing shared access to these tools helps the DOT and the local governments to analyze and validate road inventory data needed to meet federal and state reporting requirements. A shared data management application such as WISLR allows flexibility and local control of data reported back to WisDOT from the local governments. An application such as WISLR provides a single portal for query, analysis, and update of road inventory data, making it easier for all users of the system.

**Case Study: California DOT (Caltrans)**

Primary Sources: Caltrans staff in the GIS Services Branch, Office of Data Analysis and Geospatial Information Systems (ODAGIS), Division of Transportation System Information; Library and History Center, Office of Business Services and Security; and Office of CADD and Engineering GIS Support, Document Retrieval System (DRS) Administration. The case study discussions focus on functional units such as the Caltrans Transportation Library and History Center, GIS, and the Document Retrieval System (DRS).

## Background

**Caltrans Transportation Library and History Center.** In the 1960s, the U.S. Bureau of Public Roads Highway Planning Program Manual was widely used as a guide for state highway agency policy and procedure development. The manual's section on reference libraries recommended the establishment of a centralized reference library at state highway agency headquarters so that basic reference materials in subject areas related to project planning and delivery would be readily available and accessible. As a result of this and to eliminate redundancy and provide broader professional library reference and research services, Caltrans' current Transportation Library and History Center was established in 1971.

The original goals for the Transportation Library included development of departmentwide library standards; development of departmentwide uniformity in administration, policy, and procedure; integration of the computerized information library services; maintenance of the selective distribution list of publications; increased use of the thesaurus terms for data storage retrieval; expanded library services; central control over publications and other printed materials; and improved level of service.<sup>21</sup>

In 1984, a formal program to preserve the department's important historical materials was established (Policy and Procedure 84-5, now Deputy Directive 58<sup>22</sup>) thus significantly expanding the scope of the library through the addition of a special collection of research materials covering the history of Caltrans and the development of the state highway system.

The Caltrans library is well connected to the NTL community and participates in national efforts to develop better transportation information systems and tools.

The Caltrans library provides many services, designed to meet departmental research needs, to its patrons. An online library catalog and digital library system are maintained in addition to an intranet site with information and links to many transportation systems.

**Caltrans GIS.** Caltrans' Division of Transportation System Information is considered the "Pathway to Information"<sup>23</sup> and includes the offices of Data Analysis and GIS (ODAGIS), Office of Highway System Information and Performance (OHSIP), and Office of Travel Forecasting and Analysis (OTFA). From the GIS perspective, information management became a priority in the early 1990s. This has evolved over time to the GIS web services available on line today.

Deputy Directive 72 established a vision for GIS within Caltrans to fully support GIS development and application and promote its integration into business processes, analysis, and information flows. Directive 72 also contains responsibilities for the various stakeholders, including upper levels of management (e.g., Deputy Directors, Chiefs, a GIS Management Committee (GMC), and District Directors).

**Caltrans DRS.** The Office of CADD and Engineering GIS Support uses state-of-the-art computers and software to share information across functional units and is responsible for the DRS. The DRS is used to archive and access selected records that serve as evidence of the organization, functions, policies, decisions, procedures, operations, or other activities of the department, including project reports, as-built plans, and survey maps.

## General Policies and Practices

Caltrans has several policies and practices related to information management—Deputy Directive DD-58-R1, Preservation of Department's Historical Materials; Archiving As-Built

<sup>21</sup>Coles, Janet. Library History. California Department of Transportation.

<sup>22</sup>California Department of Transportation. *Deputy Directive DD-58-R1, Presentation of Department's Historical Materials*. February 8, 2010.

<sup>23</sup>Division of Transportation System Information (TSI), <http://www.dot.ca.gov/hq/tsip/faqs.php>, accessed October 18, 2011.

Plans Memorandum, April 2006; Deputy Directive 101, Records Management; Deputy Directive DD-55-R2, Management of Caltrans Information Assets and Records; and Deputy Directive DD-72, Geographic Information System (GIS) Implementation. The policies generally include Deputy Directives and contain the policy, background, responsibilities, and applicability. Several policies and practices are also in the form of memoranda or incorporated into Caltrans manuals.

### *Good Practices*

**Capture.** Good practices for this phase of the information management process are broken down by operational unit within Caltrans.

***Caltrans Transportation Library and History Center.*** The Library and History Center frequently reviews internal and external web links to look for major publications and reports that would be useful for the various divisions/offices within the department. The Caltrans Library and History Center has well developed relationships with other departmental offices including Planning, Environment, and Cultural Resources, to gather information from them for distribution through the library.

It is often a challenge to obtain and collect information from the various offices because the library relies on informal relationships. The library is working with the administration to try to come up with a directive stating that the various offices shall provide information to the library, via email, if a particular office issues a publication.

***Caltrans GIS.*** Caltrans performs a significant amount of geospatial information collection and capture, with much of the data capture being project based. Linear referencing is mostly manual at this point. Staff are beginning to use mobile terrestrial laser scanning Light Detection and Ranging (LiDAR) for surveys, in combination with video, for operations and for extracting asset information.

On the back end of *Caltrans Earth*, their implementation of Google Earth Enterprise (GEE), Caltrans developed some automated data collection/capture processes that harvest transportation assets and other data from standard databases, convert the data to GIS formats, and publish the data to Caltrans Earth via GIS web services, and make the same data available to the GIS Data Library.

**Administer.** Good practices for this phase of the information management process are broken down by operational unit within Caltrans.

***Caltrans Transportation Library and History Center.*** The Caltrans Library uses an integrated library system—SydneyPlus—to manage information and serve as a repository (store publications and photographs). Historical digital photo collections are managed via CONTENTdm®. A data management plan—the Collection Development Plan—is available related to the library.

Caltrans Transportation Library and History Center and U.C. Berkeley's Institute of Transportation Studies Library form an integrated library system that can store information, including agency publications. The NTL is used for storage and TRID for indexing. The Institute of Transportation Studies Library at the University of California, Berkeley, is responsible for submitting records of Caltrans reports to TRID, including indexing reports with the TRT. The Berkeley library also deposits electronic copies of reports in institutional repositories. Although some reports have been sent to the NTL, electronic copies of all reports have been submitted to the University of California's institutional repository, Merritt, for the past 2 years. Historical digital photo collections for the Caltrans districts are stored and managed via a Digital Collections Management System (DCMS), CONTENTdm®.

The Caltrans library provides reference and research services and preserves important current and historical departmental information (e.g., publications, documents, maps, photographs,



and artifacts). This is critical to several functions and departmental staff, including project planning and delivery, the legal division, and public affairs. The History Preservation group has a policy that they have the opportunity to review any documents that staff wants to destroy, particularly because Caltrans is still strongly paper focused. In the future, the group wants to scan and archive project documents, photos, and change records, using the capabilities of new technologies. This will result in a paradigm shift.

The Caltrans website and internal intranet site include links to various tools, standards, and publications. A wealth of information, with intuitive web interfaces, is available to Caltrans staff and the public. The library catalog is not available on the web because some of the information is sensitive. Intellectual property protection is also a consideration.

**Caltrans GIS.** Much of Caltrans' data has a spatial component (e.g., post-mile, coordinate) so staff are working on a GIS interface to the DRS, as well as GIS improvements to other departmental data management systems. Specifically, management plans would be an outcome of the spatial data infrastructure project. For 10 to 15 years, much of the GIS information has been maintained in a central GIS library in Oracle/ArcSDE. Information can be accessed and viewed via GIS software. Caltrans Earth provides access for the public and internal use. Linear referencing data is stored in a FileMaker Pro database. FME records change based on linear referencing information updates. Caltrans' recent program review reports that information management was an element ranked high for needed improvements. Moving forward, Caltrans is looking at web services, development of standard formats, and use of the cloud for GIS information. A focus on the use of the cloud is expected to help enforce consistency and data standards. The core GIS Data Library is maintained and backed up regularly. Preservation of data in central databases supported and maintained with the other departmental information assets is a significant challenge given the diffuse nature of GIS usage in Caltrans. Preservation of temporal transportation data is being addressed on the back end in databases, but a full selection of tools for easy retrieval of temporal data has yet to be implemented.

Caltrans has begun using more web server and visualization-based tools with open standards for data, replacing existing file-based data distribution and vendor-specific server GIS platforms. The data and imagery collected and presented in Caltrans Earth can be retrieved for use in both traditional GIS analysis and for geo-visualization (e.g., Google Earth & Maps or ESRI platforms) and reference. Caltrans implemented Caltrans Earth as a place to display data in a format easy to use and understandable by non-GIS trained staff. Furthermore, this tool is available both internally and externally to Caltrans partners and the public. The goal is to get access to GIS information users directly to meet their needs. There is limited security on GIS data because most GIS data is not sensitive. Caltrans is starting to work with the State's Geospatial Information Officer (GIO) to build a GeoPortal to help discover data better and Caltrans is providing the metadata to support it; however, it just lists data; it cannot be searched or downloaded.

**Caltrans DRS.** The tool used for DRS is called Falcon/DMS. There is a collaborative effort between the IT Office and the Office of Business Services to manage this information. The districts provide documents for the DRS, including as-built plans, survey maps, and permits, although, according to current policy set by the Chief Engineer, only as-built plans must be archived in the DRS. There are also several other types of records (e.g., permits, surveys, deeds, hydraulics information, and photographs). Most of what is contained in the DRS now relates to capital outlay services (COS). There has been recent interest for DRS to include other types of documents (e.g., traffic safety reports handled by the Division of Transportation System Information that is not part of COS). To aid in managing information, the DRS uses naming conventions for each category, usually based on a unique number (e.g., project number). Project



information is stored in the DRS. Microfilm copies are also made as a backup. The archival format used for most Caltrans documents is TIFF files. This is the preferred format because of the higher image quality. The TIFF format includes digital information about the information in the image. Other file types (including PDF and JPEG) are accepted if TIFF is not available. A records management unit works with the state archives and designated staff from each division and district. Records centers are located at the regional and district levels; there is also a state records center for all of state government, not only Caltrans.

**Metadata.** The Caltrans library uses the MARC 21 metadata standard. CONTENTdm® uses the Dublin Core ANSI standard metadata from ISO 19115 Geographic Information. Exhibit 2-2 presents the sample metadata for Park and Ride data. The DRS requires some metadata and is system dependent. Each has core elements, then some specific to a category. The metadata for the DRS is established by the Data Administrator and the DRS Steering Committee, usually for search and retrieval purposes.

**Data Governance and DBP.** Each division/office generally has governance responsibility for data and information in its area because much of Caltrans' information is stored in silo

### Exhibit 2-2. Sample metadata for park and ride.

**prk\_rde**

**Metadata:**

- [Identification Information](#)
- [Data Quality Information](#)
- [Spatial Data Organization Information](#)
- [Spatial Reference Information](#)
- [Entity and Attribute Information](#)
- [Distribution Information](#)
- [Metadata Reference Information](#)

---

*Identification Information:*

*Citation:*

*Citation Information:*

*Originator:* Division of Traffic Operations, Caltrans (DTO)

*Publication Date:* 20100728

*Title:*

prk\_rde

*Geospatial Data Presentation Form:* vector digital data

*Other Citation Details:*

update 20091215

*Online Linkage:* \\svcgis6\gishq\gisdl\gis\shapes\geo83\state\prk\_rde.shp

*Description:*

*Abstract:*

The PRK\_RDE feature class is a point feature class representing park and ride lots 'administered' by California Department of Transportation, derived from the HQ Traffic Operations Park and Ride database.

*Purpose:*

This dataset is intended to represent a spatial model of the items described herein. It is part of the California Department of Transportation's GIS Data Library. The purpose is to show locations of commuter park and ride lots in California.

*Supplemental Information:*

The location of data points contained in this layer was determined by using the location description of individual park and ride facilities in the database. The most points were placed with geocoding in ArcMap, using GDT as the reference layer. The rest were digitized with the help of Mapquest, GDT addresses or Thomas Bros. paper maps. (District 3 and District 4 submitted shapefiles created by District GIS staff). In December 2009, it was updated.

*Time Period of Content:*

Done Internet 95%

Source: GIS Data Library,  
<http://www.dot.ca.gov/hq/tsip/gis/datalibrary/gisdatalibrary.html#dataindex>,  
[http://www.dot.ca.gov/hq/tsip/gis/datalibrary/metadata/ff\\_prk\\_rde.shp.xml](http://www.dot.ca.gov/hq/tsip/gis/datalibrary/metadata/ff_prk_rde.shp.xml), accessed September 10, 2012.

**Exhibit 2-3. Data governance goals and objectives.**

#	Goal Title	Goal Description	Objectives
1	Leadership	Champion data solutions to ensure accountability and increase the value of data assets	Promote data governance within Caltrans Communicate data-related changes to all COIs Monitor progress and ensure accountability of data governance tasks and projects
2	Quality	Oversee efforts to provide high-quality data that is accurate, clear, and easy to access	Establish a data quality assurance program Increase the accuracy and clarity of data Improve accessibility of data
3	Prioritization	Prioritize efforts to address data gaps and needs	Establish clear priorities to address data gaps and needs Communicate priorities to Caltrans business units
4	Cooperation	Facilitate cross-organizational collaboration, data sharing, and integration (break down barriers between business units, reduce data silos)	Increase opportunities for data sharing Eliminate data silos and other barriers Ensure business units know the identities of business owners and data custodians Ensure business owners know the identities of the COIs
5	Flexibility	Encourage creative and innovative solutions to data needs	Identify innovative data solutions throughout Caltrans Communicate innovative solutions to business owners and COIs

Source: State of California Department of Transportation. *Transportation System DBP*. September 29, 2011.

systems and built for specific use within a particular division or office. There are many different business owners (i.e., managers of the business units that own data) and data custodians (i.e., Caltrans staff who are recognized data experts for their functional division/office). A few years ago, Caltrans realized there were too many reconciliation efforts and too much duplication of data. As a result, a Transportation System Data Governance Board was formed with representation from various divisions and districts. The mission of the board is to ensure that Caltrans creates and maintains reliable transportation system data accessible to Caltrans and its partners.<sup>24</sup> In 2011, Caltrans developed a Transportation System DBP. The DBP (1) contains a structure for the Transportation System Data Governance Board (Board) to address departmental transportation system data issues and problems, (2) defines roles and responsibilities, (3) develops the data-related processes and architecture, and (4) provides a timeline and an approach for implementing data governance. Exhibit 2-3 presents the data governance goals and objectives.

**Categorization Schemes—Caltrans Transportation Library and History Center.** The Caltrans library adheres to standard national descriptive cataloging policies and practices as developed by the American Library Association and the Library of Congress.

<sup>24</sup>State of California Department of Transportation. *Transportation System Data Business Plan*. RFO# TSI DPA-0003. September 29, 2011.

**Exhibit 2-4. GIS data library data index.**

DATA INDEX		
<u>Transportation</u>	<u>Physical Environment</u>	<u>Political/Administrative</u>
→ <u>Road Base</u>		<u>Cultural</u>
→ <u>Bus/Transit</u>		<u>Imagery</u>
→ <u>Rail</u>	→ <u>Soil and Vegetation</u>	
→ <u>Freight/Ports</u>	→ <u>Elevation</u>	→ <u>Map images (geo-referenced)</u>
→ <u>Airports</u>	→ <u>Hydrography</u>	
→ <u>Projects</u>	→ <u>Shoreline</u>	
→ <u>California Interregional Blueprint (CIB)</u>		
<u>Other Infrastructure</u>		

Source: GIS Data Library, <http://www.dot.ca.gov/hq/tsip/gis/datalibrary/gisdatalibrary.html#dataindex>, accessed September 10, 2012.

**Categorization Schemes—Caltrans GIS.** There are two GIS data libraries, one for the recently implemented Caltrans Earth and the other for the GIS Data Library. Both are generally mode-based with subcategories. Exhibit 2-4 presents the GIS Data Library Index as an example.

**Performance Measures—Caltrans Transportation Library and History Center.** Performance measures are also a key aspect of managing the Caltrans Transportation Library. Using a standard tool to calculate library cost benefit, the library estimated a positive ROI. This was based on average user salary and use of library materials, versus library salary and materials costs, and factoring in the time it would take the user to obtain the materials elsewhere.

**Performance Measures—Data Governance Plan.** Sample performance measures are also included for consideration in the data governance plan. The Data Governance Board will be responsible for developing performance measures related to each objective. Exhibit 2-5 lists objectives and sample performance measures contained in the data governance plan.

**Exhibit 2-5. Sample list of objectives and performance measures.**

#	Objective	Performance Measure
1	Ensure business units know the identities of Business Owners and Data Custodians	Increase # and/or % of data elements with identified owners/custodians that are available to business units
2	Increase the accuracy and clarity of the data	Reduction in # and/or % of inaccurate data elements and data elements without clear definitions
3	Communicate data-related changes to all COI	Development and implementation of a data communication plan approved by the Board
4	Eliminate data silos and other barriers	# of data silos identified, and # of solutions implemented to address data silos

Source: State of California Department of Transportation. *Transportation System DBP*. September 29, 2011.

Much of Caltrans' information is stored in silo-type systems, built for the specific use of a particular division, office, or district. A committee is working toward a more enterprisewide approach to the storage and management of Caltrans data.

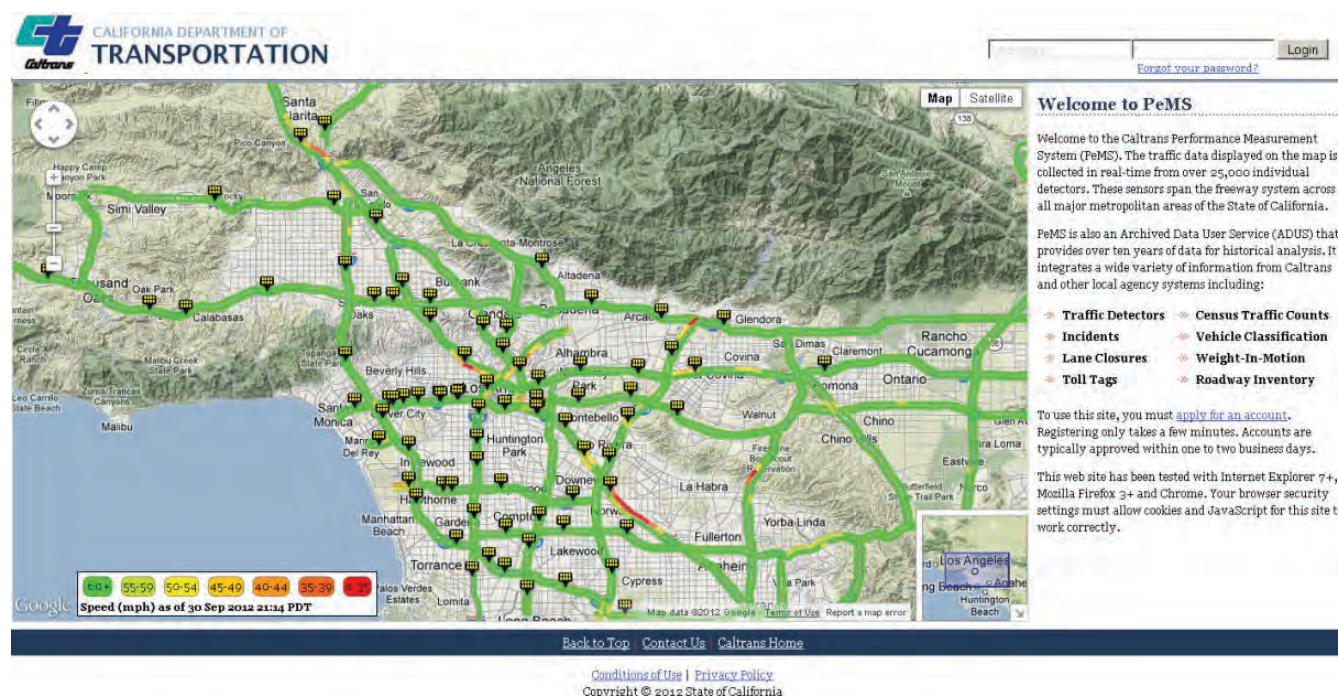
**Caltrans DRS and Other Information Systems.** The DRS allows the user to retrieve information by county, route, and post-mile. There is a similar but separate system to store and retrieve information for bridges. Caltrans also maintains a Commercial Wholesale Web Portal that enables commercial and media Information Service Providers (ISPs) to request and receive traveler information generated by the department. Information available to ISPs includes vehicle speed and volume traffic data, closed-circuit television (CCTV) video images, changeable message sign (CMS) traffic messages posted on electronic variable message signs, lane closure system (LCS) data, and roadside weather information systems (RWIS) weather and fog information. Much of this information is disseminated via Caltrans' QuickMap website at <http://quickmap.dot.ca.gov/>

Some of this information is also stored, disseminated, and/or archived in the Caltrans Performance Measurement System (PeMS). Exhibit 2-6 presents the opening screen for PeMS, which includes up to 10 years of data for historical analysis and includes traffic detector data, incidents, lane closures, toll tags, census traffic counts, vehicle classification, Weigh-In-Motion (WIM), and roadway inventory.

### Challenges

- Caltrans is a large department with numerous offices and districts, each with their own siloed data, information, and systems.
- Technology is not the challenge; issues are typically institutional or administrative.
- Getting agreement on policies, procedures, standards, priorities, and consistent resources.
- Getting top-level support for long-term storage, management, and access to information is critical to success. It is often a matter of education on the benefits.
- Trying to get GIS to fit within IT policies and standards.

Exhibit 2-6. Caltrans PeMS.



Source: Caltrans PeMS Website, <http://pems.dot.ca.gov/?redirect=%2F%3Fnode%3DState>, accessed September 30, 2012.



- Institutional differences between GIS and other offices/divisions – different approaches to information management and business processes.
- Usability problems exist with the some existing data management systems, because of lack of metadata and digital image standards. The metadata and the way the data is organized affect the ability to easily retrieve and use information.
- The need to share and integrate data has reached a critical point. As such, Caltrans is exploring options to manage information at an enterprise level.

### Next Steps

- Caltrans staff are investigating a tool called ArchiveIT and crawling tools to better manage and store web-based information resources.
- Continuous improvements to linear referencing and crowd sourcing for collecting and compiling data.
- Continue to explore cloud opportunities.
- Implement enterprisewide information management for DRS.
- Improve system design – focus on project rather than document type.
- Continue to expand Data/Information Management (Governance and Plans).

### Lessons Learned

- Understanding of the business process and prepare systems based on data-driven decision making.
- Embrace IT processes and procedures and work in collaboration with IT.
- Develop information management systems that are as user friendly as possible.

### Sources

1. Coles, Janet. Caltrans.
2. Feinberg, Harold. Caltrans.
3. Briseno, Coco. Caltrans.
4. Downey, Tremain. Caltrans.
5. Appleton, Jim. Caltrans.
6. Caltrans. *Deputy Directive DD-58-R1, Preservation of Department's Historical Materials*. February 8, 2010. <http://www.dot.ca.gov/hq/esc/CHPC/documents/DD-58-R1.pdf> (accessed 9-10-12).
7. Caltrans. Deputy Directive DD-72, Geographic Information System (GIS) Implementation. July 25, 2002.
8. Land, Richard. *Archiving As-Built Plans Memorandum*. State of California Department of Transportation. April 7, 2006. <http://www.dot.ca.gov/hq/oppd/design/m040706.pdf> (accessed 9-10-12).
9. State of California Department of Transportation. *Transportation System Data Business Plan*. RFO# TSI DPA-0003. September 29, 2011. <http://www.dot.ca.gov/hq/tsip/datagovernance.php> (accessed 9-10-12).
10. State of California Department of Transportation. *Project Development Procedures Manual, Chapter 15 – Final Project Development Procedures*. March 4, 2010. [http://www.dot.ca.gov/hq/oppd/pdpm/chap\\_pdf/chapt15.pdf](http://www.dot.ca.gov/hq/oppd/pdpm/chap_pdf/chapt15.pdf) (accessed 9-10-12).
11. State of California Department of Transportation. GIS DATA LIBRARY. <http://www.dot.ca.gov/hq/tsip/gis/datalibrary/gisdatalibrary.html> (accessed 9-10-12).
12. State of California Department of Transportation. Caltrans QuickMap Website, <http://quickmap.dot.ca.gov/> (accessed September 30, 2012).
13. State of California Department of Transportation. Caltrans Performance Measurement System (PeMS) Website, <http://pems.dot.ca.gov/?redirect=%2F%3Fnode%3DState> (accessed September 30, 2012).



### Case Study: Illinois DOT (IDOT)

Primary Sources: the Policy and Research Center (PRC) Library at the Illinois DOT (IDOT), supported by input from the Information Technology Division, the Research Unit, and the Records Management Unit at IDOT.

#### *Background*

Information management at IDOT is shared among many functional units within the agency, including, but not limited to, information technology (IT), library, research, and records management units as well as various producers of raw data (e.g., GIS and traffic safety).

The IDOT PRC Library supports the overall mission of the department by providing IDOT employees with information needed to perform daily business functions, including providing assistance with customized professional research. The PRC Library manages transportation information in the form of publications produced by the DOT, including storage and dissemination of transportation information to IDOT employees, other state employees, the public, and consultants.

Managing transportation information at IDOT is governed by policies, including the directive requiring that all IDOT-produced reports must be provided to the PRC Library as the central clearinghouse for IDOT publications. The PRC Library also deposits IDOT publications (in electronic format) to the Illinois State Library (ISL) in compliance with the State Library Act (15 ILCS 320) and the General Assembly Organization Act (25 ILCS 5/3.1). To promote the efficient and timely delivery of information, the PRC Library also participates in TKNs at the regional level (Midwest TKN) and the national level, resource sharing, and use of current technologies.

#### *Good Practices*

IDOT demonstrates good overall experience in each of the framework processes, with particular expertise in the capture and administration of transportation information.

**Capture.** The capture of IDOT research reports is coordinated by the IDOT research unit (Bureau of Materials and Physical Research) working closely with the University of Illinois at Urbana-Champaign, Illinois Center for Transportation (ICT). The ICT publishes IDOT-sponsored research reports and provides links via their website to full-text versions. The IDOT website also provides an interface for these reports through the research unit's web page, along with any internally produced publications (e.g., physical research reports). The ICT also submits electronic versions of these research reports to the TRID Database and checks the submittal option, thereby authorizing the NTL to store the documents long term electronically. The PRC Library participates in the capture of IDOT publications through the IDOT Document Deposit Program as mandated by Illinois law. Additionally the IDOT library keeps one hard copy of all IDOT publications for easy access by employees and archival purposes.

**Administer.** The PRC Library uses various technology tools to manage IDOT's collection. The PRC Library manages its collection through the use of an online card catalog, accessible to



both internal and external patrons searching for a particular publication. Searches may be conducted by employees at their own computers, and resources are sent to employees statewide in all nine district offices as well as the central office.

In addition, many IDOT publications, including most research reports, are submitted to TRB's TRID Database. The database is searchable by keyword, title, person, organization, conference, identifier, or index term.

The IDOT PRC Library collection includes information in the following categories: Civil engineering and design (including exam prep resources); materials and physical research; environment and sustainability; traffic safety and engineering; legal, policy, and procedural; records and Knowledge Management; management and strategic planning; and general office skills and IT training.

For records management, a retention schedule stating what the record is, its earliest date, and volume information, is kept to cover each record series. Recommendations for retention period are included with each record.

In addition to library cataloging and records retention, a data dictionary is used to maintain standards across databases for naming conventions. IT ensures that all data is standardized and documented.

There is no enterprise data management plan in place. Data management is partially addressed with datamarts. For datamarts to be constructed, data standardization is required. IDOT is working toward data management objectives by standardizing disparate sources of data. For example, data is being gathered from project management and financial systems, and contract identifiers are being standardized to link the information.

SharePoint is used as a document management and collaboration system. For example, MS Word documents have metadata elements attached to them. SharePoint is a metadata-based system. There is a metadata working group working to define metadata for Electronic Records Management (ERM).

The PRC Library serves as a central clearinghouse for all IDOT-produced publications, specifically for publications used across departments or provided externally to the public or consultants/contractors. Hard copies are archived in the physical library collection and electronic versions are provided to the ISL, where they are archived and maintained permanently by the staff at the state library. The link to the public interface for the archive system is <http://ediillinois.org/ppa/index.html>. Additionally, specific IDOT publications, primarily research reports, are submitted to TRB's TRID Database where the links are provided and the information is stored, although not the actual electronic file.

