OKLAHOMA TRANSPORTATION CENTER

Economic Enhancement through Infrastructure Stewardship

Data and Information Integration Framework for Highway Project Decision Making

M. Phil Lewis, Ph.D., P.E.
"David" Hyung S. Jeong, Ph.D.
Asregedew Woldesenbet

OTCREOS11.1-30-F

Oklahoma Transportation Center
2601 Liberty Parkway, Suite 110
Midwest City, Oklahoma 73110

Phone: 405.732.6580
Fax: 405.732.6586
www.oktc.org

DISCLAIMER

TECHNICAL REPORT DOCUMENTATION PAGE

1. REPORT NO. OTCREOS11.1-30-F	**2. GOVERNMENT ACCESSION NO.**	**3. RECIPIENTS CATALOG NO.**
4. TITLE AND SUBTITLE Data and Information Integration Framework for Highway Project Decision Making		**5. REPORT DATE** August 21, 2013 **6. PERFORMING ORGANIZATION CODE**
7. AUTHOR(S) M. Phil Lewis, “David” Hyung Seok Jeong, and Asregedew Woldesenbet		**8. PERFORMING ORGANIZATION REPORT**
9. PERFORMING ORGANIZATION NAME AND ADDRESS Oklahoma State University Civil & Environmental Engineering 207 Engineering South Stillwater, OK 74078		**10. WORK UNIT NO.** **11. CONTRACT OR GRANT NO.** DTRT06-G-0016
12. SPONSORING AGENCY NAME AND ADDRESS Oklahoma Transportation Center (Fiscal) 201 ATRC, Stillwater, OK 74078 (Technical) 2601 Liberty Parkway, Suite 110 Midwest City, OK 73110		**13. TYPE OF REPORT AND PERIOD COVERED** Final June 2012 – June 2013 **14. SPONSORING AGENCY CODE** University Transportation Center

15. SUPPLEMENTARY NOTES
University Transportation Center

The research team consisted of members from Oklahoma State University and Iowa State University.

16. ABSTRACT

This report presents a three-tiered framework to integrate data, information, and decision-making in highway projects. The study uses the Juran’s Triple Role concept and context graph to illustrate the relationship between data, information, and decision-making. The study discusses the complexity of data and information flow through a conceptual 3-D data flow diagram. The evolution of data and information integration in highway agencies is also outlined in the report. Potential methodologies in extracting information from raw qualitative and quantitative data are summarized. The developed framework was applied to three case studies; a) construction daily work reports, b) preconstruction cost estimation, and c) pavement management. In each case study, the three-tiered framework was applied and the current and ideal level of data and information integration for key decisions were identified and areas for improvement were also identified.

The developed framework will guide DOTs on how to generate and place right information and knowledge in the hands of decision-makers. The implementation of the framework will empower engineers to make informed and justifiable decisions, and lead to the improved accountability of project development and management. In addition, it will allow the active utilization of currently existing databases and justify the continuous and growing data collection efforts by DOTs. Furthermore, DOTs will be able to measure their performances and develop an advanced data collection and information/knowledge generation plan to support key decisions which historically were not well supported with information and data.

17. KEY WORDS Data collection, data and information integration, highway project, decision making	**18. DISTRIBUTION STATEMENT** No restrictions. This publication is available at www.oktc.org and from the NTIS.		
19. SECURITY CLASSIF. (OF THIS REPORT) Unclassified	**20. SECURITY CLASSIF. (OF THIS PAGE)** Unclassified	**21. NO. OF PAGES** 125 + covers	**22. PRICE**

SI (METRIC) CONVERSION FACTORS

Approximate Conversions to SI Units

Symbol	When you know	Multiply by	To Find	Symbol
		LENGTH		
in	inches	25.40	millimeters	mm
ft	feet	0.3048	meters	m
yd	yards	0.9144	meters	m
mi	miles	1.609	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm
ft^2	square feet	0.0929	square meters	m^2
yd^2	square yards	0.8361	square meters	m^2
ac	acres	0.4047	hectares	ha
mi^2	square miles	2.590	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.0283	cubic meters	m^3
yd^3	cubic yards	0.7645	cubic meters	m^3
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.4536	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams	Mg
		TEMPERATURE (exact)		
°F	degrees Fahrenheit	(°F-32)/1.8	degrees Celsius	°C
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.448	Newtons	N
lbf/in^2	poundforce per square inch	6.895	kilopascals	kPa

Approximate Conversions from SI Units

Symbol	When you know	Multiply by	To Find	Symbol
		LENGTH		
mm	millimeters	0.0394	inches	in
m	meters	3.281	feet	ft
m	meters	1.094	yards	yd
km	kilometers	0.6214	miles	mi
		AREA		
mm^2	square millimeters	0.00155	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.196	square yards	yd^2
ha	hectares	2.471	acres	ac
km^2	square kilometers	0.3861	square miles	mi^2
		VOLUME		
mL	milliliters	0.0338	fluid ounces	fl oz
L	liters	0.2642	gallons	gal
m^3	cubic meters	35.315	cubic feet	ft^3
m^3	cubic meters	1.308	cubic yards	yd^3
		MASS		
g	grams	0.0353	ounces	oz
kg	kilograms	2.205	pounds	lb
Mg	megagrams	1.1023	short tons (2000 lb)	T
		TEMPERATURE (exact)		
°C	degrees Celsius	9/5+32	degrees Fahrenheit	°F
		FORCE and PRESSURE or STRESS		
N	Newtons	0.2248	poundforce	lbf
kPa	kilopascals	0.1450	poundforce per square inch	lbf/in^2

ACKNOWLEDGEMENTS

The authors would like to thank Oklahoma Department of Transportation (ODOT) Pavement Management Division, Mr. William Dickinson, Roadway Division Engineer and Mr. Kirk Goins, construction database management team, Mr. Brad Hartronft, and Ms. Ann Wilson for sharing their experience, support in understanding various ODOT database system, data and information utilization and highway decision-making process.

DATA AND INFORMATION INTEGRATION FRAMEWORK FOR HIGHWAY PROJECT DECISION-MAKING

Final Report

June 30, 2013

M. Phil Lewis, Ph.D., P.E.
Assistant Professor (Principal Investigator)
School of Civil & Environmental Engineering
Oklahoma State University
Stillwater, Oklahoma

"David" Hyung Seok Jeong, Ph.D.
Associate Professor (Co-Principal Investigator)
School of Civil, Construction & Environmental Engineering
Iowa State University
Ames, Iowa

Asregedew Woldesenbet
Graduate Research Assistant
School of Civil, Construction & Environmental Engineering
Iowa State University
Ames, Iowa

Table of Contents

List of Tables

List of Figures

Executive Summary

A well-documented data plan with clear data flow, efficient data management system, appropriate tools and approaches to store, extract, and manage information and knowledge, would enhance the potential user's decision-making process. Currently, data is being considered as the new "oil" of our era. Leading organizations in various industries are utilizing sophisticated qualitative and statistical analysis method through the support of information technology to improve the use of information available to managers. These organizations and enterprises utilize available data to support their decisions in terms of planning their program, designing their activities, setting their priorities and measuring their performance to improve the level of service, satisfy their customers and increase their profit. To address this important discovery and reach the level where other organizations have reached, the highway industry must undergo a paradigm shift that emphasizes toward the integration of data, information and decision-making processes.

This study develops a three tiered data and information integration framework that can ultimately support various decisions over the life-cycle of highway projects. The study uses the Juran's Triple Role concept and context graph to illustrate the relationship between data, information and decision-making. An ideal framework is developed that analyzes gap in terms of the current level of data usage to effectively utilize data and information in supporting highway decisions. The study uses a conceptual 3-D data flow diagram to represent the flow data and information in highway decisions. In addition, it discusses potential methodologies in extracting information from raw qualitative and quantitative data. The study evaluates the framework through three case-studies; preconstruction cost estimation, construction daily work reports and pavement management.

The developed framework will guide DOTs on how to generate and place right information and knowledge in the hands of decision makers. It will empower engineers to make informed and justifiable decisions, and lead to the improved accountability of project development and management. The framework will allow the active utilization of currently existing databases and justify continuous and growing data collection efforts by DOTs. In addition, it will allow DOTs to measure their current performances and develop an advanced data collection and information/knowledge generation plan to support key decisions which historically were not well supported with information and data.

1. Introduction

1.1 Overview

The advancement in digital technology, data collection methods and centralized information and/ or database management systems has contributed to a rapid and economic data and information generation. Many strategic business decisions are utilizing these technologies along with management philosophies and data analysis methods to extract statistically reliable information and knowledge to support their daily business decisions. For instance, the credit card industry analyzes a credit card holder's spending behavior and his/her demographic statistics to adjust the customer's interest rate and identify any fraudulent activity. The medical industry is actively utilizing health care records and clinical data acquired during patient care to obtain patient's optimal health, drive new medical discoveries and preventive measures. The retail industry utilizes customer's purchase habit to design coupons, plan store layout, and attract new customers accordingly to maximize profit. Even basketball statistics are analyzed to identify key matchups in upcoming games (Bhandari et al. 1995).

The emergence and applications of knowledge management (KM) tools, and knowledge discovery in database (KDD) approaches such as data mining (DM) techniques, decision support systems, artificial intelligence, machine learning and business intelligence tools, knowledge bases (KB) and expert systems are used to effectively extract pattern, information and knowledge from a vast amount of data. In addition, there has been significant development and advancement through knowledge representation models such as ontology-based systems for active knowledge sharing and customized knowledge retrieval using an integrated data and information system. These techniques and applications are utilized in conjunction with management philosophies such as concurrent engineering (CE), lean construction, business process re-engineering (BPR), total quality management (TQM), supply-chain management (SCM) and just-in-time production to improve decision-making (Bjork, 1999).

1.2 Problem Statement

Transportation agencies and state DOTs collect a huge amount of highway project data ranging from roadway inventory to pavement condition data during the life cycle of highway projects.

These data are collected and stored as mandatory federal requirement and/or asset management programs to avoid multiple data entry and manage highway projects. However, the return on investment is not significant as compared to the amount of resources (time and money) spent on the data collection efforts. For example, Oklahoma Department of Transportation (ODOT) invests at least $600,000 annually to collect statewide pavement condition data and each resident engineer spends approximately two to three hours a week to develop daily work reports (DWRs). Most importantly, little effort has been placed in extracting information and knowledge from these data to support decision-making. Sometimes, potential users of these data such as highway project schedulers, estimators, and managers do not even know what type of data are available or how to access these data to support their decisions. Some of the reasons that can contribute to the poor usage of data and information to support decision-making might be due to a) minimal recognition or interest in using these data in the context of supporting various decision making processes during the life cycle of a highway project; b) lack of in-house resources and capabilities to analyze the data; c) insufficiency of data for any meaningful analysis or the data not stored in a standard and digital format for analysis; and/or d) not well-defined procedure and mechanism to extract, process, and analyze the data and generate usable information and knowledge to assist highway project decision makers. Therefore, there is a need to assess the current usage and develop a system to integrate data and information to support decision-making.

1.3 Research Objectives

The ultimate goal of this study is to develop a data and information integration framework for highway project decision making. In order to accomplish this goal, the following objectives are set:

1) Identify data and information required to support decision making
2) Identify the current level of data collection and management efforts
3) Develop a data and information integration framework
4) Perform case studies to apply the framework and analyze the gap between the current data utilization and ideal data and information integration to support highway agencies' decisions.

1.4 Scope of Study

Multiple highway decisions are made every day from the planning phase to the operation and maintenance phase and from the strategic and/or corporate level to the project or division level to effectively manage projects and attain transportation goals. Since the study cannot address all decisions, the scope of this study is limited to the technical aspects of three case studies over the life-cycle of highway projects: a) preconstruction cost estimating at the preconstruction stage, b) utilization of construction daily work reports generated during the construction stage and c) pavement management decisions at the operation and maintenance stage (Figure 1-1). In addition, this study classifies data into four broad categories to address the basic components of a project management; cost (budget), quality, time (schedule) and safety and environment.

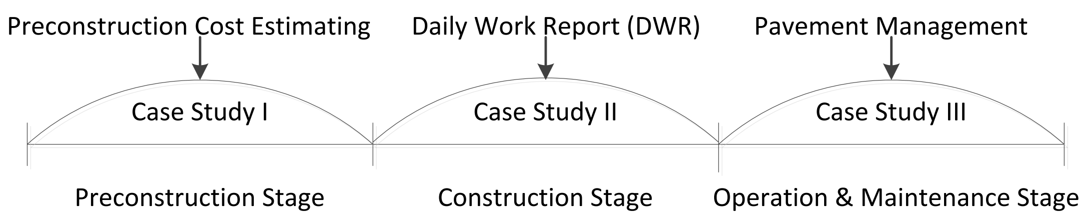


Figure 1-1 Scope of Study

1.5 Research Methodology

This study is divided into three sections to address the objectives of this research as shown in Figure 1-2. Primarily, review of prior studies, meetings and interviews with highway division decision makers and database mangers are conducted to identify current academic and industry practices, identify decisions and required information. Next, a data-information-decision making integration framework is developed to map data, information and knowledge that supports decision making. In addition, case studies are conducted on three major decision making processes to apply the framework and analyze the current database systems. Finally, a gap analysis report is prepared to compare the current data and information utilization with an ideal data collection and information framework.

1.5.1 Literature Review

The literature review focuses on prior research and relevant studies on identifying data, information and decision-making integration in the transportation industry. It investigates key decisions and various databases managed by state Department of Transportations (DOTs) during the life cycle of a highway project development. The study also discusses potential data analysis methods and approaches that can be used to generate information and support the decision-making process.

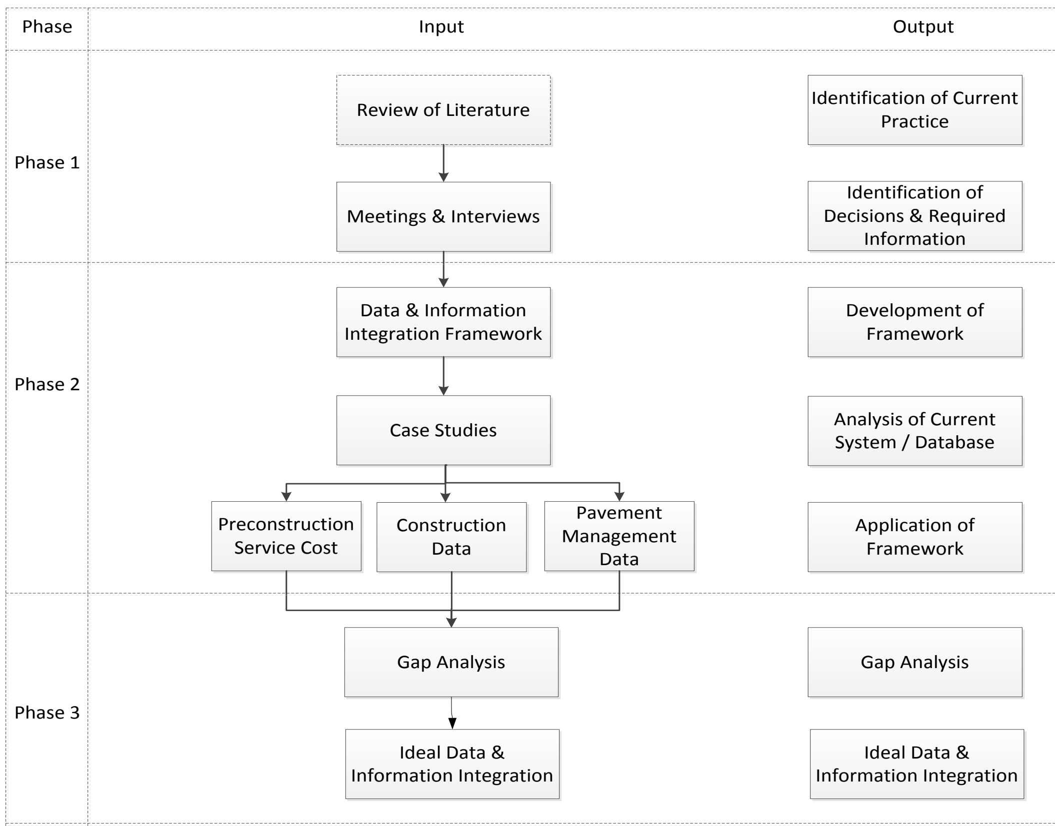


Figure 1-2 Research Methodology Flow Chart

1.5.2 Identification of Decisions and Required Information

A series of meetings and interviews is conducted with Oklahoma DOT (ODOT) and central office highway division database managers and decision makers to identify the current level of data collection and management efforts at different stages of a highway project. These meetings and interviews are used to identify the needs and requirements of data and information in terms of availability and usage to support highway decision-making.

1.5.3 Data and Information Integration Framework

Based on the collected data and information, the study introduces a three-tiered data-information-decision making integration framework to support highway project decisions. It maps the data attributes identified from the various databases to improve the quality of data and information usage to support highway decisions. The study presents a conceptual 3-D data flow diagram (DFD) to illustrate the complexity of highway decision-making. It discusses information generation using appropriate data analysis methods to extract information and knowledge.

1.5.4 Case studies

This research uses three case studies to apply the three-tiered framework developed in the previous section. It will evaluate the data-information-decision-making integration in ODOT. The study presents a gap analysis using an input/output matrix to illustrate current data usage and possible implementation of data utilization. The study uses an ideal data collection and information integration framework to map three types of paths; a) an active path, a path from data to information to decision making which is currently in active use by DOTs, b) an inactive path, a path from data to information to decision making which is currently inactive and c) a non-existing path, a path in which data is not available to generate required information to support a specific decision and the information extraction method is not known

1.5.5 Recommendations and Conclusions

In this section, key findings and contribution of the research are summarized. A guideline is outlined to effectively utilize data in extracting information and support highway decisions. This section summarizes the conclusions drawn from this research and recommends future studies required to integrate data, information and decision-making as an extension of this research.

1.6 Report Organization

This report is organized into four chapters. Chapter 2 summarizes prior studies and approaches conducted in utilization of data and information to support decision-making. Chapter 3 introduces a three-tiered data, information and decision making integration framework. A case study that demonstrates the application of the developed framework is discussed in Chapter 4. A gap analysis of the data and information utilization is also included in this section. The final chapter summarizes the findings, contribution, and future studies of this research.

2. Review of Literature

This chapter discusses previous studies conducted and related contents in data, information and decision-making in the transportation industry. It summarizes decision-making processes practiced by highway agencies, key decisions made in project development, and various databases managed by state Department of Transportations (DOT) and highway agencies. In addition, potential data analysis methods and approaches that are currently used to generate information and support decision-making processes are presented. The study also discusses data and information integration and interoperability efforts practiced by various industries.

2.1 Introduction

Today, active utilization of data in extracting information and knowledge and supporting the decision-making process is receiving considerable attention from users ranging from engineers and managers to, software developers and/or consultants. Data is being considered as the new "oil" of our era. Efficient management and business decisions are supported through well documented data, clear data flow, better data management system, advanced technology, tools and/or approaches to store, manage and enhance potential user's daily decision-making process. Davenport and Harris (2007) argue that the frontier of data usage in making decisions is shifting drastically in such a way that high-performing enterprises are building their competitive strategies around data-driven insights that in turn generate better business results. Leading organizations are 'competing on analytics' by utilizing sophisticated qualitative and statistical analysis through information technology to improve the use of information available to managers (Kennerley and Mason, 2008).

These organizations and/or enterprises utilize available data to support their decisions in terms of planning their program, designing their activities, setting their priorities and measuring their performance to improve their service, satisfy their customers and increase their profit. For example, the credit card industry analyzes a credit card holder's spending behavior and his/her demographic statistics to adjust the customer's interest rate and identify any fraudulent activity. The medical industry is actively utilizing patients' health care records and clinical data acquired during patient care to obtain patient's optimal health, drive new medical discoveries and preventive measures. The retail industry utilizes customer's purchase habits to design coupons, plan store layout, and attract new customers accordingly to maximize profit. Even basketball statistics are analyzed to identify key matchups in upcoming games (Bhandari et al. 1995).

This rapid data demand and growth has led to the development of programs, digital data storage technologies, database management systems (databases and data warehouses), applications and business intelligence tools including data mining, knowledge discovery in database (KDD) approaches, and ontology-based frameworks to effectively organize, extract pattern and obtain information and knowledge from the vast amount of data that is available in these organizations/enterprises. This advancement is improving employees' productivity; increasing data quality and accuracy; reducing the difficulty in data collection; enhancing active knowledge sharing, better communication and information retrieval to support decision-making process. Therefore, proper mechanisms should be applied in terms of collecting, storing, understanding and converting the valuable pieces of data into information and knowledge to support key decisions.

2.2 Information Hierarchy

The primary step in utilization of data is to understand the relationship between "data", "information", and "knowledge" as these terms are sometimes found to be overlapping and are used interchangeably depending on the context. Information hierarchy also known as 'data-information-knowledge-wisdom (DIKW) pyramid', 'wisdom hierarchy', or 'knowledge pyramid' is one of the fundamental models found in the information and knowledge management literature to illustrate the structural and functional relationships between data, information, knowledge and wisdom (Rowley, 2007). The basic assumption of the DIKW model is that data can be used to generate information; information can be used to generate knowledge; and knowledge can be used to generate wisdom. A typical information hierarchy based on the level of use and maturity level is shown in Figure 2-1.

Data

The English Merriam-Webster dictionary defines "*data as an evidence used as a basis for reasoning, discussion or calculation*". It is a collection of facts derived from measurements and/or observations. The word "data" comes from the Latin word datum (singular form) which means "to give". Data can include numbers, words, figures or images that requires the necessary process or organization to make it meaningful or answer specific questions. At this stage, data value is negligible unless it is not converted to a usable form or information. In this study, “data” is referred to raw data collected from highway projects and stored in data repository or databases.

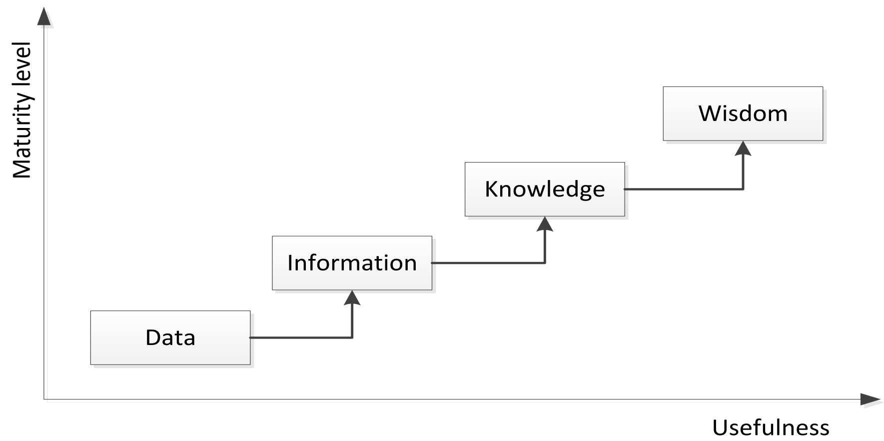


Figure 2-1 Information Hierarchy (Ackoff, 1989)

Information

Information can be defined as “*an intelligence or findings obtained from investigation, study or instruction; or a quantitative measure of the content of data*” (Merriam-Webster dictionary). Information can be either a direct form of data which does not require any form of change or a combination of one or more data that is processed, structured and manipulated to increase users’ understanding. It is an organized data that adds value to a user in providing answers as to who, what, where and when types of questions (Ackoff, 1989). In this study, information is represented by performance measures and/or outputs that can be used to support decisions as a result of analysis performed to data.

Knowledge

Although there is not a single agreed definition, in epistemology, knowledge is characterized by justification, truth and belief. The English Merriam-Webster dictionary describes “*knowledge as the fact or condition of knowing something with familiarity gained through experience or association or acquaintance with or understanding of science, art or technique*”. At this stage, observation of patterns and understanding information is well-perceived. In this study, knowledge is considered based on the facts acquired through experience and data analysis outputs (information) to support decisions and form judgments.