The records management unit is still primarily focused on storing paper copies. The state archives require paper storage, unless a document is to be kept more than 10 years. At that point, it is put on microfiche. Plans are also put on microfiche and a statement of their importance and value. Once a record is microfiched, it becomes a legal record. The reel of microfiche, a legal document, is then sent to the state archives.

In IT, digital storage occurs on four primary platforms: electronic database (on premises); electronic file system (on premises); SharePoint (on premises); and externally hosted data systems (off premises).

Long-term storage (and archiving) of IDOT publications is handled by transmitting electronic copies of the material to the ISL for storage. Standards used for archiving information include the requirement to deposit each department publication with the PRC Library. This is in support of the ISL Act (15 ILCS 320) and the General Assembly Organization Act (25 ILCS 5)

and in support of providing employees with easy access to documents as needed. Additional long-term storage electronically is provided through the submission of electronic documents to the NTL.

The PRC Library has begun to identify and digitally preserve (digital imaging) historically significant hardcopy publications and documents in the collection. Once imaging is completed, these digital documents will be submitted to the ISL's electronic digital archives <http://www.idaillinois.org/cdm/>

Distribution of IDOT research reports is coordinated by the IDOT Research unit (Bureau of Materials and Physical Research), working closely with the ICT to ensure capture and access of IDOT-produced research reports for all requestors. The widespread access to these reports is provided through the TRID Database. The TRID Database does not store documents but provides links to documents, which are typically stored elsewhere, in this case, electronically on the ICT website and in hard copy at the PRC Library and the research unit and NTL.

Multiple methods are available for distributing/disseminating publications in the PRC Library. Two examples of how the library uses tools to share information and make it usable include the internal intranet site hosted by the PRC Library on a SharePoint platform and the monthly e-newsletter, which the library disseminates to all IDOT employees directly via email. The e-newsletter includes a list of recent library acquisitions, features information on available transportation-related webinars, offers explanations of available vendor databases and agencywide membership benefits, and describes services and resources available through the ISL to IDOT employees.

The IDOT PRC Library provides enterprisewide access to various electronic databases and subscriptions, including AASHTO standards, *ACI Manual of Concrete Practice*, AISC Steel Solutions Center, *AREMA Manual*, ASCE Research Library and Civil Engineering Database, ASCE Online Journals (all ASCE journal titles online – 33 titles plus Proceedings), ASTM Standards, *Chicago Manual of Style*, and Illinois Land Sales Bulletin.

The library also allows patrons to sign up for table of contents alerts (provided electronically via email) for journals or subscriptions of interest. This facilitates dissemination of very digestible briefs, allowing for further investigation if interested.

For IT, datamart is a preferred means of data dissemination because it allows users to specify what they want. Data is diverse and information is delivered in a wide variety of forms.

### Challenges

For IDOT, certain challenges are associated with managing information within an organization, including (1) lack of standard business processes for collection/storage/dissemination of data and information across the organization, (2) lack of resources or technology tools to facilitate better information management, and (3) the existence of disparate sources of data/information, which may reside in silo information systems. Often these systems were developed to address a particular need in a specific business area, without concern that the system may duplicate data collection of specific types of data to support the system.

At IDOT, one challenge is the absence of an overarching information management policy. Without an agencywide strategy tying together all the elements of information management, individual approaches to information and/or data by various functional groups in the organization tend to exist.

Specific challenges for the PRC Library include failure by the producers of IDOT publications to submit documents to the library per the required document deposit program as governed

by the State Library Act and General Assembly Organization Act. Compliance stands at less than an estimated 50% of total publications. An additional challenge is the posting of agency policies and procedures on the intranet site only, which is not accessible outside of IDOT. Often other DOTs will ask to view IDOT policy when they are revising or creating policy for their own agencies. These policies must be accessed, saved, and provided individually by request—a time-consuming process. If these policies were available publicly, or in a secured, shared site for all DOTs, searching and using this information would be easier and less onerous for the requestor and the provider.

Constrained resources also pose challenges. The research unit does not have a records retention policy because of staffing constraints. No employees on site have the background and skill to manage photographs and graphics. Each photograph needs to be scanned, which is time consuming. A steering committee is in the final phases of setting digital imaging policy.

Given these constraints, the research unit hired the University of Illinois to administer the research program. Prior to 2006, most records were kept in email files. More recently, an intern created an access database and populated it with pieces of information basics about the research, which was an improvement. Now, the University of Illinois is developing a web-based database of projects and everything associated with projects is tied to a record.

Another challenge, particularly for IT, is the rapid pace of technology change. Some IT strategic initiatives are to modernize systems and remain current. This challenge also arises when trying to bridge the gap between IT (electronic data management) and records management (typically paper copies). The challenge is compounded when considered in light of identifying correct pieces of data relevant to decision makers.

### *Next Steps*

Ongoing IDOT initiatives to address these challenges and advance the state of the practice for information management include

- A metadata working group;
- A pilot project to test docuware, using letting and as-built plans;
- A business process mapping standards working group to define who has to see information, what approval points are needed, when information is declared a record, and so forth.

### *Lessons Learned*

- Participate in knowledge sharing groups where resources and information are shared among many. The IDOT Library participates in various Library and Information Management professional groups, including the Midwest Transportation Knowledge Network (MTKN) and the Special Libraries Association (SLA), Transportation Division (which provides access to TRANLIB, an active listserv of transportation libraries nationally and internationally). The IDOT Library is a member of the Transportation Library Connectivity and Development Pooled Fund Study TPF 5-237—a group of DOT libraries pooling federal research funds to collaborate on group projects and share good practices and research.
- Always have an elevator speech ready and ensure that upper-level management understands the work and champions information management. Convincing someone to work toward an information management governance policy is a huge step for the agency.
- Funding is important—maintaining paper records is expensive. Designation of funds to address electronic file management issues is important in meeting future information management needs.



### Case Study: Minnesota DOT (MnDOT)

Primary Sources: the Office of Transportation Data and Analysis, Office of Administration, and the Office of Enterprise Technology, commonly referred to as MN.IT Services at MnDOT.

#### Background

Starting in 2000, information management became a priority at the agency with MN.IT Services at MnDOT having primary responsibility for this function. To improve management of information and data, MnDOT developed and implemented a DBP in 2011. As part of the DBP development, MnDOT defined a vision and mission for their data and information programs (see Exhibit 2-7)

Seven key principles were identified for managing all future MnDOT data and information systems investments:

1. Data will be managed as state assets.
2. Data quality will fit its purpose.
3. Data will be accessible and shared as permitted.
4. Data will include standard metadata.
5. Data definitions will be consistently used.
6. Data management is everybody's responsibility.
7. Data shall not be duplicated.

Additional policies and standards for managing data and information at MnDOT are being developed or revised from existing standards to mature data governance at MnDOT, in accordance with these seven guiding principles. The plan focuses on (1) data governance, (2) GIS, and (3) data gaps and needs.

The data governance framework used at MnDOT includes oversight by a data governance council (composed of MnDOT's senior leadership—the division directors). The data governance council provides leadership and strategic direction. Domain stewards provide management decision making and ensure that strategic directions are implemented. The data subject area stewards are the front-line decisionmakers for data.

The ultimate goal is to create a culture where the business understands roles and responsibilities for data stewardship, management, accountability, and governance. A business data

#### Exhibit 2-7. MnDOT's vision and mission.

*MnDOT Vision for managing data and information*

*All MnDOT business decisions are supported by reliable data.*

*MnDOT Mission for managing data and information*

*To provide reliable, timely data and information easily accessed, shared for analysis and integrated into MnDOT's decision-making process.*

catalog, published in September 2012, serves as the system of record for data subject area stewardship roles.

Specific information regarding strengthening GIS business support and defining a strategic plan for GIS are also provided in the DBP, because MnDOT has identified provision for strategic direction for GIS so as to address business needs for geospatial data as a high priority. Because of the importance of GIS data in department operations, MnDOT has designated a GIS steering team, composed of managers, who have decision-making authority over GIS at MnDOT. This group is responsible for prioritizing investments in GIS technology, establishing an enterprise framework for GIS, identifying how GIS could and should fit into business processes, and advancing the state of the practice.

### *Good Practices*

**Capture.** Data capture at MnDOT occurs through many sources, including the following:

1. Traditional Business Applications (Mobile and Desktop)
2. Remote Sensing (Aerial and Satellite)
3. Global Positioning System (GPS)
4. Mobile Scanning (Light Detection and Ranging (LiDAR))
5. Automated Vehicle Location (AVL)
6. Sensor (e.g., AADT/RWIS)

Each of these sources involves unique technology and processes for the collection and Quality Control (QC) of data. Together, these methods enhance MnDOT's ability to capture the data necessary to support business needs.

The management of data and information at MnDOT was strengthened by the adoption of their DBP in 2011. The DBP and data governance at MnDOT help to establish the principles, standards, policies, procedures, and technology for collecting, processing, and disseminating data and information.

Standards for metadata elements were established by the Data Governance Work Team, prior to implementation of Data Governance in 2011. The metadata standards were formally adopted by the Business Information Council in November 2009. The mandatory metadata elements and definitions, identified in Exhibit 2-8 are based on the Dublin Core Metadata Element Set and the Minnesota Recordkeeping Metadata Standard. These elements should be applied at the table level, at a minimum. Ideally, they should be applied at the column level, based on the customer or business need.

MnDOT's approach to enterprise architecture addresses three major business challenges: (1) managing the increasing complexity of information technology systems; (2) delivering business value from those systems; and (3) planning new IT solutions to meet business needs. To deliver maximum business value, IT must plan and focus its investments on projects and initiatives aligned with the strategic direction of the business.

One of the most critical types of data used to support all business at MnDOT is location data. The department is beginning to develop and implement a new Linear Referencing System (LRS) to support managing location data. The new LRS will identify a few methods by which someone can spatially locate their data. Everyone will then be required to use the defined set of options for locating data. This will streamline and increase the efficiency in the capture and use of this type of data. MnDOT is using business process improvement (e.g., LEAN-Kaizen) in conjunction with technology improvements to ensure sustainable data management practices. LEAN-Kaizen focuses on two components: (1) the LEAN refers to a method that eliminates

**Exhibit 2-8. Metadata element schedule.**

Element	Definition	Table Level	Column Level
<b>Title</b>	The name given to the entity.	X	X
<b>Point of Contact</b>	The organizational unit that can be contacted with questions regarding the entity or accessing the entity.	X	
<b>Subject</b>	The subject or topic of the entity, which is selected from a standard subject list.	X	
<b>Description</b>	A written account of the content or purpose of the entity. Accuracy or quality descriptions may also be included.	X	X
<b>Update Frequency</b>	A description of how often the record is updated or refreshed.	X	
<b>Date Updated</b>	The point or period of time, when the entity was updated.	X	
<b>Format</b>	The file format or physical form of the entity.		X
<b>Source</b>	The primary source of record from which the described resource originated.	X	
<b>Lineage</b>	The history of the entity; how it was created and revised.	X	
<b>Dependencies</b>	Other entities, systems, and tables dependent on the entity.	X	

waste and (2) Kaizen is a Japanese word that means continuous improvement. LEAN-Kaizen helps organizations get rid of waste and continuously implement good practices.<sup>25</sup>

MnDOT is moving to SharePoint (and may use parts of the Electronic Document Management (EDM) for their document management system) to manage information by creating internal project and team websites, sharing and managing information, publishing reports for better decision making, collaborating on projects, and applying retention and disposition rules to content.

MnDOT incorporates both a data approach and a content approach for managing information, by using metadata and a business data catalog.

The categorization of transportation data supports the organized, methodical management of data systems to assist decision making at MnDOT. The categorization scheme for transportation data used by MnDOT uses the following data domains: finance; human resources; planning, programming and projects; recorded events; spatial; supporting assets; infrastructure; regulatory; business stakeholders and customers. These domains support MnDOT's strategic objectives of safety, mobility, leadership, innovation, and transparency.

The repositories, libraries, and technology tools used at MnDOT provide information across the department to support all business areas. MnDOT stores data and information in the EDMS for standard documents (e.g., Word, PDFs, images), and in Project Wise for Microstation CADD files, to ensure that the data and information can be found. Business records are stored in accordance with an approved retention and disposal schedule. MnDOT is revising its records retention

<sup>25</sup>[http://asq.org/training/lean-kaizen-a-simplified-approach-to-process-improvement\\_KAIZEN.html](http://asq.org/training/lean-kaizen-a-simplified-approach-to-process-improvement_KAIZEN.html).



schedule. The principle behind the new approach to records retention is similar to the principle behind domain stewards: organize the understanding of data into logical bundles.

Long-term storage (and archiving) of data and information are needed for data, such as travel data. Historical travel data is used for travel demand modeling and for forecasting future transportation needs, which guides MnDOT in decision making about investments in the transportation system. The primary standard for archival of information is to keep only what is needed to support analysis functions.

Data and information having permanent historical value are transferred to the Minnesota State Archives when business use is complete, in accordance with the retention schedule.

One of the most important functions of information management is to ensure that the right information gets to the intended audience on the appropriate device and under the right circumstances. The components of successful dissemination of information include layout/design of the information, how and where the information is published, what types of transformation technologies are used to provide the information in multiple formats, what type of security technology is used to protect critical information or private information from being distributed, and methods and procedures used to distribute the information. These components are discussed further as follows:

- **Layout/Design.** MnDOT is committed to presenting information to users by providing a consistent presentation layer for information products. Common formats are used for information presented over the web, for reports, and, to a lesser extent, for applications.
- **Publishing.** MnDOT publishes information to users through various technologies with an emphasis on using the web. The information is often consolidated in data warehouses and presented to users as products. The publishing environment can include other applications and published reports.
- **Transformation Technologies.** The primary transformation technologies often include Extract Transform and Load (ETL) processes to change data to information. Recently, some ETL processes have been replaced with web services to present the data in a more interactive and flexible format to meet the needs of applications and individual use.
- **Security Technology.** MnDOT security policies and procedures were implemented and have been evolving since the mid-1990s with the implementation of enterprise anti-virus, firewalling private and semi-public networks from publicly accessible networks, web content filtering, and securing physical infrastructure. More recent industry standard technologies have provided automated operating system patching and vulnerability scanning and threat analysis. MnDOT has conducted a study to classify its data, based on the state's Data Practices Act, and uses secure technologies to protect that data in transit. In addition, MnDOT requires secure authentication to non-public systems, implementing server minimum-security standards, the use of firewalling, and monitoring for compliance to security standards and policies.

MnDOT distributes much information through the website; however, searching for information on this website is not intuitive and additional assistance may be needed from MnDOT staff for locating specific types of data and information.

MnDOT also maintains a GIS base map and a spatial data warehouse for purposes of finding/retrieving location information when needed. Staff are working on a modification to the data warehouse to make it easier to provide data in various formats (e.g., publishing spreadsheets for traffic volume data as well as producing maps). Staff are noticing a significant increase in the number of hits on the website for people using spreadsheets.

Staff are also working to develop better marketing tools to tell how data is used within MnDOT. An example would be to explain how planning data is used in the community or to

support a corridor planning story. The story could include information about pavement type, number of crashes, and so forth within a corridor. This will help the DOT and local community determine the best use of funds for improvements in a given corridor.

MnDOT is piloting the use of business intelligence (BI) tools from Oracle to facilitate searches for data and information and would like to do a pilot project for locating data on snow and ice operations, bridge, human resources, financial data, and materials used in the Maintenance areas. MnDOT also established an ArcGIS web service for individual business areas to manage their data and has established a Minnesota Duluth flooding emergency preparedness website to provide critical information to the public and first responders during times of flooding.

### *Challenges*

At MnDOT, there is no single portal method to search for data. The existing website is helpful for users familiar with what is available on the website and how to navigate it; however, the website is challenging for the casual user to find specific data or information.

The costs to collect, manage, and store data versus the business value the data provides is increasingly difficult to quantify.

### *Next Steps*

Next steps include establishing measureable goals, strengthening the alignment with other initiatives (e.g., Safeguarding MnDOT, and risk management), enterprise architecture, partnership with others stakeholders (e.g., counties, DNR), and improving staff's analytical skills.

In addition, implementing recommendations documented in the MnDOT DBP will help support the commitment to manage data and information as department assets:

The infrastructure preservation recommendations set the stage for implementing an organizational approach to asset management and for addressing critical transportation infrastructure data gaps and needs. Recommendations to improve asset management include developing an asset management framework for MnDOT consistent with FHWA and AASHTO good practices (so as to promote asset management good practices in capital investment, maintenance, and system design); and determining which assets are high-priority candidates for being part of a department asset management system.

For transportation assets not currently inventoried or tracked, a way to assess the benefit and ROI for establishing inventories to track age, condition, and other attributes is needed.

Traveler safety recommendations cite the need for better data on local road characteristics and more enhanced safety data analysis tools. MnDOT identified enhancing the completeness and accuracy of roadway centerline data as the highest priority for enhancing data for traveler safety.

The mobility recommendations identify the need for research and resources to collect potentially new data to address increasing interest in multimodal accessibility, reliability, and person-throughput questions. Recommendations for improving delivery of mobility data include increasing the data and information available on traveler behavior and mode choice and determining the performance metrics, data, and information needed to implement a programmatic multimodal approach for addressing mobility, travel time reliability, and accessibility for all system users.

The financial data recommendations address the need for enhanced information on lifecycle costs, ROIs, and data for evaluating service delivery options. Recommendations for improvements in delivery of information on financial data include initiating an agencywide effort to define business needs for financial data beyond that being tracked to meet internal financial control obligations, continuing to support the implementation of the Statewide Integrated Financial Tools project, and continuing efforts to improve project cost-estimating practices and processes.

BI recommendations highlight the value of departmentwide solutions for improving data availability, integration, and analytical capabilities. Recommendations include establishing a BI Program, establishing a BI Steering Committee, establishing a scorecard with criteria for determining if business data is ready for BI, and rolling out the use of BI in three to four business areas that have varying complexity of data reporting/analysis needs.

Enterprise architecture recommendations provide an opportunity to look at how all information systems might fit together to reduce data redundancies and create operational efficiencies. Recommendations include completing a comprehensive assessment of MnDOT's information system architecture, and using the results of the assessment to identify opportunities for improving efficiencies and interoperability between existing and proposed information systems.

The data governance recommendations lay out a comprehensive series of steps for clarifying data stewardship roles and responsibilities and for setting standards and policies to reduce redundancies and promote data quality and reliability. A recommendation is also made for developing a data catalog and a thorough assessment of departmentwide information system architecture to identify opportunities for integration to reduce redundancies and promote efficiencies. Other recommendations include establishing a Data Governance Board to formally adopt data governance principles on behalf of MnDOT and incorporate them into policies, standards and processes; revising existing policies (e.g., stewardship, development, data security, database recovery, and data retention); and developing additional policies needed to implement data governance at MnDOT.

The GIS recommendation sets the stage for business process, data governance, and organizational changes to achieve desired objectives. Recommendations include undertaking a formal assessment of available GIS data and determining what is essential or core data, and where there are gaps and needs; establishing a GIS steering team to guide and direct the significant investment the department has in GIS architecture to support a broader business decision-making purpose than the project-driven architecture currently in place; and creating a GIS Business Support Unit consisting of GIS professionals to help users with the production of maps and analytical needs beyond desktop business support tools.

### *Lessons Learned*

- Use governance with flexibility.
- Integrate governance with other initiatives. This helps the business offices/units to realize how the exercise of good data and information management practices can safeguard the agency against many potential risks, not just those identified by a particular business area.
- Use a balanced approach for governance with a bottom-up and top-down approach to help convince business areas of governance needs in support of overall agency strategic needs.
- Business offices/units must lead. Business must put resources toward good data management practices and should hold other business areas accountable—data management is not IT's responsibility.
- Focus on data for decisions at all levels. Ensure that data is useful and usable. Remember the purpose of data and for what it is used.
- Establish standards that are nimble.
- Establish policies that allow for transparency and innovation in information management.
- Data management is not the end, it is the means to other ends (e.g., providing information needed by the public, policymakers, and snowplow drivers).
- Establishing governance takes time, but, it is worth the effort.
- One does not always need more data, just better access, better tools, and better understanding.

Data and information are not valuable unless they contribute to quality of life. The data and information managed by the DOT should be used to identify the current and future investment needs for MnDOT's transportation network. Future improvements to MnDOT's

transportation system will help to improve the quality of life for citizens by providing a well-maintained, safe traveling environment for the public.



### **Case Study: Mississippi DOT (MDOT)**

Primary Sources: MDOT staff and MDOT SharePoint documentation.

#### *Background*

Information management, as MDOT calls ECM, became a priority for the agency in 2008. MDOT began looking for a solution to scan, search, and retrieve its hundreds of documents. MDOT was selected to participate in the Microsoft Rapid Development Plan with the beta versions of SharePoint 2010, FAST, and other Microsoft software. The MDOT Chief Information Officer at that time was the leader of this massive effort. MDOT then decided to invest in the Microsoft Server platform including SharePoint server, Performance Point, and Microsoft SQL Server to provide enterprise search, content management, and BI capabilities to its 3,500 employees and the public.

MDOT does not have a formal group responsible for governance of data and information. However, the MDOT Information Systems Division has standard operating procedures (SOPs) that govern their security, procedures, and processes, as well as how application systems are developed and how information and records are archived. There are state laws for archiving historical information and each office in the DOT has a set of rules/procedures for how long they retain data and documents and how long they are required to make it available on request. They follow several state and federal legislative mandates regarding open records requests for data and information.

MDOT has a SharePoint Governance Plan that contains policies, roles and responsibilities, and processes to guide, direct, control, and support SharePoint. This governance plan supplements MDOT's Information Systems SOPs with the SOPs being the source to use if there are conflicts. Governance is tightly controlled in areas of substantial public external or internal exposure or potential litigation issues.

MDOT developed the MDOT Content Management Library Standards to establish a set of standards for implementing SharePoint Document Management. The standards must be followed by all divisions, District Offices, and Project Offices planning to store content in SharePoint and address the following elements: content types; document library; standard properties; custom properties; security; managed terms; and managed content types.

Exception requests to the standards will be reviewed by the Information Systems Division (ISD), SharePoint Governance Team (SGT), and Managed Metadata Service Administrator (MMA).

The categorization scheme for most MDOT documents is based on the assigned project number. The project number is the common metadata in most of the stored documents. MDOT uses SharePoint's Content Type to categorize documents. Each document has its own set of

standard and custom properties and must be identified as a specific kind of document type, and, for each document type, a content type must be defined.

Information Management is an evolving systematic process at MDOT.

### *Good Practices*

**Capture.** MDOT began capturing documents using a product called KnowledgeLake but has since changed capturing tools to Kofax.

MDOT captures financial and project data in the Financial Management System, Maintenance Management System, Construction Management System, and Equipment Management System. Data is then presented in SharePoint in a centralized project center. Documents are captured with Kofax and presented along with the data in the project center.

MDOT follows FHWA-defined Federal rules and guidelines for collection of data. Data capture requirements are generally defined on a project-by-project basis. MDOT is moving from keeping paper documents to electronic storage, with several offices migrating from paper to electronic documents. SharePoint is used as the primary document management system to capture and store certain documents in MDOT. However, other systems in specific offices are used to capture and store documents (e.g., roadway design documents and as-built plans).

**Administer.** The SharePoint environments are managed by two teams, a strategy team and a tactical team. The strategy team provides strategic insight and direction for the portal and is made up of a balance between IT and business, and central versus district. The tactical team supports the directives of the strategy team and consists of three sub-teams—operations, support, and development. MDOT uses a SharePoint Governance Plan to define the roles, responsibilities and tasks, permissions, and required skills for each team member.

Initial project/team collaboration is done on a SharePoint collaboration site. As projects are completed, the data on the collaboration site is moved to a SharePoint document site. This is the repository for official content. Metalogix StoragePoint is the leading remote binary large object (BLOB) storage and archive solution for SharePoint. StoragePoint quickly installs natively into SharePoint Central Administration, enabling MDOT to manage SharePoint BLOBs transparently as well as capture and consolidate file share content from a single SharePoint interface.

A SharePoint document library is a special list used to store documents and is based on a content type labeled as document. When a new document library is created, several columns are immediately available such as Title, Author, and Date. A document library can easily be configured to support multiple content types and allow storage of multiple types of documents in a single library.

Within SharePoint, content can be classified as either structured (based on a specialized form or SOP often in the form of a report) or unstructured (e.g., generic content such as letters or email, typically does not conform to a specific format). Content contains metadata and is stored in a document library defined to support the metadata using list or library columns.

The content of the document to be stored is reviewed and, where possible, categorized into distinct document types (note—it is easy to confuse a true document type from a document sub-type which can usually be handled by an additional Type field). Each defined document type should be determined and identify what metadata (properties) should be associated with each document type, being sure to include the standard properties.

MDOT has a standard operating procedure for document retention. Based on the document retention guidelines, the data is managed accordingly.



MDOT is moving away from keeping paper documents and moving to electronic storage. SharePoint is used as the primary document management system to store certain documents in MDOT. However, other systems in specific offices are used to store documents (e.g., roadway design documents and as-built plans). In-house written document storage systems are used to store and retrieve documents. There is also a legacy system called ApplicationXtender that has 200+ applications written to index, store, and retrieve documents. Eventually, the plan is to merge all documents electronically into the SharePoint system.

In SharePoint, content can be classified as Collaborative or Historical in nature.<sup>26</sup> A collaborative document is one that has not been finalized or met final approval, or is being worked on by a group or team. A document that has met final approval and may not be further modified is classified as historical (e.g., legal documents). MDOT SharePoint has two major portals available for document management: (1) the Collaboration Portal (MDOTCOLLAB), and (2) the Document Center (MDOTDOCS).

MDOTCOLLAB is used as the starting point for collaborative documents for MDOT organizations and teams. Sites have been provided for each division and the URL is based on a three-character representation of the division (e.g., the Collaboration Portal for information systems is <http://mdotcollab/divisions/isd>).

MDOTDOCS is used to store historical documents. Each division is provided with a site and the URL is based on a three-character representation of the division according to the Organizational Structure of MDOT. This is a public area for storing documents or archival in place of paper documents. MDOT maintains three other SharePoint portals (web applications):

- <http://mdotportal> – MDOT’s public intranet containing information useful to MDOT as a whole and typically read-only for employees.
- <http://mymdot> – The MySite deployment and online directory contain employee information and space for an individual’s documents.
- <http://sp.gomdot.com> – MDOT’s public Internet site.

Exhibit 2-9 presents MDOT’s SharePoint Governance Model.

Storage of information is also managed via quota plans, allowed file types, maximum file sizes, and deletion and backup/restore times. Exhibit 2-10 presents this information for collaboration and WCM.

State laws are in place for archiving historical information. In addition, each office in MDOT has a set of rules/procedures for how long they retain data and documents and how long they are required to make information available on request. Retention policies are enforced in the Document Center portal in SharePoint. MDOT’s data classification and retention policy is as follows:

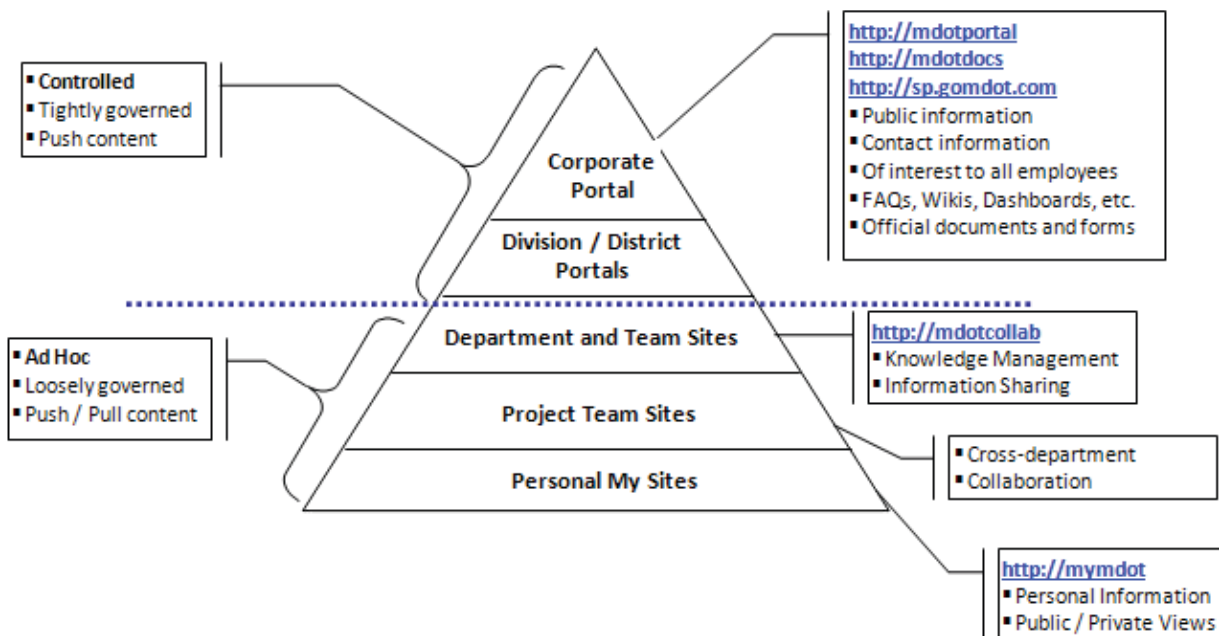
MDOT will establish and maintain systems and guidelines for the retention of electronic data and metadata, especially electronic correspondence that balance legal and operational requirements with the need for privacy.

**Data Classification.** Data classification provides a high-level guideline for understanding and managing data and information in regard to their level of confidentiality and sensitivity. Classification of data is critical in defining and implementing the correct level of protection for MDOT information. Such classifications are established for the secure generation, collection,

<sup>26</sup>Phillip C. Warren. *MDOT Content Management Library Standards – Requirements and Standards*. Mississippi Department of Transportation, January 1, 2010.



**Exhibit 2-9. Governance model.**



Source: SharePoint Governance Plan. Mississippi DOT. July 26, 2010.

access, storage, maintenance, transmission, archiving, and disposal of data. MDOT categories of data include financial, construction, contract, traffic, and email. This data is classified as follows:

- Financial Data – Limited access which applies to conditions and limitations of its access
- Construction Data – Public access which applies to no limitations of its access
- Contract Data – Public access which applies to no limitations of its access
- Traffic Data – Public access which applies to no limitations of its access

***Retained Electronic Data (hereafter, Electronic Data).*** Any and all information, in whatever format, collected and/or stored by MDOT. This includes, but is not limited to, financial, construction, and MDOT user information. This could include database records, web pages, and

**Exhibit 2-10. Storage management.**

Storage Type	Quota Plans	Allowed File Types	Max File Size	Sites Deleted	Backup / Restore
Collaboration	1GB, 5GB, 10GB	PDF, Word, Excel, PowerPoint, SilverLight, Windows Media Video	500 MB	On request by Owner, after 90 days of inactivity, unless otherwise agreed	On request by Site Owner – any 15-minute interval in previous 10 days
Web Content Management	1GB, 5GB, 10GB	PDF, Word, Excel, PowerPoint, SilverLight, Windows Media Video	100 MB	On request by Site Owner	On request by Site Owner – any 15-minute interval in previous 10 days

Source: Mississippi Department of Transportation, September 2012.

emails to name a few. It does not include information not stored, for example, live phone calls. Retained electronic data is sometimes referred to as Content.

**Metadata.** This is information about content, not content itself. For example, records relating to GIS information. That is, not the content of GIS, but only the information about GIS, such as coordinates. Another example, records relating to web site visits. That is, not the content of those web sites, but data related to the access, when it took place, what Internet addresses were involved, etc.

**Purpose.** The Electronic Data Retention Policy is intended to help employees determine what information sent, received, or gathered and its associated metadata, if any, should be, or is being, retained and for how long. All users should familiarize themselves with the retention areas that follow this introduction.

This policy is not intended to cover operational retention that occurs through backup and/or mirroring of systems. That form of retention is an incidental byproduct of disaster preparedness and is not intended to function as a substitute, in any way, for required retention or data shredding. Information Systems will not attempt to recover data, files or other information from backups or mirrors unless required to do so by law or the MDOT administration. Likewise the purging of email put in electronic trash bins is an operational matter, not a retention issue.

This policy is not intended to address retention of data in systems not operated directly by MDOT ISD and not under the control of MDOT ISD. This policy does apply to systems and services operated by MDOT divisions outside of ISD. It also applies to the retention of data where MDOT ISD has control in outsourced situations.

The retention policies for data in the Financial Management Division are determined by the Financial Management Division. However, in no event shall the retention in those systems be less than this policy stipulates.

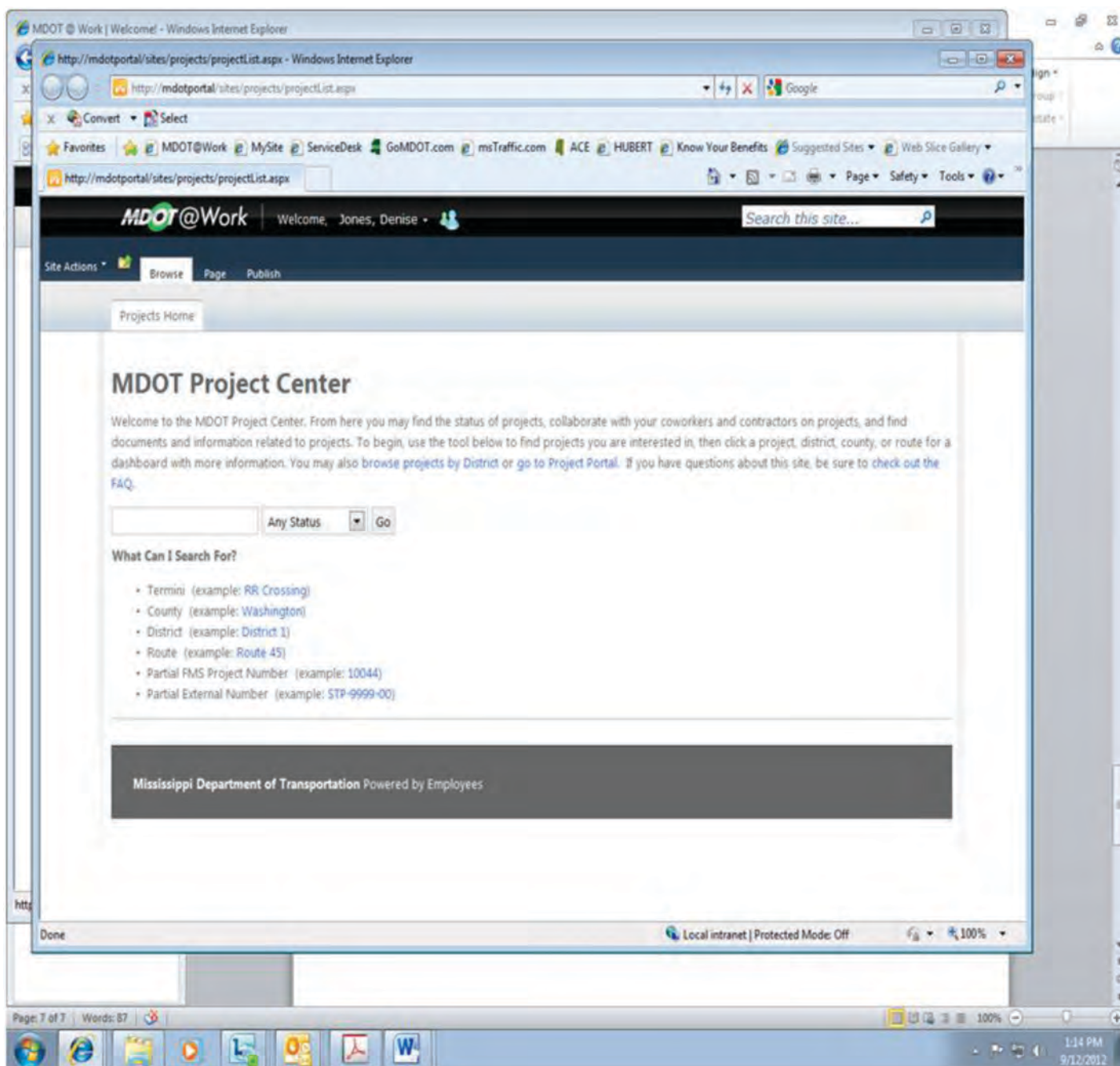
**Categories of Data and Corresponding Retention Periods.** All of MDOT's Electronic Data is categorized, each with distinct retention guidelines:

- Email Correspondence (1 year)
- Files saved on MDOT's SAN (Storage Area Network) device (1 year)
- Web sites visited (30 calendar days)
- Quarantined Spam (10 calendar days)
- Helpdesk tickets (2 years)
- MDOT Executives, Division Directors, District Engineers (Hard Drives backed up weekly)
- Financial Data (no expiration)
- Construction Data (no expiration)
- Contract Data (no expiration)
- Traffic Data (not retained)

MDOT started with a Google application for supporting searches and they now use Microsoft Search and Microsoft FAST search for searching the SharePoint archive. Searches are also done using ApplicationXtender. MDOT is migrating documents from ApplicationXtender to SharePoint, using Knowledge Lake for the migration.

Retrieval of most documents is done using SharePoint, which supports comprehensive searches. Keywords are required when contents are stored in SharePoint to ensure that transportation information is findable and accessible.

Exhibit 2-11 presents a sample project search screen and Exhibit 2-12 shows a search page for environmental documents.

**Exhibit 2-11. MDOT project center.**

Source: Mississippi Department of Transportation, September 2012.

With respect to GIS information, MDOT is trying to move away from using static maps and instead present online interactive versions of their maps. They are using GPS coordinates from their databases to locate traffic counters and traffic segments and to display these items on Google maps. MDOT is experimenting with several options for developing interactive map functions (including Google apps, and ESRI ArcGIS server and Intergraph Web Map). MDOT primarily uses Intergraph for data development projects, but also uses ESRI tools for desktop applications to support hydraulic or soil analysis.

**Exhibit 2-12. Environmental document center search screen.**

Edit	Name	CombinedProjectNumber	County	Route	ScanUser	Termini	DocumentDate	ApprovalDate	EnvProjectType	EnvSubDocumentType
		106145/101000	Wilkinson	Turkey Creek Road & Crooked Creek Road	bgray	RPL BRIDGE TURKEY CREEK RD WILKINSON CTY	9/12/2012		STATE AID	SUPPORTING DOCUMENTATION
		104866/801000	Harrison		rvincen	Dedeaux Rd. From Three Rivers Rd. to SR 605. Gulfport MS	9/11/2012	2/24/2011	LPA	DOCUMENT
		106467/101000	Newton	Road A- Pine Bluff Rd. Road B- Decatur-Conehatta Rd. Road C- Decatur-Littlerock Rd.	ctaylor	Safety Enhancements Pine Bluff Rd Newton	7/13/2012	7/13/2012	STATE AID	DOCUMENT
		106428/301000	Coahoma	61	ajohnson	GT-057 Upgrade a crossing surface on US 61 near Lula (DOT # 301 749Y)	9/6/2012	9/6/2012	MDOT	DOCUMENT
		106480/101000	Benton	72	ajohnson	US 72 near MS 7 DOT # 297 7157 MS Central RR	9/6/2012	9/6/2012	MDOT	DOCUMENT
		105778/301000	Covington	49	bgray	US 49 @ Kola Road in Covington County	9/7/2012		MDOT	SUPPORTING DOCUMENTATION

Source: Mississippi Department of Transportation, September 2012.

Given that one of MDOT's goals is to enable enterprise search to allow easy access to information, it is important to ensure that only those results the current user is authorized to see are returned. The SharePoint security model provides security at many levels. SharePoint provides support for a lightweight directory access protocol (LDAP) allowing integration with Exchange Server and Active Directory (AD). During security assignment, the user may use the Global Address List to locate and select AD Users, AD Groups, or SharePoint Groups. Group permissions are based on the default permission levels: full control, design, contribute, read, limited access, and view only. To aid in security administration, the SharePoint security model provides

inheritance by default. Current site-level security permissions are applied to the list when a new list is created. It can also be configured to no longer inherit from the parent. Security can be assigned at the site collection, site, list, and item levels.

### *Challenges*

Challenges associated with the implementation of SharePoint included the requirement that the entire system implement at one time. It would have been preferable to prioritize implementation of the system, by implementing higher priority components first and others later. Other challenges included the agency's adoption of digitizing and sharing their documents. Digital signatures are another hurdle that MDOT is trying to overcome so as to electronically sign and route documents.

### *Next Steps*

MDOT will continue the current course and schedule to capture paper documents and processes and move to digital storage and retrieval of all MDOT documents. In the governance plan, Information Services identified the following three key areas for future development: extranet collaboration, mobile worker, and BI. It is the responsibility of the IS Project team to seek opportunities to improve business processes, align activities with the goals of the business, identify possible synergies between divisions and departments, and reduce inefficiencies and duplication.

### *Lessons Learned*

Information management will be ongoing and may never be fully complete. Using a more agile project management method is more beneficial than the standard project management method. One must be able to adapt and produce solutions that best fit the agency. It is also important to work at a manageable pace when working with the agency's end users and business owners.

### *Sources*

1. Jones, Denise. Mississippi Department of Transportation.
2. SharePoint Architecture Guide. Mississippi Department of Transportation. September 21, 2009.
3. Ken Slay, Denise Jones, Mario Fulan, and Daniel Root. SharePoint Governance Plan. Mississippi Department of Transportation. July 26, 2010.
4. Warren, Phillip C. *MDOT Content Management Library Standards – Requirements and Standards*. Mississippi Department of Transportation. January 1, 2010.



## **Case Study: Virginia DOT (VDOT)**

### *Background*

Although various information services have been in existence and available to staff for decades, information management became a priority at the agency starting in 2003. The primary responsibility for this function resides with the KM office. To improve management of information and data, VDOT established a Business Governance Board in 2011 to provide policies and procedures for the use of the intranet. This governance board makes all decisions regarding training, SOP,



communications, requested functionality, and upgrades. The chair of the governance board (also the KM Officer) serves on the advisory board for SharePoint 2010 (the platform on which the intranet is built). Examples of decisions made by the governance board are as follows:

1. All current corporate documents (e.g., policies, procedures, manuals, and reports) will be stored in one place with required metadata.
2. Prior versions of corporate documents will be moved to an archive (read-only and searchable) when a new version is uploaded.
3. Reports are used to alert appropriate staff if the rule regarding linking to documents (in lieu of copying and reposting documents) is not being followed.

The Corporate Document Metadata is available on VDOT's intranet referred to as Inside VDOT. VDOT has also developed a taxonomy with descriptions for Construction Project Metadata. This effort began in February 2010 as a pilot project with a team of stakeholders from the districts, including construction engineers, who helped to design the intranet site and helped to refine the metadata for construction projects. The success of the Construction Projects site has attracted other business areas, such as Maintenance and Operations, to develop similar sites for their areas.

Performance measures are useful for tracking how well specific programs are performing with respect to pre-defined objectives, measured by a set of standards, known as metrics. One of the metrics used to track how well VDOT is doing with respect to information management pertains to an assessment of the ROI for implementation of information systems and/or improvements in business processes. For example, metrics indicate that the implementation of the Project Record Keeping System (PRKS) has a conservative ROI of about \$500,000 (based on a 1 hour/week savings per inspector). In addition, resource sharing across districts in the right-of-way and utilities function eliminates the need to hire consultants with a conservative ROI of \$1.4 million cost avoidance.

### *Good Practices*

**Capture.** VDOT relies on the partnership between the district and division offices to capture information about ongoing and proposed projects in those offices. This information is posted to the intranet by the various offices. A workspace is assigned to the offices (usable with permission only) to develop and post documents. All minor versions of the documents are kept in the workspace with major versions published to the corporate document center. The intranet is the primary tool used to aid in the capture of information and data.

**Administer.** VDOT created a Knowledge Management Office in 2003 dedicated to sharing and retaining critical organizational knowledge and managing transportation information throughout the department. VDOT has also designated staff from each district to serve as content managers for maintaining information in SharePoint about specific programs in their districts. In support of sharing knowledge and information needed for transportation-related business functions, VDOT also uses what are referred to as communities of practice (COPs). The COPs are critical in creating networks and sharing knowledge, which often results in development of new tools and also changes and improvements in business and workflow processes. As an example, the Construction Quality Managers COP helps to address the critical need for well-trained experienced inspectors and for QC in the field.

VDOT also uses a tool referred to as Knowledge Mapping to help identify areas of need for succession planning and to build networks to support succession planning.

Indexing and categorization are also used by the VDOT research library in the online catalog and by the KM office to organize information on the intranet so that it is retrievable. The index



for the intranet is a metadata set developed by the business advisors for use in the corporate document center. One example of a specialized metadata set is for construction project team sites.

The construction project team sites include the following metadata: Name – filename of document; Title; Construction Document Descriptor (Contract Administration, Contractor Survey, Roadway, Structure and Bridge, etc.); District; UPC # (the UPC number for the project); and Route # (the route number for the project).

One of the successful tools used in other DOTs for managing information and data is the use of a DBP. DBPs traditionally provide guidance, policies, procedures, and standards for the collection, use, and dissemination of data and information for an organization. VDOT has implemented a DBP for the System Operations (SO) Directorate to provide a framework for making decisions about what data to acquire, how to get it, and how to make sure it is providing value commensurate with its cost (VDOT SO Data Business Plan, June 2008).

Specific information addressed in a DBP often includes the standards for the use of metadata within the organization. Metadata is used to determine what data and information is needed to respond to inquiries and ad hoc requests for transportation data and information. At VDOT, a view of the metadata, which is usually stored in SharePoint, helps to ensure that the right information is provided to the right audience. The use of taxonomies (or classification/grouping of information) also contributes to more effective management of transportation information. VDOT has a taxonomy established in their KM system that helps in the effective access and distribution of information.

VDOT uses Microsoft Office SharePoint 2010 as its primary storage medium and KM system to retain critical corporate knowledge and information and to make this information available to staff across the agency in support of daily business operations.

Library collections and services are also used to provide information to patrons, both internal and external to VDOT, although internal customers are the primary focus. The library hosts an online catalog accessible both within and outside of VDOT. In addition, the library provides access to several information resources for business and engineering functions with electronic full-text accessible across the agency; however, because of licensing restrictions, those resources can only be made accessible to VDOT employees.

Several types of information management processes are used to retain and preserve critical corporate knowledge at VDOT, including collaboration, categorization, WCM, records management, workflow/business process management, and document management. Each process is described in the paragraphs below.

**Collaboration.** Managing transportation information at VDOT relies on collaborative efforts across several offices, including the Office of Knowledge Management, the VDOT research library, the Administrative Services Office, the Communications Department (previously Public Affairs), and all other business offices at VDOT (including the divisions and districts). Each office plays a vital role in providing data and information for assimilation and distribution of transportation information to support VDOT business operations and requests from external customers (e.g., the governor's office, legislature, local governments, and the public).

**Categorization.** The categorization scheme for transportation data used by VDOT is a result of the use of SharePoint 2010 as their KM system. VDOT considered using the TRT, which has a total of 9,425 terms and for which the primary purpose was as a common indexing language for transportation research for indexing records in the TRID Database, which covers all modes of transportation and all disciplines. However, considering its specificity, scope, and scale, the TRT is not a perfect fit for a DOT SharePoint portal and does not translate well for VDOT's business

purposes. Therefore, VDOT developed a taxonomy for classifying and grouping information. Even though a taxonomy template is provided with SharePoint, VDOT developed a specific taxonomy for use with the agency's KM system. This taxonomy, described in the agency's Corporate Document Metadata, includes the following types of information: Name (i.e., filename of document); Title; Document Descriptor (e.g., Contract, Budget, Facilities, and Form); Document Types (e.g., Vital Record); Document Owner; Date Approved; Date to be Reviewed; Division; District; Region; and Enterprise Keywords. (Keywords entered by the user enhance search and filtering and metadata consistency and reuse.)

**Web Content Management.** Web content management is a vital component supporting the interface between VDOT and the general public. The web section of the Communications Department, which is responsible for the web, has a staff of three people (a web content manager, webmaster, and web-technician) who maintain the department's website and content. In addition, other staff in remote offices maintain the web pages for nine divisions. District staff may be granted limited access to handle their parts of the web content.

The responsibilities of the web content manager include overseeing what is published on the Internet and determining who is the intended audience of the web content. The goal is to write the content in the Associated Press (AP) style and the style established by VDOT to make the content as easy to understand as possible. The intended audience includes the news media, the public, and elected and appointed officials.

**Records Management.** The Library of Virginia oversees all records management of VDOT as a member of the Records Oversight Committee (ROC). Within VDOT, the Records Management function is performed by the Administrative Services Office, in coordination with 55 other Records Coordinators in the divisions and districts. The records analysts in the Administrative Services Office coordinate with these 55 personnel as well as the Library of Virginia on the archive of records on behalf of VDOT.

The definition of records includes any and all types of records (e.g., Road design plans; Contracts; Project documents; X-rays of bridge wells; and Other types [classified as records]).

Approximately 70-80% of these records are still on paper and include C- and D-size plans and x-ray films. These records are housed in the VDOT divisions and districts and may include information on databases and on servers. Audit reports (in MS Word format) are also kept in a shared folder and Records Management staff has access to shared drives and network drives to manage the records. There are also databases used to manage the legal documents.

If districts or divisions cannot locate records (e.g., Right-of-Way, Maintenance, or Construction records), records management staff in the Administrative Services Office can access the Inventory and Storage database to determine if records are stored off site. The individual records are managed using hand-held scanners and bar-coding devices for every publication. As records are boxed up the box is assigned a bar-code number which is then entered into the Inventory and Storage database and can be subsequently located as needed. Metadata is also entered into the database.

An electronic form is used to check records in and out in accordance with established workflow processes. InfoPath 2003 is used for the check-in/check-out process. Coordination will continue between VDOT offices as the department prepares to migrate to SharePoint as its primary enterprise records management system.

**Workflow/Business Process Management.** VDOT uses SharePoint 2010 to support and enhance the management of workflow and business processes by including a section in SharePoint

containing information about lessons learned from previous projects. These lessons learned provide helpful guidance to staff in managing current projects.

**Document Management.** VDOT uses SharePoint 2010 as its enterprise document management system. This system provides a quick way for all VDOT staff to search several different types of internal and external content simultaneously. This includes searches on VDOT documents, policy, forms, and people libraries. The VDOT Library's online catalog and subscription databases help users find research reports, books, articles, and standards from transportation publishers such as TRB, ASCE, ASTM, AASHTO, ITE, other DOTs, and University Transportation Centers, as well as overseas publishers and state and Federal agencies in the United States (e.g., U.S. DOT and FHWA).

The VDOT website is also used for storing manuals, forms, and so forth for external users.

**VDOT Research Library.** Since 1954, the VDOT Research Library has collected, organized, preserved, and provided access to published documents relevant to VDOT patrons. By progressing from maintaining printed versions of materials to providing electronic versions as well for many documents, the library has expanded collections beyond the limitations of its size and budget and still found time to provide various modern information services. This was accomplished while taking into consideration the changing expectations of patrons (considering what people do, where they do it from—both physical location and computing platform—what types of information they need, and in what formats). This includes leasing agencywide access to commercially available databases that contain full-text online technical reports, e-journal content, and e-books, and then targeting this information for VDOT's library patrons. The library acts less as a gatekeeper to information and more as a facilitator by finding out what types of information are needed most and then identifying which vendors can provide databases with links to the necessary information in ways consistent with VDOT IT security and commercial authentication requirements.

VDOT uses a federated search tool called VDOT OneSearch as a single search engine across 15 to 20 databases so that the user can perform searches from a single screen and retrieve relevant results from all databases on that screen. The library has been proactive in responding to the shift in access versus ownership and in recognizing the barriers to access of physical versus virtual collections. For example, in fiscal year 2012, for every book checked out of the library's physical location (including those shipped), VDOT patrons accessed 16 items from their desktops through self-service online access. Formats for online material included HTML, PDF, MP3 (audio), and FLV (audio/video).

In addition to collections, the library also provides various information services to VDOT patrons including interlibrary loans, scanning and delivering documents from hard copy to digital format (document delivery), literature search and synthesis (reference and research services), and acquisitions/collection development to support specific developmental objectives of the agency or its divisions. To promote and educate users on resources, the library has used a combination of in-person and virtual techniques. Library staff has conducted site visits to districts and divisions and has presented at the meeting of the Commonwealth Transportation Board. Because of limitations on time and travel, the library has also engaged in synchronous and asynchronous user education sessions, conducting how-to webinars and producing a series of simple instructional videos for demonstrating database usage and promoting library services and collections.

To ensure long-term storage and archival of critical content, VDOT is establishing an archive that will allow the agency to preserve critical institutional information. The archive responsibility will be part of the records management function at VDOT and will house both archived

corporate documents and team information, which will house all documents and emails pertaining to a specific project. The team site will be open to partners to contribute to directly. Once the project is completed, the site will be archived. VDOT also maintains a historical archive at the Library of Virginia and the VDOT research library collects VDOT publications comprehensively.

Standards are in place to determine what types of material are archived, both internally and externally, in what types of formats, and for how long. All internal documents are on permanent retention in multiple formats. These include the final published documents, but not draft documents.

One of the most important functions of information management is to ensure that the right information gets to the intended audience on the appropriate device and under the right circumstances. The components of successful dissemination of information include layout/design of the information, how and where the information is published, what types of transformation technologies are used to provide the information in multiple formats, what type of security technology is used to protect critical information or private information from being distributed, and methods and procedures used to distribute the information.

**Layout/Design.** The layout and design of information distributed on the VDOT website is determined by the web content manager using state web design guidelines and any department guidelines, as well as Federal Section 508 requirements. Every attempt is made to keep paragraphs to one or two sentences using precise and clear statements.

**Publishing.** Publishing of information on VDOT's website is managed through the Communications department, which works with staff (including cartographers, videographers, and graphic artists) to present information using visualization (e.g., pictures and maps) to make the information more interesting and understandable. The basic guideline is that whatever is not sensitive or proprietary can be published on the Internet. Published material is generally in a PDF format to facilitate reading of the material using Adobe Reader as the free reader software.

**Transformation Technologies.** The primary format for distribution of information from the web is PDF format. Although much of the original source content may reside as an MS Excel spreadsheet, MS Word document, or MS PowerPoint presentation, these documents are typically converted or transformed into a PDF format before being published on the website. This is because Adobe Reader is more prevalent than Excel Reader which keeps VDOT in compliance with Section 508.

**Security Technology.** VDOT has technology to control access to staff responsible for maintaining the WCM system, records management functions, and the SharePoint system (ECM system). The types of security include controlled access through secure logins. For instance, in the case of the website, the content can be accessed from outside the firewall using a secure login. VDOT also uses Software as a Service (Saas) provided by Crown Peak.

The Knowledge Management office routinely reviews the metadata in SharePoint for accuracy to ensure that the right content gets to the right audience, prior to distribution and dissemination of information.

Procedures and tools also need to be used to ensure that transportation information is retrievable (findable and accessible) in a timely manner. In support of this, VDOT uses FAST search for the intranet. VDOT uses multiple search scopes, including one that will search on selected external sites such as FHWA.

The library uses interlibrary loan functions and subscription databases to obtain and distribute documents and information as needed. The documents may be in a hard copy (original or photocopied) or in electronic formats such as PDF or HTML. One of the databases—Business Books Summaries—is like Cliffs Notes and can be used to locate particular titles. Most library use happens on line with the patron engaging in self-service as opposed to having to travel to the library to be served physical objects by library staff. Other types of information, such as datasets, are also available in data files from various VDOT offices. The InsideVDOT intranet is also accessible for locating information for internal users while external users can access the VDOT website.

### *Challenges*

- Lack of executive-level support for information management, because of lack of understanding of the importance of data and information management;
- Lack of support from the staff for a governance framework to direct the management of information, because of a lack of understanding of the staff's critical role in implementing and supporting governance at the agency;
- Lack of IT support for secure access/authentication into online resources, especially those hosted outside the firewall, and policy restrictions on sites that are allowed to be visited, not to mention software that monitors time spent on line, which can have a chilling effect on use; and
- The existence of disparate sources of data and information, which may reside in what are referred to as silo information systems. These silo information systems often were developed to address a particular need in a specific business area without regard for whether or not the system might duplicate data collection of specific types of data to support the system.
- Challenges associated with information management practices are as follows:
  - Understanding what VDOT employees want and how to provide it to them;
  - Responding to the continuous change in work force at VDOT by marketing the available services provided by the VDOT library;
  - Consistency is needed in identifying policy documents because they may take several forms;
  - Better metrics are needed to see how much usage is occurring for library holdings; and
  - Need to get the highest quality publications at the best prices and get this information integrated with the online catalogs to let everyone know what is available.