Wisdom

The highest level of knowledge management hierarchy is wisdom. Wisdom is the ability to utilize knowledge using thorough realization, deep understanding and experience of terms, events, and circumstances acquired through time. The English Merriam-Webster dictionary describes "*wisdom as a combination of knowledge, insight and judgment through accumulated philosophic, scientific learning, good sense and discerning qualities and relationships".*

Decision

At this level, decision is represented by the application of knowledge and wisdom to promote business judgment, gain competitive advantage and visualize long-term goals and consequences. A decision is "*a final product of the specific mental/cognitive process of an individual or a group of persons/organizations to arrive at certain conclusion*" (Kennerley and Mason, 2008). Decisions can be as simple as yes or no and choosing among alternatives to complex analysis of finding relations to obtain a more reliable and justifiable result. In this study, decision is drawn by highway managers' to execute highway projects based on the knowledge obtained from analysis, experience and critical thinking process to meet transportation goals.

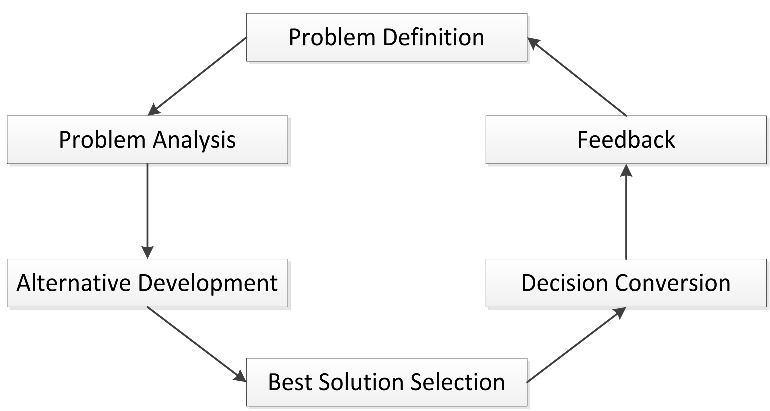


Figure 2-2 Decision-Making Process (Drucker, 1955)

Decision-Making

Decision-making is the process or act of making final judgments or selection based on available alternatives to attain a certain level of required goals and/or objectives. Decision-making can be defined as "*the action of carrying out or carrying into effect*". It is a reasoning process that can range from rational and formal to irrational and informal method based on explicit or implicit knowledge. Although various studies suggest different types of decision-making steps or procedures, Drucker (1955) classified the key elements of scientific decision-making into six steps from the management point of view: a) defining/identifying a managerial problem, b) analyzing the problem, c) developing alternative solutions, d) selecting the best solution out of alternatives, e) converting the decision into action, and f) ensuring feedbacks for follow-ups (Figure 2-2).

Similarly, the European commission of project management and stakeholder guide map also recommended a six stage process as a general principle from transportation decision-making perspective (EU, 2000). This include problem definition, option generation, option assessment, formal decision-making, implementation and monitoring and evaluation. Neely and Jarrar (2004) proposed a performance-planning-value-chain (PPVC), a six step process to improve decision-making; hypothesizing from data utilization perspective; collecting; analyzing; interpreting; communicating; and making informed decisions. In this study, decision-making refers to a process which involves the use of one or more raw data collected at different stages of a highway project life-cycle and generating information and knowledge by applying appropriate analysis that tends toward supporting the selection, judgment and execution process of highway project management. Some good examples of transportation decision include pavement treatment selection, project selection, contract time determination, etc. Figure 2-3 shows a decision-making process from data utilization perspective adapted from Neely and Jarrar (2004).

Figure 2-3 Decision-Making Process

However, it is important to note that this linear decision-making process can be repetitive, parallel, or cyclical that might trigger a second process depending on the project phase and approaches used. May (2003) classified transportation decision-making approaches as vision-led (dependent on individual vision), plan-led (dependent on professional planners based on set of procedures), objective-led (achieve high-level objectives and identify problems and barriers to be addressed) and consensus-led (based on active involvement of various stakeholders to reach agreement at each stage). This study uses a combination of these approaches to discuss prior studies and current practice of data, information, and knowledge utilization in the highway project decision-making processes.

2.3 Key Decisions during Highway Project Development

In a typical highway project, different decisions are made by different highway divisions at various levels across the life cycle of a project to enhance pavement life, improve quality, reduce cost and increase public safety for the road user and transportation agency. For example, transportation asset management decisions can be made at strategic level, network level, and project level based on asset management perspective (Flintsch and Bryant 2006). Strategic level deals with decisions made by higher level officials such as the commissioner, director and/or governor in setting policies, identifying objectives, resource allocation and utilization. Network level deals with decisions such as determining the overall scope of an agency's needs, budget allocation and transportation planning. The network level is further broken-down into program level and project selection level. The program level deals with the overall highway program in terms of setting capital improvement plan, development of programs by metropolitan planning organization (MPO) and rural planning agencies (RPO) while the project selection level deals with prioritization of projects and pavement treatment selection. Project level decisions involve design and management of projects to meet the work plan assigned by higher levels to meet the agency's goal. Decisions can range from selection of pavement type and cost estimating, to safety improvement and traffic control. The project level, also called field or operational level is involved in providing tools to aid in optimization of actual work accomplished.

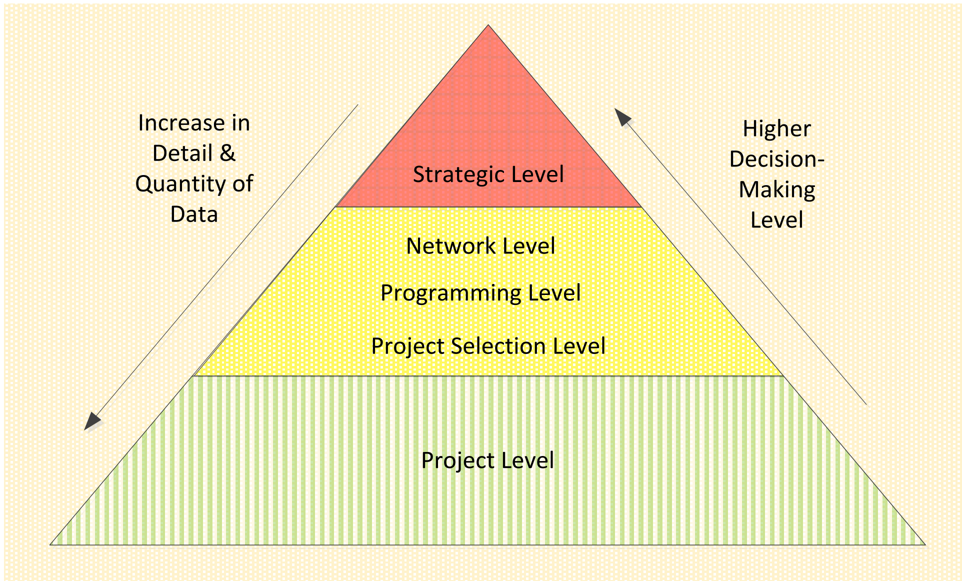


Figure 2-4 Decision-Making Hierarchy (Flintsch and Bryant 2006)

Although there exists an overlap between these decision-making levels, a typical decision hierarchy is shown in Figure 2-4. It is important to note that as decision-making level increases, the level of detail and granularity of data decreases. In this hierarchical decision-making framework, there are various stakeholders and project participants who are involved in the planning, design, construction and maintenance phase who play major roles in the management and decision-making of highway projects. Table 2-1 shows stakeholders involved in highway decision-making at different levels. These stakeholders range from the public who uses information to evaluate the overall program or performance to commissioners who use highway project data and information to set policies. In the strategic and network level, organizations such the Federal Highway Administration (FHWA) use data and information to develop guidelines and performance measurement tools to improve decision-making of highway agencies and state DOTs. Research partners such as Transportation Research Board (TRB) utilize data and information to enhance software program development and promote best practices. At a program level, program managers such as capital improvements, MPO and RPO are involved in administration of program funds, resource allocation and overall project development. Various project managers and division engineers are responsible for the selection of a design alternative, traffic control, contractor selection, etc. at project selection and project decision-making levels.

Table 2.1 Decision Makers

Decision Level	Stakeholders	Decision Makers	Decisions / Information Use
Strategic & Network Level	Commissioner	Transportation Board / Committee / Legislature, etc.	Set policy, identify project objective, allocate budget, etc.
	Regulators	US DOT/FHWA, AASHTO, etc.	Develop guidelines, performance measure, assessment tools, etc.
	Partners	TRB, FTA, NHTSA, RITA, etc.	Develop software program, decision support tools, best practice, etc.
Program Level	Program Managers	Capital Improvement, Local Governments, MPO RPO, STIP, etc.	Administer transportation funds, determine priorities, project development, resource allocation, etc.
Project Selection & Project Level	Project Managers	Pavement Manager, Right-of-Way, Environmental, Bridge Manager, etc.	Project selection, treatment selection, safety improvement, traffic control, contractor selection, etc.
	Division Engineers	Planner, Scheduler, Designer, Superintendents, Operation & Maintenance Engineers, etc.	Type of pavement selection, track project progress, identify contract time, etc.
	Public	Consultants, Suppliers, Contractors, Public, Academician, etc.	Performance evaluation

The basics of transportation decision-making commences with identifying opportunities to improve the transportation system for the user through transportation planning (FHWA, 2012). Transportation planning starts with setting goals and visions based on critical factors such as population growth economic changes, transportation needs, public input, etc. Then, the visions and goals are translated to 20 year or long-range transportation plan (LRTP) based on available alternatives to make effective decisions that meet the goals and funds. The state DOTs and MPOs develop a short-range (4-year) improvement plan under Transportation Improvement Program (TIP) and Statewide Transportation Improvement Program (STIP) respectively. During these two phases, states and MPOs make various decisions such as estimating reasonable project costs, evaluating and prioritizing strategies to match the budget.

Once the transportation plans are set, the next step is project development or project-planning process. This phase involves decisions such as identifying projects, establishing program objectives, evaluating potential projects and developing final programs'. During this stage, various criteria are used to support these decisions over the life cycle of highway projects. In a typical project, different DOT divisions' (bridge, pavement, local government, etc.) inputs and recommendations are considered for examining current condition in their respective disciplines. Public input, district office's input, capacity analysis, economic development, etc. are taken into account to analyze and prioritize projects. As project objectives, system analysis, estimated project costs, revenue projections, and etc. are also analyzed. Once program objectives are set, highway divisions evaluate potential projects based on technical factors or information such as safety (crash history), traffic (growth, volume) and pavement condition.

This step serves as a bridge between the transportation planning process and the actual project design, construction and operation. It also includes detailed investigation of individual projects with regard to the location (right-of-way acquisition), function, and assessment of environmental impact (NEPA process). Based on the investigation, final plans and project development certification are developed. Then, projects are let and construction of the project begins. Once implementing the operation or construction of the project is completed, the project performance is evaluated as final step. This step is used as a feedback to check transportation visions and goals. A sample comprehensive project development process of local public agency (LPA) for federal-aided projects adopted from the Iowa Department of Transportation is shown in Appendix A.

An example of a key decisions in the transportation planning process, the environmental planning or NEPA process (National Environmental Policy Act) is shown in Figure 2-5. This decision-making process is aimed at avoiding damage and minimizing social and environmental impacts that can be caused by a highway project. It includes meeting air quality standards, preserving historic archeological places, avoiding hazardous wastes, reducing noise pollution, reserving endangered species, keeping water quality, etc. This process has to be in compliance with federal regulations and permits such as Section 404 permits of the Clean Water Act, the Farmland Protection Policy Act, Clean Air Act, and National Pollutant Discharge Elimination System (NPDES), etc.

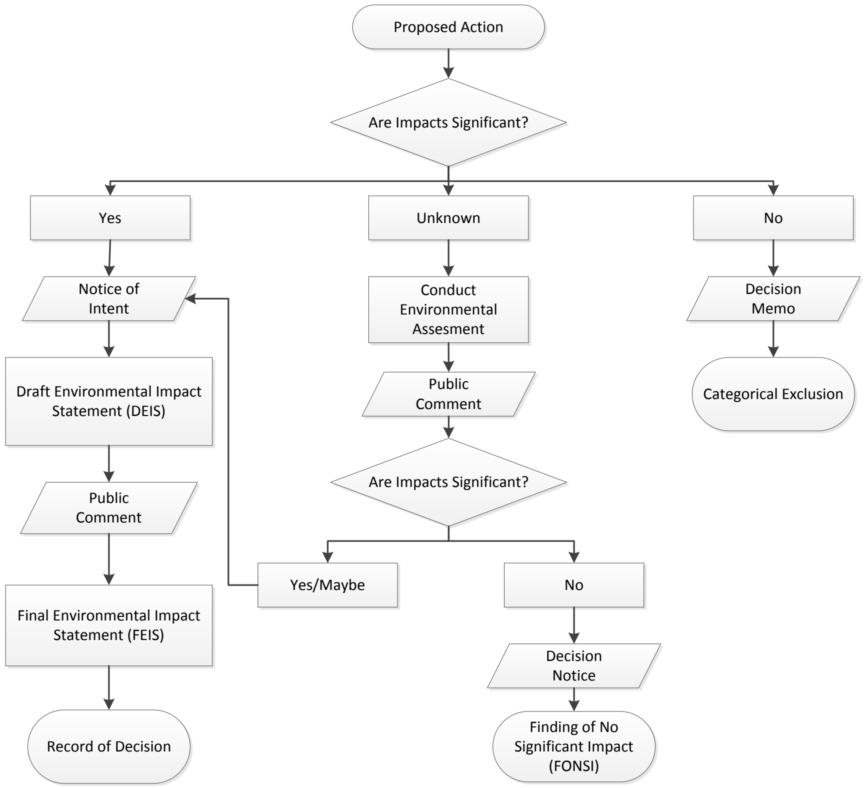


Figure 2-5 NEPA Document Decision-Making Process (US EPA, 2012)

Figure 2-6 illustrates a summary of some of the decisions over the life-cycle of highway projects along with information to support it. Some of the decisions made in the operation and maintenance phase include identification of maintenance needs, allocation of funds, selection of treatment strategies, prioritization and selection of projects. These decisions require information as input to support the decision-making process. These information/ analysis/ performance measures include sufficiency rating, needs study, cost/benefit analysis, and life-cycle cost analysis, etc. The pieces of information in turn depend on data such as pavement history, annual daily traffic (ADT), and pavement condition, etc. Similarly, some of the decisions made during the construction phase include contract time determination, selection of construction method, traffic control, etc. Production rate estimation (quantity of work installed per day), project progress tracking (actual cost per planned cost), safety analysis (number of accidents per project), cost (cost per mile, percentage of construction cost etc.) are some of the required information to support construction decisions, whereas daily work reports, reported quantities are the data inputs for this set of information. The selection of contractors (vendor analysis), contract analysis, procurement strategy and award of contracts are supported through bid monitoring and evaluation, and unit price information in the bidding phase. Capacity analysis, traffic analysis, and environmental assessment are part of the information utilized in the selection of design alternatives, cost estimation, and overall engineering decisions during the planning and design phase.

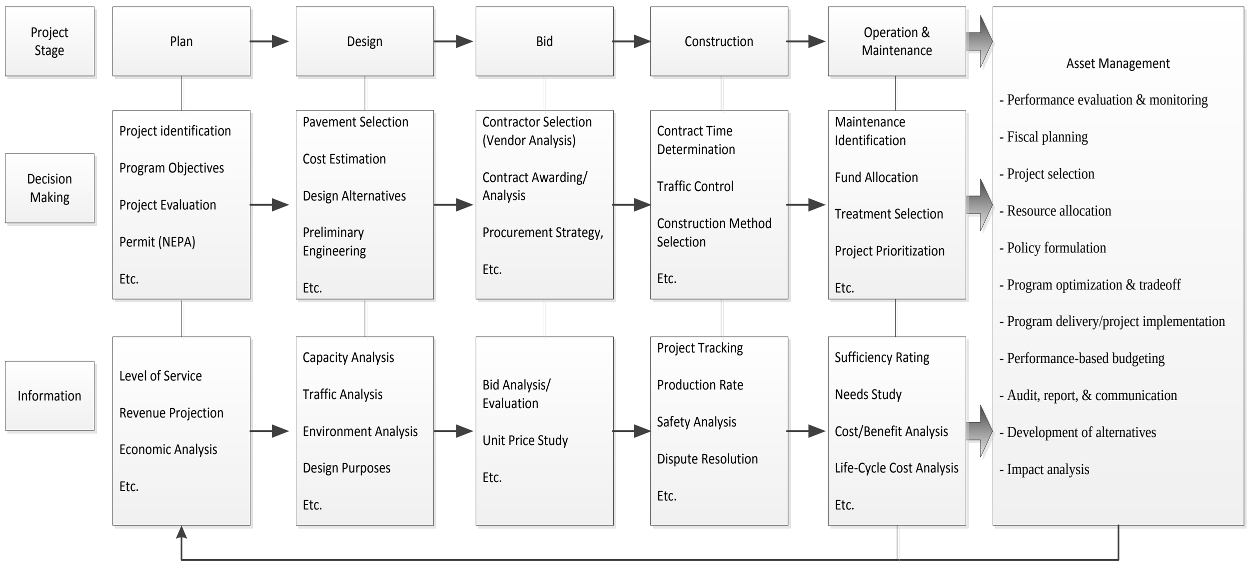


Figure 2-6 Highway Project Decisions

However, it **i**s important to note that although the decision-making process could be a good guideline of the overall process, it does not fully illustrate the types of data and information that is required and analysis methods to support each decision. For instance, the NEPA process does not show what performance measure or information is used to make “yes” or “no” decision for environmental impacts to be considered significant. There is also an overlap of information and decisions over project phases. For example, although sufficiency rating and needs study are used primarily in the planning phase, the operation and maintenance phase also considers the analysis for their maintenance projects. Decisions such as project selection and project prioritization lie in the planning as well as in the operation and maintenance phases. In addition, it should be noted that most of the higher level decisions (fiscal planning, policy formulation, etc.) are heading towards the asset management program which is responsible for the decisions in the planning and also operation and maintenance phase. Therefore, there is a need to differentiate the data and information needs for each type of phases and the level of decisions along with the necessary data analysis methods to integrate and make effective decisions.

2.4 Data Integration Effort

The overlap of data and information at various levels requires a smooth flow and integration effort that utilizes knowledge management tools and applications to support decisions over a project life cycle. Data integration is defined as "*the method by which a multiple data set from a variety of sources can be combined or linked to provide a more unified picture of what the data mean and how they can be applied to solve problems and informed decisions that relate to the stewardship of transportation infrastructure assets*" (Flintsch and Bryant 2006). Integration can be viewed from two perspectives, a) merging information from different sources and b) preparing the information in a usable and accurate form to the various end users. Integration allows organizations or agencies to have access to complete data and information on a timely manner with high accuracy, consistency and clarity, reduced duplication, greater accountability and easier communication. It helps in acquiring a comprehensive and coordinated system which enhances program development with lower data acquisition and storage cost. In addition, integration of data will act as a knowledge base. In this study, integration is viewed from data preparation in converting it to information to support highway decision-makers. However, prior works performed from both perspectives are summarized in this section.

2.4.1 Data and Information Use Integration Efforts

Many studies have been conducted to develop different models and various approaches have been implemented to enhance integration efforts to make use of data in the construction industry. One of the prominent approaches is the development of advanced database systems. Halfawy et al. (2005) developed a prototype model-based system for bridge design processes called bridge core model (BCM) using ISO STEP standards. The model involved the use of an integrated data model and a centralized project data repository based on four perspectives; structural design, structural analysis, construction scheduling and cost estimating by implementing an object-oriented database management system (DBMS). Although the study addressed the use of a standard data model in integrating data from the design perspective, it does not specifically show the utilization of data and information in supporting design decisions. Froese (1992) advanced the use of computer aided project management (CAPM) through development of integrated standard object-oriented data models for the architecture, engineering and construction (AEC) industry. The study demonstrates a system to integrate computer aided design and drafting

(CADD) program with various applications such as estimating, plan-generation and scheduling through a shared object-oriented database. The model consisted of three elements; development of data model, domain model, and project model. Yu et al. (1999) later developed a computer-integrated facilities management (CIFM) framework that is supported by the facilities management classes (FMC) and the industry foundation class (IFC). The study was able to show the use of project data in design, engineering and construction phases in the facilities management (FM) phase. This concept has a potential application in the transportation industry where various data are generated in the early phases of a project that can support decisions in later phases of operation and maintenance.

However, these integration approaches lack extensibility in terms of having limited semantic representation that might be difficult to make changes at later stages and these representations do not support multiple views from multiple domains due to predefined schema (O'Brien et al. 2000 and Rivard and Fenves 2000). Another noticeable approach or methodology that is utilized in the data and information integration is the use of semantic technologies such as ontology. Ontology is defined as "an explicit specification of conceptualization" (Gruber, 1995). It can be considered as a more advanced knowledge representation model that is used as mediating medium between information that is stored in different departments' and data sources (CAD systems, enterprise resource planning (ERP) applications, database, etc.) with the user. Shen (2004) developed a tree-structured model to handle the data-heterogeneity problems encountered in construction projects based on ontology to address the problem of limited semantic representation. The model addresses the integration problem of design knowledge, cost and schedule data. Wang et al. (2011) also showed an ontology-based approach to context representation and reasoning for managing context-sensitive construction information using case studies in the building industry. However, most of the ontology frameworks are in conceptual stage and actual application on real projects is still on progress.

Another potential approach in integration efforts of data usage is the work performed in classifying and organizing project documents to effectively utilize them as information. Caldas et al. (2002) showed the machine learning method such as support vector machines (SVM) algorithm can be an efficient system in automated classification of construction text documents based on related project components. Caldas et al. (2005) later developed a text information integration model (TIIM) to integrate project documents in the AEC/FM industry. The study focused on providing a semiautomatic support for integrating unstructured data types (text-based documents) such as contract documents, filed reports and change orders. Other studies which addressed integration efforts from active utilization of data and information include Soibelman and Kim's work (2002), who presented the data preparation process to generate construction knowledge through knowledge discovery in databases (KDD). The study showed the potential use of KDD to identify causes of construction activity delays using U.S. corps of Engineers construction management database (Resident Management System). Soibelman et al. (2004) later developed "data fusion" methodology to bridge historical databases and data analysis techniques as part of construction management knowledge discovery. Bao and Zhang (2010) presented the principles and methods of analyzing data using a decision support system architecture based on data warehousing through structured query language (SQL) implementing, optimizing performance, and data mapping. Boulicaut and Jeudy (2010) also presented constraint-based data mining approach using inductive queries to generate, manipulate and apply patterns.

Overall, these studies address integration from four perspective; a) data collection and storage and database development (project database approach), b) data interoperability of software systems and compatibility issues aspect (standardization model approach), c) document management and classification approach, and d) data analysis perspective for effective communication and facilitating the business process. However, these studies typically focus on specific elements or business processes or project phases such as design or construction to efficiently share and exchange project data within a specific discipline/division. In addition, most of the integration efforts focus on the vertical or the building industry as compared to horizontal or the transportation industry. Moreover, they under-emphasize the importance of data integration efforts from the transportation agency's or owner's decision-making point of view and pay little attention to effectively convert data into information and knowledge over a project life-cycle. These studies do not address the relationship between data, information, knowledge and decision-making of highway projects. Although it is difficult to address every decision made over the life–cycle of a highway project, this study focuses on the integration of data, information and decision through a three-tiered mapping and data analysis efforts of three potential decisions made across project life-cycle.