### *Next Steps*

- VDOT will continue to evaluate the types of information needed from a high-level perspective and will continue to coordinate deploying this information to the extranet.
- VDOT will continue to use InsideVDOT and SharePoint to make publications usable for the customers.
- The library will continue in its role as facilitator among the IT Division, publishers of material, vendors, and patrons to identify the types of materials and publications needed.

### *Lessons Learned*

- Capture lessons learned from previous (project) experiences;
- Document information from previous projects and make it available across the agency (such as through the use of a KM system);
- Use consistent indexing terms for quickly locating and accessing the repository of lessons learned; and
- Do not try to do everything at once. Focus on what services can be provided by each specific office responsible for information management whether it is the library at the DOT, the Knowledge Management office, or the Administrative Services Office.





## Case Study: Washington State DOT (WSDOT)

Primary Sources: WSDOT staff from the Office of Research and Library Services within the Strategic Planning Division, and the Office of Information Technology.

### Background

**WSDOT Library.** The WSDOT library serves as a repository of agency publications and provides library services for agency employees and contractors, including reference support, circulation, interlibrary loan, and news alerts. Information management exists in many forms, across multiple departments within WSDOT and has been a priority of the agency for more than fifty years, starting with the establishment of a technical library in 1968, the Highways Library as a branch of the Washington State Library (WSL). In 2003, internal management of the library was reassigned from Finance & Administration to the Research Office with staff-level increases and online resources purchased in 2004. The library collection contains more than 40,000 items in multiple formats, and the library provides access to engineering- and transportation-specific database products. The library's mission and focus is to identify, locate, and obtain information and materials for WSDOT employees and contractors statewide. The library is open to the public for in-house use of materials and the public and employees of other state agencies may borrow materials.<sup>27</sup>

**Summary Enterprise Information Management Initiatives at WSDOT.** In 2011, the WSDOT Office of Information Technology prepared a Summary Overview of Enterprise Information Management Initiatives at WSDOT from 1998 to 2006. As part of this, several papers, reports, and studies were compiled to provide a brief history of information management initiatives. Common issues cited under this summary included duplication of data, the need for more integrated systems, and the information systems being siloed because of the autonomous structure associated with WSDOT's business units.

In 1999, a Data Council committee of WSDOT mid-level managers from diverse business areas organized to address data access and documentation.

The purpose of the Data Council is to place accountability for data resources in the hands of departmental employees with a stake in its quality and protection. The Data Council will provide a forum for the Data Stewards, Data Creators, Data Users, and Data Resource Management to work cooperatively in the best interests of the department.<sup>28</sup>

The Data Council goals involved providing leadership and guidance to

1. Maximize the consistency of data definitions and values throughout the department,
2. Minimize the cost of collecting and maintaining accurate data within the department,

<sup>27</sup>History of the Washington State Department of Transportation Library, August 2005.

<sup>28</sup>Gordon Kennedy. A Summary Overview of Enterprise Information Management Initiatives at WSDOT from 1998 to 2006. WSDOT Office of Information Technology. October 4, 2011.



3. Encourage sharing of accurate and timely data throughout the department, and
4. Promote data as a departmental asset.<sup>29</sup>

The Data Council was successful in creating a metadata repository and establishing data stewardship as a business function. The Council also helped to acquire legislative funding in 2001 for development of the WSDOT Data Catalog.

In August 2005, the Document Management and Workflow Work Group (DMWG) formed and its mission was to

Develop a high-level strategic plan and policy to guide future implementation of EDM, workflow, imaging, web-based content, and electronic signature technologies for all organizations within WSDOT.<sup>30</sup>

In late 2005, several reports were prepared assessing several critical WSDOT information systems. The need for consistent, sharable information was a key theme. A challenge to this is that the existing core applications are highly siloed so they are not able to easily or quickly share information. The recommended solution included

. . . the implementation of a central project repository that would contain all related information and provides collaboration tools that promote sharing of project information.

The recommended solution . . . includes the capture and storage of complete, consistent, sharable information about WSDOT-owned and related . . . assets.<sup>31</sup>

In June 2006, a report by the Statewide Program Management Group (SPMG) identified problems with conflicting data and lack of confidence in the integration of data. An outcome was the development of the Project Management Reporting System (PMRS) in 2011, which provides the agency with a broad set of software tools to integrate schedule management, contract management, cost management, earned value management, and electronic content management to better support management and delivery of capital transportation projects.

WSDOT is discussing how to facilitate enterprise information management, including streamlining data collection, cost efficiency, clarifying authoritative sources, and improving findability.

### *Responsibility*

Many offices and individuals throughout WSDOT are responsible for information management because it is spread over different divisions and districts. A few of note include

- Strategic Planning Division—Office of Research & Library Services and Statewide Traffic & Collision Data/Geographic Roadway Data Office;
- Information Technology Division—Data Management Services, which includes Data Resource Management (implementation and operations of databases) and Information Resource Management (design of and reporting from databases);
- Communications Division—Internal and public website management; and
- Enterprise Risk Management Division—Records Management Office.

Business units also maintain independent information resources.

Within WSDOT's Enterprise Risk Management Division, the WSDOT Records Management Office manages the records program; coordinates the safeguarding of records, records inventory,

<sup>29</sup>Gordon Kennedy. A Summary Overview of Enterprise Information Management Initiatives at WSDOT from 1998 to 2006. WSDOT Office of Information Technology. October 4, 2011.

<sup>30</sup>WSDOT. *Improving Information Accessibility*. August 7, 2006.

<sup>31</sup>Gordon Kennedy. A Summary Overview of Enterprise Information Management Initiatives at WSDOT from 1998 to 2006. WSDOT Office of Information Technology. October 4, 2011.

retention, destruction and release of public records; provides central records services for engineering documents; and administers the distribution of plan sheets and office copies of contracts.

### *Good Practices*

**Capture.** Each information resource manages the capture of information independently. Requirements around types of information may be perceived differently, although policies and standards apply and guide their capture. These types of information include data, images, documents, and websites. Any data captured or information created is a record of the department. Some of these records are also published.

Efforts to capture WSDOT's information have included

- As-builts were scanned into the Records Management Information System (Accordé database) by Engineering Records.
- The WSDOT library obtains and maintains print monographs and serials and electronic and digital publications. Librarians send email reminders to senior managers requesting they send necessary documents to the library.
- Photographs (digital and historic print) are being captured, maintained, and accessed using a wide variety of media and methods by Engineering Records, communications, the Aerial Photography lab, Geotechnical Services, and regional project offices.

**Administer.** Information at WSDOT exists in various forms (e.g., data, images, documents, and other information resources) and formats (e.g., paper, websites, database, and electronic). Records must be secured and access must be controlled by agency procedures. In addition, WSDOT publications must be submitted to the library so that access to these items is available to employees and the public, as appropriate.

WSDOT's IT Manual includes a standard on data management practices (400.01) with principles for managing data as a departmental asset. The manual includes standard definitions of procedures, roles, and responsibilities.

**WSDOT Policies, Practices, and Governance.** WSDOT has several IT standards and policies associated with information management:

- Executive Order E 1037.01, Electronic Information Management, . . . directs employees to protect and secure electronic data and information. Electronic data and information are valuable agency resources whose management is critical to the success of the department's mission. Electronic data and information are records subject to the same laws and rules as records that are not electronic.<sup>32</sup>
- WSDOT's OIT Standard 400.01 Data Management Practices defines management practices for the planning, design, creation, and maintenance of data at WSDOT. This standard complies with and supports the Washington State Technology Services Board Policy No. 401-S4, Information Technology Security Standards, Securing Information Technology Assets and employs the principles described in the FHWA's Asset Management Primer (U.S. DOT 1999) and the Data Management Association's Guidelines to Implementing Data Resource Management (DAMA, 2001). Data includes records of any kind stored on electronic media under the jurisdiction of WSDOT.
- WSDOT's OIT Standard 400.04 *Database Design and Data Modeling* defines the methods and techniques to be used in the design of WSDOT computer databases. These standards are intended to meet WSDOT's information needs, ensure consistent meaning and usability across all business functions, speed application development, reduce maintenance, protect data integrity, and prevent data redundancy.

<sup>32</sup>WSDOT Executive Order E 1037.01 Electronic Information Management. September 19, 2011.

All state agencies within Washington are also required to comply with the Policies & Standards—State Technology Manual. Within this manual the Office of the Chief Information Officer (OCIO) established standards and policies for the consistent and efficient operation of information technology services throughout state government. WSDOT meets OCIO standards, within staffing and resource requirements.

In addition, the *WSDOT Traffic Manual*, Chapter 11, “Traffic Engineering Records Management,” contains guidelines for managing traffic engineering records. State agencies are required to submit publications to the WSL. WSDOT used to have an internal policy requiring collection of agency publications but this was removed in an update of the Administrative Manual. A new policy is planned but has not yet been written.

**Categorization Schemes.** The WSDOT Library catalogs information using various approaches. For example, library collection materials, including agency publications, are categorized using Library of Congress subject headings and classified according to the Dewey Decimal System. The library indexes WSDOT research reports on the Internet using TRT terms and author-provided keywords. The report abstracts follow a basic metadata schema also used for describing the agency manuals produced by Publication Services.

Similar to other information systems within WSDOT, image collections exist throughout multiple departments and in many forms. Two WSDOT image collections have advanced indexing tools: the aerial photography collection and Stellent image system within the Records Management Division. The 2010 *Identification Needs in Developing, Documenting, and Indexing WSDOT Photographs (WA-RD 731.1)* research report assesses the current state of archiving and indexing images at the agency and recommends next steps toward developing an agencywide image documentation and access system.

WSDOT believes that, in general, taxonomies add some structure and improve information organization and retrieval within the business community they serve, although some could use improvement (e.g., such as adding user warrant, consistent tagging, search index optimization, and better relating of categorization schemas). Multiple categorizing schemes are needed because WSDOT is a complex and diverse organization. There are well-managed taxonomies for subsets of information, but they are not coordinated to look like a unified system to information seekers. An enterprise strategy on taxonomy could be beneficial, with proper indexing.

**Metadata.** Metadata varies by information resource. Geospatial data sets are documented in compliance with the FGDC’s Content Standard for Digital Geospatial Metadata. WSDOT also uses ISO 19115-2003 for geographic information metadata. ISO 11179 has been used by WSDOT for formulating definitions, specifically Part 4, Formulation of Data Definition, and Part 5, Naming and Identification Principles. Technical metadata is maintained for almost all databases supported by the IT Division. Business metadata (e.g., definitions, purpose, and meanings) is maintained by data stewards with assistance from the WSDOT Data Catalog Administrator.

WSDOT uses an in-house-designed metadata repository, WSDOT Data Catalog, to provide information about the data WSDOT owns. The WSDOT Data Catalog tells what data is available, where data is located, what it means, who collects the data, who is responsible for the data quality, and how to get access to the data. It was based on the Decision Package Metadata Repository for Cataloging Departmental Data for the 2001–2003 biennium. The data catalog is meant to reveal data as a departmental asset; create a unifying logical structure for all WSDOT metadata as a framework for building a comprehensive inventory of data elements; create a single access point for finding all metadata, including making the best use of existing metadata stores by downloading from them or linking to them; use Data Catalog to classify Data by its

Security Level (e.g., Public, Sensitive Confidential, and Confidential with Special Handling); and be a tool to manage definitions of agency business terms through a stewardship process.<sup>33</sup>

As of February 2012, the data catalog holds metadata (physical inventory) for

- 2,536 databases (including Production, Staging, Development, and Quality Assurance [QA] Environments);
- 103,783 Tables (including Temp, Staging, and Views);
- 1,510,234 Data Elements (including Development, Testing, Production, and Staging);
- All SQL and Mainframe systems managed by WSDOT OIT, Bridge Office, Regions, and Washington State Ferries; and
- ArcSDE databases (GIS).

In addition, the data catalog has an evolving categorization of business terminology that links to physical data metadata managed through a governance process with business stewards.

However, many regional and business systems throughout the department are not managed by OIT and thus not documented in the data catalog (e.g., estimated to be thousands of Excel spreadsheets and hundreds of Access databases). These data resources are not documented, but some may play vital roles in their business areas or regional offices.

The WSDOT Data Catalog content is accessible via either Browse or Search. The taxonomy is structured according to Business Topics as shown in Exhibit 2-13. It is based on the functions and important topics to WSDOT, with the exception of the Geospatial facet, which is based on ISO 19115 Topic Categories. A user selects a data subject from the Business Topics as illustrated in Exhibit 2-14. The data elements, tables, and databases are surrogate records of the objects (metadata). The schemas for each database are periodically harvested and loaded into the WSDOT Data Catalog, thereby creating the physical metadata inventory.

The data catalog search relies on the object name or term search in a definition entry (e.g., business term, database, table, or data element name). Search uses basic SQL commands for keyword search. It does not have fuzzy logic capability to retrieve like terms or near spellings—this limits obtaining a full list of relevant results. Exhibit 2-15 shows the Data Catalog Advanced Search.

WSDOT Data Catalog includes Common or Business Metadata, Shared Business Metadata, Data Subject Metadata, Characteristic Metadata, Characteristic Variation Metadata, Physical Metadata, Shared Physical Metadata, and Database Metadata.

In some cases, the terms for the data catalog are based on authoritative sources such as the *Manual on Uniform Traffic Control Devices* (MUTCD); ANSI Model Minimum Uniform Crash Criteria (MMUCC); ANSI D16.1-1996 Manual on Classification of Motor Vehicle Traffic Accidents, ANSI D20-2003: Data Element Dictionary for Traffic Records Systems; and AASHTO and TRB publications. Some terms are based on WSDOT-produced engineering publications and, occasionally, the TRT was used to verify terminology. The *Federal Enterprise Architecture Data Reference Manual* was also used to help frame terminology.

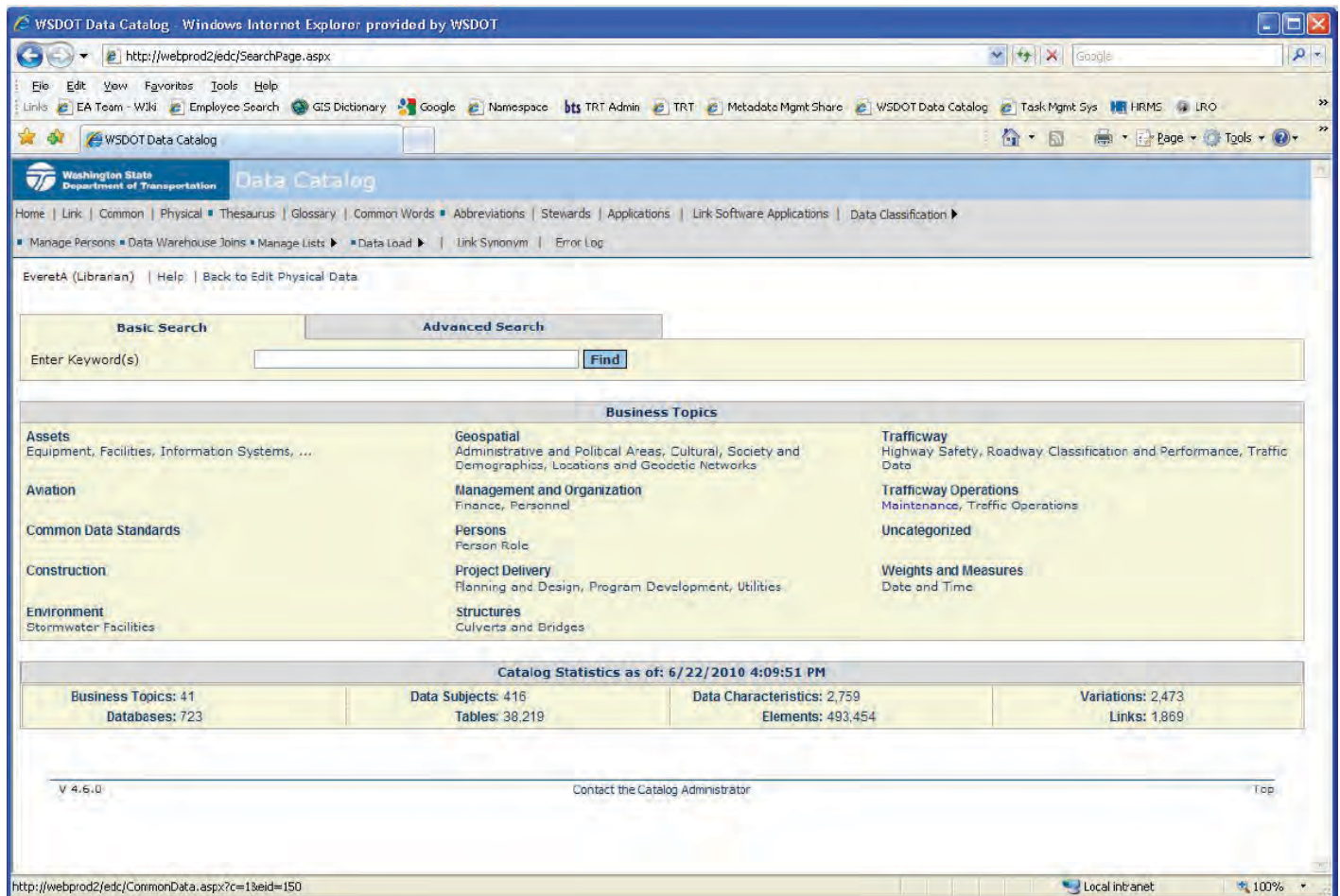
For new software design projects, the Data Catalog Librarian is involved from the beginning of a project to work with the business unit, data modelers, and programmers to determine the best terminology for the database, tables, and attributes within the application. The Librarian also helps the team define the data dictionary with Data Stewards making final approval of the content object's definition. Terms that may affect more than one business area will be developed or refined through mutual agreement and consensus. Not only is term selection important, but developing the meaning or concept of a content object is critical because it forms the basis of a common vocabulary for communications throughout the department.

---

<sup>33</sup>WSDOT, Office of Information Technology. WSDOT Data Catalog FAQ.

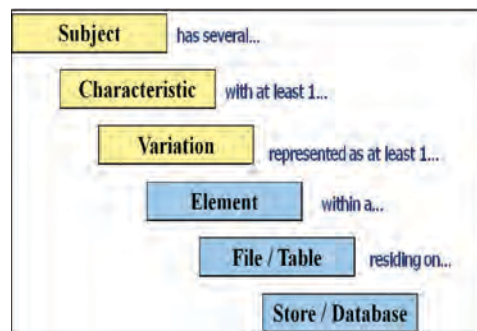


Exhibit 2-13. WSDOT Data Catalog (metadata repository).



Source: White Paper: Indexing Metadata Repository Records and Data Sets with the TRT, December 6, 2010.

Exhibit 2-14. WSDOT Data Catalog data subject structure.



Source: White Paper: Indexing Metadata Repository Records and Data Sets with the TRT, December 6, 2010.

**Exhibit 2-15. WSDOT Data Catalog advanced search.**

Source: White Paper: Indexing Metadata Repository Records and Data Sets with the TRT, December 6, 2010.

WSDOT recognizes that the primary intent for the development of a data catalog was addressed but there are opportunities to improve with new tools, streamline the taxonomy, and expand to a broader user community for acquiring needs and requirements. OIT suggested some future improvements including a more efficient connection with the agency's BI tool (Interactive Reporting Studio). The intent here is for the WSDOT Data Catalog to be the sole source for definitions for the agency and provide definitions to other applications starting with Oracle EPM Reporting and Analysis Studio (formerly Hyperion Interactive Reporting Studio). In addition, the data catalog would be improved with navigation allowing access from more than one hierarchy and the use of descriptive metadata such as Dublin Core.

**File Formats/Naming Conventions/Terminology.** WSDOT makes extensive use of Microsoft Office file formats for business documents, although there is no agencywide naming-convention standard. Business systems of longstanding importance to the agency (e.g., financial, personnel, collisions, traffic, and program management) are in mainframe systems using ADABAS and VSAM file formats. Business systems developed over the past 20 years use SQL Server databases. FileMaker Pro databases are used for forms-based and small-office tools. SharePoint is used in many situations for team collaboration. Engineering design drawings are developed in Bentley Microstation format and also maintained as paper prints. Many geospatial databases are stored in ESRI's geodatabase format.

Data resources developed with assistance of the IT Division follow standard naming conventions. The SQL Server database follows a published naming convention. Standardized terminology and categorization schemes typically exist within specific information resources but not one across the enterprise.

WSDOT's IT Manual 400.04 Database Design and Modeling Standard contains design practices, model format and layout, terminology, and naming conventions.

Information resources are stored in various locations throughout the department. Information storage includes but is not limited to physical documents and electronic resources. Physical documents are stored in

- The WSDOT library—The State of Washington requires that state agencies submit two copies of print publications and one copy of electronic publications to the state library (RCW 40.06.030). The Portable Document Format (.pdf) is the preferred format for electronic publications. Copies are also requested for the agency library. In addition to agency publications, the WSDOT library holds print monographs and serials, electronic journals, and digital publications.



- The plans vaults.<sup>34</sup>
- Digital and historic print photographs are maintained and accessed using a wide variety of media and methods by Engineering Records, communications, the Aerial Photography lab, Geotechnical Services, and regional project offices.
- Records are stored in offices until no longer in active use and then stored off site according to their retention schedule.
- Individual business units and employees retain copies of paper reports and documents, although there are efforts to migrate content to electronic format and reduce redundant storage.

Electronic resources are stored in

- The WSDOT Data Warehouse, which stores integrated agencywide digital data about finances, construction, collisions, traffic, and projects. A convenient user tool provides access to these resources.
- The GIS Workbench, which provides a user-friendly access point to geospatial data resources.
- The WSDOT Data Catalog, which is a comprehensive listing of data resources supported by the IT Division.

Electronic documents are stored in

- Office-specific network folders. Documents to be shared agencywide are stored on folders accessible across the agency—this appears to all system users as the Corporate W: drive.
- A OneDOT SharePoint environment, which also permits data to be stored and shared across the agency.
- Plans and as-builts, which are in the Records Management Information System. Project offices are expected to use the Livelink document management system for project information.

Access to and preservation of information is determined by state and federal laws and guidance pertaining to records, publications, Internet, and data management. Although these practices and disciplines overlap, a systematic and holistic approach to information management across these information sources does not yet exist. The practices help with managing information but WSDOT recognizes that further alignment of information management strategies would facilitate better management and access to information resources.

Records are sent to archives when they no longer need to be accessed on a regular schedule and are retained until their scheduled destruction. If records are identified as having significant historical value, they are preserved in perpetuity.

The WSDOT Traffic Manual contains a chapter on Traffic Engineering Records Management that includes a section on preservation. Public records must be retained for the minimum time period identified in the retention schedule (see Exhibit 2-16), which is based on WSDOT and Secretary of State guidelines.

BI environments (the Data Warehouse and GIS Workbench) are actively supported by business data stewards. IT staff committed to supporting the technologies work actively with data stewards to develop and maintain reports, respond to changing needs, and ensure proper access permissions.

WSDOT employs elements of KM to help organize and make findable information that users need, when they need it. The WSDOT library does this by cataloging information resources

---

<sup>34</sup>Plans vaults are rooms where design drawings are stored. These are typically in a physically secured room and stored in large, metal flat files. These document collections are vital historical reference materials for engineers when planning new construction or maintenance and repairs. Plans vaults exist in every regional office, the ferry division, and WSDOT headquarters.

**Exhibit 2-16. WSDOT public records retention schedule.**

General Correspondence	30 days
Interagency Agreements (MOUs)	6 years
Invitations to Bid and RFPs	6 years
Legal Issues (does not include litigation)	6 years, followed by archival
Personal Services Contracts	6 years, essential records
Program Development or Event History Files	5 years, followed by archival
Project Files <sup>1</sup>	6 years, followed by archival
Public Complaints	3 years
Purchase Contract Support Materials	6 years, essential records
Q & Low-Cost Enhancement Project	6 years, followed by archival
Design Documentation <sup>2</sup>	
Request Legislation and Bill Files	2 years, followed by archival
Staff Meeting Minutes	2 years
TRACTS Records	Infinite
WAC Rules and Traffic Regulations	6 years, followed by archival
Working Files	2 years

<sup>1</sup>Includes Critical Accident Location (CAL), Critical Accident Corridor (CAC), and Intersection Accident Location (IAL) analyses conducted by the Traffic Offices.

<sup>2</sup>The *Design Manual* recommends 75 year retention in a permanent retrievable file for Design Documentation Packages on construction projects.

Source: WSDOT Traffic Manual M 51-02.06, June 2011.

into a searchable online catalog in which users can locate relevant and available library materials; providing desktop access to selected information and knowledge resources via the WSDOT Library website on the department's intranet; providing value-added search tools and services such as literature searches and Daily News Clips delivered by email; indexing WSDOT research reports using the TRT; and advising business units on collections maintained by their offices.

In terms of specialized information dissemination services, the WSDOT Library creates products to meet specific information needs of specific users including WSDOT staff, legislators, state and local government stakeholders, and the public. Email news alerts are sent out daily or weekly to hundreds of subscribers on current topics, including high-occupancy travel (HOT) lanes/congestion pricing, public-private partnerships, and infrastructure stimulus spending. The library intranet site includes toolkits that provide helpful information on subjects, including the state's Plain Talk initiative and the annual session of the Washington Legislature.

WSDOT distributes geospatial data via a centralized site maintained by the OIT, the WSDOT GeoData Distribution Catalog. This data is used by WSDOT's transportation partners, government entities, schools, private businesses, and the general public. The WSDOT GIS community promotes interagency data exchange and resource sharing with the data being available for free. Data is provided in ESRI shapefile, georeferencedjpg, and multiresolution seamless image database (MrSID) formats. WSDOT's geospatial data is developed in coordination with other state agencies through the Washington Geographic Information Council (<http://geography.wa.gov/wagic/>).

The Statewide Travel & Collision Data Office (STCDO) within the Strategic Planning Division includes the Collision Data & Analysis Branch, the Travel Data & Analysis Branch, and the Management & Administrative Support Branch. Travel data for state highways, highway

performance monitoring, functional classification, and collision data for all roads in Washington State is collected, processed, analyzed, and disseminated by STCDO.

State standards require data-sharing agreements for sensitive data distributed outside the agency. The WSDOT Administrative Contracts office works with IT and business staff to develop proper external data-sharing contracts. Internal data sharing is documented with data-sharing agreements in the form of memoranda signed by the data stewards providing and receiving the data.

### *Challenges*

Information and data management tactics are mostly invisible to transportation professionals delivering the agency's core mission. The deliberate management of information and data as enterprise assets has recently become of interest, as some business leaders have experienced the costs of redundant, disorganized, undiscoverable, and poorly constructed data resources.

The focus of information management tends to be on databases because there are documented practices and standards. Practices and standards for text, spreadsheets, and image resources are managed by types of information providers (e.g., library, records, Internet) and by unique business units with fewer documented practices and standards.

There is lack of common understanding about the meaning of data, information, and knowledge. Understanding the distinctions and relationships would facilitate design of a user-based information management strategy and clarify roles and responsibilities.

Information literacy is low. This includes understanding about current information standards and practices, quality resources, scope and limitation of information resources, copyright/intellectual property considerations, costs of information management, security, and the responsibilities of individual employees in each of these.

Response by senior managers is uneven—many recognize the problem and understand the costs and opportunities involved but the relative priority of developing a solution varies. So long as business units enjoy autonomy in their budgeting and strategies for technology and information management, they will continue to develop information systems focused on the needs of their organization. This will continue to produce redundant, disorganized, undiscoverable, and/or poorly constructed data resources.

Lack of an enterprise information architecture and data governance that ensures compliance with information management standards and modern practices makes it difficult for business units to contribute to an enterprise information resource. Although there is growing interest in enterprise information management, there is no governance structure to guide development or manage implementation.