2.4.2 FHWA Integration Effort

The Federal highway Administration (FHWA) provides a set of guidelines of key activities in conducting data integration as part of an asset management program. In the guideline, it has put five major tasks that need to be considered along with elements to consider in the data integration process (Figure 2-7). The first step in data integration process is analyzing user requirements, understanding the business process, and identifying the characteristics of existing systems (perform a requirement analysis). The second step of this process involves the recognition of the relationship and mapping of data and process flows by identifying the inputs and outputs required in modeling the process. This is followed by identifying the types of data storage or management systems by defining alternatives, evaluation and selection of database architecture, identifying the risks and the level of efforts and time required for integration. The final step in this process is the development and implementation of the chosen strategy in terms of computer programming, software/hardware setup and database testing. FHWA’s integration procedure is a fundamental step in collecting and storing the right types of data in terms of addressing users’ requirement. However, the study fell short of addressing data analysis efforts that need to be utilized in converting collected data into information and knowledge in supporting key decisions.

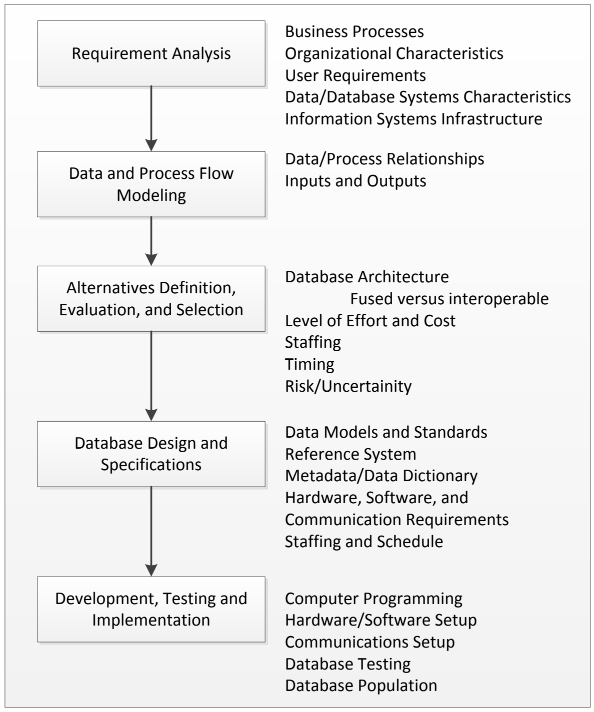


Figure 2-7 Data Integration Process (FHWA 2010)

The FHWA and the American Association of State Highway and Transportation Officials (AASHTO) develop software programs/modules under the transportation software management solution (Trns.port) program to convert data into information in supporting business decisions across the highway project life-cycle. Some of the programs designed to support project cost estimation during the preconstruction phase include CES (Cost Estimation System), Trns.port Tracer, Trns.port Preconstruction System and Trns.port Estimator. These programs are used for preparing cost estimates ranging from conceptual to final engineers estimate. For instance, CES is a full–range estimation system used in preparation of parametric, cost-based and historical bid-based cost estimates, while Trns.port Estimator is a standalone system that is used in the preparation of detailed cost estimates, supporting cost-based and historical bid-based estimation. On the other hand, Trns.port Tracer serves as a parametric cost estimating tool for planning and budgeting transportation projects at pre-design and preliminary design phases (AASHTO, 2009). Trns.port Preconstruction is a project and proposal system used in developing PS&E estimate and managing projects in terms of creating proposals, schedule letting and contract awarding. Trns.port Preconstruction application uses a standalone spreadsheet called PES Worksheet (SAPW) for project data entry and editing information.

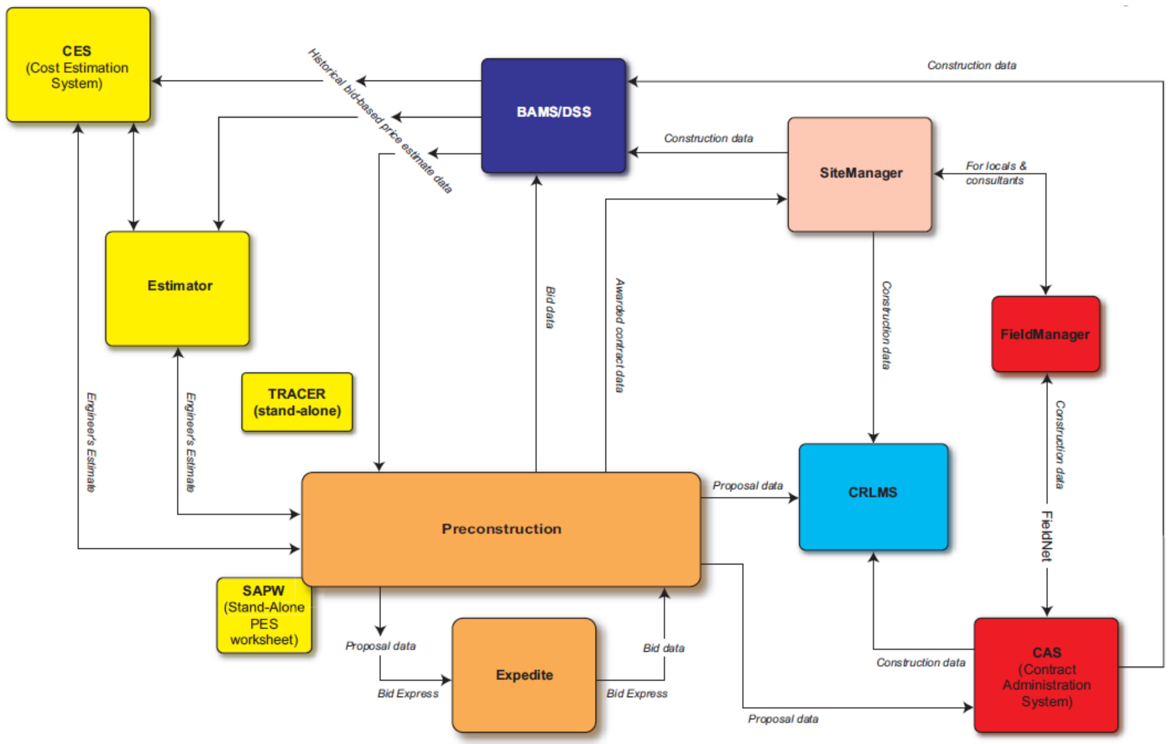


Figure 2-8 AASHTO Trns.port Programs (AASHTO, 2009)

Another program that is developed by AASHTO is BAMS/DSS (Bid Analysis Management System/Decision Support System). BAM/DSS is a relational open architecture historical database and bid analysis software that is used for bid monitoring and evaluation, vendor analysis, contract analysis, item price estimation, collusion detection and as-bid to as-built analysis (AASHTO, 2009). BAM/DSS supports Trns.port Preconstruction, CES and Estimator, while CES and Estimator can exchange information with Trns.port Preconstruction. Other programs that are currently in service or under development include PES (Proposal and Estimates System), that is used in preletting phase of a bid; LAS (Letting and Awards System), for helping highway agencies to advertise and process proposals, track proposal holders, review bid information, and award contracts; Expedite, an electronic bidding system for providing secure online communications for bid items; CRLMS (Civil Rights and labor Management System); CAS (Construction Administration System), for managing contract information from award to final payment; SiteManager and FieldManager, for construction management system and as-built or field management system during construction phase respectively (AASHTO, 2009). Figure 2-8 shows an overview of AASHTO Trns.port Programs.

Figure 2-9 Annual Costs of AASHTO Modules (FHWA, 2012)

The aforementioned modules and/or software programs can be utilized as business intelligence tools and enterprise-wide database systems to support highway agency decisions at various levels. These programs are specifically designed for highway agencies and their design consultants. Although some of the programs are PC-based or standalone programs, they are capable of exchanging information and are interlinked with agency's websites and other cost databases. However, there is an overlap between these programs that might have the opportunity of duplicating data. In addition, the total cost of maintaining these modules exceeds $500,000 per year in which a highway agency may implement one or two of these modules, but not all (Figure 2-9). In addition, the amount of resources (cost and time) to acquire the module and train personnel's might even surpass an agency's annual funding. Furthermore, the compatibility of some of the modules with other programs are limited.

2.4.3 Data Interoperability Efforts

The second perspective in data integration efforts is interoperability. Currently, there is a huge demand and need for the development of an optimized software program and system to assist various tasks in the life cycle of construction projects. But most programs and systems either act as standalone software programs or are created to address specific and individual needs of smaller, discrete divisions rather than being overall organization or enterprise-based programs. Various construction industry participants have addressed the issues related with the need for data interoperability, data and information integration in the construction industry. However, a limited number of organizations are working towards the development of standardized data and information integration systems to improve the exchange of information between software applications (data interoperability) in the construction industry.

One of these organizations is, International Alliance for Interoperability (IAI) or currently known as Building Smart Alliance. Build Smart is an international organization aimed at developing "a universal basis for process improvement and information sharing in the construction and facilities management industries" (IAI, 2007). Build Smart has developed an Industry Foundation Class (IFC) model that provides an information exchange framework that consists of building components, work processes and interactions for the AEC industry. Another organization working towards data interoperability is Machinery Information Management Open Systems Alliance (MIMOSA). MIMOSA is involved in the creation of open information standards for life-cycle asset management in the military industry sector which later focused its work on facility management (Johnson, 2004). MIMOSA has developed data and information exchange standards to capture various process industry data which consists of manufacturers, asset inventories, system components, condition status, and associated work orders (Johnson, 2004). The National Institute of Standards and Technologies (NIST) and Fully Integrated and Automated Technologies for Construction (FIATECH) also developed a guide for information sharing throughout the lifecycle of a project which targeted the process industry. The guide focuses on the use of interoperable data standards, specification of exchange requirements, information quality management, and information retention policies (Fallon and Palmer 2006).

2.4.3.1 Building Information Modeling (BIM)

One of the interesting developments in data integration and interoperability of systems experienced within the last two or three decades is the advancement of Building Information Modeling (BIM). The National Building Information Model Standard (NBIMS) Project Committee defines BIM as "A *digital representation of physical and functional characteristics of a facility. A BIM is a shared knowledge resource for information about a facility forming a reliable basis for decisions during its life-cycle; defined as existing from earliest conception to demolition".* AIA (2007) defines BIM as *"*a *model-based technology linked with a database of project information*".

BIM often referred as Virtual Building Environment (VBE) or Virtual Design and Construction (VDC) can be considered as the next generation tool as it is changing the way projects are designed, constructed, operated and managed in the AEC industry. Beginning from the traditional project delivery methods such as Design-Bid-Build, the AEC industry has progressed through the years from 2D drawings using paper format (blue print) to 3D and 4D visualization using electronic files and advanced project management, and currently to BIM and/or life cycle management (Eastman et al. 2011). Some of the main advantages of BIM over traditional systems or software programs include faster and more effective processes, better design, controlled whole-life costs and environmental data, automated assembly, better customer service, and lifecycle data (Azhar et al. 2008, Smith and Tardif 2009, Eastman et al. 2011). A BIM can be used for activities such as fabrication of shop drawings, code reviews, forensic analysis, facilities management, cost estimating, construction sequencing, conflict, interferences and collision detection (Azhar et al. 2008).

2.4.3.2 Construction Operation building Information Exchange (COBIE)

Traditionally, once a project is completed and turned over to the client, all project documents and necessary paper works are submitted to the facility management team in either paper forms and/or digital CDs for maintenance and operating purposes. Usually, these documents are not used in the operation and maintenance (O&M) phase and require a lot of time and effort to input these data as part of facility management systems. However, Construction Operations Building Information Exchange (COBIE) has allowed facility management to take advantage of the advancement in information technologies and be equipped with data and information integration system. Some of these data and information (referred as ‘as-built’ data) incorporated in COBIE include room lists and area measurements, material and product schedules, construction submittal requirements, construction submittals, equipment lists, warranty guarantors, and replacement parts providers (NIBS, 2012).

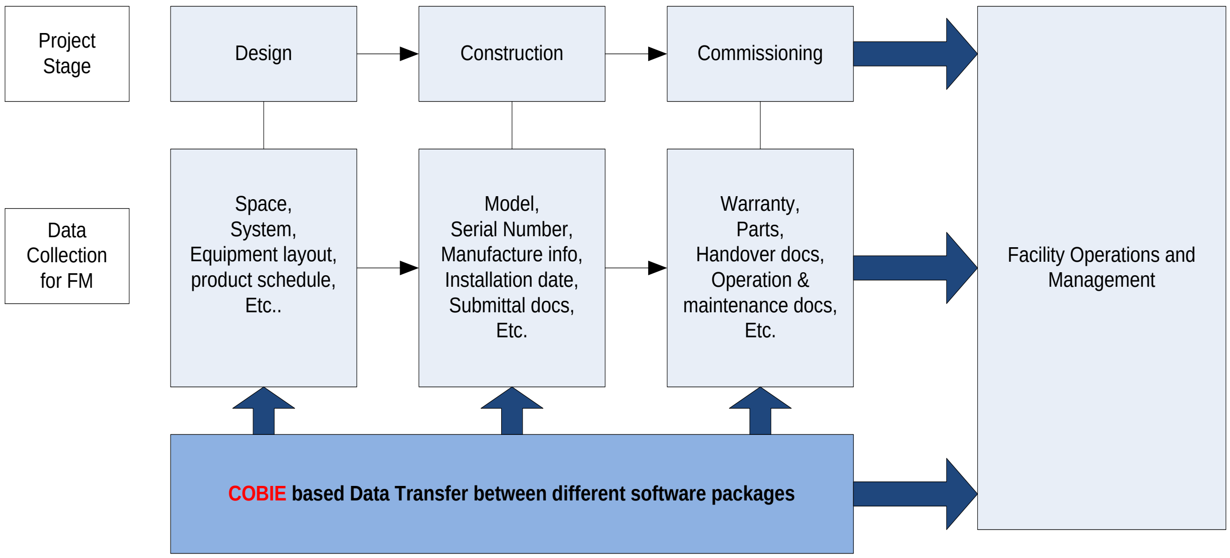


Figure 2-10 Role of COBIE as a Structure for Information Transfer

COBIE is an internationally accepted standard information specification developed under the National Institute of Building Sciences (NIBS) for the purpose of life cycle capture and delivery of information to support the operation and maintenance (O&M) or facility management phase of building projects (NIBS, 2012). COBIE is also known as Industry Foundation Class (IFC) Facility Management Handover Model View Definition (MVD). In COBIE system, data and information are acquired at five different stages during a project execution; early design stage, construction documents design stage, construction quality control stage, product installation stage, and system commissioning stage (Figure 2-10). It provides a responsibility matrix for project participants to agree on their role in supplying the required 'as-built' data and information throughout the project stages. During these stages, data and information are required to be provided by various project participants including the designer with space layout, system list, types of equipment, and the location of named equipment; the constructor with equipment make, model and serial number, manufacturer literature, warranty and replacement parts information; and the commissioning agent with the job plan data associated tools, training and equipment requirements (COBIE, 2007).

2.4.3.3 Construction Industry Standards

Interoperability is defined by BIM (2011) as "the ability to pass data between applications, and for multiple applications to jointly contribute to the work at hand". There exists a number of internationally accepted data standards and models used in organizing data and schema language. The two most commonly used data models used in the construction industry are the Industry Foundation Class (IFC) and CIMSteel Integration Standard Version 2 (CIS/2). The ISO-STEP international standards developed these two models to support product and object model exchanges within different industries. Before the introductions of these data models, interoperability relied on file-based exchange formats which are limited to geometry such as Drawing Exchange Format (DXF) and Initial Graphic Exchange Specification (IGES) and Application Programming Interfaces (APIs) (BIM, 2011). Table 2-2 shows types of exchange formats in the construction industry.

Table 2.2 Type of Exchange Formats in the Construction Industry (BIM, 2011)

Type of Format	Exchange Format
Image Formats	JPG, GIF, TIF, BMP, PNG, RAW, RLE
2D Vector Formats	DXF, DWG, AI, CGM, EMF, IGS, WMF, DGN, PDF, ODF, SVG, SWF
3D Surface and Shape Formats	3DS, WRL, STL, IGS, SAT, DXF, DWG, OBJ, DGN, U3D PDF (3D), PTS, DWF
3D Object Exchange Format	STP, EXP, CIS/2, IFC
XML Schemas	AecXML, Obix, AEX, BCxml, AGCxml
Game File Formats	V3D, X, U, GOF, FACT, COLLADA
GIS formats	SHP, SHX, DBF, TIGER, JSON, GML

For instance, COBIE is developed to be compatible with various software packages and file format which makes it easier to be viewed by the various project participants in different phases. For instance, designers can use it as Industry Foundation Class (IFC) file format, while constructors can use a spreadsheet to upload equipment information. Some of the formats supported by COBIE include STEP Physical File Format, ifcXML, SpreadsheetML, and Open Document XML. This data and information integration specification will not only reduce paper works, but also lower the time, effort, and cost associated with managing a building facility and enhances effective communication and management. COBIE can even be incorporated in Building Information Model (BIM) so that information can be compiled and exchanged in an effective and efficient manner through a facility life cycle.

2.5 Department of Transportation Databases

Efficient management and business decisions are supported through well documented data, clear data flow, better data management system and/or technology, tools and/or approaches to store, manage and enhance potential users' daily decision making process. As much as data can be useful, proper caution should be taken in terms of collecting, managing, understanding, utilization and selecting the right type of tool and application as it might lead to wrong decisions and judgments. Failure to address poor quality data and information can lead to wasting resources; failure to provide quality services and unsuccessful overall policy and management (Audit Commission, 2007). Currently, various departments of transportation (DOTs) including the Oklahoma DOT has implemented or utilized various databases and data management systems. These databases and management systems range from in-house spreadsheets to commercially available programs using manual to automated data collection systems across project phases. Some of these databases include GripLite or highway inventory, Contract Fee Proposal Spreadsheet, Proposal and Estimates System (PES), Letting and Awards System (LAS), SiteManager, FieldManager, FieldBook, and Project Management System (PMS). A brief summary of current DOT databases is shown in Table 2-3.

i. Highway Inventory

Typically, an enterprise-wide GIS database system is used to collect and manage highway inventory data utilized by the system planning and research division. This GIS system (e.g. Geographical Resource Internet Portal Lite or GRIPLITE Mapping System) is an intranet-only portal that consists of three modules, roadway, bridge, and traffic inventory. The roadway inventory database includes number of lanes, width, functional class, right of way, route classification, railroad crossing, control type, and terrain area type for all the road sections. The bridge inventory stores bridge design data along with as-built or construction data that includes bridge span length, width, length, inspection report, bridge characteristics and features. Traffic related data such as average annual daily traffic (AADT), signals, lights, traffic control, crash statistic, etc. are stored in the traffic inventory For example, ODOT's highway inventory system is accessed through its website at http://192.149.244.31/griplite/index.htm. Typically, most of these data attributes are collected manually by the design team for design data input and construction inspector or project engineer administering the project for as-built data. A semi-automated collection system is used for data attributes such as annual average daily traffic.

Table 2.3 Current DOT Databases

Division/ Source	Database	Type of Data	Sub-Elements	Collection Method
System Planning/ Research	Grip lite/ Highway Inventory	Roadway Inventory	Functional Class, Right of Way, Route Classification, Terrain Area Type, right-of-way, railroad crossing, etc.	Manual / Semi-Automated
		Traffic	Average Annual Daily Traffic (AADT), signals, lightings, traffic control, crash statistic, etc.	
		Bridge Inventory	Bridge span, width, length, load limit, inspection reports, etc.	
Preconstruction	In-house Spreadsheets	Preliminary Engineering Cost Data	Engineering hours, number of sheets, etc.	Manual
Bidding	PES/ LAS	Contract Documents	Bid information, award contracts proposal holders, advertisement, pre-bid, etc.	Semi-Automated
Construction Division	SiteManager	Construction Data	Daily work report, reported quantity, material, change order contractor payment etc.	Manual
Pavement Management	Pavement management System (PMS)	Pavement History	Pavement surface type, thickness, composition, etc.	In-house - Automated
		Distress Data	Longitudinal Cracking, Transverse Cracking, Patching, Spalling, Fatigue, etc.	Consultant - Roadware
		Friction Data	Average Roughness, Ride, Average Rut etc.	In-house
		Other (structural)	Deflectometer (FWD), ESAL	In-house Roadrater

ii. Preconstruction Database

Data collected during the during the design phase are either stored in individual computers or a database depending on the agency. For instance, ODOT engineers in roadway, bridge, right of way, and surveying divisions have developed in-house contract fee proposal spreadsheets for the purpose of negotiating contracts with consulting firms for projects outsourced. This spreadsheet is developed based on the estimated work efforts (engineering man hours and hours per mile) along with its associated costs to prepare the set of plans from the preliminary stage to the final plan preparation. The spreadsheet consists of a cross tab of seven main plan development activites, a detailed list of tasks and sub-tasks along with a skilled labor category. ODOT engineers use this spreadsheet to estimate and match the work efforts required by each engineer for each task based on the amount of sheets required for each task and project length by comparing it with similar, previous highway projects. Data are partially stored electronically on each division engineer's personal computer, but the majority of the project data are stored in hard copies (paper format) as part of the engineering contract data.

iii. Bid Database

Bid information, award contracts, proposal holders, advertisement, pre-bid, etc. are stored in AASHTO's Trns.port programs of PES (Proposal and Estimates System), LAS (Letting and Awards System). In addition, highway project information such as bridge data, maps, programs, reports, and bid documents are available on respective DOT websites (e.g. http://www.okladot.state.ok.us/). It is important to note the maintenance of this system is done through manipulation of data items using an interface with Oracle database. In addition, some of data collected and stored through these databases are managed by the asset management program.

iv. Construction Database

DOTs store construction project data through one or a combination of AASHTO's Trns.port construction contract administration software programs such as SiteManager, FieldManager and/or FieledBook. For instance, the SiteManager is a multi-tier architecture construction management tool developed for the purpose of data entry, tracking, reporting, and analysis of contract data during the construction phase of a highway project. SiteManager consists of six basic functions to view and store highway construction project data; contract administration, daily work reports, contractor payments, change orders, material management, and civil rights management systems. These functions allow data and information acquisitions such as materials and equipment, project location, job-site conditions, construction pay items, reported quantity, and weather conditions. Currently, 16 states have the license to operate SiteManager to avoid repetitive data entry and manage contract data during the construction phase. The ODOT utilize SiteManager as a primary construction database.

Currently, the ODOT SiteManager contains a database of more than 1,500 previously completed and ongoing construction projects since 2002 which includes DWRs of highway construction projects along with information such as project location, job-site conditions, construction pay items, duration, delays, reported quantity, resources used (crew, equipment and machineries) and weather conditions. In addition, ODOT uses the Trns*port SiteXchange program for transfer of subcontract information with prime contractors. Typically, ODOT sends contract file along with vendor file to the contractor. The contractor fills the contract and vendor information that contains subcontract vendors, items, item quantities and item prices into the SiteXchange Subcon program and submits it to ODOT construction division for approval.

v. Pavement Database

Pavement condition assessment data are stored in the pavement management systems (PMS). Typically, the PMS consists of three divisions, pavement history, distress data and friction data. Pavement history database is used to store previous treatment applications in terms of pavement surface type, thickness, composition, and treatment cost. Pavement distress data or structural aspect consists of longitudinal cracking, transverse cracking, patching, spalling, fatigue, etc. Average roughness, friction, ride, average rut, etc. are included in the functional aspect or pavement friction category. In addition, various non-destructive evaluations test data included in the PMS are falling weight deflectometer (FWD) for checking structural adequacy, profilometer to report the smoothness or IRI and rut depth, and friction for evaluating the skid resistance and ESAL (equivalent single axle load) for pavement design. The functional data contains one record for each 100^{th} mile of roadway surface condition. Structural data mainly contains one record per 100^{th} mile of structural layer information while the analysis part contains condition indexes on structurally homogeneous segments or control sections of roadway. For instance, the ODOT collects 8000 miles of data every 100^{th} mile or 800,000 records annually that consists of 65 data fields that is 52 million pieces of data (Calvarese, 2007). The IDOT GIS database consists of over 300,000 records that portray highway structure and history. These collected pavement data along with highway inventory data can be primarily utilized for decision making of needs study, sufficiency rating report, maintenance strategy, and selection of the type of treatments required for rehabilitation projects.

It is important to note that some DOTs primarily utilize dTIMS™ pavement management software program to develop inventory of physical assets and perform a life cycle cost analysis to support their pavement management plan. dTIMS™ allows generation of projects by year, recommended treatments by project and year, and overall summaries of condition, backlog, treatment cost, and treatment length (IDOT, 2012). The program is set up to provide assistance to districts in the selection and prioritization of rehabilitation and reconstruction projects. It also has capabilities in doing a network level analysis of the condition of the system and the funding levels needed to maintain status quo or improve the system. However, it has been used in a limited extent for both of these purposes.

2.6 Data Analysis Methods

Once data is collected and it should be properly stored, analyzed and managed to be utilized for its intended purpose. In this case, the purpose is to extract data and convert it into information and knowledge to support decisions. This purpose is enhanced through the use of relevant and appropriate data analysis methods. Data analysis is a process of applying statistical, logical and/or analytical techniques to describe, illustrate, evaluate, measure and infer data. Data analysis can be applied in the form of descriptive explanations, performance metrics, predictive modeling, and optimization techniques for use in reporting, developing a common platform, making strategic and optimal business decisions. A typical data analysis includes inspection, cleaning, transforming and modeling of data with the aim of extracting useful information, suggesting conclusions and supporting decision-making (Ader, 2008). In this study, the process of data analysis is described through data mining approach.

Today, the advancement in data analysis tools and techniques through the support of information technology has led various industries to take advantage in promoting their business and enhance their organization's performance. Ittner and Larker (2006) reported that there is growing evidence in the greater use of effective analysis tools to deliver better financial performance. Kennerley and Mason (2008) argue that the problem in converting data to information and knowledge to support decisions is not the lack of tools and applications. Various tools, techniques and applications have been developed or practiced by multi-disciplines ranging from social science to information technology to analyze, interpret and visualize data. Some of these tools include knowledge management (KM) tools, and knowledge discovery in database (KDD) approaches such as data mining (DM) techniques, decision support systems, artificial intelligence, data warehousing, machine learning and business intelligence tools, and knowledge bases (KB). These tools and techniques have been utilized in conjunction with management philosophies such as concurrent engineering, lean construction, business process re-engineering, total quality management, supply-chain management and just-in-time production to improve decision making (Bjork, 1999).

Knowledge discovery in databases (KDD) and/or data mining (DM) is one of the powerful techniques used to extract relationships, patterns and knowledge from large databases. KDD and DM includes a combination of approaches and concepts such as machine learning, artificial intelligence, pattern recognition, visualization, statistics and database systems. KDD can be defined non-trivial extraction of implicit, previously unknown, interesting, and potentially useful information from data (Fayyad et al. 1996 and Chen 2001). A typical KDD involves six basic processes: problem statement and goal definition, identifying the sources (collection and understanding), preparing the data, build and train a model, validate the model, and implementation. In a typical KDD process, it should be noted that data preparation and cleanup is the most important, difficult and time consuming process which takes up a large portion of the overall process. Cabena et al. (1998) showed that 60% of the time in KDD process goes to the effort put in preparing the data to obtain clean and relevant data for conducting analysis. Data can be collected in impromptu manner such that important data might be missing or wrongly recorded. In this step, various approaches such as transformation, imputation, cutoff values, and removal of outliers is conducted. Therefore, the quality and details of available data should be properly assessed as it might lead to wrong decisions. Classification and examples of data analysis methods are further discussed in the next chapter.

3. Data and Information Integration Framework

This chapter discusses the methodology and concepts used in this study. It presents a three-tiered approach in integrating data, information and decision-making. It classifies various data attributes that are utilized in highway decision-making process into four major classifications. A conceptual 3-D data flow diagram (DFD) and prototype data-information-decision making integration framework to support highway decisions are also presented.

3.1 Evolution of Data and Information Integration

The advancement in digital technology, data collection methods and centralized information and/ or database management systems has contributed to a rapid and economic data and information generation which facilitates better communication, management, decision-making and dissemination of information and knowledge. This study classifies the evolution and advancement process of data, information, and knowledge integration using four broad categories to better understand the ultimate goal of this study. These include data collection method, process, approaches or methodologies utilized and type of data storage system. Figure 3-1 illustrates three types of generations in knowledge management and information flow using four approaches in the transportation industry.

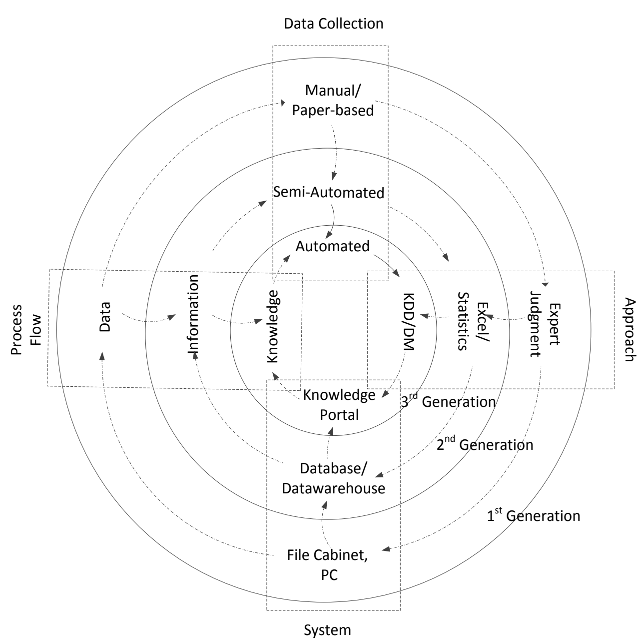


Figure 3-1 Golden Circle of Knowledge Management for Highway Industry

3.1.1 1st Generation

The traditional process of collecting raw data has been in practice for the last couple of decades for the purpose of keeping records, communication, sharing information, reporting and resolution of disputes. Typically, this raw data is collected manually in paper-based format, where a project participant might fills a report or an incident in a construction site (e.g. daily work report) or sign a contract document. These data are usually kept in a storage room or a file cabinet unless a dispute arises or need to make contractor payments, etc. In some cases, these data are stored in personal computers (PC) and database systems that are utilized by a decision maker (either a project manager, engineer or designer) at a certain level of a project phase. At this stage, these decision-makers utilize their expertise and/or engineers' judgment to make decisions. This stage of collecting raw data manually using paper based documentation through utilization of expert judgment in decision-making is considered as the 1st Generation in knowledge management for this study.

3.1.2 2nd Generation

The transition and interpretation of raw data into valuable information has led to the emergence of 2nd Generation in the transportation as well as other industry sectors. This generation features the implementations of computer applications such as excel based spreadsheets, MS-Access and advanced database systems through paper-based/semi-automated data collection mechanisms which allow the storage and generation of information for planning, estimation and management of highway projects. In this generation, although the use of manual data collection still persists, the use of automated systems using mobile technologies such as smartphones, camera, tablet-based PCs, geographical position system (GPS), and notebooks have greatly influenced the data collection method. The advancement in database administration programs and data warehouses has also improved the way data are stored and managed, which provides easier access to the project participants. In addition, the use of analytical techniques such as statistical methods, artificial intelligence and decision-support tools such as excel spreadsheet have greatly influenced the knowledge management progression. In this generation, the wide acceptance of these advanced systems and programs has affected the way project participants make decisions. The collection of data in semi-automated manner, storing and managing data in more advanced database systems, and extracting information through the utilization of statistical methods and advanced information generation tools is considered as the 2nd generation of knowledge management.

3.1.3 3rd Generation

The emergence and applications of advanced knowledge management (KM) tools, big data analytics algorithm and knowledge discovery in database (KDD) approaches such as data mining (DM) techniques, decision support systems, machine learning and business intelligence tools, knowledge bases (KB), expert systems and ontology-based frameworks has led various industry sectors into a new era or the 3rd Generation. These techniques are being utilized in conjunction with management philosophies such as concurrent engineering, lean construction, business process re-engineering, total quality management, supply-chain management and just-in-time production (Bjork, 1999). In this era, the data and information is converted to knowledge and utilized to support various decisions. Data are collected through automated systems. The advancement of cloud computing has revolutionized how data can be stored, accessed and communicated. This 3rd Generation is greatly improving decision-making and information system management in terms of capturing data, storing, organizing, sharing, integrating, communicating, extracting pattern, information and disseminating knowledge in a more efficient manner.

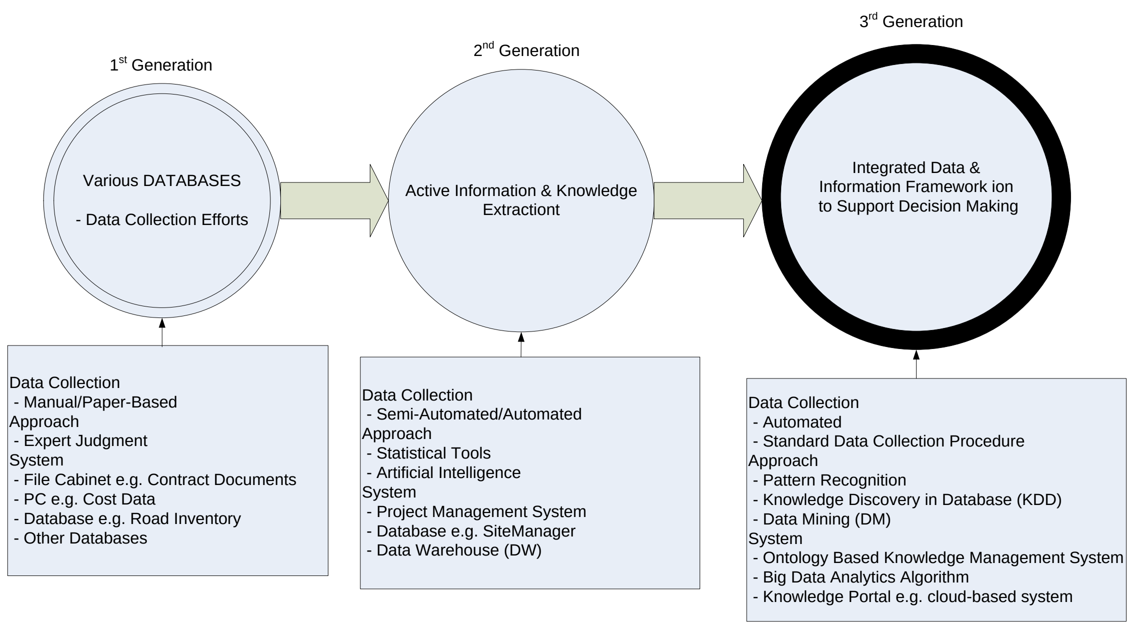


Figure 3-2 Evolutions of Data and Information Integration for Highway Agencies

However, most transportation agencies and DOTs are behind other industry sectors in actively taking advantage of data utilization for their transportation objectives. The evolution of data and information integration of state DOTs is illustrated in Figure 3-2. Many DOTs are in the transitional stage from 1st generation to 2nd generation since they are currently collecting and storing various types of data and putting a limited amount effort in extracting information and knowledge from the data. The Manual and/or paper-based type of data collection and use of file cabinets and PCs for data storage is fairly shifting towards semi-automated data collection system and advanced database system respectively. The primary source of decision-making which is expert judgments is also fairly supported through the use of statistical tools and excel spreadsheets. However, in fully active 2nd generation, DOTs would be able to benefit from scientific methods such as data mining (DM), artificial intelligence and project management systems to the collected data in order to extract various information and knowledge to support the decision-making process. In addition, data should be acquired in automated manner and knowledge portals and management systems such as ontology-based frameworks and should be developed and supported with big data analytics algorithms, pattern recognition and knowledge discovery in database (KDD) tools to extract information and support highway decisions (3rd generation).

3.2 Three-Tiered Approach

Various decision are made daily by various project participants from the planning phase to operation and maintenance phase and from strategic and/or corporate level to project or division level. The success of these decision-making processes is highly dependent on the type of data collected and effective conversion into information and knowledge. This study uses Juran's "triple role" concept to define the highway project decision-making process from data utilization perspective. In a typical process, every party in a process plays three principal roles: supplier, processor and customer (Juran, 1988). These three roles apply at every level of a construction process through the various parties involved (Jang et al. 2003, Oberlender 2000). At a project level, an owner's need is used as an input or supplier to an engineer who designs or processes the project and provides the plans and specification as an output for bid preparation department in a design process. Similarly, the bid analyst team is the customer of the designer (supplier) that uses the plans and specifications as input in preparing the bid documents and setting selection criteria (processor) for contractors' use in construction in a bid letting process. This procedure continues in a similar trend in the construction, maintenance, planning and project development processes.

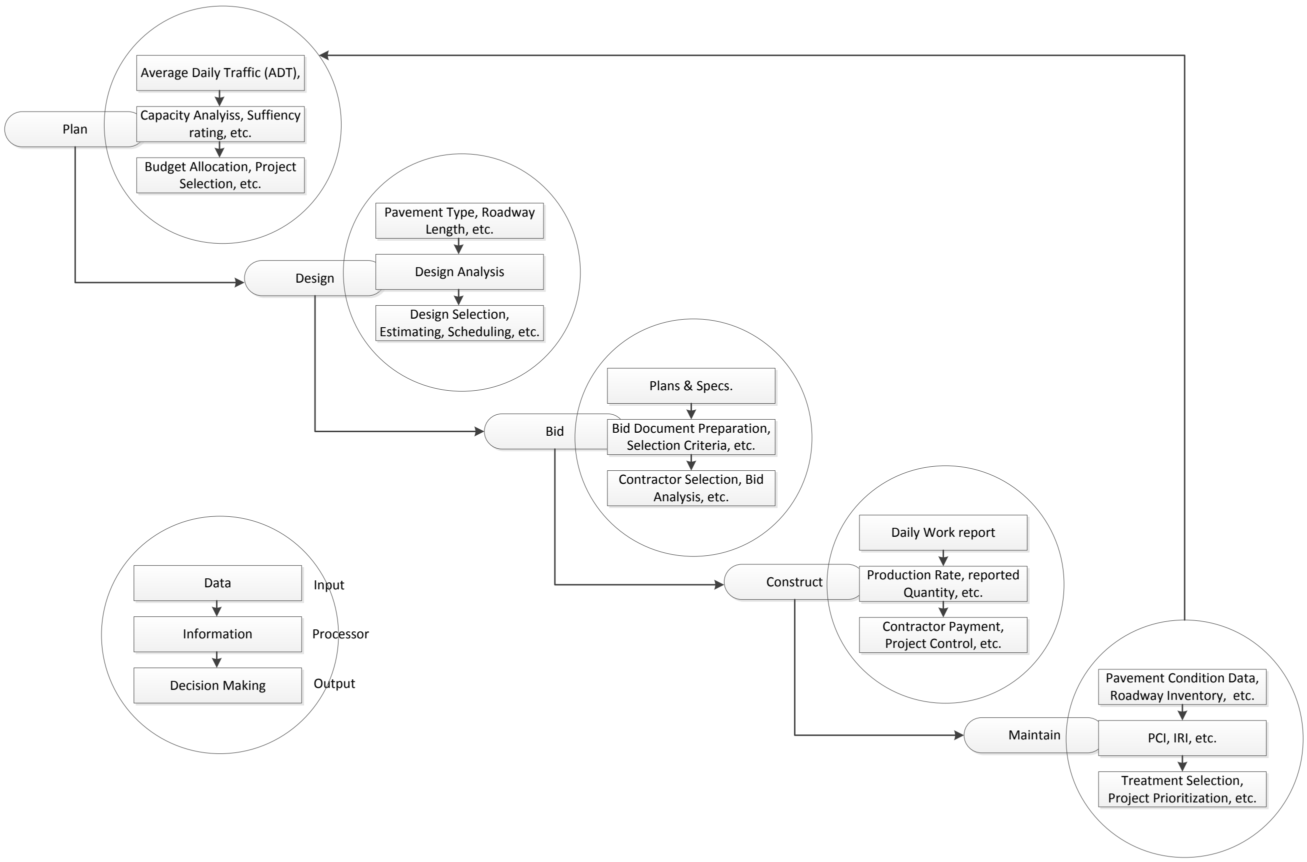


Figure 3-3 Triple Role Concept in Highway Decision Making Process

This triple role concept can further be broken-down within project phases to support specific decisions. Typically, a decision making process involves three steps; a) collecting raw data, b) converting raw data into information and knowledge and c) utilizing the information to support decisions. In this study, raw data is considered as a supplier to meet the customer or decision-maker's need, while information is used as a processor to convert raw data into usable form, whereas decisions are considered as final output. For instance, the use of roadway inventory data such as roadway length, number of lanes, and average daily traffic (ADT) can be used to perform traffic analysis or capacity analysis that can serve as information in roadway planning, design and/or alternative selection process. Contractor's payment or project control (customer) can be supported through the estimation of production rate and earned value management (processor) based on reported quantity, unit price and (supplier). Figure 3-3 shows an example of a three-tiered approach of the decision-making process in a highway project life-cycle using Juran's triple role concept.

However, it is important to note that the highway decision-making process does not necessarily follow a linear path. There exists an overlap and/or multiple usages of data and information in various decision-making processes. The raw data collected or the information analyzed in one phase can be used as data and/or information in other phases along the project path for managing active projects or retrieved after project completion for forecasting future projects. This implies that data and information can serve as a supplier, processor or customer at acertain stage of a highway project life cycle. This approach is further discussed using a data flow diagram in the next section.

3.3 Data Flow Diagram

Understanding the actual flow of data and information can help identify existing problems or obstacles to improve data and information quality, standard and use in support of various decisions during the life cycle of highway projects. Typically, a data or information flow model is a two dimensional approach used to visualize flow of data within a system and understand the relationship between internal and external data processes in an organization based on the type of work performed. Various approaches and techniques are utilized to model processes and map the flow of data and information. Some of these techniques and languages include system context diagram (SCD), unified modeling language (UML), structured analysis and design (SADT), integrated definition for functional modeling (IDEFO), enterprise data model, architecture interconnect diagram, and problem diagrams (Lee et al. 2007 and Trbelsky and Sacks 2010).

One of the techniques that is utilized in the early stages of a structured system analysis and design method to model the flow of data and information is data flow diagram (DFD). Data flow diagram models a system as a network of transformations lined by paths of data (De Marco, 1978). A DFD is a diagrammatic illustration represented by four components; an entity for input/ output that is characterized by a rectangle; a system or process denoted by a circle; a file or data store designated by parallel lines, and data flow between entity, system and database represented by an arrow based on De Marco conventions. DFD could be a good source to identify source and visualize the overall process of activities within a system with regard to the type of data input and output, data storage and flow of data. DFD is an important technique for modeling a system's high-level detail by showing how input data is transformed to output results through a sequence of functional transformations (Lee Vie, 2001). DFD is an advanced form of flow charts and pseudo-codes with respect to program designs.

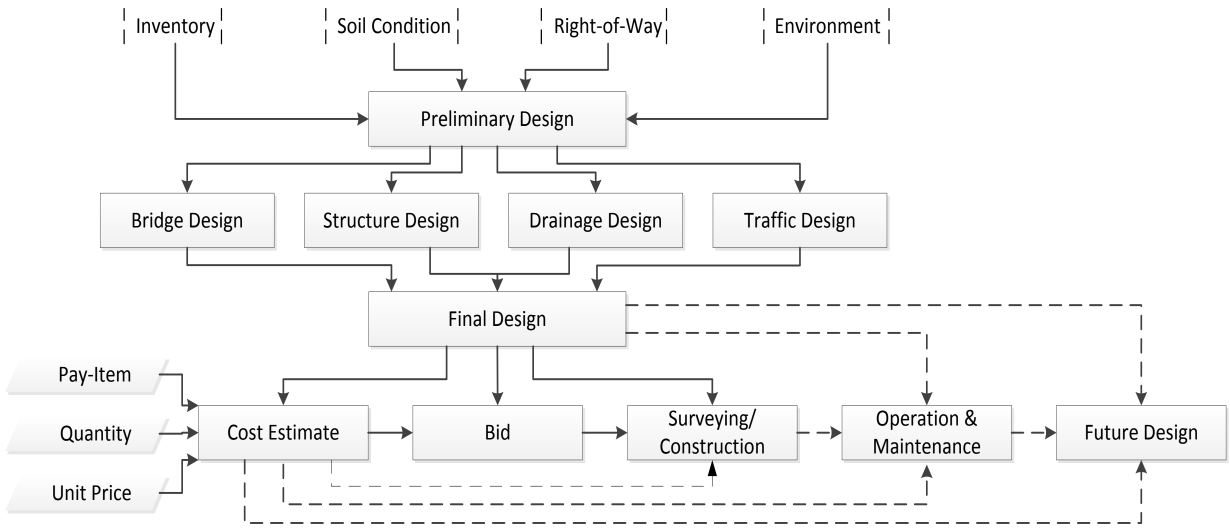


Figure 3-4 Highway Design Data Flow Diagram

Figure 3-4 shows a simplified data flow diagram example for highway design. Inventory data, soil condition, and environmental data are used by a roadway designer in setting alignment and preliminary design of overall pavement. Design developed during this phase is in turn used by other divisions such as bridge engineers, soil engineers, right-of-way experts, drainage specialists and traffic analysts. The design information is utilized by the respective divisions for further analysis to support their decisions and develop new information to be passed back to the roadway designer for final roadway design. Once design is finalized, the roadway engineer provides the plans to the cost estimator, bid analysts and surveyors for cost estimation, analysis of bid and construction preparation respectively. The estimator uses the plans to extract information such as quantity of work and pay-item activities along with cost data and unit price data to prepare a reliable cost estimate. Later, this design and cost estimate data are used in the construction phase, operation and maintenance and design phases of future projects.

However, DFD does not show the sequence of operation or timing of the processes associated with the data flow. In the construction industry, current models and approaches used in mapping flow of data and information focus either on the total work process of an organization or a particular business procedure, but not from a decision-making standpoint. In addition, existing data and information flow in highway agencies is complex in nature and difficult to track as project data and information are passed in sort of woven or spider-web like structure. Furthermore, the data and information can be processed in either an iterative, cyclical, parallel or linear nature. Therefore, the ability to track the process of data, information and decision making and identify the level of quality of these data and information is critical in managing highway projects. This study presents a conceptual time-valued 3-D DFD to show the process of data and information flow along highway project phases from decision-making perspective.

Based on the triple role concept, the principal component in information and knowledge retrieval to support decisions is data. Raw data (D) from one division can be a source for conducting various analysis or information (I) to a single or multiple divisions. These pieces of data and/or information are in turn used as knowledge or output to support various types of decision-making (DM) across the project life-cycle. Figure 3-5 shows the conceptual time-valued 3-D data flow diagram to illustrate the data and information flow between highway divisions. The diagram shows an X-Y-Z plane to represent the database, project phase and decision-making level respectively. The x-axis illustrates the type of database that is utilized in extracting the data and converting it to usable form. The y-axis determines the type of decision made in which project phase while, the z-axis defines the level of decision made and the type of data and information required in a project life-cycle. Arrows are used to represent data and information flow between divisions. Raw data (D_i) can be transferred from one division to multiple divisions ($Division_i$), where some of these data are returned as information (I_i) or passed to the next process or division as information for supporting specific decision.

In this model, a division is used as a processor to convert the data into information or utilize the data and information obtained from various disciplines for analysis purposes in supporting key decisions. Data and information path can either be available or existing, but not utilized or non-existing. Different colors are used to differentiate the different usage of data in the decision-making process across a project life-cycle. However, this flow of data and information diagram creates a many-to-many relationship which makes it difficult to track the exact path of data and information flow. Therefore, the relationship between data, information and decision can be effectively described using a context concept or algebraic expression. The formal context uses a triple set (D, I, DM) consisting of two sets D and I and a relation $DM \underline{C} D X I$. The members of I are called objects, while the members of D are called attributes. An object $I_i \in I$ has an attribute $D_i \in D$, if$(D_i, I_i) \in DM$.