There are no negative consequences for poor information management. It is difficult to assess the actual costs associated with ineffective information management practices and limited findability and access to information. This makes it more difficult to justify improvements to information management.

Website management is distributed across business units and practices to preserve web content vary. This has led to loss of web content when sites are removed.

Sometimes employees express frustration and concern about search capabilities and inability to find information resources. This is, in part, because there is inconsistent categorization across information resources, independent taxonomies are often not linked together, and some of the information sources are not indexed. The metadata schemas for WA-RD reports and manuals was a step in improving some findability, as was the project Search page <http://www.wsdot.wa.gov/projects/search/>. Proper indexing, consistent tagging, improved search

engine optimization, and taxonomies that have been vetted by the user community using card sorts, focus groups, and so forth could help improve search results. An enterprise common vocabulary/thesaurus would also help improve information management.

### Next Steps

- In August 2012, an IT board was formed at the direction of the WSDOT Secretary of Transportation. The intent of the group, as tentatively understood now, is to be a decisive body backed by policy to discuss IT project prioritization. This might provide enterprise leadership on some of the agencywide data management challenges identified herein.
- A statewide project to move all time reporting and leave management to an IT solution is now underway. This presents WSDOT, along with other state agencies, the opportunity to reconsider how employees interact with information resources.
- A policy on personal devices (bring-your-own-device or BYOD) could help modernize business processes, including some aspects of information management.
- A plan for asset data management is being developed.
- Discussions continue about governance and early steps for broader information management.

### Lessons Learned

- Understand the organization's business needs for data and information.
- Communicate to bridge gaps in awareness and understanding among managers of information resource types and between business units and information management professionals.
- Understand that content is different from the technology system that it resides in. The information content may be used by business units other than the primary creators/users. The content may also be needed decades later. Managing content so that it can address this business need requires attention. Library science has demonstrated successful strategies for managing content across diverse user groups and extended time periods.
- We all create and use information daily. We are all responsible for managing our portion of information responsibly. Having clear principles and intent and strategy to carry them out is critical to success.
- It is beneficial to have staff resources and knowledge that consider the big picture for information management. It is also valuable to have staff who can investigate experiences, guidance, and standards associated with information management to determine usefulness and usability.
- Delay will not improve the situation.

### Sources

1. Leni Oman, Office of Research and Library Services, Director
2. Kathy Szolomayer, WSDOT Library, WS Head Librarian
3. Gordon Kennedy, OIT, Data Management Services, Information Resources Manager
4. Andy Everett, IT Data Catalog Administrator
5. Endicott-Popovsky, Barbara, and Simon, Michael. Center for Information Assurance and Cybersecurity Information School; University of Washington. *Identification Needs in Developing, Documenting, and Indexing WSDOT Photographs*. WA-RD 731.1. Washington State Department of Transportation. February 2010.
6. Everett, Andy. White Paper: Indexing Metadata Repository Records and Data Sets with the Transportation Research Thesaurus, December 6, 2010.
7. Federal Geographic Data Committee. Metadata Quick Guide. April 2006. <http://www.fgdc.gov/metadata/documents/MetadataQuickGuide.pdf> (accessed September 12, 2012)
8. Federal Enterprise Architecture Program. The Data Reference Model, Version 2.0. November 17, 2005. [http://www.whitehouse.gov/sites/default/files/omb/assets/egov\\_docs/DRM\\_2\\_0\\_Final.pdf](http://www.whitehouse.gov/sites/default/files/omb/assets/egov_docs/DRM_2_0_Final.pdf) (accessed September 12, 2012)

9. History of the Washington State Department of Transportation Library. August 25, 2005.
10. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) 11179, Information Technology—Metadata registries (MDR). <http://metadata-standards.org/11179/> (accessed September 12, 2012)
11. Kennedy, Gordon. A Summary Overview of Enterprise Information Management Initiatives at WSDOT from 1998 to 2006. WSDOT Office of Information Technology. October 4, 2011.
12. Library Collections, Records Management and Archives – WORKING DRAFT.
13. Office of the Chief Information Officer (OCIO). State Technology Manual STANDARD NO. 141.10, Securing Information Technology Assets. April 10, 2012. <http://www.ofm.wa.gov/ocio/policies/documents/141.10.pdf> (accessed September 12, 2012)
14. WSDOT OIT. WSDOT Data Catalog FAQ.
15. WSDOT. Traffic Manual M 51-02.06. Engineering and Regional Operations. June 2011. <http://www.wsdot.wa.gov/publications/manuals/fulltext/M51-02/Chapter11.pdf> (accessed September 12, 2012)
16. WSDOT OIT. IT Manual, Standard 400.01 Data Management Practices. June 1, 2011.
17. WSDOT OIT. IT Manual, Standard 400.04 Database Design and Data Modeling. June 6, 2011.
18. WSDOT Executive Order E 1037.01 Electronic Information Management. September 19, 2011.
19. WSDOT Executive Order E 1029.00 Information Technology Asset Management. August 6, 2004 reviewed August 15, 2010.
20. WSDOT. Managing Knowledge at WSDOT – The Information You Need When You Need It. August 2007. <http://www.wsdot.wa.gov/NR/rdonlyres/7BEA6684-B483-4B20-8A18-44ACE7CF7417/0/ManagingKnowledge.pdf> (accessed September 12, 2012)
21. WSDOT. Interagency Data Sharing Agreement.
22. WSDOT. Improving Information Accessibility. August 7, 2006.
23. WSDOT. WSDOT Library. September 2010.



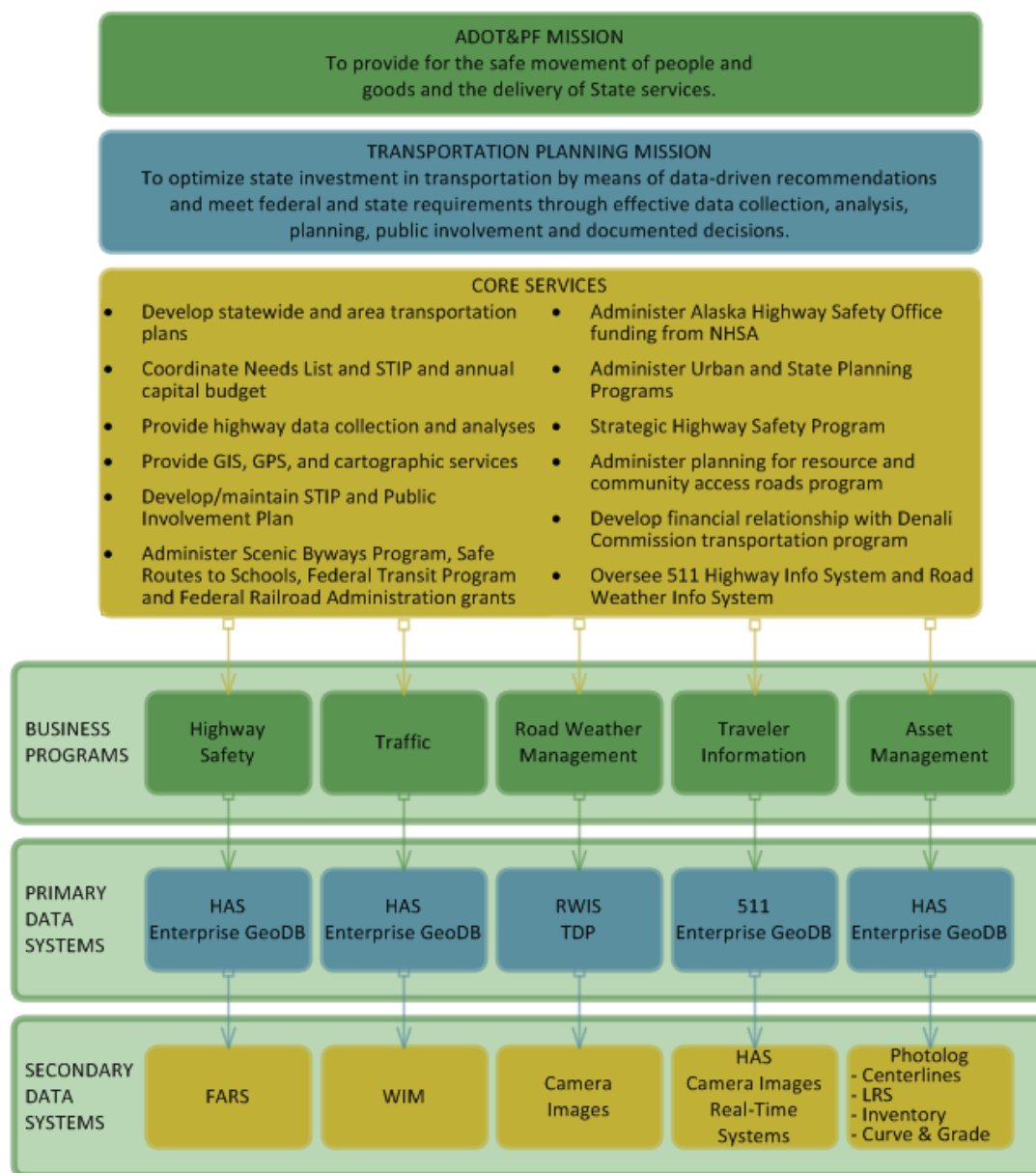
### **Example: Alaska Department of Transportation and Public Facilities (ADOT&PF)**

The ADOT&PF is a good practice example of how the implementation of a DBP helps to improve managing data and information systems. The DBP incorporates policies, standards, methods, and tools, which are needed to ensure that quality data and information are provided to decisionmakers in a timely manner. A framework was developed to demonstrate how the primary and secondary data systems support the Transportation Planning Mission and the overall Mission of the department. Exhibit 2-17 illustrates this relationship.

The DBP was developed as a multi-year, multi-phase effort, which began in 2005. Exhibit 2-18 illustrates the phased approach for development of the DBP, including a Data Action Plan during Phase 3 to guide the implementation of the DBP.



**Exhibit 2-17. ADOT&PF’s mission and core business programs.**



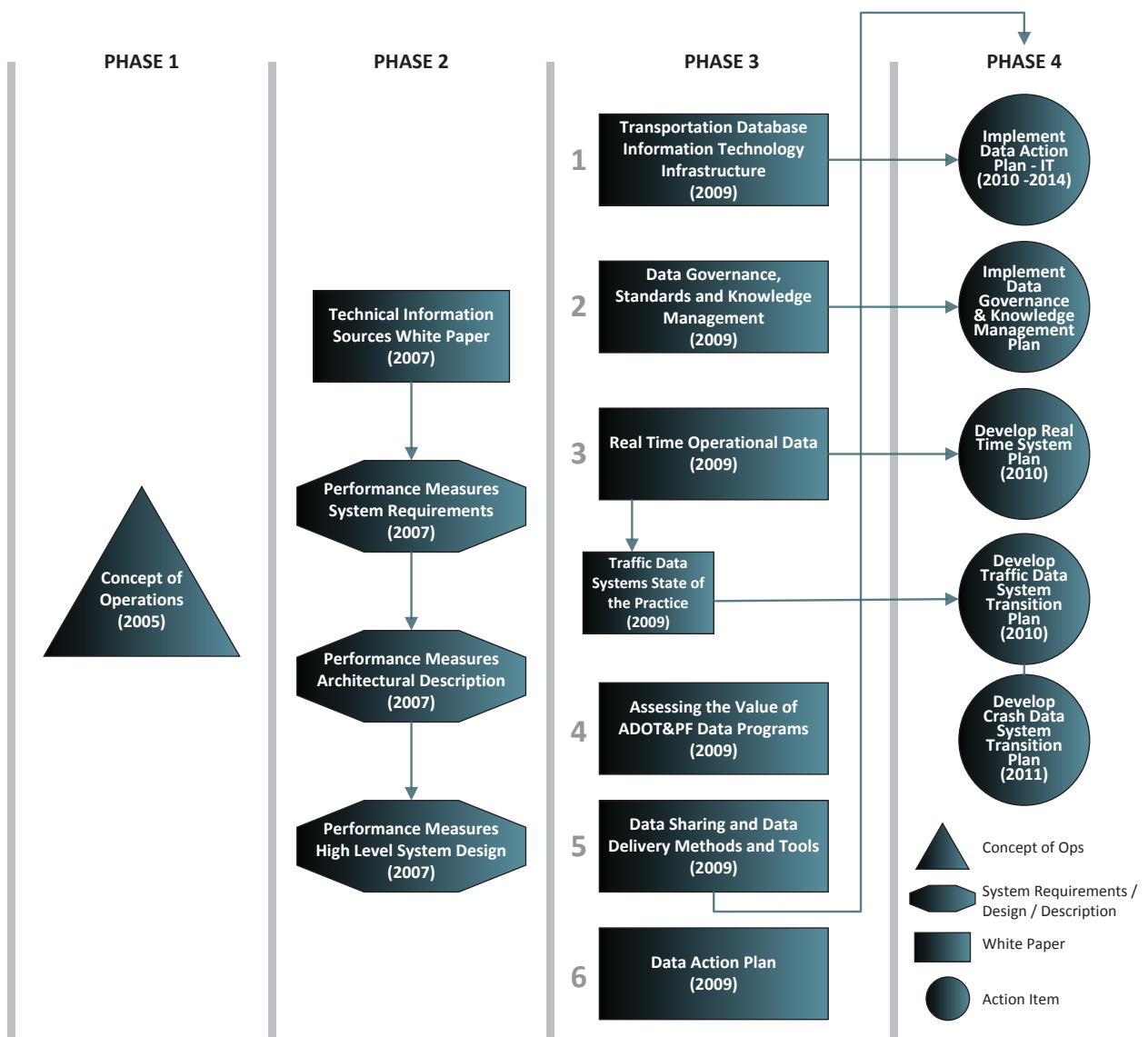
Establishing the Data Action Plan allowed ADOT&PF to focus on the highest priority data and information needs first, as part of its data management program. The Data Action Plan includes recommendations for implementation of technology, processes, and standards.

Notable outcomes at the Program Development Division and the department because of the implementation of the DBP include the following:

- The document puts into perspective the key role served by the Program Development Division in the overall department operations. If data and information are not produced in a timely and efficient manner, the job of servicing the huge customer base is not well done.
- The work flow process diagrams in the DBP are very important because they identify where the division could be more efficient. The Community of Interest (COI) diagrams



Exhibit 2-18. ADOT&PF's phased DBP development.



are extremely important because they illustrate how extensively the customers rely on data provided by the division. The considerations of the needs of the COIs are part of the decision-making process when planning for hardware/software upgrades to improve delivery of data and information to them.

- The DBP also helped the staff to more clearly see their role in providing critical data and information to support their business areas.

Other direct changes that came out of the DBP are that staff now speaks in terms of the five program areas and addresses them as such. It keeps things in perspective, especially as the division transitions from the existing Highway Analysis System that is not GIS-based into a GIS environment.

The department still has ongoing efforts to establish an enterprise data collection plan and an enterprise policy and procedure for data governance, which will incorporate some recommendations from *NCHRP Reports 666* and *706*.

### Lessons Learned:

Use of a DBP at ADOT&PF helps the agency to

- Establish goals for data programs,
- Assess agency data programs,
- Establish a data governance framework, and
- Ensure proper use of technology/tools to support data management.

Once implemented, a DBP needs to be reviewed regularly (at least annually) and updated to reflect any changes in policies, standards, business processes, technology tools, or organizational realignments that may affect management of data programs and information at the DOT.

Establishing COIs (for transportation data programs) and continuing outreach to the COIs helps the DOT to make appropriate decisions on developing/implementing new hardware/software solutions to support data programs relied on by the user community.

Use of a DBP helps everyone in the DOT to understand their role in providing high-quality data and information in a timely and efficient manner for all stakeholders.



### Example: Arizona DOT (AZDOT)

The Arizona Memory Project provides access to the primary sources in Arizona libraries, archives, museums and other cultural institutions. Visitors to the site will find some of the best examples of government documents, photographs, maps, and objects that chronicle Arizona's past and present.

The project enables Arizona cultural institutions to exhibit digital collections online and serves as a central online repository for access to digital content representing the history and culture of Arizona. The Arizona Memory Project was launched in March 2006 and has been granted the Arizona Centennial 2012 legacy project designation by the Arizona Historical Advisory Commission.

The project includes more than 95 exhibits with more than 67,000 full-text searchable digital objects (but it still only represents a small percentage of materials held by contributing institutions). The project adds, on average, 1.5 collections per month. Collections include photographs, maps, state and federal documents, and oral histories and video representing over 50 Arizona museums, libraries, historical societies, and schools.

In 2010, over 10,000 promotional fliers and cards were distributed to partner repositories and the public. Each month, Internet statistics are distributed to partners indicating the number of web visits per collection. Monthly statistics also show that the project receives over 1.2 million unique views per year.

The project is an Open Archives Initiative-compliant digital library and is a registered data contributor on OAIster.org. This project is made possible in part by a grant from the U.S.

Institute of Museum and Library Services to the Arizona State Library, Archives and Public Records under the provisions of the Library Services and Technology Act.

#### Lessons Learned:

- Market the type of information available from the source agency to the widest audience possible, using newsletters, fliers, meetings, and the Internet.
- Maintain a central location or web portal as a shared space to store multiple types of information provided from many sources. A single portal with easy access also enables the quick search and retrieval of information by the user community.
- Use WCM software to document statistics on the use of the website (daily, weekly, monthly, annually). This enables the sponsoring agency to make adjustments in the content displayed on the website as needed. In the case of transportation information, this is particularly useful for supporting emergency management operations by providing additional information on evacuation routes or road closures during times of emergencies.

All Arizona agencies are required by law to have a records management program in place to systematically link business processes to records in paper or electronic format to

- Capture or create (record) the information necessary to support and document the process.
- Ensure that the records are accessible (can be located) as long as they are needed.
- Retain records as long as they are needed to support the entire process (including reference after the transaction that generated the record is completed). (Those retention periods are defined on a records retention schedule developed specifically for the agency or on a general schedule issued by the Library and Archives Records Management Division.)
- Ensure that the records are protected from unauthorized alteration or loss.
- Dispose of records properly, either by destruction or transfer to an archives.
- Balance the costs of records programs (including the costs of programming sophisticated recordkeeping functionality into an electronic recordkeeping system—ERS) against the value of and risks associated with the records.

Laws and regulations often dictate retention periods for specific types of records or place requirements on the content or nature of those records. Agencies must be able to produce all records relevant to litigation. Records that are relevant to pending or current litigation must be preserved, even if the retention period has passed.

The Records Management Division offers workshops and consults with government agencies to ensure that the agencies have an effective and efficient records management program in place and operates a center for storing inactive records pending disposal. One of the Division's most important functions is to work with agencies to develop records retention schedules indicating how long record series must be kept. These retention schedules are located at [http://www.azlibrary.gov/records/documents/pdf/all%20-%20management\\_ocr.pdf](http://www.azlibrary.gov/records/documents/pdf/all%20-%20management_ocr.pdf). For example, this document states that all paper publications produced by a public body (including brochures, pamphlets, newsletters and other published reports) must be retained for a year after superseded or obsolete, unless otherwise specified in this retention schedule and after sending two copies to ASLAPR, Law and Research Library, State Documents.

Disposition is the final chapter in the records lifecycle, resulting in destruction of the records or their permanent, archival retention. Arizona laws establish a process that determines which records are to be destroyed and how long those records must be kept before destruction, as well as which records must be kept permanently in the state archives. These laws apply to all records, regardless of format. The ability to demonstrate that records were disposed of legally and routinely is a critical defense against charges of spoliation or tampering with evidence in the case of litigation.

### Lessons Learned:

- Policies regarding preservation of records in paper or electronic format should be instituted by law and/or supplemented with department policies at the DOT. Such laws or department regulations for the proper retention and/or disposal of records are critical in defending the agency against charges of spoliation or tampering with evidence in the case of litigation.
- Policies and laws regarding retention/disposal of records should be made available to all DOT employees who may be responsible for managing records, such as reports, documents, photos, digital images, and videos, that may contain information that can be used in litigation or other types of inquiries.

As federal or state laws or department policies governing such procedures change, the records retention/disposition procedures used at the DOT should be updated as needed and re-distributed to the appropriate department personnel.



### Example: Colorado DOT (CDOT)

CDOT illustrates how data governance is used to support managing performance data and information at the department. CDOT recognized that several benefits could be derived from the implementation of a DBP and data governance framework for performance data, based on similar experiences at other transportation agencies. These agencies have demonstrated success in using governance to improve management of their core data programs, which are used to support decision making across the organization. CDOT, therefore, developed such a plan targeted for performance measures in 2011 to improve

- Identification of priority performance measures,
- Identification of critical data for those measures,
- Data quality assurance and control methods,
- Interfaces with a data management system (for performance measures), and
- Consistency in use and reporting of data and to minimize the burden of reporting.

Many steps were involved in developing the data governance plan for CDOT.

Step 1: Assessment of CDOT's Strategic Planning process and its Performance Management (PM) program. This evaluation included identifying existing performance measures and making recommendation for improved or new performance measures.

Step 2: Identify data sources used to support the performance measures.

Step 3: Develop a Data Catalog to document the data source for each performance measure, data business owner, stakeholders, applicable data definitions, data formats, and IT & Business SMEs. Documenting this information in a Data Catalog format helps to identify the core data systems and roles and responsibilities of both IT and Business units for managing data programs (in this case, data used for performance measures)

Step 4: Evaluate other good practice examples for applicability to CDOT.

Step 5: Evaluate state of data governance at CDOT to assess data needs and gaps and develop a data governance plan to address those needs/gaps.

Step 6: Recommendations for Data Governance:

- Create a formal data governance structure with an Oversight Committee and Data Governance Working Group;
- Develop a Data Governance Charter;
- Complete a Data Inventory/Assessment;
- Develop and adopt a Data Governance Procedure, which includes a framework, glossary of terms, description of roles/responsibilities, and identification of stakeholders and business data owners and IT sources needed to maintain the critical data systems;
- Perform a Risk Assessment of Data Programs; and
- Implement Ongoing Data Management and Governance at CDOT (develop a Data Governance Manual, formalize and document data standards, develop a communication plan to market the impacts and benefits of data governance to the department).

Lessons Learned:

- Following a multi-step approach for implementing governance similar to the approach at CDOT ensures that assessments of existing programs are completed and needs and gaps are identified so they can be addressed as part of a DBP using data governance.
- As part of the data governance framework, CDOT also identified good practices in the use of technology/tools used to display data and information pertaining to performance measures.
- Data governance at CDOT provided enabled them to coordinate a large number of stakeholders through the use of stakeholder groups organized around specific data programs used to support performance measures.



### Example: Georgia DOT (GDOT)

GDOT is a good example of how a Data Governance framework and the use of BI tools help to manage data programs at the DOT successfully.

GDOT has an official Data Governance Policy that sets the IT Policy for Data management strategy for GDOT. [ . . . ]. The purpose of GDOT's data management strategy is to provide data solutions based on standardization and QC for core reference data, which is the fundamental business data in GDOT (GDOT Publications Policies & Procedures, Policy 8075-2-GDOT Data Management Strategy, September, 2009). The IT Director/Chief Information Officer (CIO) is responsible for enforcing this policy.

GDOT also has a Data Resource Management Steering Group, composed of Business Unit Data Stewards (BUDS) and the Information Technology Data Executive Administrator – IT Application Administrator (ITAA). This group is empowered to make decisions about changes to information systems at the DOT. They meet about twice a year and are charged with handling any issues presented to them. A flowchart was developed to outline the roles and responsibilities

of those persons/offices responsible for administering data governance at GDOT. This document defines the roles for several functions including

- Business Unit Executive Data Sponsor;
- BUDS;
- Business Unit Data SME;
- GDOT CIO;
- Information Technology Data Executive Administrator/ITAA;
- Data Management Team Leads;
- Database Administrators, Database Developers, and Data Architects;
- IT Security Architect;
- Data Resource Management Steering Group; and
- Data Architecture Work Group (DAWG).

All electronic documents are stored in SharePoint, which provides for change control on all documents, security controls, and attaching metadata to documents so that they can be queried and easily searched. Microstation Project Wise is used to centralize and store all new design plan documents not stored in SharePoint. Project Wise will be used to perform change control on these documents and will serve as the data store for the TIF images of design plans.

GDOT will also implement intelligent searching, which includes stamping documents with tags and metadata to indicate the content of the document and to make recommendations on where to begin searches for particular information, including photos and videos. Because of GDOT's adopting an enterprise approach to storing data and information and using GIS technology, software development has been reduced from years to weeks, and the same size staff can now serve six times the number of customers.

#### Lessons Learned:

- Implementing Data Governance has helped GDOT establish policies, roles, and responsibilities, which can be clearly communicated to all staff responsible for managing data and information at the department.
- GDOT is benefiting from the partnering of business units with the IT office through the Data Resource Management Steering Group. This group is empowered to make decisions about changes in information systems at the DOT. This helps ensure that business needs are considered as part of any enhancements to application systems.
- Software development time has been drastically reduced because of the use of BI tools within a GIS environment.
- Use of BI tools also helps to ensure that critical documents are secured and properly stored in systems such as SharePoint and Project Wise, with metadata and tags to identify the purpose of the documents. This helps ensure that the correct information can be located to respond to requests.



#### Example: Iowa DOT

Traditional methods for storing digital photographs, images, or videos within a content repository require users to import the digital files and manually key the index information that



correlates to the files to make them available to all other system users via document querying. Typing, reviewing, and editing the required index information can be error prone, time consuming, and tedious.

With the software being applied at Iowa DOT, PixLogic, index information for digital photographs, images, or videos will be indexed automatically, and that index information will be passed to e.Power CME, Iowa DOT's ERMs. This will allow users to store digital photographs, images, and videos in the repository without having to index files manually. The architecture of the technology being considered will, at a system level, compare images against a visual repository of shapes and colors stored in the image database libraries and automatically analyze the content of the images or video frame by frame. The software enhancement will be able to identify objects such as trucks, cars, and people, and compare the images to the objects on file to determine accuracy of match. Images may then be queried visually based on how similar the objects are to the comparison image provided by the user; matching images are presented to a visual hit list for the user to review.

#### Lessons Learned:

- Automated indexing software for cataloging digital photos, images, or videos, such as that used at Iowa DOT, can be implemented at other DOTs to eliminate time-consuming, error-prone manual processes associated with indexing.
- Enhancements can also be made to the indexing software as business needs dictate. These include the need to distinguish between images and objects that appear to be trucks, or cars, or pedestrians.
- Automated indexing software can also be used to provide easier access to a larger user base seeking photo and video information. The indexing information about photos and videos can be transmitted to other enterprise records management systems to facilitate retrieval of the photos and videos from their original repositories.



#### **Example: Kansas DOT (KDOT)**

KDOT's realization that social media has become part of the way the government connects to the governed led staff to establish the KTOC, as well as accounts on Twitter, YouTube, Flickr, Facebook, blogging and more, to get messages to Kansans.

KDOT staff established a very efficient way for Kansans to access information and communicate with the DOT. In 2009 and 2010, KDOT staff established at least ten Twitter accounts that attracted more than 4,200 followers. They posted some 13,500 tweets on specific road conditions, live coverage of events, public meetings, legislative debate, safety messages, and so forth. The number of followers and tweets has grown every day.

KDOT staff also posted videos (including a video news release) on YouTube that have been viewed thousands of times; hundreds of photographs have been viewed on Flickr; and KDOT is involved in several Facebook pages related to projects or safety.

KDOT staff established KTOC as a forum for transportation discussions between not only KDOT and citizens/stakeholders, but between those in that latter group. They have used KTOC to post important information about project selection, project scoring, and so forth, and have encouraged the members (more than 1,000 ) to participate in the dialog. KTOC also hosts blogs and, during the 2010 Put the Brakes on Fatalities Day campaign, Governor Parkinson and U.S. Transportation Secretary Ray LaHood submitted blogs written specifically for KTOC.

#### Lessons Learned:

DOTs can use social media venues such as Twitter, YouTube, Facebook, Flickr, and blogs to provide information to a wider audience interested in transportation-related information. This may include the general public or stakeholders who use the data and information in their daily business operations.



#### **Example: Louisiana Department of Transportation and Development (LaDOTD)**

LaDOTD has developed an Enterprise GIS system that provides access to geospatial information and analysis across the entire organization. LaDOTD staff has created an environment where GIS is available to all professional staff as a tool to perform work. The system provides both internal and external access to the organization's information assets, including data, images, maps, and other useful formats. The system can be accessed by any user, regardless of their level of GIS expertise, and online and classroom training is available to advance users up the continuum. The entire user base is supported by a professional IT-GIS staff that is part of, and supported by, the agency's existing IT infrastructure.

To facilitate retrieval, there is a copy of ArcGIS Explorer on everyone's desktop within LaDOTD. Internal staff can access web services and display any LaDOTD project on a map. They can also view the data associated with the project and download it to a geodatabase. Access is limited to internal users to protect sensitive data. Examples of internal data include project-specific data, videolog data, data to support other internal applications, and utility relocations (i.e., permits on highways).

They have also developed several GIS applications available for external use:

- The LaDOTD Projects Application, which allows contractors and other users to view where projects are, turn on legislative districts, and select and download data.
- Hydrowatch, which displays United States Geological Survey (USGS) stream and U.S. Army Corps of Engineers (COE) river gauges. Users can view a location and hyperlink directly to USGS or U.S. Army COE websites to view detailed stream data.
- The Louisiana Emergency Management Plan was also developed based on LaDOTD GIS data. They have developed multiple maps to support emergency management activities during hurricanes, including pre-storm information delivery, GIS weather data on storm

progression, road speed impact, a search and rescue operational grid and other maps to support these activities, flooded street elevations, and damage assessment activities.

Lessons Learned:

- Using BI tools such as GIS applications can provide multiple types of integrated transportation data in a map-based visual format to a wide audience (using a central database repository for geospatial data).
- Use of GIS applications requires that the user community have access to the technology tools (such as ArcView or ArcGIS Explorer) needed to query, retrieve, and analyze the data. For more experienced users, an enterprise GIS system can be used to integrate data layers from other areas in a DOT to support typical DOT operations, including asset management, emergency management, travel monitoring, and safety. Use of GIS also offers the capability to perform a visual QA/QC on data and correct errors as needed.



### Example: Maine DOT

Maine DOT is a good example of how GIS technology is used to manage, access, and share location information. Maine DOT uses location data to extract information from separate databases, because all transportation assets have location as a common attribute. GIS is the primary tool used to manage and access location information.

Several years ago, Maine DOT launched a Map Viewer web-GIS application. This system allows all employees to view location-based data and to link data from map features to other existing web applications (on the internal intranet). This enables employees to access the department's most up-to-date data and documents, including inspection images, inspection reports, capital project information, ROW plans, street view, and traveler information.

Map Viewer also can enable employees to use location as a basis for communications, by capturing the current map session (i.e., the current extent, layers, background and any drawings/labels) and producing an outgoing email message with a URL that would connect anyone else to the same map session (even if they have never used Map Viewer). A link to a customized session of Map Viewer is available to the public at <http://www.maine.gov/mdot/mapviewer/>.

The types of data that can be displayed on the map(s) include Roads (Principal, Minor, and Other Arterials, Major/Urban and Minor Collectors, and Local roads), Highway Features, Maine DOT Projects, Property Information, Survey Control Points, Traveler Information, Water Features, and Boundaries. Map Viewer allows the user to save a map as a PDF file.

Maine DOT provides this publication for information only. Reliance on this information is at user risk. It is subject to revision and may be incomplete depending on changing conditions. The department assumes no liability if injuries or damages result from this information. This map is not intended to support emergency dispatch. Road names used on this map may not match official road names.



### Example: Nevada DOT (NDOT)

EOS.Web Express is the library management system used at NDOT. The library uses EOS.Web express for capturing and searching for information. Created exclusively for organizations that do not require all the functionality of an enterprisewide solution, EOS.Web Express is an economical solution for small to mid-sized special libraries. It is a fully web-based system, built on Microsoft.NET technology and the SQL Server relational database. This system is cost-effective, easy to use, and has industry-leading 24/7/365 live chat client support services.

Initial capture of information from publications includes cataloging the documents, whether they are in electronic or hardcopy form. Available documents, prior to 1992, are not yet in electronic form. The library is working to gather information produced by the DOT and to put this information into a single location. The document or report will be cataloged into the library catalog and a web link established from the library's website to the electronic document.

The EOS.Web Express system is used to query the library catalog for available documents and to do Google searches for various reports and information. NDOT retains both electronic and printed copies of reports and information produced by them and TRB, but they do not archive reports from other states. The library also keeps historic information and periodically sends documents over to the State Archive Library for retention.

#### Lessons Learned:

- Select an information management technology solution that is right for the agency or organization. For the NDOT library management system, an enterprisewide solution was not needed. NDOT was able to implement a cost-effective solution using EOS.Web that provided the functionality needed for managing the library's collection of transportation materials.
- Use other repositories and archives (outside the sponsoring agency) as available for retention of historically significant or otherwise important documents, reports, maps, photos, and digital images to supplement the collections maintained at the DOT's library and/or business units.



### Example: New York State DOT (NYSDOT)

NYSDOT is a good example of how a structured governance framework for Information Technology is used to ensure IT investments continue to provide meaningful business value,

whether measured in economic value or customer satisfaction. The governance process also helps NYSDOT to maintain a technology environment that is rational, sound, and continuously aligned with achieving NYSDOT's desired business outcomes. Through the IT Governance Policy, transparent, formal, and visible decisions are made regarding the selection and prioritization of IT investments. Furthermore, this Policy helps to ensure that resources are effectively applied to initiatives that are most closely aligned with strategic goals and objectives.

IT Governance at NYSDOT is executed by the Executive Council and five Division Councils:

- ACR and Legal Council & External Relations,
- Engineering Council,
- Operations Council,
- Policy and Planning Council, and
- Administrative Services Council.

IT Governance participants include the CIO, the ITD Project Management Office (PMO), ITD Customer Relationship Managers (CRM), Division IT Liaisons, Division Council Representatives, Regional IT Managers, and the Requestors of IT Investments.

NYSDOT also has an IT Strategic Plan, which describes the IT direction for the department for 2008–2012. The Plan includes Mission and Vision statements for the IT Division and documents a technological strategy for achieving the agency's goals.

The Mission and Vision statements, which are typically components of a governance program read as follows<sup>35</sup>:

**ITD Mission Statement:** The Information Technology Division provides high-quality IT solutions in a secure, reliable, cost-effective manner to facilitate the delivery of a safe, efficient, balanced, and environmentally sound transportation system for New York.

**ITD Vision Statement:** The Information Technology Division will strive to deliver best practice technology solutions that are recognized by customers and peers. Through our focus on collaboration and teamwork, our employees will receive great satisfaction as they are challenged to deliver high-quality results every day.

The NYSDOT IT Strategic Plan also states that the NYSDOT Information Technology Division Strategic Plan is focused on five key goals as illustrated in Exhibit 2-19. Each goal indicates a direction, desired improvement, or success measure for organizational performance in a key area. Achieving these goals will enable the realization of the organization's vision. Every goal is supported by initiatives, which represent the ITD's primary activities. Collectively, these goals and initiatives constitute the strategy for ITD: A plan for moving forward that creates value for NYSDOT.<sup>36</sup>

**Lesson Learned:**

- Establishing a well-defined governance structure for IT empowers an agency to make appropriate decisions regarding current and future technology investments in support of agency goals and objectives.
- Documenting the governance policies in a format such as a Strategic Plan (e.g., IT Strategic Plan) provides a platform to educate all persons/offices on the expectations and goals and objectives for use of governance within the agency.
- Establishing mission and vision statements for governance at the agency helps to set the direction for current and future policies, practices, and standards to be used in managing

<sup>35</sup>NYSDOT 2008–2012 ITD Strategic Plan, <https://www.dot.ny.gov/programs/?repository/2008%20DOT-ITD-STRATEGIC-PLAN.pdf>.

<sup>36</sup>NYSDOT 2008–2012 ITD Strategic Plan.

**Exhibit 2-19. NYSDOT's IT division strategic goals.**

data programs to support business needs. Mission and vision statements for data programs are separate from agency mission and vision statements.



### **Example: North Carolina DOT (NCDOT)**

Capture of library materials at NCDOT is referred to as acquisitions. This is the life blood of the library. There are several streams of acquisition of materials as indicated below.

The primary acquisition stream is TRB (e.g., Transportation Research Records (TRRs) and transportation specialty publications and NCHRP reports).

Another acquisition stream is NCDOT-generated documents and research reports. The library also purchases books as needed.

NCDOT's library is considered a sister library to the state library and has web access to query that catalog and to exchange materials without requiring the exchange to be handled as an interlibrary loan.

In short, the NCDOT library capture policies include

- Purchasing books as needed and tapping into other streams such as TRR and other national cooperative research reports;
- Providing funding support to TRB and NCHRP to obtain information; and
- Having query access, via the web, to the state library system.



## Lessons Learned:

- Use all sources available for adding to the collection of transportation-related materials (e.g., reports, documents, maps, photos, digital images, and videos) maintained and managed by the DOT. This includes providing links to external sources of information or purchasing additional collections of materials as needed. The sources of additional information and materials may be provided and/or used by the DOT's divisions, offices, and user communities, such as metropolitan planning organizations (MPOs), transit authorities, law enforcement, and other state and local transportation agencies.
- Enter into sharing and exchange of information agreements or arrangements with state and national organizations to obtain copies of transportation-related publications, which will enhance the collection of materials maintained at the DOT.

The NCDOT Librarian is responsible for the physical preservation and repair of documents housed in the library, as needed.

NCDOT's library is starting to modernize the existing catalog system into a web-based system. Staff are working in conjunction with the state library to use the catalog system called Voyager. The NCDOT library also has the traditional library stacks for storing documents, including printed reports from TRB, NCHRP, and AASHTO. The NCDOT library also has an electronic version of all publications available from these organizations. This information is available via the following web link: <http://ncdot.org/~research>.

NCDOT also has a Board of Transportation and the library has been required to keep a hard copy of the proceedings and meeting minutes of the Board, since 1955. This is how the library began at the DOT. The responsibility for maintaining these records of the Board of Transportation is still part of the library's ongoing duties. The NCDOT library is planning to preserve the documents in an electronic format. Library documents will be stored in SharePoint in the future for internal use, and access to the catalog and electronic documents will be available through the Internet. This will include research reports and documentation related to the business of the Board of Transportation.

## Lessons Learned:

- DOTs should consider using newer available technologies for cataloging and indexing transportation information maintained by the libraries at the DOTs.
- EDMS such as SharePoint are an option that can be considered for establishing a central repository of information, including documents related to transportation information. Several DOTs, including North Carolina (future) and Virginia (current) are using or plan to use SharePoint to facilitate management of documents and information.
- EDMS facilitate easy querying and access to information through categorization and taxonomy schemes within the EDMS.

**Example: Ohio DOT (ODOT)**

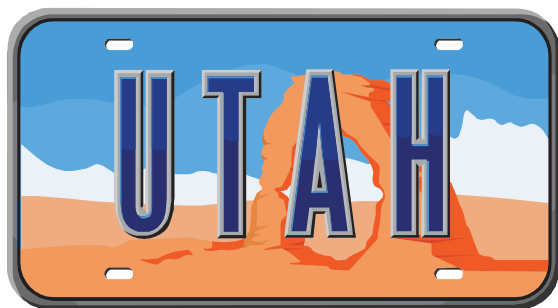
ODOT is a good example of how the use of an Information Technology (IT) Governance framework benefits ODOT in managing data and information across the department. ODOT

established an IT Governance Council, composed of members of executive management (including Division Directors and Deputy Directors). The council is responsible for approving any application development. If a business unit wants to develop a system, the unit must submit a project plan to the council, along with a description of the requirements and how the system relates to the department's core strategic initiatives. If the project scope changes significantly or if the complexity of the project changes, it must be reviewed by the council for continued approval to proceed. In the case of the Office of Technical Services, the Office Administrator is responsible for approving access to various applications supported by that office, including traffic information systems. Other business units are designated as the responsible office for maintenance of data in their particular areas, such as bridge and pavement data. These offices would typically be described as Data Business Owners within a Data Governance framework.

Data stewards also are responsible for maintaining different types of data (e.g., traffic monitoring data). The traffic data stewards are supported by the IT office, which is responsible for maintaining servers and mainframe data and application systems. However, responsibility is shared by the IT office and the Traffic Monitoring section. In this case, a local Access database is used to store traffic data, with the responsibility for maintaining the servers residing with the IT office. While the Access database is maintained on IT servers, the staff in the Traffic Monitoring section develops Net apps to maintain the data.

#### Lessons Learned:

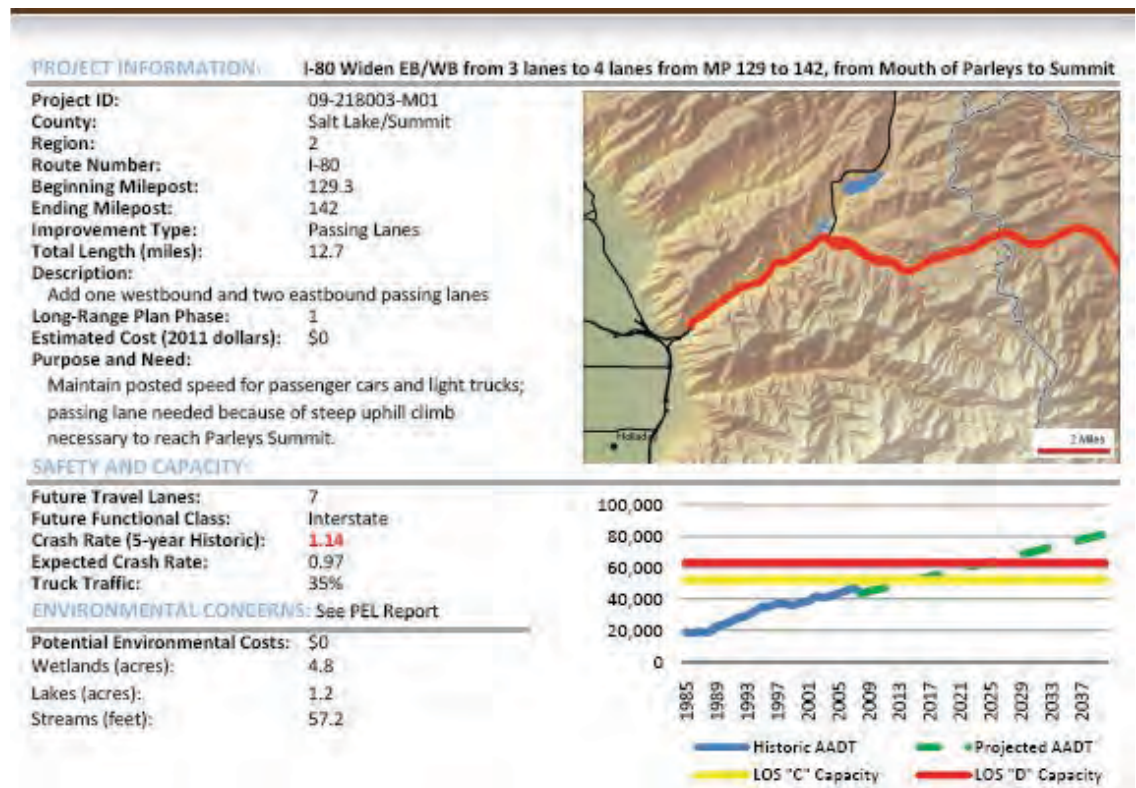
- Implementing a Governance Council or Board helps the organization designate the persons/offices responsible for establishing policies, procedures, and standards for the collection and use of data and information within the organization.
- The Governance Council is a high-ranking authority responsible for overseeing initiatives to develop or enhance existing application systems, to meet the needs of the organization, with consideration of the overall goals and objectives of the agency.
- Establishing clear roles and responsibilities for data governance within the organization helps everyone to understand how each job contributes to the efficient management of data and information.
- Responsibilities for governance should be shared between the IT office and the business units. The business units should retain responsibility for management of specific data and information needed to support their business needs and the IT office should ensure that hardware and software infrastructure is in place to support the business application systems.



#### Example: Utah DOT (UDOT)

UDOT is a good example of how technology tools can be used to benefit the management of data and information at the department. UDOT has developed an interactive web application using GIS tools for planning and analysis to provide data and information in an easily accessible and understandable format. The application, known as uPlan, facilitates synchronizing planning efforts with other state agencies, local governments, federal agencies, utility companies,

Exhibit 2-20. UDOT planning network.



Source: uPlan Summary, October 2010.

and within UDOT's many departments.<sup>37</sup> Data is compiled from various sources and is displayed on an interactive map. Reports can also be generated, which are linked to the geospatial display of the data. Data is also stored in a geographic reference center that allows for peer-to-peer access to data and for establishing data ownership.

Exhibit 2-20 illustrates how uPlan is used to display integrated data for project planning purposes. This exhibit includes the following types of information for a proposed widening project on I-80:

- Project-specific information,
- Safety and capacity information,
- Environmental impacts, and
- Historic and projected AADT and Level of Service (LOS) Capacity estimates.

Benefits of using uPlan include the following<sup>38</sup>:

- Pavement and bridge reconstruction plans are being coordinated with long-range capacity needs to synchronize the timing of various projects.
- Identified safety concerns can be coordinated with critical freight corridors to adjust the priority and timing of various highway improvements.
- UDOT Planning is working with each UDOT Region to assist in gathering and coordinating multiple smaller highway improvements into larger improvements so as to create the lowest construction impact to the traveling public.
- Similarly, larger transportation plans are being segmented into programmable projects that provide the highest benefit/cost in difficult economic times.

<sup>37</sup>uPlan Summary, October 2010.

<sup>38</sup>Planning with uPlan, Matt Rifkin, P.E., InterPlan, Intermountain ITE 2010 Technical Paper.

#### Lessons Learned:

- Providing access to integrated datasets through a single portal shortens the time to evolve data into information (and ultimately into knowledge) for use in decision making.
- Integration of data in a similar manner illustrated in uPlan can result in higher benefits and lower costs by eliminating redundant data collection processes and saving time in data analysis by using GIS tools for geospatial analysis.
- Use of integrated data helps facilitate cooperative working relationships among departments and improves understanding of issues and needs regarding data across the department. uPlan provides a way for one office to demonstrate a particular need or issue to other departments by displaying a particular type of data using the uPlan GIS tools.

The Utah DOT library requires the following information to be provided at the completion of research:

- A final research report,
- Information for the library newsletter about the research,
- A video explaining the purpose of the research, and
- Participation in a roundtable meeting to explain the final results of the research.

The library staff coordinates communication of relevant research to the appropriate offices within UDOT. They come in contact with considerable information daily and the information is forwarded to the appropriate offices either via email or printed version. This is usually a one-way transmittal and this type of documentation is not expected to be returned to the library. Information received is cataloged, and every month a list of publications is sent to the various offices to let them know what has been received.

The library is using alternative methods described in *NCHRP Synthesis of Highway Practice 610* to communicate research to use options available on the web and is using videos to make finding information an interactive experience and to make the research usable. They are also now using YouTube to provide access to information in a more usable way.

UDOT Research Division reports are posted on the UDOT website. Copies are also saved on CD-ROM and filed in the UDOT library. A hard copy of the report is retained in the UDOT library and three hard copies are sent to the state library, in accordance with UDOT policy. An electronic copy is also posted to Utah Government Publications Online.

Publications from other divisions within UDOT are cataloged and retained in the library in hardcopy format where they are shared with the library staff. The library also uses the newsletter and regions/division visits to advertise their services to their customers.

UDOT library materials are loaned almost exclusively to UDOT employees in accordance with UDOT policy. Private citizens are welcome to use the library and to make photocopies of materials, but may not remove library materials from the premises. Every effort is made to accommodate private requests/needs for information in other ways. Interlibrary loans are made to non-UDOT employees (e.g., consultants working with UDOT and non-Utah residents) on a very limited basis and after consultation with the librarian. The interlibrary loan option may be used for older documents in printed format only, or the documents may be in foreign countries or in other lending libraries.

#### Lessons Learned:

Libraries in DOTs can use various methods to market their services to the internal and external user community and promote the types of collections/information available through the library. The methods used for marketing services can include newsletters, Internet advertisements, and social media options (including YouTube, Twitter, and Facebook).



### Example: Wisconsin DOT (WisDOT)

WisDOT is a good example of how a data management tool, using GIS technology, is used to support decision making at the DOT and at the local government level. The following information pertaining to the data management tool, known as Wisconsin Information System for Local Roads (WISLR), is found at: <http://www.dot.wisconsin.gov/localgov/wislr/index.htm>.

WISLR is an Internet-accessible system that helps local governments and WisDOT manage local road data to improve decision making, and to meet state statute requirements. With GIS technology, WISLR combines local road data with interactive mapping functionality. The result is an innovative system that allows users to display their data in a tabular format, on a map, or both.

With WISLR, users can produce maps that show the location of road-related data and see trends that might otherwise go unnoticed. For this reason alone, WISLR aids with organized and logical assessments about local road data.

WISLR offers an advantage to local governments, who are required to report the pavement condition of roads under their jurisdiction to WisDOT. Because WISLR is a receptacle for local road information (e.g., width, surface type, surface year, shoulder, curb, road category, functional classification, and pavement condition ratings), WISLR generates the data locals need to get the job done.

Access to inventory information also aids with other tasks, such as compliance with Governmental Accounting Standards Board Statement 34 (GASB 34). This statement mandates reporting the value of local roads as infrastructure assets.

Local governments can use WISLR's querying, analytical, and spreadsheet tools to organize and analyze data. They can also update and edit their data. This combination improves accuracy for both pavement condition rating submittals and road inventory assessment.

WISLR can also be used by local governments to assist with following types of reports<sup>39</sup>:

- Municipal and county certification listings;
- Inventory listing by certification year and descriptions;
- Standard City, Village and Town (CVT) maps and functional classification maps;
- Construction and bridge reports;
- Annexation information;
- Roadway data collection information—how to measure road inventory;
- Statute 86.26, town road standards based on average daily traffic counts;
- Statute 86.302 local roads; and
- Statutory Chapters 204 and 205.

<sup>39</sup>WISLR Brochure, 2003, <http://www.dot.wisconsin.gov/localgov/docs/wislr-brochure.pdf>.



## Lessons Learned:

- Developing similar data management tools and allowing shared access to these tools helps the DOT and local governments to analyze and validate road inventory data needed to meet federal and state reporting requirements.
- A shared data management application such as WISLR allows flexibility and local control of data that is reported back to WisDOT from the local governments.
- An application such as WISLR provides a single portal for query, analysis, and update of road inventory data, making it easier for all users of the system.

WISLR is unique in that the system can enable local governments to share data directly with the DOT. This includes the type of local data needed by the DOT and road inventory and traffic data used for meeting federal reporting requirements of the Highway Performance Monitoring System (HPMS).

**Example: FHWA**

The FHWA Office of Operations is developing a DBP to improve the management of travel mobility data. This DBP addresses the needs and gaps for this program as identified in a white paper, *Data Capture and Management, Needs and Gaps in the Operation and Coordination of U.S. DOT Data Capture and Management Programs* (November, 2010). Several components of a successful data management program have been identified as part of the development of this DBP and are as follows:

- **Define roles and responsibilities** of data collectors, data managers (data business owners, data stewards, data custodians) and data users (data stakeholders, COI)
- Establish policies and procedures for **sharing data**
- **Establish data standards** for collection and reporting of data
- **Establish a risk management strategy** to secure the data and limit risks to the agency
- Invest in and implement the **technology** (hardware/software) needed to sustain and improve the data programs
- **Develop a data catalog** with data definitions and file formats used to store data and information
- **Avoid information overload** and control the amount of information being transmitted
- **Start small** during development of the data management plan and address critical needs first—“Small” could be defined geographically or by category of data
- **Build for scalability** to ensure that the system scales well in the real-world
- Determine what **data** and information need **to be stored and backed up**, and discard unneeded data
- **Identify** the data and information **users and owners** from the start
- **Make data and information available** as soon as feasible (i.e., provide real-time or pseudo real-time data)
- **Establish metadata standards** to minimize the potential for inconsistency caused by the creation or modification of metadata by numerous participants



- **Implement data security policies** to ensure security of data sets, data environments, and the hardware and software associated with both
- A thorough, well-documented and clearly communicated **Intellectual Property (IP) policy framework** is necessary to provide a clear understanding of the rules of the game with respect to licensing, patents and other aspects of intellectual property protection

Each of these points is equally applicable at a DOT for integration into existing data and information management programs or to assist in developing new programs.



### **Example: National Oceanic and Atmospheric Administration (NOAA)**

An example of an alternative lifecycle is from the NOAA, which determined that its environmental data will be based on an end-to-end data management lifecycle that includes the following steps:

1. Determine what environmental data are required to be preserved for the long term and how preservation will be accomplished;
2. Develop and maintain metadata throughout the environmental data lifecycle that comply with standards;
3. Obtain user requirements and feedback;
4. Develop and follow data management plans that are coordinated with the appropriate archive for all observing and data management systems;
5. Provide for delivery to the archive and secure storage;
6. Enable integration and/or interoperability with other information and products; and
7. Conduct scientific data stewardship to address data content, access, and user understanding.

Each of these practices is similar to those of DOTs whose data and information management have been formalized through some type of data management or DBPs. These practices include obtaining stakeholder feedback, using archive and secure storage methods, integrating multiple types of data and information, and the use of metadata and standards.



### **Example: Oak Ridge National Laboratory (ORNL)**

A report from the Oak Ridge National Laboratory (Hook, L., et al. *Good practices for Preparing Environmental Data Sets to Share and Archive*. Oak Ridge National Laboratory, 2010,

<http://daac.ornl.gov/PI/BestPractices-2010.pdf>.) indicates the following best practice experiences in improving the sharing of data and information:

1. Define the Contents of Data Files
2. Use Consistent Data Organization
3. Use Consistent File Structure and Stable File Formats for Tabular and Image Data
4. Assign Descriptive File Names
5. Perform Basic Quality Assurance
6. Assign Descriptive Data Set Titles
7. Provide Documentation

Each of the preceding seven practices (or a combination thereof) may result in producing documentation that can be included in a data catalog, which provides a centralized location for identifying data systems, data elements, data definitions and data file structures. All of these are needed by the data stewards and data custodians responsible for managing data according to established policies and standards.

## Use of the Cloud

In 2011, the National Institutes of Standards and Technology (NIST) published a definition of cloud computing that describes the essential characteristics of the cloud, possible service models that may be used to deliver cloud-based storage and services, and the deployment models that have varying degrees of connectivity outside a single organization. The basic terminology of the definitions is summarized in Exhibits 2-21 through 2-23. The NIST definitions, generally understood among IT experts and users of these systems, provide a basis for discussing the elements of cloud computing.

Cambridge Systematics conducted a survey to examine the use of web-based or cloud storage for DOT data. Exhibit 2-24 summarizes the main survey findings. Only those state agencies that responded to the survey are included in the table. The survey results indicate that many states have explored the use of cloud computing but have deemed it unsafe or inappropriate for state business. A few states have adopted “cloud first” strategies similar to that of the federal government and have deployed or are planning to deploy cloud computing strategies meeting the NIST definition.

**Exhibit 2-21. Essential characteristics of the cloud.**

Characteristic	Summarized Description
On-demand self-service	Users can automatically request and conExhibit computing capabilities without the need to contact a human at the provider side
Broad network access	Services are available over a network using various standard equipment types
Resource pooling	Resources such as memory, storage, bandwidth and processing are shared among users, and are allocated dynamically according to demand
Rapid elasticity	Resources can expand and contract in a near-instant fashion on user request
Measured service	Resources are measured for a usage dimension (such as processing capability, storage, or bandwidth) and are provided to the user allowing them to monitor use and the provider to charge for only the usage demanded

Source: The NIST Definition of Cloud Computing, 2011.

**Exhibit 2-22. Cloud service models.**

Model	Description	Example
Software as a Service (SaaS)	A service available over the network where all aspects of the provider's software are maintained in the cloud, and only a basic client interface is available directly to the user	Google Drive, Hotmail
Platform as a Service (PaaS)	The user may create applications and have control over their deployment and configuration, with the application running on a cloud-based server	Engine Yard
Infrastructure as a Service (IaaS)	The most flexible model, the provider maintains the fundamental computing resources with the client choosing operating system and networking capability	Amazon.com Web Services

Source: The NIST Definition of Cloud Computing, 2011.