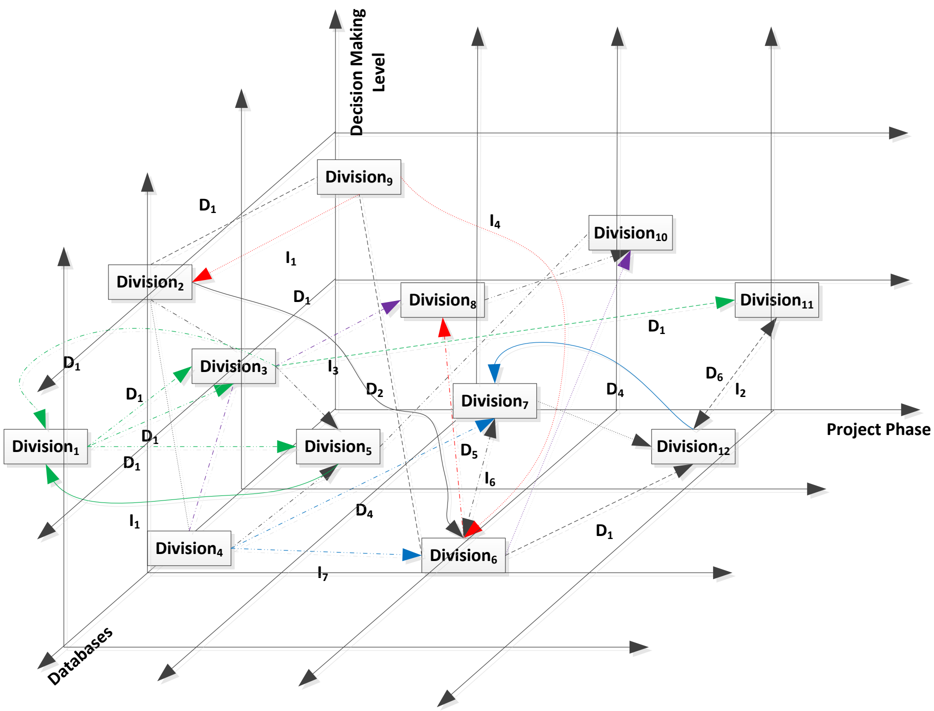


Figure 3-5 Conceptual Data Flow Diagram

Let a certain division $Div \coloneqq \; D, I, DM$ represent data, information and decision respectively in a formal context. The set of data attributes of an object $I_i \in I$ is

$$att_{Div} \; I_i \;\; = \;\; D_i \in D \quad I_i DM D_i \quad \dots \dots \dots \dots \dots \dots \dots \dots . . (1)$$

The set of data attributes shared by all objects (information) of a subset $A \; \underline{C} \; I$ is

$$att_{Div} \; I_i \;\; = att_{Div} \; I_i \quad \dots \dots \dots \dots \dots \dots \dots \dots . . (2)$$

The set of objects that have all data attributes of a subset $B \; \underline{C} \; D$ is

$$obj_{Div} \; B \;\; = \;\; I_i \in I \quad B \, \underline{C} \, att_{Div} \; I_i \quad \dots \dots \dots \dots \dots \dots \dots \dots . . (3)$$

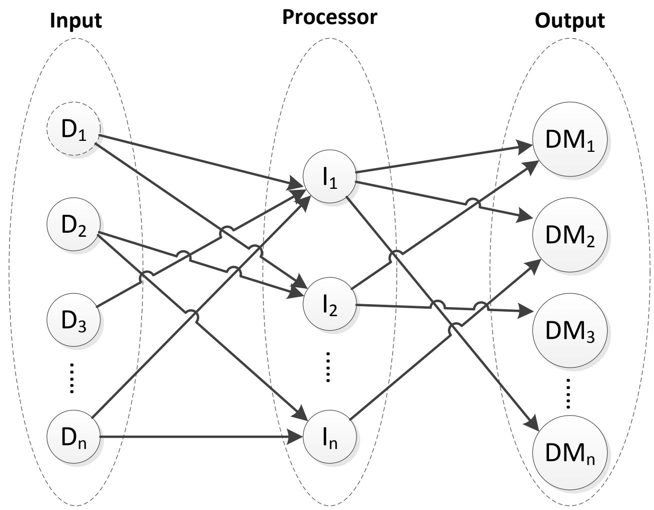


Figure 3-6 Context Graph

Figure 3-6 shows a context graph. If the attributes $D_1, D_2, D_3 \ldots D_n$ are represented as input symbols, the objects $I_1, I_2, I_3 \ldots I_n$ are processor symbols and $DM_1, DM_2, DM_3 \ldots DM_n$ are the output symbols, then:

$$D_{1,}D_{2,} \ldots . D_{n,} \rightarrow I_{1,}I_{2,} \ldots . I_n \tag{4}$$

$$I_{1,}I_{2,} \ldots . I_n \rightarrow DM_{1,}DM_{2,} \ldots . DM_{n,} \tag{5}$$

The arrow "→" is used as a connector or flow between inputs and outputs that are used to produce a certain result. Raw input data denoted by $D_{1,}D_{2,} \ldots . D_n$ symbols are used in the production of information, denoted by $I_{1,}I_{2,} \ldots . I_n$ symbols. Similarly, the information is used as input to produce or support the decision-making denoted by $DM_{1,}DM_{2,} \ldots . DM_n$ symbols. For instance, based on Figure 3-6, $D_{1,}$ D_3 and D_n are used to produce I_1, while $D_{1,}$ and D_2 are used to produce $I_{2..}$ I_1 and I_2 in turn produce DM_1. Each of these flows or transactions is assumed as a single transformation and can be decomposed into an input/output (I/O) matrix form.

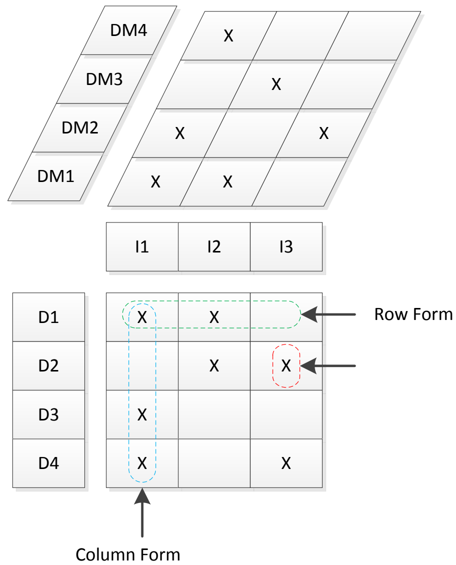


Figure 3-7 Input/Output Matrix

The many-to-many relationship in Figure 3-6 is converted into two separate I/O matrix; data/ information (D/I) and information/decision-making (I/DM) matrix as shown in Figure 3-7. The marked elements in Figure 3-7 are a subset of the corresponding relations that is used to develop each output. This matrix can be represented by four types of forms; a) elemental, b) row, c) column, and d) matrix forms. Elemental, row and column forms are primitive term forms as only matrix forms are directly decomposable (Adler, 1989). A process must consist of two inputs and two outputs to apply decomposition or apply this algebra.

a. Elemental form: $D_2 \rightarrow I_3$ where, D_2 is used to produce I_3

b. Row form: $D_1 \rightarrow I_1, I_2$ where, $D_{1,}$ is used to produce I_1 and I_2

c. Column form: $D_{1,} D_{3\,,} D_4 \rightarrow I_1$ where, $D_{1,}$ D_3 and $D_{4,}$ are used to produce I_1

d. Matrix form: $D_1, D_{2\,,} D_{3\,,} D_{4\,,} \rightarrow I_1, I_{2\,,} and\ I_3$ where, $D_{1,} D_{2\,,} D_3\ and\ D_{4\,,}$ are used to produce $I_1, I_{2\,,} and\ I_3$

These forms hold true for the information/decision-making (I/DM) matrix in terms of the elemental form and matrix form that may consist of $I_3 \rightarrow DM_2$ and $I_1, I_{2\,,} I_3 \rightarrow DM_1, DM_{2\,,}\ DM_3\ and\ DM_4$ respectively whereas the raw form and column form are interchanged. For instance, the column form takes $I_1 \rightarrow \ DM_1, DM_2\ and\ DM_4$ where the raw form takes $I_1, I_2\ \rightarrow\ \ DM_1$. This matrix form will allow one to visually identify what types of data and information are required to support a specific decision. An ideal data-information-decision-making framework and the application of the input/output matrix are explained more in the next sub-section and using case studies in the next chapter respectively.

3.4 Ideal Data and Information Integration Framework

This study uses a combination of top down approach and bottom-up approach to map the data, information and decision-making in highway projects. Based on the context graph and input/output matrix developed in the previous section, a 3-tiered hierarchical framework is developed to match and convert data into meaningful information and knowledge to support the decision-making (Figure 3-8). The framework consists of raw data (Tier I), information (Tier II) and decision-making (Tier III). The framework will integrate and map these three entities based on the decision-maker's requirement. Mapping these three-tiered components using hierarchically dependency and inclusive relationships will help identify three types of paths. The first path is an active path that indicates active use of data currently employed as information in support of decision-making. The second path is an inactive path meaning that there are currently available data but is not utilized in decision-making. However, in this inactive path, all the required data are available, so a well-defined method to transform data into information would quickly make this path an active path. The third path is a non-existing path indicating that there is not available data to generate required information to support specific decisions and information extraction method is not known.

The three types of paths are demonstrated using a solid line to represent active data usage, dotted line to represent available but under-utilized data usage and a broken line for non-existing paths. This framework is used as a basis for developing a gap analysis between current status of data management and usage and ideal data collection and information creation. The framework will be able to show what types of data should be collected, and what methods can be used to convert the data into meaningful information and knowledge to support various highway project decisions. For example, it will be able to show what types of pavement condition assessment data are used to develop deterioration curves of different pavement types, which will determine the most effective timing and treatment option for a specific pavement. In addition, it will allow the development of an active utilization plan of currently existing databases. It will also help develop a new data collection and information/knowledge generation plan to support key decisions which historically were not well-supported with information and data.

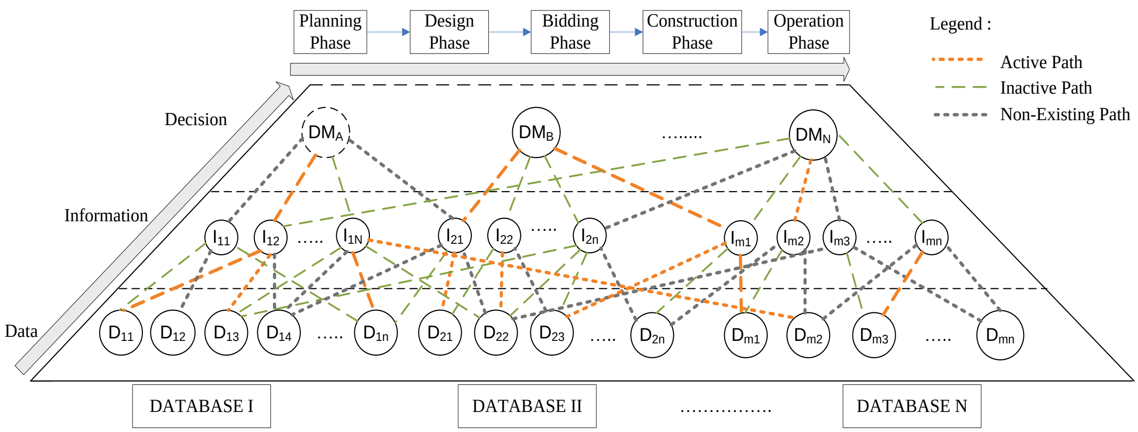


Figure 3-8 Three Tiered Hierarchical Framework

3.5 Data Attributes

In this study, data that are used to support decisions are broadly classified into four parameters to address the fundamental project management goals. These include cost (budget), quality, safety and environment and time (budget). Further classification is conducted based on the type of data and/or database utilization and decision making across the different phases of a highway project. These two classifications are performed due to the overlapping effect of data that can be utilized for various types of decisions across a project life-cycle. Data can range from unstructured data types such as e-mails, public input, reports, etc. to structured data such as average daily work report (AADT) or highway type. This classification mainly addresses semi-structured and structured data types that can be utilized in decision-making. In this study, unstructured data types that can be used to support decision-making are shown using structured data format.

3.5.1 Cost

Data attributes that contribute to finishing highway projects on budget or spending public money efficiently or enhancing economic growth are categorized under the cost module. These cost related data can be direct or indirect costs that are associated with available budget, quantity of work performed, unit prices, travel costs, preconstruction cost estimate, treatment cost, construction cost, etc. These data play part in measuring the performance of projects; assess if projects are performed according to budget, quantity estimates are accurately billed, ensures if cost estimates met the requirements, payments are made according to contracts, and cost of changes are fairly priced (Nzekwe, 2010). Overall, they support decision in terms of allocating funds, negotiating costs, allocation of resources, tracking project progress, contractor payment, and selection of effective treatment.

3.5.2 Quality

Types of data that are used to measure or meet the functional, structural, legal requirements and quality related issues and conduct highway projects according to a standard specification are categorized under the quality parameter. These data may include changes (design errors, change orders, material testing during construction, etc.), roadway inventories (pavement type, pavement thickness, highway system, etc.), pavement condition data (rutting, faulting, cracking, etc.) and pavement history (type of treatment, composition, material, etc.). Quality data are used to measure performances, assess reliability, minimize reworks and conform to standards. They are utilized in the decision making of designing pavements, treatment selection, project prioritization and managing assets.

3.5.3 Safety & Environment

Traffic data (AADT, Profile, etc.), accident rates, vehicle information, driver information, crash location, weather, etc. are categorized under safety parameter. For example, safety data address questions of crashes as to where, when, why, who and how crashes have occurred to perform crash analysis /reports, increase work zone safety, improve roadway and bridge design and assess risks in selection of alternatives. In addition, right-of-way, environmental documents such as NEPA can be included in this category to account for the social, economic, environmental, and political aspects of a highway project.

3.5.4 Time

Data attributes that constitute in project scheduling, contract time determination, engineering hour estimate (resource allocation), treatment selection and production rate identification to facilitate on-time delivery, allocate reasonable time, and select appropriate alternative fall under time module. Time data can range from time sheet (engineering hours, number of plans or sheet production) and reported quantity to pavement age and/or treatment life.

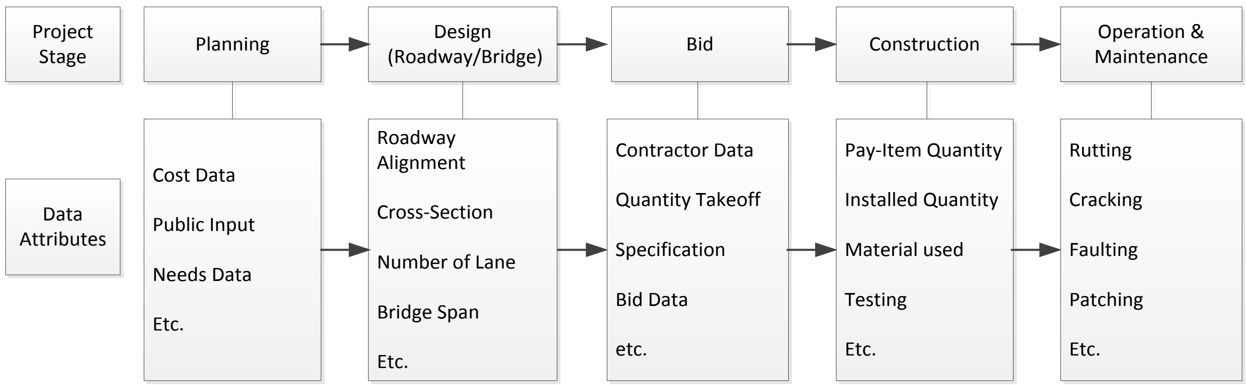


Figure 3-9 Highway Data Attributes Based on Project Phases

An alternative method of classifying data attributes based on data collected during highway project phases is shown through an example in Figure 3-9. Planning data also incorporate cost data, budget availability and public input, etc. During this phase, most data are used from the various divisions for planning purposes. These data include pavement condition data, construction data and roadway inventory data. Design data ranges from physical data inventories such as surveying (alignments, cross-section, subsurface and super elevation, etc.) and roadway components (shoulders, curbs, sidewalk, drainage pipes, etc.) to bridge elements (geometry, span, length, material, load data, etc.). Data attributes in the bidding phase include contractor data, quantity takeoff, specification, etc. Construction materials, equipment, crew size, reported (installed) quantity, and tests that represent the construction aspect of highway projects fall under construction data. Data that measure the performance of pavements such as rutting, cracking, patching, spalling, fatigue, etc. constitute operation and maintenance data.

3.6 Information Generation Using Data Analysis Methods

In this study, information acts as a medium or “processor’ between data and decision based on the triple role concept. Information is critical at this stage as managers and potential users are looking towards extracting valuable inputs that can support their decisions. In addition, it will identify important data requirements that can be changed into a usable form. Information can be performance measures, metrics and/or data that can add knowledge to the user. Information can be generated from raw data by applying appropriate data analysis techniques. These techniques can range from simple descriptive statistics to inferential statistics (regression models) and application of artificial intelligence tools such as neural networks. These analysis techniques in turn are used to optimize selection process, obtain reliable estimate, increase safety, improve quality of work and enhance the overall decision making.

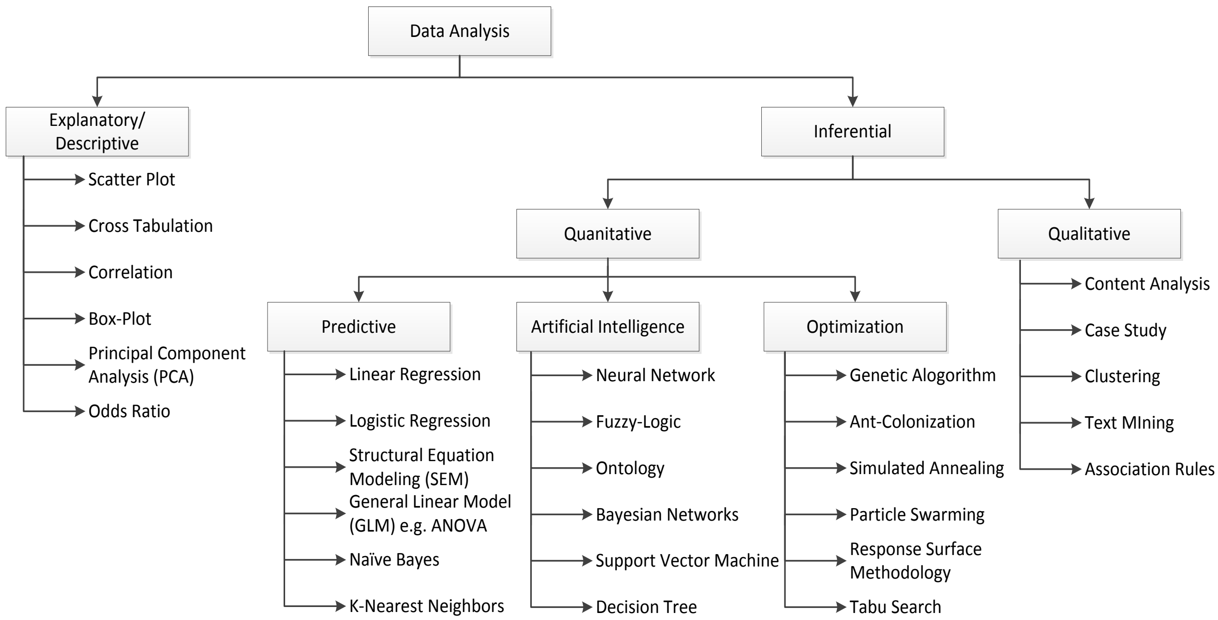


Figure 3-10 Data Analysis Methods

Data analysis methods can broadly be classified into explanatory and inferential analysis based on type of data usage from statistics standpoint. Explanatory or descriptive statistics deals with understanding of the data, identifying correlations/relationships between data, and calculating threshold values like average, minimum and maximum values. Common graphical techniques used in this type of analysis include a scatter plot, a box-plot, a cross-tabulation, a correlation, and a principal component analysis (PCA), etc. Inferential analysis deals with drawing conclusions and identifying patterns from a set of observational or sample data. It can perform tasks such as classification, estimating, prediction, affinity grouping, and clustering. Inferential analysis can be divided into qualitative and quantitative analysis for the purpose of inducing decisions. Typically, qualitative data analysis deals with semi-structured and unstructured data types like textual data. Content analysis, clustering, market basket analysis and text mining, etc. are some of the qualitative data analysis methods. For instance, Ng et al. (2006) utilized clustering analysis for text mining to assess facility conditions. Abdollahipour (2012) has used association rules to categorize the pavement treatment types that are frequently used in Oklahoma DOT rehabilitation projects. Quantitative analysis can further be divided into predictive modeling, and artificial intelligence and optimization techniques. Predictive modeling primarily includes parametric approaches such as regression models (linear, logistic, etc.), structure equation modeling (SEM), general linear model (GLM), etc., while artificial intelligence incorporates neural network, fuzzy-logic, ontology, decision tree and vector machines, etc. However, both predictive models and artificial intelligence have been used mainly to estimate cost, time (schedule or duration), resources and productivity, etc. Optimization techniques are also utilized as data analysis methods in supporting decision-making processes that require multiple criterion and/or tradeoff analysis. It has been used in determining the optimum number of piers and span length in bridge design, pavement treatment selection and resource (equipment) management. Optimization techniques include various algorithms such as genetic algorithms, particle-swarm optimization, ant-colonization, and tabu search, etc. Figure 3-10 shows a classification of data analysis techniques.

Table 3-1 summarizes various studies that were conducted to implement qualitative and quantitative analysis methods to develop models in supporting decision makers using DOT project records. Examples of qualitative analysis include different models developed by state DOTs in performing highway project outsourcing analysis. The Arizona DOT model has a function based on qualitative judgments to rank the potential for outsourcing. Some of the factors used in this weighted score analysis method include strength of competitive market, quality of service, control, risk of contracting out, legal barriers, political resistance and impact on public employees. The Pennsylvania contractibility model (CONTRAS) uses a 1 to 5 rating system that considers factors such as unit cost comparison, degree of labor intensity, existence of critical time constraints, contractor availability, work volume, planning difficulty level, requirement of special equipment or skill and amount of inspection required (Wilmot et al. 2002). On the other hand, Oregon DOT and Wisconsin DOT adopt a Balanced Scorecard approach that has a qualitative and cost index to decide as to perform design in-house or favor outsourcing (Rogge et al. 2003 and Eger et al. 2002).

Examples of quantitative analysis include work done by May (2003) who developed regression and Bayesian deterioration models for predicting pavement distresses using Mississippi DOT pavement management system in allocating resources and selection of an effective rehabilitation strategy. Woldesenbet and Jeong (2012) developed a data-driven component based prediction models for estimating preliminary engineering (PE) costs of roadway projects. The study showed the use of data mining techniques to develop decision tree and regression models based on ten years of historical project records from the Oklahoma DOT. Similarly, Williams et al. (2012) developed a regression model for estimating the engineering hours of capital improvement projects for the New York State DOT. However, the model was a general model which makes it difficult to allocate resources and negotiating costs for specific divisions. Weisbrod and Backwith (1992) developed an economic simulation model called REMI (Regional Economic Model Inc.) to evaluate the development impacts of highway investment using Wisconsin DOT (200 mile four-lane highway project) as a case study. The study showed how economic benefits of highway projects can be estimated and the estimates can be used for benefit-cost analysis to support policy decision-making.