**Exhibit 2-23. Cloud deployment models.**

Model	Description
Private cloud	Built and used for a single organization, may exist on or off organization's premises
Community cloud	Built and used by a community of users with shared concerns
Public cloud	Resources are available to the general public
Hybrid cloud	A combination of two of the above models, working together in a way that allows applications and data to flow between them

Source: The NIST Definition of Cloud Computing, 2011.

**Exhibit 2-24. Uses of the cloud by state.**

State	Cloud Options
ADOT&PF	ADOT&PF does not use the cloud for applications, but is exploring opportunities for use of the cloud, including storing/using imagery and other background layers to support the Department's GIS applications.
AZDOT	AZDOT is beginning to use ArcGIS online to share information. The business data resides locally, but the data for the web is uploaded to a cloud-based system. In the future, it may become overwhelmingly cost effective to transition primary data servers to the cloud, but at this time, it is too much of a drastic and perceived risky step from the perspective of the IT group.
CDOT	CDOT indicated in an independent survey that they have plans to store data in the cloud.
ConnDOT	<p>Bentley Systems is hosting the following set of data on a private cloud on behalf of ConnDOT:</p> <ul style="list-style-type: none"> <li>Contract Plans</li> <li>Contract Specifications</li> <li>Supplemental Documents associated with contract advertising</li> <li>Some legacy plans for completed projects</li> <li>Bridge Data for 6000+ Bridges</li> <li>Potential data for future storage includes: <ul style="list-style-type: none"> <li>Signal plans and associated data</li> <li>Shop and working Drawings</li> <li>Project Design Data (High Value Data – Models)</li> </ul> </li> </ul> <p>ConnDOT Planning section is also seeking to use EXOR and an application for crash data in the near future.</p> <p>The City of Hartford priced out options for use of the cloud, but determined that they are too small to make it cost effective.</p>

**Exhibit 2-24. (Continued).**

State	Cloud Options
Delaware	<p>The State of Delaware hosts a private cloud intended to eliminate the need for an extensive set of server replacements. The state is also moving toward use of the public cloud subject to important restrictions:</p> <ul style="list-style-type: none"> <li>The state retains full ownership of the data.</li> <li>The data is not allowed to reside offshore.</li> <li>The provider must encrypt all non-public data in transit to the cloud.</li> </ul> <p>In the event of termination of the contract, the Service Provider shall implement an orderly return of State of Delaware assets and the subsequent secure disposal of assets.</p>
FDOT	<p>FDOT is exploring cloud-based options for both server and storage provisioning. The State of Florida is undergoing data center consolidation. Most state agencies with data centers are being required by statute to consolidate their data centers into a couple of State-run hosting facilities. At FDOT, the data center was moved to a shared hosting facility in November of 2011. The district data centers are scheduled to move in FY 2014–2015. The State-run hosting facilities operate on a complete cost recovery model and bill the agencies for data center services. They incur the costs and FDOT pays the bills.</p> <p>Based on FDOT’s preliminary analysis of cloud service providers on Florida’s State term contracts, they are seeing substantially lower costs for cloud-based provisioning of servers and storage as compared to what the State-run hosting facilities are charging. FDOT plans to do a proof of concept model using Microsoft’s AZURE infrastructure to provision some servers to support their BoxTone implementation to confirm the viability of a cloud-based solution.</p> <p>Should that prove successful, FDOT’s CIO plans to lobby the State-run hosting facilities to move more towards being a broker of cloud services rather than standing up infrastructure within the facility. Although all of the state’s infrastructure may not be able to be cloud-provisioned, the CIO believes that a substantial portion of it can be with significant cost savings.</p>
GDOT	Georgia DOT is not using the cloud but is investigating moving in this direction. In particular, the cloud will potentially be used for Videolog and traffic data processing.
IDOT	Illinois DOT indicated in an independent survey that they store data in the cloud.
Iowa DOT	Iowa DOT indicated in an independent survey that they store data in the cloud.
Michigan DOT	<p>Michigan DOT is using a private cloud available for the State of Michigan referred to as “MiCloud.” Using this private cloud avoids most of the security issues associated with off-site, external provider storage and provides the benefit of much higher data transmission speeds because of its physical location within the state, linked with statewide high-speed transmission network capabilities. The state also believed that shaping an external service to be compliant with the state’s security policies would destroy any cost advantage of using the external cloud.</p> <p>The state cloud storage option has been used to deliver photolog data to DOT employees in the Lansing area. This process seems to be working well. The DOT is also exploring having their current pavement condition monitoring vendor create a site to serve the business need for monitoring pavement condition.</p>
Montana DOT	Montana DOT indicated in an independent survey regarding use of the cloud that they store data in the cloud.
New Hampshire DOT	New Hampshire DOT indicated in an independent survey that they have plans to store data in the cloud.
North Dakota DOT	North Dakota DOT indicated in an independent survey that they have plans to store data in the cloud.
State of Utah	The need to close data centers to save costs is what drove Utah to deploy a private cloud for use by Utah State agencies. This private cloud is similar to the private cloud deployed in Michigan. Utah has also embraced and adopted policies for a hybrid cloud, which combines elements of a private cloud and publicly available storage and computing resources. The State’s Technical Architecture Review Board has approved a standard approach for dealing with external cloud computing providers and integrates those external services with private cloud elements when it makes sense from a cost and data security standpoint. To date, Dropbox, Evernote, SugarSync, and Google Docs have been approved as data storage and cloud-based service providers.
HI, MI, NV, NY, OH, PA	Several DOTs indicated in an independent survey that they do not store data in the cloud nor do they have any plans to do so in the future.

Delaware, Michigan, and Utah have been at the vanguard of adopting cloud computing policies and installations and can serve as models for other states wanting to move toward a broader adoption of private, hybrid, and public clouds to better address agency information technology needs. They are discussed in detail below.



### Example: Delaware

The State of Delaware has established some rigorous policies and procurement guidelines for use of cloud computing resources. Along with hosting a private cloud intended to eliminate the need for an extensive set of server replacements, the state also is moving to use the public cloud (subject to some important restrictions).

Delaware considers the following elements to be non-negotiable in its agreements with external providers:

- The state retains full ownership of the data.
- The data is not allowed to reside offshore.
- The provider must encrypt all non-public data in transit to the cloud.
- In the event of termination of the contract, the Service Provider shall implement an orderly return of State of Delaware assets and the subsequent secure disposal of assets.

In addition to these broad principles, Delaware (1) provides a detailed set of specific terms and conditions that must be present in any contract with a cloud-based provider and (2) has integrated cloud computing into its information security policy with an established process for approving contracts with external providers.

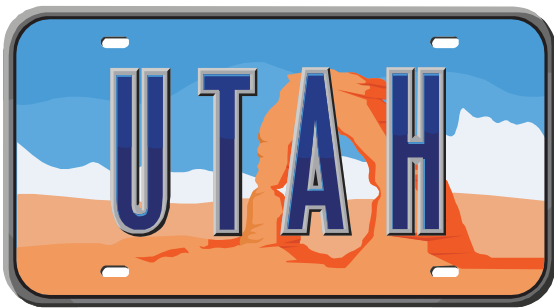


### Example: Michigan

In Michigan DOT's case, the use of a private cloud (known as MiCloud) allows the DOT to avoid many of the security issues associated with offsite, external provider storage and has the added benefit of much higher data transmission speeds because of its physical location within the state linked with statewide high-speed transmission networking capabilities. Much like using a private provider, MiCloud allows agencies to sign up for storage capabilities using a

self-service intranet website and the costs are established up front based on the precise amount of storage required. MDOT charges \$0.01167/GB/day to MiCloud users, which is over 80% cheaper than commercially available storage solutions. Most users never have to interact with an actual human to set up, reconfigure, or change their hosting options, with all common customer service requests available directly to the user and handled automatically by the system's software.

The state's IT leadership considers the standard terms and conditions that accompany offsite data storage and hosting agreements to be unconstitutional because of their common need to indemnify the provider. This was a driving reason behind the state pursuing its private cloud strategy. The state also believed that shaping an external service to be compliant with the state's security policies would destroy any cost advantage of using the cloud. The state has built the system using an Open Virtualization Format, which will allow the state to readily engage with external hosting providers should the state choose to move in this direction. The state intends to maintain a central link between the statewide MiCloud and external service providers, in part to resolve any technical, legal, security, and policy issues at a single point of contact and remove the need for agency IT staff to handle these matters.



### Example: Utah

The State of Utah is another leader in deploying a private cloud for state agencies to use in an automated fashion. The state provides the full spectrum of SaaS, IaaS, and Platform as a Service (PaaS) provisioning. A primary driver behind Utah's decision to pursue a cloud strategy was the need to close data centers to save costs; the state took 38 existing data centers and consolidated them into 2 centers. This action resulted in (1) a one-time capital cost savings by eliminating the need to replace aging equipment, (2) documented massive energy savings (enough to power 978 homes for 1 year), and (3) a reduction in staffing needed to provide direct customer service to users.

Utah has also embraced and adopted policies for a hybrid cloud, which combines elements of a private cloud and publicly available storage and computing resources. The state's Technical Architecture Review Board has approved a standard approach for dealing with external cloud computing providers and integrates these external services with private cloud elements when it makes sense from a cost and data security standpoint.

The state's "Policy on Use of External Service Providers for Data Storage" dictates that the state must explicitly approve external providers used for storage of state information, and the provider must pass a departmental security review. Furthermore, the state must have a contract with the provider that specifically addresses central management, use, storage of, and deleting of sensitive data. To date, Dropbox, Evernote, SugarSync, and Google Docs have been approved as data storage and cloud-based service providers under the policy. The policy requires users to explicitly disclose the use of external services so that the state may access them at any time, and it restricts their use to that of non-confidential, non-sensitive state data.



## Conclusions

Part 2 has presented key elements of transportation information management practices of selected DOTs and other organizations. For various reasons, practices vary widely among the DOTs. The scale and scope of relevant information to be managed differs among states, as do the institutional structure of the agencies and the functional characteristics of the systems they manage. Resource availability, leadership interest, and other factors may lead some agencies to pursue their development of transportation information management practices more aggressively.

The cases and examples presented here are not intended to suggest that there is a set of “best” practices to be recommended for universal application. Rather the experience described here may be useful to agencies less advanced in one or another aspect of information management. Nevertheless, experience shows that certain practices are likely to help many agencies improve their management of transportation information and that further development of such practices can advance transportation system management generally. These are the subject of Part 3.

# Making Improvements in DOT Transportation-Information Management Practices

Part 3 defines the common elements of data and information within DOTs, proposes a model of the processes used for information management, discusses stakeholders' roles in information management, and presents information DOTs can use to capture, administer, and retrieve information, as well as govern transportation data and information activities.

## Improving Information Management in a Transportation Agency

When managing information, DOT professionals often use different definitions for the same term—stakeholders (e.g., IT, transportation data managers, library scientists, and data users) have different backgrounds and use different terms for daily functions. For a DOT to better manage information, common definitions must be established. Gaining a clear understanding of meanings from different points of view is important in establishing the definitions. Ultimately, stakeholders share the goal for information management—to improve findability, searchability, learnability, and organization of transportation information.

This report defines the terms below specifically as follows:

- **Data** is a representation of facts, concepts, or instructions in a formalized manner suitable for communication, interpretation, or processing by humans or computers.<sup>40</sup>
- **Data management** is a set of processes and techniques related to collecting and managing data so that data can be used to accomplish tasks effectively.
- **Information** is data and documents that have been given value through analysis, interpretation, or compilation in a meaningful form.<sup>41</sup> Information can be in electronic and physical form.
- **Information management** is used to describe the organization of and control over the collection and management of information from one or more sources, and the processing, delivery, and sharing of that information to users. Information management addresses how an organization develops policies, practices, technologies, and other resources and capabilities to manage information as a resource, including the management functions associated with creating, collecting, processing, transmitting, disseminating, using, storing, and disposition of information, both automated and non-automated.<sup>42</sup> A simpler version of the definition is from AIIM (The Global Community of Information Professionals): “Information management is the collection and management of information from one or more sources and the distribution of that information to one or more audiences.”<sup>43</sup>

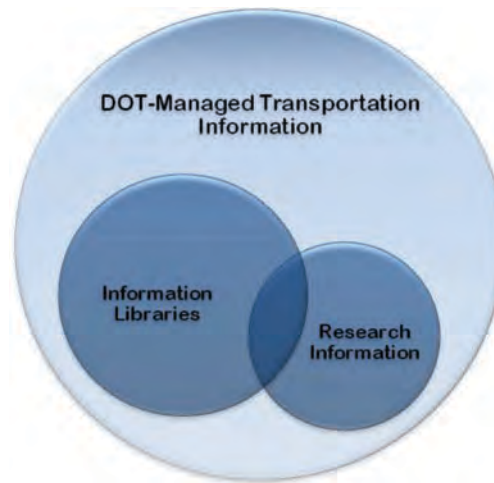
<sup>40</sup>ANSI/ANSII TR2-1998, Glossary of Document Technologies.

<sup>41</sup>[http://www.arma.org/standards/glossary/index.cfm?id\\_term=246](http://www.arma.org/standards/glossary/index.cfm?id_term=246).

<sup>42</sup>Caudle and Marchand, 1989, *Managing Information Resources*.

<sup>43</sup><http://www.aiim.org/what-is-information-management>.

**Exhibit 3-1. DOT-managed transportation information.**



### **Types of Information**

One of the challenges with managing DOT information is that literature or research results within libraries constitute only a small part of the whole of DOT-managed information (see Exhibit 3-1). Part 3 of this report focuses on the larger scope of all data and information at DOTs.

DOTs are responsible for many reports, documents, records, and data that come in various forms (e.g., GIS files, databases, design drawings, printed documents, spreadsheets, microfilm, maps, photos, video, and raw traffic data). One component of managing information involves preparing an inventory of the types of information at the DOT. Because information and data have been evolving rapidly over the past decade, the contents of any inventory are likely to be in flux. An inventory can be categorized in many different ways. Various types of DOT information—categorized by project, condition/performance, research, administration, and inventory/status—are as follows:

- Project Information
  - Engineering Diagrams
  - Planning Studies
  - Technical Reports
  - Environmental Reports
- System Condition/Performance-Related
  - Traffic Data
  - Safety Data
  - Pavement Condition
  - Bridge Condition
  - Performance Reports
  - Forecast Data
- Research
  - Data Collected Throughout Research
  - Completed Research Reports
- Administration
  - Financial/Budget
  - Contact Information
  - Annual Reports

- Inventory
  - GIS Data and Maps
  - Asset Inventory Databases
  - Roadway Log
  - Classification
- Truck Information
  - WIM Truck Data
  - Permit Data
  - Oversize/Overweight Restricted Roadways and Bridges

This list is only an example and is not comprehensive. Another categorization approach is to classify according to site, project, network, and administration. Regardless, DOTs should define types of data based on their situations and consideration of agency priorities and legislative, federal, and other requirements. Agency priorities, structure, and requirements change, so the classification structure should be flexible and allow for change.

Exhibit 3-2 lists representative types of information by function, includes a description of the information formats, and indicates how transportation information, created as part of common DOT business functions, includes data or data-generated information formats such as CAD, GIS, and other computer-generated graphics.

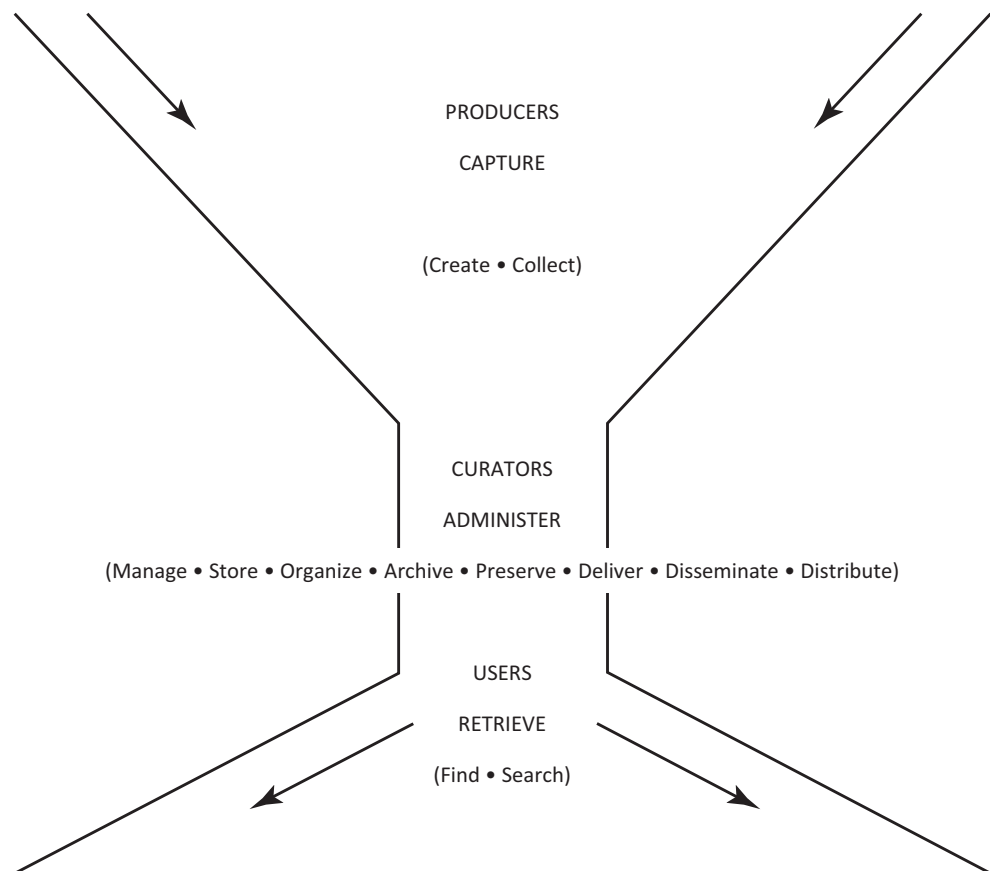
### Information Management Processes at DOTs

The ANSI/AIIM/ARMA TR-48-2006 ECM components were used as the starting point for establishing the processes involved with information management at DOTs. This concept, depicted in Exhibit 3-3, consists of information capture, administration, and retrieval and is described below:

- **Capture** represents capture from the TR-48-2006 ECM components (e.g., defining content, information capture, and content harvesting). Capture is used to encompass creating, observing, measuring, collecting, and so forth.

**Exhibit 3-2. Types of transportation information by function.**

Functions	Information Types	Formats				
		Documents	Data Tables	CAD	GIS	Graphics
Project Information	Engineering (e.g., Drawings)					
	Specification					
	Performance Test					
	Planning Study					
	Technical Reports (e.g., Materials Research)					
	Environmental Report					
System Condition/Performance	Traffic Data					
	Safety Data					
	Performance Report					
Research	Research Report					
Administration	Financial					
	Contact Information					
Inventory	GIS Data					
	Asset Inventory Database					

**Exhibit 3-3. Information management framework.**

- **Administer** encompasses the policies, procedures, organizational structures, rules and the like that an agency uses to direct and govern how transportation information is stored, preserved, and delivered to users.
- **Retrieve** is an iterative process that generates new content through analysis of the original resource or through identification of new data, information, or knowledge by users. Retrieve is used to encompass searching, finding, interpreting, retrieving, and using.

ECM activities are shown in blue and records management activities are shown in light blue. All stakeholders, regardless of where they fit into a DOT, are active participants in the information management process and may be producers, curators, or users of information and data at various stages of the business processes. As shown in Exhibit 3-3, producers are generally responsible for information capture; curators administer information; and users retrieve information.

Concerns for each step of the process are as follows.

- **Capture**
  - Human- versus application-created information and data
  - Data types and technologies
- **Administer**
  - Tools and techniques for managing information, including governance, moving content around an organization, and monitoring performance (indexing, input designs, and categorization)
  - Data Management Plans
  - Metadata

- Where to put content and find it again (repositories, library services, technologies)
- Storage Management (collaboration, web and Database Content Management, Records Management, Workflow/Business Process Management, and Document Management)
- Long-term storage and archival of critical content
- Archive types
- Standards
- How to get the right content to the right audience on the right device and under the right circumstances (layout/design, publishing, transformation technologies, security technology, and distribution/dissemination)
- **Retrieve**
  - Retrieval in a findable and accessible manner and format

## Stakeholder Roles in the Information Management Process

Stakeholders include information managers with specific responsibility for information capture and administration functions: information professionals, IT professionals, or data managers. Additional stakeholders include DOT GIS offices, all levels of management, transportation professionals in local and regional transportation agencies, federal transportation-related agencies, transportation research organizations, and academic institutions. Stakeholders can also include those doing business with DOTs and other transportation-related agencies; NGOs; and citizens who have an interest in or need for or want to know about transportation information.

DOT stakeholders can be categorized as producers, curators, or users within the information management process. For example, a DOT GIS office typically would produce GIS data tables and content. DOT librarians typically would curate DOT information such as research reports. Curators could also include IT professionals or other information managers. Users are the internal or external consumers of the information. In some cases, a stakeholder can both produce and use a data set. For example, the planning division may produce and use traffic data.

Assigning roles to each of the stakeholders is an important aspect of information management; stakeholders need to be able to see how they fit into the process to focus their energies to improve information management within their DOT. DOTs may find it useful to produce tables to indicate user roles and responsibilities for the activities associated with each function in the information management process.

Successful management of transportation information at a DOT relies on coordination among people (stakeholders/users), processes, and BI tools within a well-defined and structured information management program. The following chapters discuss how transportation information management can be accomplished at a DOT, whether it is just starting to implement information management or it is trying to improve information management practices.

## General Approaches to Improving Information Management

This section presents general information that DOTs can use to establish and maintain effective transportation information management practices see (Exhibit 3-4). Each approach includes actions for how to implement it, as well as examples gleaned through research conducted for NCHRP Project 20-90 as described in detail in Part 2; some of the approaches reflect research from prior NCHRP projects, such as *NCHRP Report 666: Target-Setting Methods and Data Management to Support Performance-Based Resource Allocation by Transportation Agencies*. DOT staff responsible for capturing and administering data and information to support business needs will find this report useful in improving the management of transportation information within their agencies.



**Exhibit 3-4. General approaches for improving information management.**

APPROACH
Establish agency goals for improving management of data and information
Establish policies and procedures within a governance framework to define the roles and responsibilities of the business units and librarians in managing data and information for the agency
Demonstrate the ROI for investing in methods and tools to improve the management of transportation data and information
Develop data business plans (DBPs) as an option
Share data and information
Use the Information Management Lifecycle and apply the appropriate competencies for each stage of the lifecycle
Use improved technology tools to help the business units and information managers manage, store, and retrieve information for all customers in an easily understandable format
Use formal communication channels to support the sharing and exchange of information among business units, information managers, and internal and external customers

General approaches that a DOT can undertake to establish and maintain effective transportation information management practices follow, along with examples. The general approaches reflect investigation of current information management practices and case studies of the information management needs and activities of selected DOTs.

- **Establish agency goals for improving management of data and information.** This approach strengthens the alignment of data programs in support of core business functions from an enterprise perspective. Goals for improving management of data and information within business units should also be established to align data programs and functions of the unit. Within the context of a DOT, business units refer to the divisions, districts, and offices responsible for conducting the business functions of the agency. These functions include, but are not limited to, management of assets and resources and providing information technology services to support the agency and the core business functions of planning, design, construction, maintenance, and operations. For an example, review the case study for MnDOT in Part 2.
- **Establish policies and procedures within a governance framework to define the roles and responsibilities of the business units and librarians in managing data and information for the agency.** Policies should provide agency managers and staff with clear direction on how to manage the transportation information as an integrated effort on behalf of the user community and stakeholders. For an example, review the material for IDOT and NCDOT in Part 2.
- **Demonstrate the ROI to the agency for investing in methods and tools to improve the management of transportation data and information.** Business unit managers should continue to demonstrate the ROI to senior managers for how new technology, standards, policies, and procedures can improve the information management at the agency. Ways to demonstrate ROI can include tracking usage statistics at a library or requests for information, which can be used to justify procurement of additional materials such as documents and hardware and software to meet user demand. For examples, review the material for VDOT and WSDOT in Part 2.
- **Develop a DBP to help DOTs capture, store, archive, preserve, and disseminate transportation information.** For DOTs struggling with limited resources to coordinate the processes

identified in the information management process, outsourcing is always an option. For details, review the material for ADOT&D, CDOT, and MnDOT in Part 2.

- **Share data and information.** Sharing data and information with the agency can pay large dividends in business value. For an example, see the material for ORNL in Part 2.
- **Use an information management process.** Using an information management process helps to ensure that data and information are managed based on feedback from the user community. For an example, see the material for NOAA in Part 2.
- **Use improved technology tools to help business units and information managers manage, store, and retrieve information for all customers in an easily understandable format.** Technology tools can be used to increase efficiency and reduce time in the collection, storage, search, retrieval, and dissemination of information at a DOT. Many DOTs use the MS SharePoint system to manage, store, and retrieve transportation information to respond to inquiries. Other tools are also available. For examples, review the material for IDOT, MDOT, NCDOT, and VDOT in Part 2.
- **Use formal communication channels (such as transportation knowledge networks [TKNs]) to support sharing and exchange of information among business units, information managers, and internal and external customers.** There are many excellent sources of transportation information from both internal and external communities. DOTs can maximize their ability to provide needed information to their users by taking advantage of all available sources of transportation information and incorporating links to this information through a web portal or other means. For an example, review the case study for IDOT in Part 2.

## Approaches Tied to Functions of the Information Management Process

This section provides approaches specific to the functions of the information management process: capture, administer, and retrieve as illustrated in Exhibit 3-5. The statements were formulated based on research for NCHRP Project 20-90.

### Capture

- **Require the use of metadata, define the metadata standards to be used, and provide examples of the types of metadata to be used at the DOT.** Metadata must be included with any type of transportation data and information managed by a DOT. The use of metadata helps to ensure that the right information is used to respond to inquiries. Metadata must be maintained as data sets and data definitions or use of data and information may change when information systems are replaced or enhanced. For examples, review the material for MnDOT and Caltrans in Part 2.
- **Establish agency policies for the collection and capture of transportation information.** Agencies using established policies and practices for the collection and capture of information from all available sources typically benefit from the reduction of duplicate data/information collection efforts and clear direction on how, where, and when information is to be collected. Use of collection and capture policies often results in more efficient use of resources, which allows funds to be redirected to other critical tasks such as analysis or assisting with research for particular transportation projects. For examples, review the material for IDOT and NCDOT in Part 2.
- **Use both data and content management practices for managing information.** DOTs should establish practices or develop plans that address the management of data as well as content. Content includes that which is contained within the DOT's library and that which is managed

**Exhibit 3-5. Summary of suggested approaches tied to information management process.**

Process	Approach
<b>Capture</b>	Require the use of metadata, define the metadata standards to be used, and provide examples of the types of metadata to be used at the DOT
	Establish agency policies for the collection and capture of transportation information
	Use both data and content management practices for managing information
	Use technology tools to facilitate the capture of data and information
	Use a single portal for access to and distribution of information
	Develop relationships internally with business units and externally with peer DOTs and other national and international transportation agencies in order to capture, share, and exchange data and information
<b>Administer</b>	Establish governance policies for managing data and information across the agency
	Develop policies for archiving of data rooted in the business with consideration for privacy of data and information
	Develop formal records retention policies and procedures to support preservation of important current and historical documents for a DOT
	Establish a DBP to help manage data and information
	Follow a multi-step, multi-phase approach for implementing data management programs
	Participate in a peer group
	Find a champion
	Assign resources and appropriate competencies
	Establish Mission and Vision statements for managing data and information at the DOT
	Establish categorization schemes for data and information at the DOT to ensure organized, methodical management of data and information
	Use semantic schemes
	Use authoritative glossaries and vocabularies
	Use taxonomies for classifying and grouping transportation information at the point of storage
	Use state libraries and national transportation libraries to supplement the archives maintained by the DOT's libraries and business units
	Use digital preservation
	Use BI tools to facilitate integration and sharing of data in an easily understandable format
	Use multiple types of BI tools to improve management of data and information
	Provide content (data and information) in an electronic format to enable delivery to the widest audience possible
	Select methods for distribution of information and data from the best available technology tools, including the web and social media
	Use technology
	Use the cloud for storing large datasets
	Use taxonomy management tools
	Use visualization tools (e.g., maps) for delivery and dissemination of data and information
	Use market information management techniques, such as a newsletter to advertise the types of transportation information available from the DOT to the user community
	Use the "Tell the Story" marketing technique to explain how data is used at the DOT
	<b>Retrieve</b>

by business units. Addressing the need for improved practices in both areas facilitates quick access to information when needed by the user community. For examples, review the material for MnDOT and VDOT in Part 2.