Table 3.1 Examples of Quantitative Information Generation Methods

Studies / Criteria	Saito et al. (1991)	Weisbrod & Backwith (1992)	May (2003)	Wilmot and Cheng (2003)	Molennar (2005)	Nassar et al.(2005)	Woldesenbet, and Jeong (2012)	Williams et al. (2012)
Data Source	Indiana DOT	Wisconsin DOT	Mississippi DOT	Louisiana DOT	Case Studies	Illinois DOT	Oklahoma DOT	New York State DOT
Method used	Statistical Methods	Simulation	Bayesian, Regression	Statistical Methods	Monte Carlo Simulation	Regression	Decision Tree, Regression, Neural Network	Regression
Type of Project	Bridge	Highway	Pavement	Bridge & Highway	Highway	Highway	Roadway	Highway
Critical Factors	Region, bridge type, deck area, substructure area, age, functional class, component condition index	-	-	-	-	Initial planned cost, complexity, percent of bridge and roadway project	Project length, location, project type, highway type, route type, contractor type, let year, construction cost	Construction cost, number of sheets, number of lanes, project length

Molennar (2005) showed that preliminary cost estimates should be represented by a range of estimates using Monte Carlo Simulation in developing a probabilistic cost estimation system under various scenarios. The study presented a Cost Estimating Validation Process (CEVP) based on nine case studies. The process was intended to better understand the risks associated with mega highway projects for a more transparent assessment of uncertainty. Saito et al. (1991) utilized statistical tools in developing cost estimation models for bridge replacements using six years of 280 historical project data obtained from Indiana DOT (INDOT). The study concluded that adding component cost models is better than a total bridge cost model. The model addressed eight explanatory variables including region, bridge type, deck area, substructure area, age, functional class, component condition index, and completed work to explain the response variables, component cost, subtotal costs, and unit cost, respectively. Nassar et al. (2005) applied a regression model to estimate design costs of consulting firms based on 59 highway projects obtained from Illinois Department of Transportation (IDOT). Other studies include a fuzzy approach by El Wakil and Zayed (2012) to predict work task duration for a construction process. It is important to note these methodologies and studies are examples of potential data analysis or information generation methods.

4. Case Studies

This chapter presents three case studies; pavement management, preconstruction service cost, and daily work report or construction data to illustrate the ultimate benefit of the three tiered data-information-decision framework discussed in the previous chapter. The study uses two previously completed research projects and one active project as part of the case studies in terms of identifying the different decisions made, databases utilized and mapping the data and information flow. In addition, the study develops a gap analysis and an ideal data-information-decision making integration framework. A questionnaire survey and a series of interviews were conducted to identify key decisions, databases, data and information utilization in highway decision-making process and perform a gap analysis (Appendix B).

4.1 Case Study 1 – Construction Daily Work Reports (DWR)

4.1.1 Construction Field Data Collection Process

A typical construction data collection process begins when a construction project is awarded to a construction firm. Contract data is created in PES/LAS (Proposal and Estimates System/ Letting and Awards System) by the office of contracts and passed to the construction division. Once construction of the project commences, field inspectors and/or superintendents keep track of item quantities placed during the course of a project, document actual site conditions, and record any important activities in the project. These pieces of data are passed to a person in the office who compiles the data if there is more than one inspector per project making entries for all contracts or the same person enters the data into a contract administration software program. In Oklahoma DOT, inspectors make a note of the work activities in their note book and input the data into a contract administration program, SiteManager upon returning to the office on the same day or at the end of the week. Once data such as daily work reports (DWR) are recorded on a standalone workstation database, they are transferred to the server's database. The process automates the transfer of data from the server to and from standalone models.

4.1.2 Database

State agencies utilize various contract administration programs such as SiteManager, FieledManager and FieldBook developed by the Federal highway Administration (FHWA) to collect and manage construction data. For instance, currently 16 states have the license to operate

SiteManager to avoid repetitive data entry and manage contract data during the construction phase. The Oklahoma DOT is one of the state agencies that utilizes SiteManager as its primary contract administration software program. SiteManager is a client-server based system that is interlinked through web that can be accessed across the various construction offices across the state. SiteManager consists of six basic functions or modules to view and store highway construction project data; contract administration, daily work reports (DWR), contractor payments, change orders, material management, and civil rights management systems. These functions allow data acquisitions such as materials and equipment, project location, job-site conditions, construction pay items, reported quantity, change orders and weather conditions. Currently, the ODOT SiteManager contains a database of more than 2000 previously completed and ongoing construction projects collected since 2002. This study utilizes one of the prominent function that is stored in the ODOT SiteManager, daily work reports (DWR) as one of the case studies. SiteManager daily work report (DWR) consists of major tabs or folders that include DWR info, contractors, contractor equipment, daily staff, work items, and force accounts. In addition, it consists of diary, diary adjustments, history, process list, reference tables and various templates.

a. DWR Info: This tab is used to create new daily work report and collect daily activities. It incorporates contract id, inspector name, date, highest and lowest temperature, and morning and afternoon weather conditions. The tab has three options to indicate if there is no data, work was installed, contractor was on site, and daily staff was on site. Accidents, delays, instructions and general remarks are also recorded in this tab. Figure 4-1 shows a screen shot of the DWR info tab.

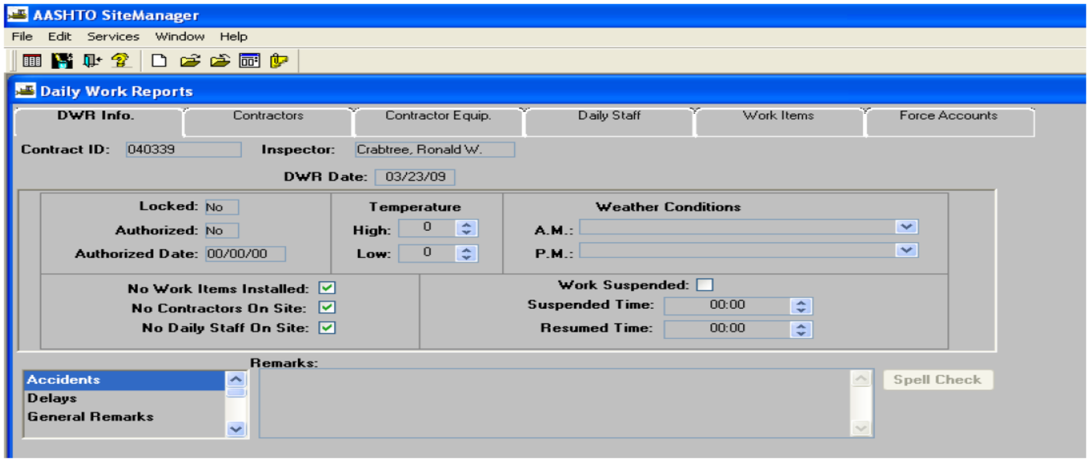


Figure 4-1 Screenshot of DWR Info Tab

b. Contractors: This tab is used to enter daily data on the contractor, supervisors, and variable labors (by personnel type) along with the number of hours worked. The personnel type and supervisor names are attached to a vendor master list as a reference table where the superintendent is able to easily pick the variables from a list. Figure 4-2 shows a screen shot of the contractor tab.

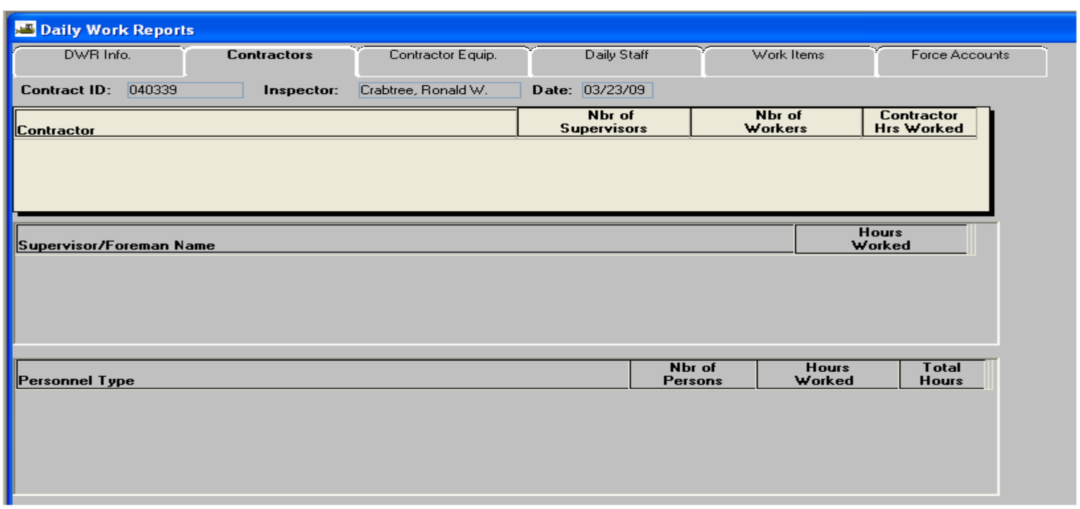


Figure 4-2 Screenshot of Contractor Tab

c. Contractor Equipment: This tab documents the daily presence and use of equipment on the job site for the selected contractor along with the number of hours each unit of equipment was used. The equipment is associated with a vendor's list as a reference table with the unit of equipment ID assigned to it. Figure 4-3 shows a screen shot of the contractor equipment tab.

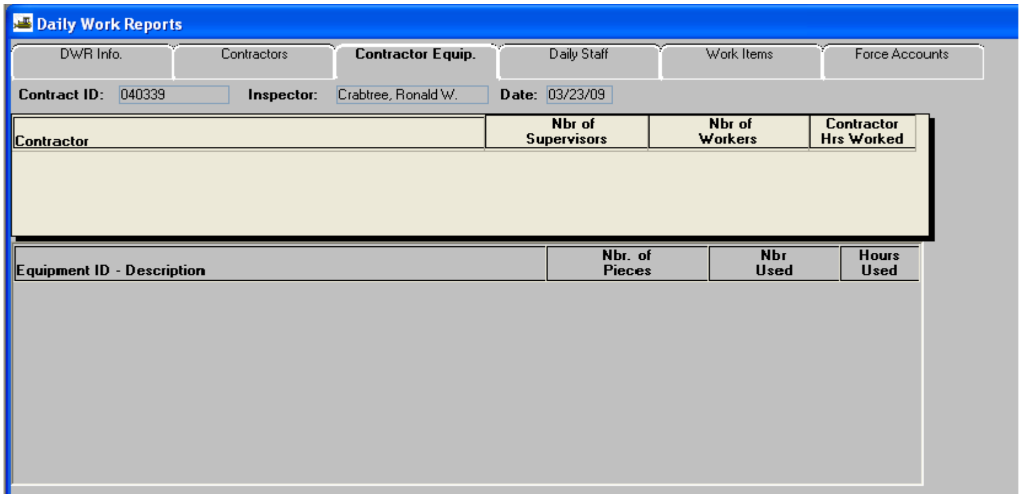


Figure 4-3 Screenshot of Contractor Equipment Tab

d. Work Items: This tab is used for adding daily usage data for installed work items. It includes the quantity of work reported to date, quantity installed to date, quantity paid to date, bid quantity, unit price, the station, offset and distance. In addition, it is important to note that the Oklahoma DOT does not use some of the SiteManager information tabs such as diary adjustments, daily staff/staff member and force accounts. Figure 4-4 shows a screen shot of the work item tab.

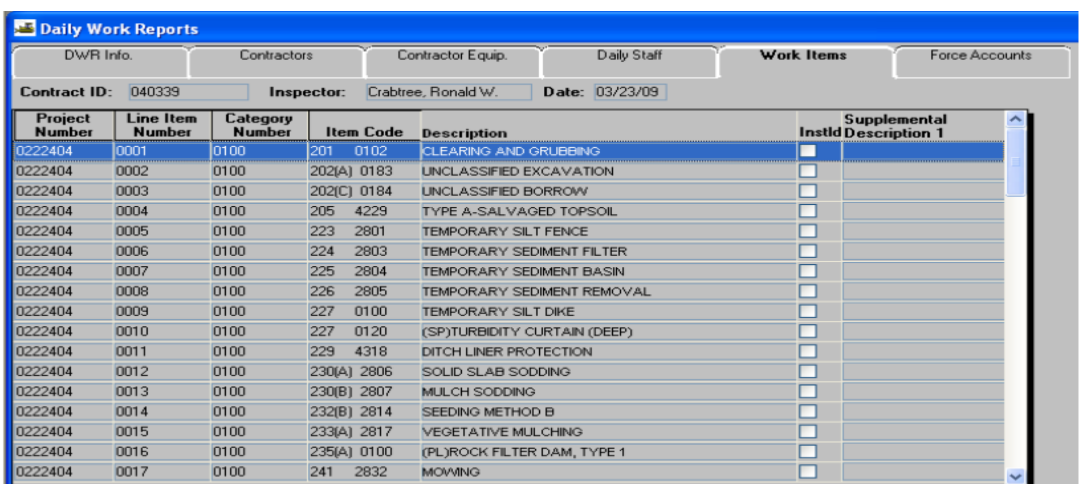


Figure 4-4 Screenshot of Work Item Tab

When data or information is requested, SiteManager, as well as excel queries are used to pull data from the database. Overall, this database is digital formats and documents all quantities which result in contractor payment. It keeps track of working days charged and creates weekly working day reports that are sent to the contractor and also contractor evaluations among others. However, there are data that are still displayed on paper and its intended use is limited.

4.1.3 Current Utilization of DWR Data

At a project level, daily work report (DWR) of highway construction projects is one of the primary data collected during the construction phase where superintendents are the sole data suppliers. Current usage of this DWR data is analyzed in Table 4-1. Based on the analysis, the majority of existing DWRs are either used for reporting (I_3) or claim analysis (I_2) for resolving disputes that may arise between the client (DOT) and the contractor. Project work item or activities along with the amount of placed quantity are the most utilized data in terms of making contractor payments (I_1), tracking schedule (percentage completion) based on the working days charged, (I_4) and reporting the weekly working days. However, these data have the potential to be utilized in information generation such as estimation of production rate, determination of contract time, safety analysis, material analysis, equipment management, and resource allocation.

Table 4.1 Current Utilization of DWR Data

Type of Data	Data Attributes	Data	Description	Data Type	No Use	Current Use			
						Contractor Payment	Dispute Resolution	Reporting	Percentage Completion
						I1	I2	I3	I4
DWR Info	Contractor ID	D1	ID 000001-100000	Numeric : Ordinal				X	
	Inspector Name	D2	Last and first name	Character :				X	
	Date	D3	xx/xx/xxxx	Numeric : Ordinal				X	
	Low Temperature	D4	Temp. °F	Numeric : Interval			X	X	
	High Temperature	D5	Temp. °F	Numeric : Interval			X	X	
	AM Condition	D6	Sunny, windy, cloudy, etc	Character :			X	X	
	PM Condition	D7	Sunny, windy, cloudy, etc.	Character :			X	X	
	Work Suspended	D8	Time AM/PM	Numeric : Ordinal			X	X	
	Work Resumed Time	D9	Time AM/PM	Numeric : Ordinal			X	X	
	Humidity	D10	-	-	X				
	Precipitation	D11	-	-	X				
Contractor	Contractor	D12	Name	Character :			X	X	
	Subcontractor	D13	Name	Character :			X	X	
	Supervisor	D14	Foreman, superintendent, etc.	Character :				X	
	Personnel	D15	Laborer, concrete finisher, etc.	Character :			X	X	
	Supervisor Hourly	D16	Number of Hours	Numeric : Interval			X	X	
	Personnel Hourly	D17	Number of Hours	Numeric : Interval			X	X	
	Supervisor Number	D18	Count	Numeric : Interval			X	X	
	Personnel Number	D19	Count	Numeric : Interval			X	X	
Equipment	Equipment Type	D20	Backhoe loader, bulldozer, etc.	Character :			X	X	
	Equipment Hours	D21	Number of Hours	Numeric : Interval			X	X	
	Equipment Number	D22	Count	Numeric : Interval			X	X	

Type of Data	Data Attributes	Data	Description	Data Type	No Use	Current Use			
						Contractor Payment	Dispute Resolution	Reporting	Percentage Completion
						I1	I2	I3	I4
Work Item	Project Number	D23	ID 0000001-1000000	Numeric : Ordinal		X	X	X	X
	Work Item	D24	Clearing & grubbing,	Character :		X	X	X	X
	Placed Quantity	D25	Count	Numeric : Interval		X	X	X	X
	Location	D26	Linguistic	Text			X		
	Station	D27	xx+xx.xx - yy+yy.yy	Character :			X		
	Offset	D28	Right, left of baseline or	Character :			X		
	Distance	D29	Distance from baseline or	Numeric : Interval			X		
Remarks	Accidents	D30	Station, time, officer, etc.	Text	X				
	Delays	D31	Linguistic	Text	X				
	Site Condition	D32	Dry, muddy, etc.	Character :			X	X	
	General Remark		Linguistic	Text			X	X	
	Instructions	D34	Linguistic	Text			X	X	
	Personnel Remarks	D35	Linguistic	Text			X	X	
	Supervisor Remark	D36	Linguistic	Text			X	X	
	Prime Contractor	D37	Linguistic	Text			X	X	
	Subwork	D38	Linguistic	Text			X	X	
	Type of Day	D39	Linguistic	Text	X				
	Unapproved Work	D40	Linguistic	Text	X				
	Visitors	D41	Linguistic	Text	X				

Currently, the DWR misses weather condition data such as humidity (D_{10}) and precipitation (D_{11}) that are critical in earthwork analysis, delay analysis, and determination of production rate that could support scheduling decisions. DWR consists of numeric (ordinal and interval), character or nominal variables and text formats. The numeric and character variables can be used in developing various prediction and estimation models. Although information generation methods such as text mining can be used to extract information and knowledge from text formats, accidents (D_{30}), delays (D_{31}), and common remarks ((D_{33} –D_{41}) such as general remark, prime contractor work, site conditions, sub-work and visitor's data that are linguistic should be standardized and converted to structured data types to effectively generate reliable quantitative or qualitative information.

4.1.4 Ideal Data-Information-Decision Making Integration Framework

Based on the three-tiered framework developed in the previous chapter, the use of data is classified into data, information and decision-making for the purpose of integration. In this case study, data is considered as all types of raw construction data collected from highway project site, while information refers to performance measures and analyzed outputs based on raw data inputs to evaluate, understand relationships and develop models to support highway decisions. Decisions incorporate utilization of higher level information and knowledge to support final selections and/or judgments. Currently, the primary use of DWR data is for assessing quantity of work for determining contractor payments, dispute resolution (claim analysis), reporting and percentage completion (I_1 - I_4). However, DWR data can be utilized in measuring performance and analyzing actual project conditions such as determining production rate, accident analysis, delay analysis, material analysis and as-built information (I_5 - I_9) which in turn can be used as information to support various decisions across a project life-cycle: resource allocation, contract time determination and maintenance (DM_1 - DM_3) during the planning phase, roadway design, bridge design, and traffic and safety design (DM_4 - DM_6) during the design phase, cost tracking, schedule tracking and material quality control (DM_7 - DM_9) during the construction phase.

i. Resource allocation (DM_1) – refers to but not limited to distribution of transportation funds or budget across projects, assignment and time allotment of skilled manpower and assignment of equipment in the planning of rehabilitation and new highway projects. During the planning phase, these decisions may be influenced by the productivity and construction as-built data of highway projects.

ii. Contract time determination (DM_2) - refers to the identification of the amount of time or duration required to complete work tasks contained in a contract document. This decision may be affected by factors such as weather conditions, material delivery, and location of project, etc. in which allocating appropriate contract time through the consideration of these factors reduces public inconvenience, minimizes cost overruns and increases safety.

iii. Maintenance decisions (DM_3) - include assessments and measuring performances in terms of prioritizing projects, selecting projects and selection of optimized treatment. These decisions heavily rely on previous pavement construction history, type of treatment utilized, as-built data and information.

iv. Roadway design decisions (DM_4) - In the design phase, decisions such as selection and evaluation of alternatives such as pavement type (concrete, asphalt or a combination), shoulder type, pavement thickness, and other geometric decisions (number of lanes, width, median type, horizontal and vertical alignment, etc.) are performed to meet the level of service of roadway.

v. Bridge design decisions (DM_5) - include selection or identification of span length, width, and number of bridges required, etc. that may be optimized based on as-built data and information, previous accident and reports.

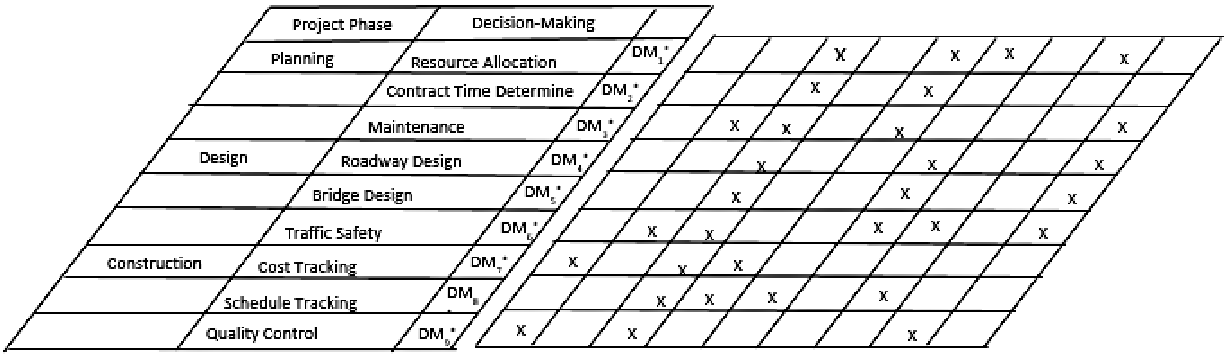
vi. Traffic and safety decisions (DM_6) - include choosing the right type of light fixtures, identifying the number of traffic signs and posts required and type of guardrail needed. These decisions may be affected by factors like the location of a project, as-built information, and traffic analysis.

vii. Cost tracking or benchmarking (DM_7) - In the construction phase, these decisions denote cost related assessments as to approve payments and make appropriate choices to finish projects on budget.

viii. Schedule tracking (DM_8) - refers to tracing actual project progress with planned schedule. It will assist a project manager in making decisions such as adding resources or working extra hours if projects are behind schedule.

ix. Quality control (DM_9) - deals with measuring the performance of work tasks according to contracts and specifications. It should be noted that these decisions (DM_1 - DM_9) are only examples of potential highway decisions made at various phases of a project that may utilize construction daily work report and do not incorporate all decisions made by highway agencies.

An input/output matrix is developed to map the relationship between these decision-making with data and information (Table 4-2). Missing data, information and decision are represented by D_i*, I_i* and DM*. Based on the matrix analysis, weather condition ($D_4 – D_9$), work item and placed quantity along with its location (D_{23} – D_{29}) and remarks of personnel, supervisor, prime contractor, sub-contract work and site condition (D_{32} – D_{38}) have the highest potential in generating information. Equipment data ($D_{20} – D_{22}$) and crew size (D_{12}– D_{19}) has relatively a fair usage in extracting information.