- **Use technology tools to facilitate the capture of data and information.** Many DOTs use GPS technology for capture of location data and integration of this data with existing GIS and linear referencing systems. Other types of technology tools can also be used to facilitate the capture and indexing of information. For examples, review the material for MDOT and NDOT in Part 2.
- **Use a single portal for access to and distribution of information.** Providing a single portal for access and distribution of information makes it much easier for the user community and stakeholders (including library patrons and typical DOT divisions and offices) to locate information when needed. This portal can be established on the agency's intranet for internal users and/or on the Internet for external users and stakeholders. The portal also can facilitate access to specific sites. For examples, review the material for AZDOT in Part 2.
- **Develop relationships internally with business units and externally with peer DOTs and other national and international transportation agencies in order to capture, share, and exchange data and information.** Many benefits are derived from fostering and maintaining working relationships with other offices within a DOT, peer states, and national and international transportation agencies. One of the most significant benefits is the ability to capture and share data and information among users. This sharing may include the exchange of data sets, research reports, and documents related to typical functions within a DOT such as planning, design, construction, maintenance, or financial planning. For examples, review the material for Caltrans and NCDOT in Part 2.

## Administer

- **Establish governance policies for managing data and information across the agency.** Governance policies delineate the rules, standards, procedures, and responsibilities for coordinating and managing data and information needed to support important business functions of the agency. In many cases, a governance policy and framework helps foster cooperation among business units and the information technology (IT) office. For examples, review the material for Caltrans, GDOT, Maine DOT, MDOT, NYSDOT, ODOT, WSDOT and NCDOT in Part 2.
- **Develop policies for archiving of data rooted in the business with consideration for privacy of data and information.** Business units at DOTs should share in the responsibility for determining what information and documentation needs to be retained in the department's archive. Business units usually are familiar with the need for historical data and information to support work in their respective business units. Participation by business units in determining what needs to be archived helps to ensure that institutional knowledge is retained as employees move to different assignments within the agency or leave the agency altogether. For examples, review the material for MDOT in Part 2.
- **Develop formal records retention policies and procedures to support preservation of important current and historical documents for a DOT.** Establishing records retention policies by department mandate or by state law is essential for ensuring that documents, records, and information are available when needed to support department functions (e.g., project planning and delivery). Archived records also may be needed to respond to litigation. For examples, review the material for AZDOT and Caltrans in Part 2.
- **Establish a DBP to help manage data and information.** DOTs typically have strategic plans that delineate the policies, standards, and procedures to guide the state in planning, design, construction, maintenance, and operations of its transportation facilities. It is equally important to have a plan that considers transportation data and information as an asset of the

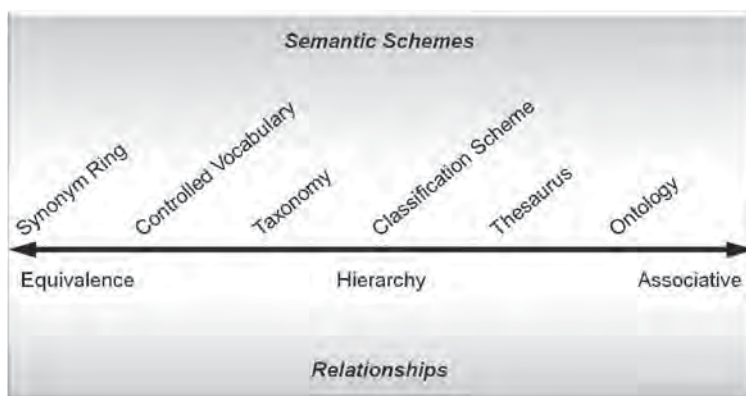
agency. Many DOTs are now developing and implementing DBPs to serve this purpose. For further information, see and review the material for ADOT&D, CDOT, MnDOT, VDOT, WSDOT, and FHWA in Part 2.

- **Follow a multi-step multi-phase approach for implementing data management programs.** See the material for FHWA in Part 2 for an example.
- **Participate in a peer group.** Participation in various professional groups responsible for managing data and information (such as state and national library associations and TRB committees) enables DOT librarians and data managers to stay informed of the most recent practices in managing data and information. These practices can be considered for incorporation into data and information management programs as needed. For examples, review the material for IDOT and NCDOT in Part 2.
- **Find a champion.** DOT information managers should always have information ready to help upper-level management understand the importance of information management and to champion its effort at the organization. It is very important to have upper management involved in setting the direction for information management at the agency. For examples, review the material for MaineDOT in Part 2.
- **Assign resources and appropriate competencies.** The issue of funding is also important. It is very costly to maintain paper records; therefore, allocating some funds to address electronic file management is necessary to meet information management needs of the future. See Part 2 for details.
- **Establish Mission and Vision statements for managing data and information at the DOT.** Mission and vision statements establish a clearly defined direction for the agency regarding the management of its data and information assets. For an example, review the case study for MnDOT in Part 2.
- **Establish categorization schemes for data and information at the DOT to ensure organized, methodical management of data and information.** Although DOT information managers may be the primary audience, all stakeholders need to be able to understand and use the categorization scheme. Stakeholders need to be able to understand and use the categorization scheme to some extent without training (e.g., categorization schemes should be as inherently easy to use as Google is to users of the web). Although training, experience, and subject matter expertise remain important and valuable for effective use of a categorization scheme, the scheme needs to be usable “out of the box.” The usability of a categorization scheme is usually measured by (1) discreteness of broad categories, (2) consistency in indexing information, and (3) consistency in finding information.

The categorization schemes identified should be aligned with agency business units. This method facilitates the retrieval of data and information by all work groups searching for information pertinent to their work unit. For an example review the material for MnDOT in Part 2.

- **Use semantic schemes.** Semantic schemes exist along a continuum based on the types of relationships that characterize the scheme. Schemes can vary from simple to complex in terms of the difficulty in making these types of relationships. See Exhibit 3-6.
- **Use authoritative glossaries and vocabularies.** A glossary is an alphabetical list of terms in a specific subject area, while dictionaries usually have a broader scope. In the context of this project, glossary and dictionary mean the same thing. Glossaries typically include a definition for each entry. Glossaries are often created to support the function of an organizational unit, a project, a policy initiative, compliance with a legislative mandate, or some other specific purpose. As such, glossaries are useful resources when building a terminology scheme for a subject area. They represent important concepts in the domain. Because glossaries have definitions, they are helpful in understanding the nuances of terms and phrases in a discipline. For an example review the material for WSDOT in Part 2.
- **Use taxonomies for classifying and grouping transportation information at the point of storage.** This ultimately enables quick search and retrieval of information in a timely, efficient

**Exhibit 3-6. Semantic schemes: simple to complex.<sup>6</sup>**



<sup>6</sup>After: Amy Warner. *Metadata and Taxonomies for a More Flexible Information Architecture*.

manner. For examples, review the material for VDOT in Part 2. Exhibit 3-7 describes facets generally applicable to all types of content.

- **Use state libraries and national transportation libraries to supplement the archives maintained by the DOT’s libraries and business units.** Using additional storage capacity of the archives available at many state libraries increases the amount of transportation information that can be provided in the form of documents, photos, videos, and so forth to the user community. Several DOTs, including Illinois and Arizona, have arrangements with their state libraries for archiving of publications produced by the DOT. Some DOTs also archive reports produced by the DOT at the NTL. Repositories can be used to store printed and electronic historical content in addition to information that may be stored in bibliographic databases (which may only include information on the authors and titles and summaries of the content and not the actual documents themselves).
- **Use digital preservation.** Most resources today are created, published, and used electronically, and the electronic version of the content is the resource of record. This may be a PDF file for a technical report, a set of database layers for GIS visualization, or a CAD file format for a blueprint. All of these are electronic files that require specialized software (and sometimes

**Exhibit 3-7. Facets applicable to all types of content.**

Facet Name	Description	Transportation Examples
Type	The type of resource	Technical Report, Data Set, Whitepaper, Drawing
Function	The purpose of the content and any business activity or function that the content is about or related to	Procurement, Human Resources, Engineering, Policy Making
Organization	The organizational unit that the content is about or related to	FHWA; TRB; VDOT
Location	The geographic location or facility that the content is about or related to	Vancouver, WA; San Antonio Federal Complex; 800 E Leigh St, Richmond, VA
Coverage	The period when the content is effective or that it refers to	January 1, 2010–December 31, 2010; September 30, 2009, 5:00 pm; Twentieth Century
Subject	Other themes that the content is about or related to	Maintenance Practices, Domestic Transportation, Liability



a particular version of the software) to be rendered human readable. Simply preserving the bits will not ensure that the files will be readable in the future. Although an electronic version of a technical report may be retrievable based on the words and phrases that occur in the file, GIS, CAD, and many other types of visual electronic content do not contain descriptions that can be used to retrieve them. Descriptive information called metadata needs to be created and linked to the file so such information can be stored and retrieved.

Because most DOT resources are being produced electronically today, digital preservation is becoming the primary focus of records management. Digital preservation, like other forms of records management, requires processes to identify, collect, classify, store, preserve, and maintain digital resources in readable format. See examples in Part 2 for illustrations.

- **Use BI tools to facilitate integration and sharing of data in an easily understandable format.** Many technology tools are available to support the integration and sharing of data and information. One of the most widely used tools is GIS. GIS allows for integration of multiple data layers, which can be displayed on a single entity such as a digital map. For examples, review the material for MaineDOT, UDOT, and WisDOT in Part 2.
- **Use multiple types of BI tools to improve management of data and information.** Other BI tools help DOTs in managing transportation data and information. These tools can include the use of knowledge management systems or electronic cataloging tools to manage content and information. Use of knowledge management systems provides mechanisms for change control and security controls and allow for attaching metadata to documents to support easy query and search capabilities. For examples, review the material for IDOT, NDOT, and VDOT in Part 2.
- **Provide content (data and information) in an electronic format to enable delivery to the widest audience possible.** Because many options are available today for accessing information electronically through the web and/or social media, DOTs should provide transportation information in an electronic format as the primary format, with printed copies as a secondary medium for delivery of information. See examples in Part 2 for illustrations.
- **Select methods for distribution of information and data from the best available technology tools, including the web and social media.** For examples, review the material for KDOT, MDOT, MnDOT, and UDOT in Part 2.
- **Use technology.** Use all types of available technology to store data and information in many formats (photos, images, videos) and to provide easy access to information through the use of web-based systems. Electronic storage systems include those used for cataloging and indexing of documents and acquisitions maintained in a DOT library. These include systems like MS SharePoint which can be used for metadata to describe data sets and information systems supported by the DOT. For examples, review the material for IDOT and NCDOT in Part 2.
- **Use the cloud for storing large datasets.** Many DOTs and state agencies are investigating or using the cloud (i.e., the Internet) to serve as a storage facility for large data sets in lieu of traditional in-house server-based storage of data. In particular, the federal government has put in place a “cloud first” policy, with an aim toward consolidating server farms and saving costs and, at the same time, improving the responsiveness of data providers to changes in agency demands. Although the federal government has mandated that its agencies move elements of their IT infrastructure to the cloud, states are under no such requirement and have decided their level of adoption of the cloud on a case-by-case basis; some states are leading the charge toward cloud-based policy and investment, while others are maintaining their role as owners of data storage infrastructure.

Some DOTs use a private cloud instead of an external cloud. One of the advantages of using a private cloud is that the state has more control over the security of the data and information stored on the cloud compared to using the external cloud available on the Internet. Some states have decided not to use the cloud because of such security concerns. For examples and details, see Part 1.

- **Use taxonomy management tools.** TMS are specialized software that enables an organization's users to view, apply, and modify a common set of controlled vocabulary lists to classify enterprise content. ECM refers to tools that organize information and knowledge, as captured in databases, documents, websites, and repositories. TMS tools are commonly associated with functions such as metadata management, indexing, and search.
- **Use visualization tools (e.g., maps) for delivery and dissemination of data and information.** This is an effective way to communicate and deliver data and information that are easily understood by most of the population. The type of information at a DOT typically displayed on maps includes emergency or planned road closures, locations of high incidents of crashes, construction zone locations, traffic counting and weighing stations, maintenance facilities, and bridges. Many visualization tools are available such as ESRI's ArcGIS, Google Maps, and Intergraph Web Map for display of geospatial data in an interactive format. For examples, review the material for Caltrans, LaDOTD, and UDOT in Part 2.
- **Use market information management techniques, such as a monthly newsletter, to advertise the types of transportation information available from the DOT to the user community.** Several DOTs use their Internet website to disseminate transportation information to the external user community. DOTs, such as Illinois, distribute monthly e-newsletters internally to market the information and cataloged materials available at their DOT libraries. For examples, review the material for IDOT in Part 2.
- **Use "Tell the Story" marketing technique to explain how data is used at the DOT.** DOTs can use this technique to tell data stories to explain how data is used within the agency. For examples, review the material for MnDOT in Part 2.

## Retrieve

- **Explore the use of available technology tools for retrieval of information.** Many technology tools are available to aid in the search for and retrieval of transportation information. These tools are often embedded in software application systems such as ArcGIS (for location data and information) or MS SharePoint (for KM systems). DOTs should investigate options for implementing search and retrieval tools that integrate well with their existing technical architecture infrastructure. DOTs should also consider their database environment and how data and information are stored (including information cataloged in the DOT's library). For examples, review the material for LaDOTD and MnDOT in Part 2.



# Glossary, Terms, Abbreviations

Common terms used by DOTs follow. Sources for the definitions are indicated where applicable. Some definitions have been modified to be more applicable to DOTs.

AADT	average annual daily traffic
AD	Active Directory
ADOT&PF	Alaska Department of Transportation and Public Facilities
API	Application Programming Interfaces
Archive	A collection of records stored for preservation purposes. <sup>44</sup>
AVL	automatic vehicle location
AZDOT	Arizona DOT
BI	Business intelligence
BLOB	Binary large object
BUDS	Business Unit Data Stewards
Business Units	Within the context of a DOT, business unit refers to the divisions, districts, and offices responsible for conducting the business functions of the agency.
BYOD	bring your own device
CAC	critical accident corridor
CAD	Computer-aided design
CAL	critical accident location
Category Value	A set of controlled values for populating a metadata element or attribute used to describe and/or manage content items.
CCTV	closed-circuit television
CDOT	Colorado DOT
CFR	Code of Federal Regulations
CIO	Chief Information Officer
Classification Scheme	A method of organization according to a set of pre-established principles, usually characterized by a notation system and a hierarchical structure of relationships among the entities. <sup>45</sup>
CMS	Changeable message sign
COE	Corps of Engineers
COI	Community of Interest
Content Collection	A set of content being managed, e.g., data, information, knowledge of various types such as documents, CAD, maps, and graphics.

<sup>44</sup>ANSI/AIIM/ARMA TR48-2006, *Revised Framework for Integration of Electronic Document Management Systems and Electronic Records Management Systems*.

<sup>45</sup>Washington State DOT, 2008, *Glossary for Findability and Controlled Vocabularies*.

Content Management	Creating, managing, and using resources such as data, information, and knowledge to support organizational activities.
Content	A facet of information, it includes data, information, and knowledge of various types, such as documents, CAD, maps and graphics. Content management is the set of processes and technologies related to the collection, management and publication of information so that it can effectively be found and used to accomplish tasks.
Controlled Vocabulary	An established list of preferred terms for populating a metadata element or attribute used to describe and/or manage content items.
Conversion	Process of changing a document to another form, e.g., paper to microform, paper to digital image, microform to digital image, etc. <sup>46</sup>
COP	Community of practice
COS	Capital outlay services
CRM	Customer Relationship Managers
CVT	City, Village and Town
Data Management	A set of processes and technologies related to collecting and managing data so that it can effectively be used to accomplish tasks.
Data	A representation of facts, concepts, or instructions in a formalized manner suitable for communication, interpretation, or processing by humans or computers. <sup>50</sup>
DAWG	Data Architecture Work Group
DCMI	Dublin Core® Metadata Initiative
DCMS	Digital Collections Management System
DMA	Data Management Association
DMWG	Document Management and Workflow Work Group
Document Management	The management of documents in order to manage, control, locate, and retrieve information in a system. <sup>1</sup>
Document	Recorded information that can be treated as a unit. The term is employed here to include narrative information plus maps, databases, technical drawings, audio and video objects, and other content saved as digital files. <sup>1</sup>
DOT	Department of transportation
DRS	Document Retrieval System
DTS	distributed terminology system
ECM	Enterprise content management
EDMS	Electronic Document Management System
Electronic Document Management	The electronic management of documents contained in an information technology system, using computer equipment and software to manage, control, locate, and retrieve information in the system. <sup>1</sup>
Electronic Records Management (ERM)	The management of electronic and non-electronic records contained in an information technology system using computer equipment and software according to accepted principles and practices of records management. <sup>1</sup>
ERM	Electronic Records Management
ERMS	Electronic Records Management System
ERS	Electronic Recordkeeping System
ETL	Extract Transform and Load
FAQ	Frequently Asked Questions
FGDC	Federal Geographic Data Committee
FOIA	Freedom of Information Act

<sup>46</sup>ANSI/AIIM TR2-1998, *Glossary of Document Technologies*.

GASB 34	Governmental Accounting Standards Board Statement 34
GDOT	Georgia DOT
GEE	Google Earth Enterprise
GIO	Geospatial Information Officer
GIS	Geographic Information System
Glossary	An alphabetical list of terms in a specific subject area—dictionaries usually have a broader scope. In the context of this project, glossary and dictionary mean the same thing.
GMC	GIS Management Committee
GPS	Global Positioning System
HOT	High-occupancy travel
HPMS	Highway Performance Monitoring System
IaaS	Infrastructure as a Service
IAL	intersection accident location
ICT	Illinois Center for Transportation
IDOT	Illinois DOT
IEPD	Information Exchange Package Documents
Indexing	A method by which terms or subject headings from a controlled vocabulary are selected by a human or computer to represent the concepts in or attributes of a content object. The terms may or may not occur in the content object. <sup>2</sup>
Information Management	An overarching term used to describe the organization and control over the collection and management of information from one or more sources, and the processing, delivery, and sharing of that information to users. <sup>47</sup> Information management addresses how an organization develops policies, practices, technologies, and other resources and capabilities to manage information as a resource. This includes the management functions associated with the creation, collection, processing, transmission, dissemination, use, storage, and disposition of information, both automated and non-automated. <sup>48</sup> A simpler version of the definition is from AIIM (The Global Community of Information Professionals): “Information management is the collection and management of information from one or more sources and the distribution of that information to one or more audiences.” <sup>49</sup>
Information Technology (IT)	The application of methods and scientific knowledge for the collection of information. IT is commonly associated with the use of electronic principles and devices for the development and implementation of electronic systems. <sup>51</sup>
Information	Data and documents that have been given value through analysis, interpretation, or compilation in a meaningful form. <sup>4</sup>
Integration	The combination of several software applications such that data can be transferred from one application to others through a consistent interface so as to better coordinate tasks and merge information. <sup>51</sup>

<sup>47</sup>Caudle and Marchand, 1989, *Managing Information Resources*.

<sup>48</sup>Caudle and Marchand, 1989, *Managing Information Resources*.

<sup>49</sup><http://www.aiim.org/what-is-information-management>.

<sup>50</sup>Turban, et al., 2008; Ward and Peppard, 2002, *Strategic Planning for Information Systems*.

<sup>51</sup>ISO 15489, *Information and Documentation — Records Management*.

IP	Intellectual Property
IR	Information retrieval
ISD	Information Systems Division
ISL	Illinois State Library
ISP	Information Service Providers
IT	information technology
ITAA	IT Application Administrator
KDOT	Kansas DOT
Keyword Match	A search algorithm used by several mainstream search engines including Google. This method returns results based on the frequency of the search term or phrase in the document. <sup>52</sup>
KM	Knowledge Management
KMGOC	Knowledge Management Governance Oversight Committee
Knowledge	Information that has been combined with experience, context, and interpretation that make it possible to understand and draw implications from both data and information. Information and knowledge are closely related, but not the same. Knowledge consists of data and information that have been organized and processed to convey understanding, experience, accumulated learning, and expertise as they apply to a current problem or activity. <sup>50</sup>
LaDOTD	Louisiana Department of Transportation and Development
LC	Library of Congress
LCS	Lane closure system
LCSH	Library of Congress Subject Headings
LDAP	Lightweight directory access protocol
LiDAR	Light Detection and Ranging
Lifecycle	The course of developmental changes through which information or an information system passes from initial creation through mature uses to final disposition or replacement. <sup>51</sup>
LOS	Level of Service
LRS	Linear Referencing System
MDOT	Mississippi DOT
MDR	Metadata Registries
Metadata Repository	“A metadata repository is a database, either physical or virtual, that contains data about data (i.e., metadata). An effective repository stores and organizes metadata, consistently and reliably describing the data available, where it is located, what it means, who collects the data, who is responsible for the data quality, and how to access the data.” <sup>52</sup>
Metadata	Data describing context, content, and structure of documents and records and their management through time. Literally, data about data. <sup>1</sup>
MMA	Managed Metadata Service Administrator
MnDOT	Minnesota DOT
MOU	memorandum of understanding
MPO	Metropolitan planning organizations
MS	Microsoft SharePoint
MTKN	Midwest Transportation Knowledge Network
NCDOT	North Carolina Department of Transportation



NDOT	Nevada Department of Transportation
NGO	Non-governmental organization
NIEM	National Information Exchange Model
NIST	National Institutes of Standards and Technology
NOAA	National Oceanic and Atmospheric Administration
NSDI	National Spatial Data Infrastructure
NTL	National Transportation Library
NYSDOT	New York State DOT
OAIS	Open Archival Information System
OCIO	Office of the Chief Information Officer
ODAGIS	Office of Data Analysis and Geospatial Information Systems
ODOT	Ohio DOT
OHSIP	Office of Highway System Information and Performance
OMB	Office of Management and Budget
Ontology	A faceted taxonomy that uses richer semantic relationships among terms and attributes and strict specification rules.
ORNL	Oak Ridge National Laboratory
OTFA	Office of Travel Forecasting and Analysis
PaaS	Platform as a Service
PC	Personal computer
PeMS	Performance Measurement System
PM	Performance Management
PMO	Project Management Office
PMRS	Project Management Reporting System
POS	Part of speech
PRC	Policy and Research Center
Precision	A measure of a search system's ability to retrieve only relevant content objects. It is usually expressed as a percentage calculated by dividing the number of retrieved relevant content objects by the total number of content objects retrieved. A high-precision search ensures that, for the most part, the content objects retrieved will be relevant. However, a high-precision search may not retrieve all relevant content objects. See also recall. Recall and precision tend to be inverse ratios. When one goes up, the other usually goes down.
Preservation	Processes and operations involved in ensuring the technical and intellectual survival of authentic records through time. <sup>51</sup>
PRKS	Project Record Keeping System
QA	Quality Assurance
QC	Quality Control
RDA	Resource Description and Access
RDF	Resource Description Framework
RDFS	RDF Schema
Recall	A measure of a search system's ability to retrieve all relevant content objects. Usually expressed as a percentage calculated by dividing the number of retrieved relevant content objects by the number of all relevant content objects in a collection. A high recall search retrieves a comprehensive set of relevant content objects from the collection. However, high recall increases the possibility that less relevant content objects will also be retrieved. See also precision. Recall and precision tend to be inverse ratios. When one goes up, the other usually goes down.

Record	Information created, received, and maintained as evidence, or information created by an organization or person in pursuance of legal obligations or in the transaction of business. <sup>48</sup>
Records Management Application (RMA)	A software system that performs electronic records management according to an accepted Electronic Records Management Systems (ERMS) standard. <sup>51</sup>
Records Management	The planning, controlling, directing, organizing, training, promoting, and other managerial activities involved with creating, maintaining, using, and disposition of records to achieve adequate and proper documentation of organization and agency policies and transactions and the effective and economical management of the organization or agency operations. <sup>51</sup>
Repository	A direct access device on which the electronic records and associated metadata are stored. Also a synonym for archive. <sup>51</sup>
RFP	request for proposal
RFQ	Request for Quotation
RMA	Records Management Application
ROC	Records Oversight Committee
ROI	Return on investment
RSS	Really simple server
RT	Related Terms
RWIS	Roadside weather information systems
SaaS	Software as a Service
SAN	storage area network
SGT	SharePoint Governance Team
SKOS	Simple Knowledge Organization System
SLA	Special Libraries Association
SME	Subject matter experts
SO	System Operations
SOP	Standard operating procedures
SPMG	Statewide Program Management Group
STCDO	Statewide Travel & Collision Data Office
Synonym Ring	A synonym ring is a set of words or phrases that can be used interchangeably for searching, e.g., “Fringe parking” and “Park and ride.”
Taxonomy	A system for identifying and naming things, and arranging them according to a set of rules into a series of classes and sub-classes.
Term	One or more words designating a concept.
Thesaurus	An alphabetical lexicon that controls synonyms, and identifies the semantic relationships among terms.
TKN	Transportation Knowledge Network
TMS	Taxonomy management systems
TransIT	transportation information technology
Transportation Research Thesaurus (TRT)	A controlled vocabulary developed for the transportation domain to index transportation-related documents. The TRT is managed by TRB and advised by the TRB Committee on Library and Information Science in Transportation’s subcommittee for the TRT. <sup>2</sup>
TRID	Transport Research International Documentation
TRR	Transportation Research Records
TRT	Transportation Research Thesaurus
TSI	Transportation System Information
UDOT	Utah Department of Transportation

UPC	Universal Product Code
URI	Universal resource identifiers
URL	Uniform resource locators
USGS	United States Geological Survey
VDOT	Virginia DOT
W3C	World Wide Web Consortium
WBS	Work breakdown structure
WCM	Web content management
WIM	Weigh-In-Motion
WisDOT	Wisconsin Department of Transportation
WISLR	Wisconsin Information System for Local Roads
WSDL	Web services description language
WSDOT	Washington State DOT
WSL	Washington State Library

*Abbreviations and acronyms used without definitions in TRB publications:*

A4A	Airlines for America
AAAAE	American Association of Airport Executives
AASHO	American Association of State Highway Officials
AASHTO	American Association of State Highway and Transportation Officials
ACI-NA	Airports Council International-North America
ACRP	Airport Cooperative Research Program
ADA	Americans with Disabilities Act
APTA	American Public Transportation Association
ASCE	American Society of Civil Engineers
ASME	American Society of Mechanical Engineers
ASTM	American Society for Testing and Materials
ATA	American Trucking Associations
CTAA	Community Transportation Association of America
CTBSSP	Commercial Truck and Bus Safety Synthesis Program
DHS	Department of Homeland Security
DOE	Department of Energy
EPA	Environmental Protection Agency
FAA	Federal Aviation Administration
FHWA	Federal Highway Administration
FMCSA	Federal Motor Carrier Safety Administration
FRA	Federal Railroad Administration
FTA	Federal Transit Administration
HMCRP	Hazardous Materials Cooperative Research Program
IEEE	Institute of Electrical and Electronics Engineers
ISTEA	Intermodal Surface Transportation Efficiency Act of 1991
ITE	Institute of Transportation Engineers
MAP-21	Moving Ahead for Progress in the 21st Century Act (2012)
NASA	National Aeronautics and Space Administration
NASAO	National Association of State Aviation Officials
NCFRP	National Cooperative Freight Research Program
NCHRP	National Cooperative Highway Research Program
NHTSA	National Highway Traffic Safety Administration
NTSB	National Transportation Safety Board
PHMSA	Pipeline and Hazardous Materials Safety Administration
RITA	Research and Innovative Technology Administration
SAE	Society of Automotive Engineers
SAFETEA-LU	Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (2005)
TCRP	Transit Cooperative Research Program
TEA-21	Transportation Equity Act for the 21st Century (1998)
TRB	Transportation Research Board
TSA	Transportation Security Administration
U.S.DOT	United States Department of Transportation