Type of Data	Data Attributes	Data	Information: Contractor Payment I1	Dispute Resolution I2	Reporting I3	Percentage Complete I4	Production Rate I5*	Accident Analysis I6*	Delay Analysis I7*	Material Analysis I8*	As-built Informatio I9*
DWR Info	Contractor ID	D1			X		X				
	Inspector	D2			X		X			X	
	Date	D3			X		X	X	X		
	Low Temperature	D4		X	X		X	X	X	X	
	High Temperature	D5		X	X		X	X	X	X	
	AM Condition	D6		X	X		X	X	X	X	
	PM Condition	D7		X	X		X	X	X	X	
	Work Suspended Time	D8		X	X		X	X	X		
	Work Resumed Time	D9		X	X		X	X	X		
	Humidity	D10*					X	X	X		
	Precipitation	D11*					X	X	X		
Contractor	Project ID	D12		X	X		X				
	Subcontractor	D13		X	X		X				
	Supervisor	D14			X		X				
	Personnel	D15		X	X		X				
	Supervisor Hourly work	D16		X	X		X				
	Personnel Hourly work	D17		X	X		X				
	Supervisor Number	D18		X	X		X				
	Personnel Number	D19		X	X		X				
Equipment	Equipment Type	D20		X	X		X			X	
	Equipment Hours	D21		X	X		X			X	
	Equipment Number	D22		X	X		X			X	
Work Item	Project Number	D23	X	X	X	X	X				
	Work Item	D24	X	X	X	X	X				
	Placed Quantity	D25	X	X	X	X	X		X	X	X
	Location	D26		X			X		X	X	X
	Station	D27		X			X		X	X	X
	Offset	D28		X			X		X	X	X
	Distance	D29		X			X		X	X	X
Remarks	Accidents	D30					X	X	X		
	Delays	D31					X	X	X		
	Site Condition	D32		X	X		X	X	X	X	X
	General Remark	D33		X	X		X	X	X	X	X
	Instructions	D34		X	X		X	X		X	X
	Personnel Remarks	D35		X	X		X	X	X	X	X
	Supervisor Remark	D36		X	X		X	X	X	X	X
	Prime Contractor Work	D37		X	X		X			X	X
	Sub work	D38		X	X		X			X	X
	Type of Day	D39					X				
	Unapproved Work	D40			X						
	Visitors	D41									

Figure 4-5 Ideal Data, Information and Decision-Making Input / Output Matrix

In terms of information, production rate determination (I_5), accident analysis (I_6), and as-built information (I_9) has a higher potential in supporting highway decisions. Decision-making like resource allocation (DM_1), maintenance (DM_3), traffic and safety design (DM_9) have high possibility of making advantage of these information. Roadway and bridge design, cost, schedule and quality control has also a good potential in utilizing extracted information and data to support future design project tracking and control.

In the development of the three-tiered framework, three paths, active, inactive and non-existing paths were described in integrating DWR data. It is important to note that current data usage matrix of I_1 – I_4 discussed in Table 4.1 are considered as active paths. The process of converting data into information from I_5 – I_9 are considered inactive as there are available data in current DOT database but are underutilized with the exception of humidity and precipitation (as part of weather condition data) that is non-existing. For example, the potential use of almost all data attributes in developing production rate estimation models except unapproved work items and visitor data is considered as inactive path. Accident Analysis is another inactive path that can be analyzed using data such as weather condition (D_3 – D_{11}) and incidents such as accidents, delays, and site conditions along with other remarks (D_{30} – D_{36}). All paths that map the information with the decision-making is considered as non-existing path as it's not currently utilized to support highway decisions. The development of a production rate (I_5) can be used in allocating resources (DM_1), determining contract time (DM_2), planning maintenance projects (DM_3) during the planning and scheduling phase and tracking schedule (DM_5) during the construction phase. Accident analysis (I_6) can be used to allocate resources (DM_1), design roadway, bridge and traffic and safety (DM_4 - DM_6).

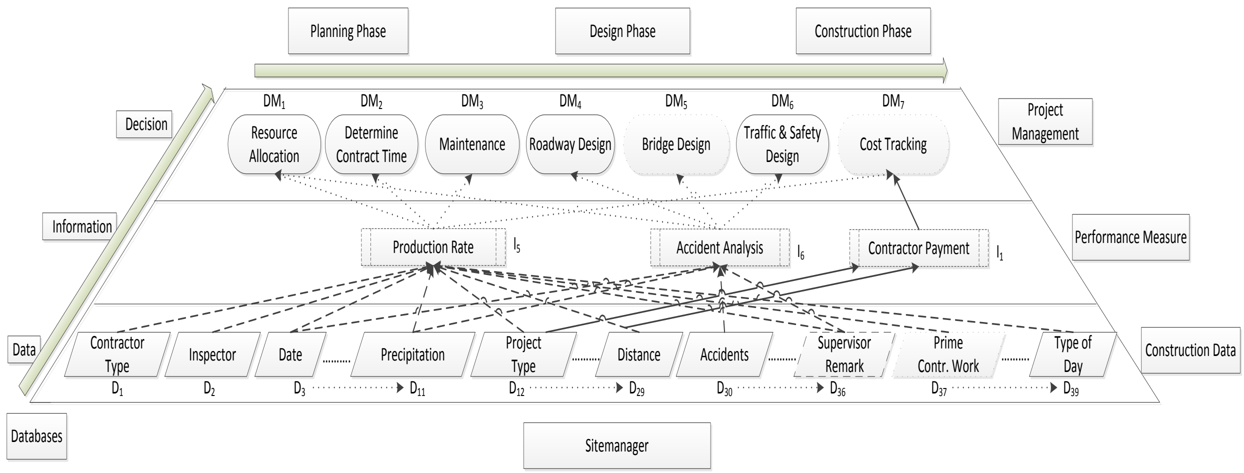


Figure 4-6 Three-Tiered Construction DWR Decision-Making Framework

Figure 4-5 shows an example of the three–tiered framework using the examples described above. The solid lines in the figure represent active paths, while the dotted broken lines represent inactive paths and the dotted lines represent non-existing paths. However, it is important to note that additional data attributes such as soil type from the material division, road system, type of route, highway type and project type from roadway inventory or contract management database and average daily traffic (ADT) from traffic inventory may be incorporated from other divisions to generate more reliable information and develop comprehensive decision-support systems. In addition, it should be noted that some of these data need major data cleanup and change in data structure to easily retrieve information and perform reliable analysis

4.1.5 Gap Analysis

A well-organized recorded dataset of completed highway projects, in which a project's progress is clearly documented, can be an excellent source for developing reliable information. These data ranges from project-level data to site-specific factors collected during construction, such as project location, job-site conditions, rainfall data, weather conditions, contractors' productivity, and other related information. Use of these data could be an effective approach in planning future highway projects. It can enhance contractor payment, track cost and schedule; predict production rate estimation and determine contract time of future projects; address environmental issues and safety measurement; allocate resource management and control work quality; asset management data collection (such as pavement operation & management); reporting and preparation of legal disputes; overall planning and designing highway project. Although inspectors in highway agencies and DOTs spend a huge amount of time in recording data (spending one to two hours per day on average), the current system does not allow a simplified generation of information. Decision-makers requirement is not well-addressed and its use is very minimal. McCullough (1997) argue that the Indiana Department of Transportation, INDOT supervisory personnel spends on average between 30%-50% in recording and analyzing field generated data.

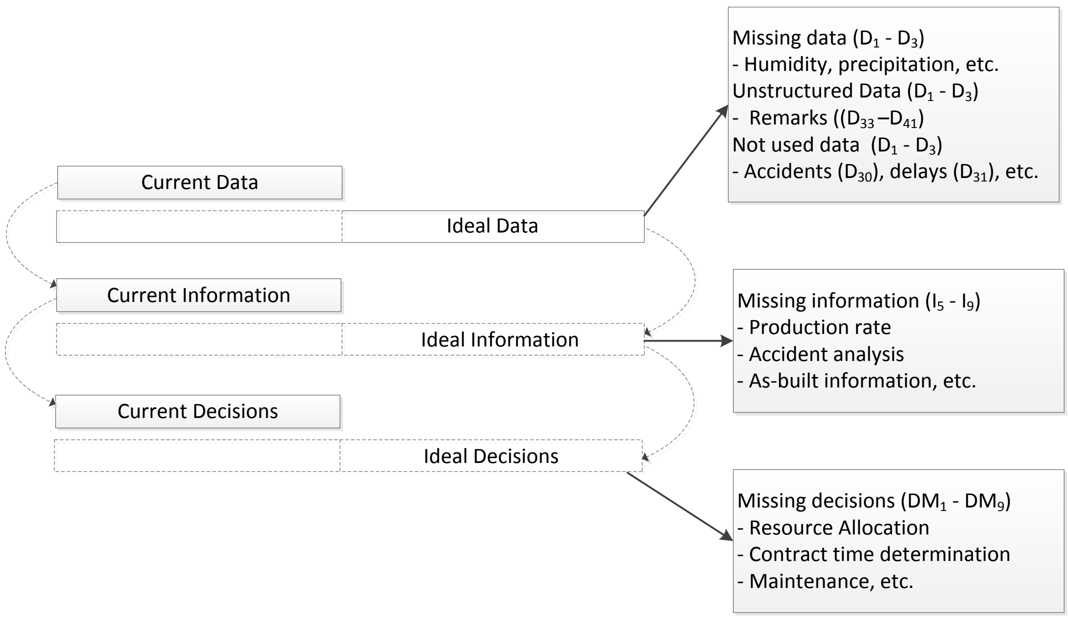


Figure 4-7 Gap Analysis

Based on context analysis, DWR are often utilized in reporting and preparation of legal disputes. Reported quantity and work item are the primary data that are utilized in contractor payments and tracking project progress. More than 35% of the DWR data are linguistic in nature (accidents D_{30}, delays D_{31}, and common remarks D_{33} –D_{41}). In addition, there are unused, missing and/or repetition of data observed in the DWR. Weather condition data such as humidity (D_{10}) and precipitation (D_{11}) are not collected. Data such as type of day (D_{39}) in the remarks section has already been described using temperature, morning and afternoon condition (D_4- D_7) in the DWR info. Although accidents include the station, time, officer who reported it, the agency, and description of the incident, these data should be allocated into separate items to perform investigation such as safety or crash analysis. This applies to location (D_{26}) of work item placed in which attributes such as structure number, location of as-built data should be separate entities in reference to the baseline or centerline that can be used in future roadway and bridge design.

Possible types of delay should be enumerated as a list to easily capture data. Figure 4.6 shows a gap analysis between current and ideal data, information, and decisions.

RPT-ID: RDWRHCON	Oklahoma	DATE: 10/25/2011
USER: awilson Wilson, Antoinetta C.	Department of Transportation	PAGE: 1 of 1

DAILY WORK REPORT FOR CONTRACT: 090614

DWR Date: 04/14/2010	Contract ID: 090614	Authorized: No	Locked: No	Paid: No
Inspector ID: cdebord	Inspector: Debord, Chris			
High Temp: 80	Low Temp: 57	A.M. Condition: SUNNY AND WINDY	P.M. Condition: SUNNY AND WINDY	
Work Suspended Time: 00:00	Work Resumed Time:00:00	No Work Items Instld: ☒	No Contrs Present: ☐	No Staff Present: ☒

Remarks: Yes		
	General Remarks	Excavation for trench, placing perf and non perf pipe, placing bedding and backfill material.
	Prime Contractor Work	Excavation for trench, placing perf and non perf pipe, placing bedding and backfill material. Located on South West side of 142
	Site Conditions	0.00

Contractor Information

Contractor ID:202912267 Contractor Name: SCHIRALLI CONSTRUCTION CORP. Hrs Worked: 40.000

Nbr of Supervisors: 1 Nbr Of Workers: 3

Supervisor/ Foreman :	Name	Hrs. Worked
	FOREMAN	10.000

Variable Labor:	Personnel Title	Qty	Hrs.Worked
	LABORER COMMON	3	10.000

Equipment:	Description	Qty	Qty Used	Hrs. Used
	Excavator, Track	1	1	6.000
	Loader, Bobcat (Skid) Wheel	1	1	4.000
	Loader, Backhoe	1	1	6.000
	Loader, Wheeled	1	1	2.000
	Truck, Crew	1	1	2.000

Figure 4-8 Typical DWR Report

In addition, while recording contractor, personnel and equipment information (D_{12}- D_{22}), contractor and sub-contractor works should be separately addressed to a particular work activity along with quantity of work performed and resources (crew and equipment) rather than only to a project. This is shown using a typical daily work report used for reporting purposes in Figure 4-7. In the report, the types of equipment, personnel and supervisor utilized for excavating trench, placing perforated and non-perforated pipes, placing bedding and backfill material are all listed in one report. As is, these data can only be used for claim analysis or reporting purposes. Moreover, potential information (I_5- I_9) and decisions (DM_1 - DM_9) that can be generated and supported using available data are missing or under-utilized.

Every highway construction project is unique with respect to project scope, location, site condition and weather conditions. These factors should be put into consideration to determine the relationship between each other and the effects of these factors on extracting potential information and decision-making such as production rate and contract time determination to make reliable estimates. The previous OTC project OTCREOS7-1-22 entitled "*Development of an Improved System for Contract Time Determination Phase – III*" showed that historical project records such as daily work reports (DWRs) along with soil data, average traffic information, and weather information collected from the Oklahoma Department of Transportation (ODOT) are potential sources in identifying critical factors that affect the production rates of major highway activities and developing production rate estimation models. These production rate prediction models are in turn inserted into the contract time determination system to more accurately determine contract time of future highway projects. The system allows managers to realistically and effectively estimate project duration and minimize the inconvenience of the general public due to unjustified lengthy duration of a project.

In the OTCREOS7-1-22 project, although data on highway construction projects are stored electronically in a database system, it was difficult to retrieve data and convert it to information due to its linguistic nature. Daily work reports were reviewed line by line to determine the quantity of work and durations for controlling highway activities. In addition, the construction crew size (equipment and manpower) were collected but they were not organized in a manner to track specific activities that can be utilized in determination of production rate or allocation of resources for future maintenance and rehabilitation projects. Previously, the ODOT used the production rate chart based on minimum, maximum and average values to estimate production rates which were not justifiable and heavily relied on the engineers experience and judgment. Therefore, a more reliable and accurate data collection system or database system should be implemented to easily retrieve information and queries. The use of the current systems such as SiteManager should be utilized beyond reporting purposes and making contractor payments. In addition, potential information generation methods and decision support systems should be applied to extract valuable information and knowledge to support critical decisions such as scheduling and contract decisions. An ideal data, information and decision-making integration is discussed in the next section.

During this active utilization of data and information, various information generation techniques can be applied to visualize and support highway decisions. Descriptive statistics such as histograms, scatter plot, pie-chart can be used to summarize reports to higher decision-makers in terms of identifying the amount of resource allocation (in terms of funding, crew and equipment), material analysis, difference in cost, and percentage of completion for project control and reporting purposes. Inferential statistics can be used to quantify the relationship between production rates and potential drivers or factors and estimate production rates (Smith, 1999, O'Connor and Huh, 2005, Chong, 2005, and Jiang and Wu, 2007). Neural Networks can be utilized in mapping environmental and managerial factors with productivity estimation and estimation of labor productivity (Chao and Skibniewski, 1994 and Abourizk et al. 2001). Optimization techniques such as simulation can be utilized to convert as-built information in bridge and highway design (in selection of pavement type, bridge span, etc.) and operation and maintenance phase (in selection of optimum treatment options, material and resource allocation, etc.). Other potential information generation methods include expert system and fuzzy set theory (Christian and Hachey 1995, El-Rayes and Moselhi, 2001 and Pan, 2005.

One of the primary reasons for the gap in the integration effort is the lack of skilled data analysts and/or experts to analyze data and convert these data into information and knowledge. There needs to be a clear path in assigning responsibilities in terms of collecting, analyzing and making decisions. For instance, certain data collection (pavement condition data), information generation (as-built information) and higher level decisions (fiscal planning, policy formulation, etc.) are heading towards the asset management program which is for the decisions in the planning and also operation and maintenance phases. Another potential reason is due to the fact that data collection efforts are part of a federal mandatory requirement to avoid multiple data entry, better communication and reporting rather than addressing decision-maker's need which lacks a well-developed requirement analysis and performance measures. The function of data should be well defined to meet the needs of the user. The user should be able to retrieve data in a timely manner to support its decision. In addition, transportation /highway agency departments focus on specific division or business process to promote their own division's need rather than developing an integrated system. Moreover, some of the collected data are still linguistic in nature, miss important data and/or are un-structured that makes it difficult for the user to quantify data to extract information. However, with incorporation of missing data (availability), active usage of data (utilization), collecting structured data (format), applying the right data analysis method (tool/technology) and use of data analyst (skilled staff) more information and knowledge can be extracted to support highway decisions. Table 4.3 summarizes the gap in terms of staff, function, time, technology, availability, format (structure), and technology or tools.

Table 4.2 Summary of Gap Analysis

Criteria	Gap
Staff	Need for data analyst or data scientist
	Need for responsible party in data collection, information generation and decision-making
Function	Need for decision-maker requirement, identifying characteristics and use
Time	Need for data and information to reach the user or decision-maker in a timely manner
Availability	Missing data and information
Format/Structure	Need for change of textual or linguistic data types, lack of standard
Individuality	Division having standalone units to match only particular needs
Technology	Need for appropriate tools and technology to extract information

4.2 Case Study 2 -Preconstruction Service Cost

4.2.1 Preconstruction Service Decision-Making Process

Various decisions are made during the early stage of planning and design phase which can impact the overall cost, schedule, performance and quality of a highway project. These decisions may include determination of feasibility options; development and budget comparison at strategic level; conceptual cost estimation and planning concepts at network level; identification of right-of-way and acquisition of permits such as NEPA at program level; selection of design work by in-house or outsourcing to consultants, consultant selection at project selection level; and design alternatives, project control and review of construction documents at project level. A typical preconstruction services decision hierarchy is shown in Figure 4-8.

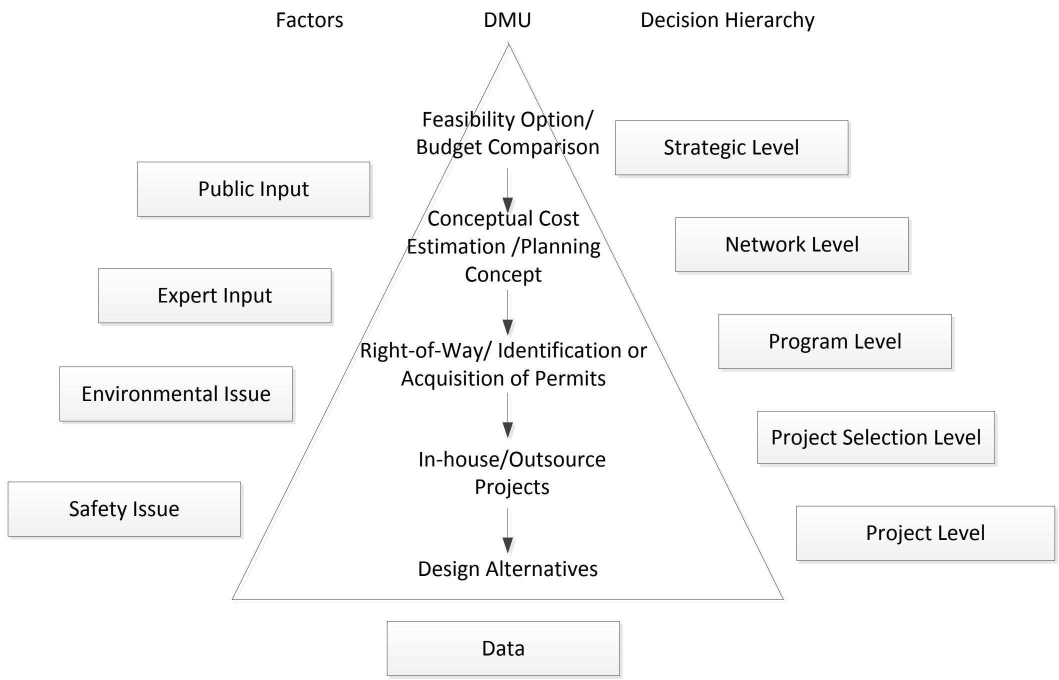


Figure 4-9 Preconstruction Service Decision Hierarchy

Of these key decisions preconstruction service or preliminary engineering (PE) cost estimation is taken as case study. Preliminary engineering cost estimation commences with the authorization of funding from federal or state agency to plan and design activities and ends with the delivery of project plans and specifications for project bid preparation. During this process, highway engineers need project data to predict reasonable preconstruction and/or preliminary engineering cost. In addition, they have to make decisions as to perform design works by in-house or outsource it to consulting firms to effectively utilize the allocated budget. Preconstruction service cost is part of project development and implementation process that involves various divisions including roadway, bridge, right-of-way, surveying, location and environment along with project management division. A typical project development process practiced by ODOT is shown in Figure 4-9.

Primarily, the planning division initiates a project by defining a project scope and conducting a conceptual estimate. The project management (PM) division will then perform the project charter process to obtain approval. Once a project is approved, surveying and environmental studies are conducted by the survey and the planning divisions respectively. Then, the roadway division develops a preliminary roadway plan and passes it to the bridge division for hydraulic analysis and setting bridge grade requirements. The bridge division then sends it back to the roadway division for setting the finished grade. Following the preparation of preliminary bridge and roadway plans, the PM division facilitates preliminary field plan review meetings with the preconstruction divisions. The division compiles logistic information and sends it to the right-of-way (ROW) and utility division. The PM division gives a four week notification prior to the preliminary field meeting.

Then, the ROW and utility division will compile and send the data to roadway and other divisions two weeks before the meeting. The preliminary field review meeting is intended to check environmental concerns, ROW needs, check alignments, verify project scope, etc. In the preliminary meeting, a total of 14 sets of plans are expected from the divisions. A preliminary estimate of earthwork and survey data sheets should also be presented in the meeting. If there is a need for new ROW, a separate meeting will also be conducted with a cost estimate. During this process, while the material division performs geotechnical study, the roadway designs the pavement. After the preliminary meeting, a draft agenda will be prepared by the PM for review and comments. It will then distribute a final agenda for a final plan field review meeting two weeks prior to the meeting. The meeting is intended to verify plan changes from previous meetings, discuss constructability issues, erosion control, construction sequence, etc. The final plan review meeting will be utilized by ODOT staff and consultants when all information is acquired. Finally, the preliminary engineering cost is estimated based on the amount of work effort put or engineering hours spent to conduct these tasks. Each division of ODOT performs its own study to meet project requirements and prepares an Excel sheet of contract fee proposal based on total hours required to develop the set of plans from preliminary stage to the final plan preparation.

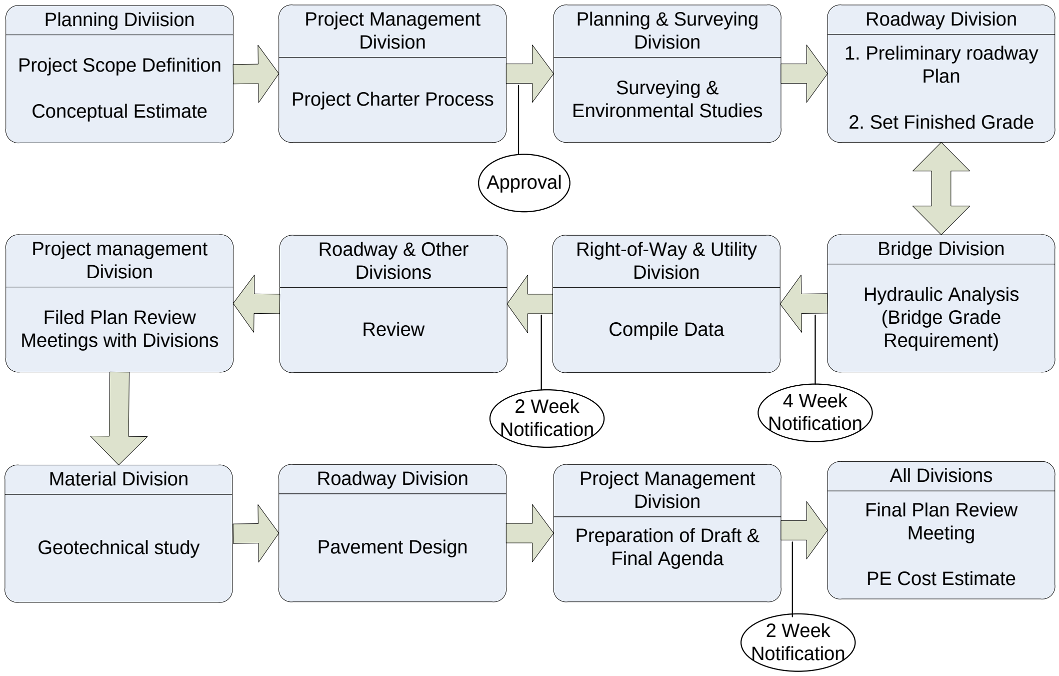


Figure 4-10 ODOT Project Development Process

Each division is responsible for making the decision on performing design tasks in-house or outsourcing to consulting firms. For instance, the environmental division performs approximately 80% of their environmental decisions by in-house design team based on the type of project and environmental clearance. Overall, the ODOT outsources approximately more than 50% of their design works to consulting firms. Design contracts are let based on initial contracts and then enter into negotiation for final contract. Consultants are compensated on hourly basis or reimbursed on actual work hours for the design services. ODOT follows the Uniform Audit & Accounting Guide developed by the American Association of State Highway and Transportation Officials (AASHTO) for procedures in examining, auditing, and reporting costs that are incurred by architect/engineering (A/E) firms for engineering and design services to comply with federal regulations. Although one of the main reasons for outsourcing is the lack of in-house resources and/or expertise to design complex projects, presence of low quality cost data; minimal recognition of existing data; absence of well-defined procedure or mechanism to convert these data into information and knowledge to assist highway engineers in estimating PE costs contribute to an increased rate of outsourcing.

4.2.2 Database

Currently, the ODOT's accounting system does not have a function to control the engineering cost estimate with regards to allocating resources (time and money) for designing highway projects. This creates an obstacle to track engineering hours and the cost associated with a particular project for in-house design as an engineer may be involved on multiple projects at a particular time. However, for a project outsourced to consulting firms, highway divisions have in-house developed excel spreadsheets, called contract fee proposal sheet to estimate the preliminary engineering costs. The spreadsheet is developed based on the amount of work effort (total engineeering hours) required to develop the set of plans from the preliminary stage to the final plan preparation for the purpose of negotiating contracts with consulting firms. The spreadsheet consists of a cross tab of seven main plan development activites, a detailed list of tasks and sub-tasks along with a skilled labor category. ODOT engineers use this spreadsheet to estimate and match the work efforts required by each engineer for each task based on the amount of sheets required for each task and project length by comparing it with similar, previous highway projects.

For instance, the tasks associated with preparing a preliminary roadway plan includes creating title sheet and location map, drafting a typical section, developing plan and profile sheets, designing drainage structures, developing finished grade line, designing super-elevation, and developing the preliminary construction sequence. These tasks are further broken down into sub-tasks to estimate the amount of engineering hours required by the skilled laborer (project manager, project engineer, senior engineer, design technician, Computer Aided Design [CAD] technician, and clerk). The sub-tasks in developing the plan and profile sheets include preparing survey files, generating horizontal alignment, generating existing ground and profile, and generating and drafting plan and profile sheets. Once engineering hours are calculated, they are multiplied by the respective labor rate to obtain preliminary engineering cost.

The Oklahoma Department of Transportation (ODOT) does not have a specific database to support its decision in estimating preliminary engineering costs except the contract fee proposal spreadsheets mentioned above. Primarily, these preliminary engineering cost data are partially stored electronically on each division engineer's personal computer, but the majority of the project data are stored in hard copies (paper format) as part of the engineering contract data.

Engineering divisions might utilize other division's database to extract data such as annual average daily traffic (AADT), route classification, terrain area type, etc. from planning and research division's roadway inventory database. A screenshot of a roadway division PE cost proposal spreadsheet is shown in Figure 4-10.

OKLAHOMA DEPARTMENT OF TRANSPORTATION
ROADWAY ENGINEERING CONTRACT COST PROPOSAL
8/28/2006

E.C. NO. ______ SWO NO. ______ JOB PIECE NO. ______
COUNTY ______ CITY ______ PROJECT NO. ______
HIGHWAY NO. ______ ORGANIZATION SUBMITTING PROPOSAL ______
PROJECT DESCRIPTION :

Labor Description	Labor Catagory: Project Manager	Senior Engineer	Project Engineer	Design Technician	CAD Technician	Clerical				
Labor Rate										
ROADWAY PLAN DEVELOPMENT	HOURS						Sub-Task	Task	Activity	
3.2 Prepare Preliminary Roadway Plans	104	406	682	680	1002	3			2877	$ 77,282
3.2.2 Create Title Sheet And Generate Location Map	2	8	16		38			64		
3.2.5 Draft Typical Section	4	22	60		108	2		196		
3.2.6 Develop Plan & Profile Sheets	92	244	386	284	520	1		1527		
3.2.6.2 Prepare Survey Files For Design	2	2	2	4	4	1	15			
3.2.6.3 Generate Horizontal Alignment	42	80	148	24	32		326			
3.2.6.5 Generate Existing Ground			4	16			20			
3.2.6.6 Generate Profile	44	140	184	48			416			
3.2.6.7 Generate P&P Sheets	2	4	20	64	164		254			
3.2.6.8 Draft P&P Sheets	2	18	28	128	320		496			
3.2.7 Design Drainage Structures	0	64	96	56	80	0		296		
3.2.7.1 Analyze Existing Drainage System		32	16				48			
3.2.7.2 Determine Size of Cross-Drain Structures		16	40	16			72			
3.2.7.3 Generate Drainage Map		16	40	40	80		176			
3.2.8 Develop Finished Grade Line	2	0	120	308	144	0		574		
3.2.8.1 Design Vertical Alignment				88	32		120			
3.2.8.2 Develop Templates			56	64			120			
3.2.8.3 Develop Preliminary Cross Sections	2		64	116	88		270			
3.2.8.5 Develop Preliminary End Areas And Volumes				40	24		64			
3.5.3 Design Superelevation	2	28	4	32	32			98		
3.7.1 Develop Preliminary Construction Sequence	2	40			80			122		
## Establish Right-of-Way Requirements	62	426	755	142	775	2			2162	$ 59,390
3.15.1 Finalize Horizonal and Vertical Alignments	12	72	72	72	72			300		
3.15.6 Develop R/W Submission Plans	44	324	653	40	697	2		1760		
3.15.6.4 Design Driveways	20	120	128	16	152	2	438			
3.15.6.5 Design Side Drains	6	24	149	8	77		264			
3.15.6.6 Finalize Roadway Drainage Structure Design	16	172	360		452		1000			
3.15.6.7 Design Special Drainage Structures	2	8	16	16	16		58			
3.18.2 Prepare Comparative Estimates for Pavement Des	6	30	30	30	6		102	102		
## Prepare Roadway Final Cross Sections	6	4	112	560	328	0			1010	$ 24,818
3.15.4 Develop Cross Sections	4	0	96	344	272	0		716		
3.15.4.1 Generate Cross Sections			20	104			124			

Figure 4-11 ODOT PE Cost Proposal Spreadsheet

4.2.3 Current Utilization of Preconstruction Data

Existing preconstruction data can be classified into cost data, contract data, functional data, design data (roadway, bridge, environmental, right-of-way and traffic) and outsourcing data. Most of these data attributes are structured data types with a combination of interval, nominal and ordinal variables. The majority of these preconstruction data are utilized in roadway design (I_1). A fair amount of the functional data (D_{18} –D_{25}) and bridge data (D_{37} –D_{44}) are used in bridge design (I_2). Almost all of the cost data (D_1 –D_9) are utilized either for reporting costs (I_4) and controlling schedule (I_5). Project length (D_{14}) and average daily traffic (D_{49}) are the most utilized data attribute. Functional data (D_{18} –D_{25}), project type (D_{10}), type of work (D_{11}) and cost data (D_1 –D_9) are relatively utilized in information generation for roadway (I_1) and bridge design (I_2). Some of the data types that are existing but not utilized include contract data such as contract let year, fund type, division, consulting firm and route type (D_{12} –D_{17} except D_{14}). In addition, no type of data with regards to outsourcing (D_{54} - D_{61}), such as work volume, time constraints, or planning difficulty level is not collected as part of the preconstruction data system. Table 4-3 shows current preconstruction data utilization.

Table 4-3 Current Utilization of Preconstruction Data

Type of Data	Data Attributes	Data	Description	Data Type	No Use	Current Use				
						Roadway	Bridge Analysis	Environmental	Cost Reporting	Schedule Contr.
						I1	I2	I3	I4	I5
Cost data	Pay Period	D1	xx/xx/xxxx date	Numeric : Ordinal					X	X
	Labor Hours	D2	Number of Hours	Numeric : Interval					X	X
	Vehicle Miles	D3	Mileage	Numeric : Interval					X	X
	Labor Dollars	D4	Amount of Dollar	Numeric : Interval					X	X
	Vehicle Dollars	D5	Amount of Dollar	Numeric : Interval					X	X
	Personal Expense	D6	Amount of Dollar	Numeric : Interval					X	X
	Unit Price	D7	Cost in Dollar	Numeric : Interval					X	X
	Plan Sheets	D8	Number of Sheets/plans	Numeric : Interval					X	X
	Type of Sheet	D9	Plan & Profile, drainage, etc.	Characters :Nominal					X	X
Contract Data	Project Type	D10	Replacement; Interchange New Construction, etc.	Characters :Nominal		X	X			
	Type of Work	D11	Bridges & Approaches; Grade & Drain, etc.	Characters :Nominal		X	X			
	Let Year	D12	Years xxxx - yyyy	Numeric : Interval	X					
	Fund Type	D13	SSP; STPY; BRFY, etc.	Characters :Nominal	X					
	Project Length	D14	z - miles	Characters :Interval		X	X	X	X	X
	Division	D15	Geographical Division 1 , 2, 3, etc.	Numeric : Ordinal	X					
	Consulting firm (CS)	D16	CS_1; CS_2; CS_3; etc.	Characters :Nominal	X					
	Route Type	D17	SH; I; US: City Street	Characters :Nominal	X					
Functional	Area Type	D18	Rural; Urban; Suburban	Characters :Nominal		X	X			
	Terrain Type	D19	Rolling; Flat	Characters :Nominal		X	X			
	Highway Type	D20	Collector; Principal arterial;	Characters :Nominal		X	X			
	Highway	D21	NON-NHS; NHS	Characters :Nominal		X	X			
	Access Control	D22	Full, Partial, None	Characters :Nominal		X	X			
	Vertical Alignment	D23	"K" values			X	X			
	Horizontal Alignment	D24	Degree of curve	Numeric : Interval		X	X			
	Superelvation	D25	X ft	Numeric : Interval		X	X			
Roadway	Pavement Type	D26	Asphalt; Concrete; Asphalt concrete; etc.	Characters :Nominal		X				
	Shoulder Type	D27	Sod; Asphalt; Concrete;	Characters :Nominal		X				
	Number of Lanes	D28	2; 3; 4; 6; 8	Numeric : Interval		X				
	Lane Width	D29	10'; 11'; 12'	Numeric : Interval		X				
	Shoulder Width	D30	2'; 4'; 5'; 6'; 8'; 10';	Numeric : Interval		X				
	Alignment	D31	Existing, New located; Offset; Parallel Lanes, etc.	Characters :Nominal		X				
	Section	D32	2D; 4D: 2L; 4L; 5O; 6L; 7L; 8L	Characters :Nominal		X				
	Typical Section	D33	Open section; Curb & Gutter; Combination	Characters :Nominal		X				
	Storm Sewer	D34	Yes (0); No (1)	Characters :Dummy		X				
	Sidewalks	D35	Yes (0); No (1)	Characters :Dummy		X				

Type of Data	Data Attributes	Data	Description	Data Type	No Use	Current Use				
	Detour	D36	Closed route; Closed signed route; phased; shoo-fly, etc.	Characters :Nominal		X				
Bridge	Bridge Number	D37	ID xxxxxxx	Numeric : Ordinal			X			
	Span	D38	X ft Y inch	Numeric : Ordinal			X			
	Sufficiency Rating	D39	0- 100	Numeric : Interval			X			
	Construction Year	D40	XXXX, built year	Numeric : Ordinal			X			
	Bridge Width	D41	X ft	Numeric : Interval			X			
	Bridge Length	D42	Y ft	Numeric : Interval			X			
	Clearance	D43	Z ft	Numeric : Interval			X			
	Load	D44	M ton	Numeric : Interval			X			
Environment	NEPA Document	D45	Linguistic	Text				X		
	Permit Type	D46	COE; OWRB; FAA; COE; OWRB; USACE; USCOE	Characters :Nominal				X		
Right-of-Way	ROW Requirement	D47	Yes (0); No (1)	Characters :Dummy		X				
	Utility Conflicts	D48	Yes (0); No (1)	Characters :Dummy		X				
Traffic	ADT	D49	Traffic Count	Numeric : Interval		X	X	X		
	New Guardrail	D50	Yes (0); No (1)	Characters :Dummy		X				
	End Treatment	D51	Yes (0); No (1)	Characters :Dummy		X				
	Highway Lighting	D52	Outside, median or no lighting	Characters :Dummy		X				
	Traffic Signals	D53	Yes (0); No (1)	Characters :Dummy		X				
Outsource	Critical time	D54	Number 1 - 10	Numeric : Ordinal	X					
	Work Volume	D55	Number 1 - 10	Numeric : Ordinal	X					
	Planning Difficulty	D56	Number 1 - 10	Numeric : Ordinal	X					
	Requirement of	D57	Number 1 - 10	Numeric : Ordinal	X					
	Amount of Inspection	D58	Number 1 - 10	Numeric : Ordinal	X					
	Degree of Labor	D59	Number 1 - 10	Numeric : Ordinal	X					
	Political reasons	D60	Number 1 - 10	Numeric : Ordinal	X					
	Quality of Service	D61	Number 1 - 11	Numeric : Ordinal	X					

Based on the analysis, the current preconstruction data usage is limited to roadway and bridge design, environmental approval, reporting cost reporting and schedule control. However, these data attributes have the potential to be used in the development of more advanced decision support system such as preconstruction cost prediction. Prior studies that focused on preconstruction management have noted the lack of predictive tools to estimate design costs (Knight and Fayek 2002). In addition, the current system does not allow agencies to perform effective analysis with respect to allocating resources (budget and skilled labor), estimating conceptual and construction cost, and bid analysis. There are still controversies and various studies conducted as to perform preliminary engineering designs by either in-house team or consulting firms to allocate appropriate PE funding. Sometimes, there are discrepancies in capturing the planning effort of highway projects that involve complex projects due to the fact that some projects are divided into major phases that span over three to ten years. Moreover, there is no model that is used to generate information or map the data, information and knowledge that would support user's decision-making.

4.2.4 Ideal Data-Information-Decision Making Integration Framework

A set of databases such as contract documents, roadway and bridge inventory, project cost data and traffic inventory can be utilized in developing an ideal data, information and decision-making path. In this case study, the raw data includes cost data, contract data, functional data, design data (roadway, bridge, environmental, right-of-way and traffic) and outsourcing data. Apart from the current usage of information listed through I_1 to I_5, estimation of engineering hours (I_6), number of sheets or plan required (I_7), cost model (I_8), traffic forecast (I_9) and project outsourcing analysis (I_{10}) are potential information that can support highway decisions. Key decisions that can utilize these information include allocating resources in terms of budget or assigning skilled labor to projects depending on the size, complexity etc. (DM_1), conceptual estimation for the purpose of "go" or "no go" decision (DM_2), and performing design in-house or outsourcing decision (DM_3) during the planning phase. During the design phase, multiple design decisions such as estimating preconstruction or preliminary engineering cost (DM_4), roadway design (e.g. selecting type of pavement (DM_5), bridge design (e.g. choosing bridge span right-of way approval, DM_6), environmental design (DM_7), traffic and lighting design (DM_8), surveying design (DM_9) and right-of-way approval (DM_{10}) decisions can be supported through the use of data and information. In addition, construction cost estimation (DM_{11}) and bid analysis (DM_{12}) for the purpose of contractor selection and negotiating costs can make use of the information during bidding phase.

i. Resource allocation (DM_1) – refers to but not limited to distribution of transportation funds or budget across projects, assignment and time allotment of skilled manpower and assignment of equipment in the planning of rehabilitation and new highway projects. During the planning phase, these decisions may be influenced by cost data, engineering hours and number of plan sheets to be developed.
ii. Conceptual estimating (DM_2) - refers to the estimation of project cost based on limited scope in making a "go" or "no go" decision as to perform the project or reject it. This decision may be affected by project data and information such as daily traffic, project type, unit price, and various design analysis such as traffic analysis and capacity analysis.

iii. In-house/outsourcing decision (DM_3) – refers to the decision made in the selection of design works to be performed by in-house or outsource it to consulting firms. This decision may be influenced by data such as cost, time constraint, project type and complexity of the project.

iv. Preliminary engineering cost estimation (DM_5) – includes estimation of preliminary engineering cost based on improved scope definition and project data acquired from preliminary survey and study.

v. Roadway design decisions (DM_5) - In the design phase, decisions such as selection and evaluation of alternatives such as pavement type (concrete, asphalt or a combination), shoulder type, pavement thickness, and other geometric decisions (number of lanes, width, median type, horizontal and vertical alignment, etc.) are performed to meet the level of service of roadway based on functional and design data.

vi. Bridge design decisions (DM_6) - include selection or identification of span length, width, and number of bridges required, etc. that may be optimized based on functional and design data.

vii. Environmental approval (DM_7) – refers to making “yes” or “no” decision in assessing environmental impacts to be considered significant or not.

viii. Traffic and safety design decisions (DM_8) - include choosing the right type of light fixtures, identifying the number of traffic signs and posts required and type of guardrail needed.

ix. Survey Design/Estimate (DM_9) - refers to decisions to estimate cost of survey, selection of resources (mobilizing equipment) and setting alignment.

x. Right-of-way approval (DM_{10}) - deals with selecting optimal alignment that may consist of relocating environments and utilities that is cost effective and convenient to the public.

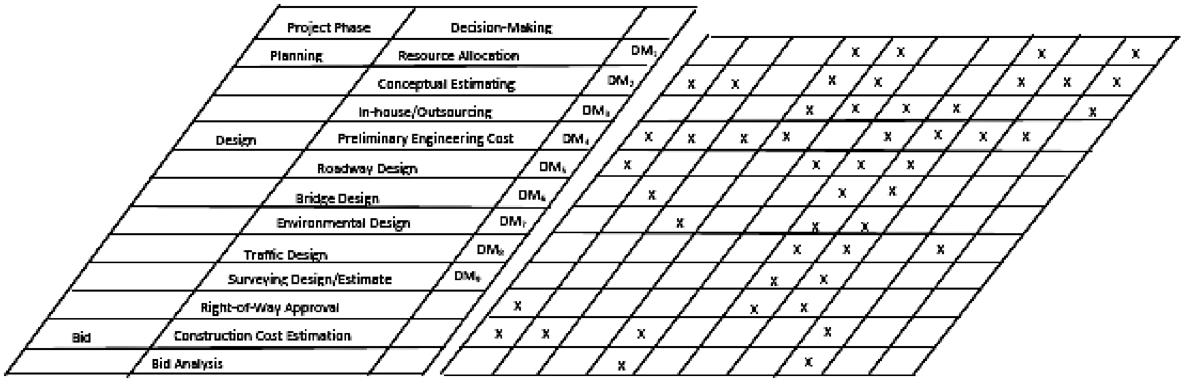
xi. Construction cost estimation (DM_{11}) – refers to detailed cost estimate based on final design data. It may be represented as bid estimate.

xii. Analysis of Bid (DM_{12}) – includes decisions as to the selection of contractors by setting criterion based on previous performance, cost data and scope of work. Table 4-5 illustrates an ideal data, information and decision-making using an input/output matrix.

Project Phase	Decision-Making		I1	I2	I3	I4	I5	I6*	I7*	I8*	I9*	I10*
Planning	Resource Allocation	DM_1				X	X			X		X
	Conceptual Estimating	DM_2	X	X		X	X			X	X	X
	In-house/Outsourcing	DM_3				X	X	X	X			X
Design	Preliminary Engineering Cost	DM_4	X	X	X	X		X	X	X	X	
	Roadway Design	DM_5	X				X	X	X			
	Bridge Design	DM_6		X				X	X			
	Environmental Design	DM_7			X			X	X			
	Traffic Design	DM_8						X	X		X	
	Surveying Design/Estimate	DM_9						X	X			
	Right-of-Way Approval		X					X	X			
Bid	Construction Cost Estimation		X	X		X				X		
	Bid Analysis					X				X		

Type of Data	Data Attributes	Data	Roadway Analysis (I1)	Bridge Analysis (I2)	Environmental Approval (I3)	Cost Reporting (I4)	Schedule Control (I5)	Engineering Hours (I6*)	Number of Sheet (I7*)	Cost Prediction (I8*)	Traffic Forecast (I9*)	Project Outsourcing (I10*)
			Information									
Cost data	Pay Period	D1				X	X	X	X	X		X
	Labor Hours	D2				X	X	X	X	X		X
	Vehicle Miles	D3				X	X	X	X	X		X
	Labor Dollars	D4				X	X	X	X	X		X
	Vehicle Dollars	D5				X	X	X	X	X		X
	Personal Expense	D6				X	X	X	X	X		X
	Unit Price	D7				X	X	X	X	X		X
	Plan Sheets	D8				X	X	X	X	X		X
	Type of Sheet	D9				X	X	X	X	X		X
Contract Data	Project Type	D10	X	X				X	X	X	X	X
	Type of Work	D11	X	X				X	X	X	X	X
	Let Year	D12						X	X	X	X	X
	Fund Type	D13						X	X	X	X	X
	Project Length	D14	X	X	X	X	X	X	X	X	X	X
	Division	D15						X	X	X	X	X
	Consulting firm (CS)	D16						X	X	X	X	X
	Route Type	D17						X	X	X	X	X
Functional	Area Type	D18	X	X				X	X	X		
	Terrain Type	D19	X	X				X	X	X		
	Highway Type	D20	X	X				X	X	X		
	Highway Classification	D21	X	X				X	X	X		
	Access Control	D22	X	X				X	X	X		
	Vertical Alignment	D23	X	X				X	X	X		
	Horizontal Alignment	D24	X	X				X	X	X		
	Superelvation	D25	X	X				X	X	X		
Roadway	Pavement Type	D26	X					X	X	X		
	Shoulder Type	D27	X					X	X	X		
	Number of Lanes	D28	X					X	X	X		
	Lane Width	D29	X					X	X	X		
	Shoulder Width	D30	X					X	X	X		
	Alignment	D31	X					X	X	X		

Figure 4-12 Ideal Data, Information, Decision-Making Input /Output Matrix I



Type of Data	Data Attributes	Data	Information: Roadway Analysis I1	Bridge Analysis I2	Environmental Approval I3	Cost Reporting I4	Schedule Control I5	Engineering Hours I6*	Number of Sheet I7*	Cost Prediction I8*	Traffic Forecast I9*	Project Outsourcing I10*
	Section	D32	X					X	X	X		
	Typical Section	D33	X					X	X	X		
	Storm Sewer	D34	X					X	X	X		
	Sidewalks	D35	X					X	X	X		
	Detour	D36	X					X	X	X		
Bridge	Bridge Number	D37		X				X	X	X		
	Span	D38		X				X	X	X		
	Sufficiency Rating	D39		X				X	X	X		
	Construction Year	D40		X				X	X	X		
	Bridge Width	D41		X				X	X	X		
	Bridge Length	D42		X				X	X	X		
	Clearance	D43		X				X	X	X		
	Load	D44		X				X	X	X		
Environmental	NEPA Document	D45			X			X	X	X		
	Permit Type	D46			X			X	X	X		
Right-of-Way	ROW Requirement	D47	X					X	X	X		
	Utility Conflicts	D48	X					X	X	X		
Traffic	ADT	D49	X	X	X			X	X	X	X	
	New Guardrail	D50	X					X	X	X	X	
	End Treatment	D51	X					X	X	X	X	
	Highway Lighting	D52	X					X	X	X	X	
	Traffic Signals	D53	X					X	X	X	X	
Outsourcing	Critical time Constraint	D54*					X					X
	Work Volume	D55*					X					X
	Planning Difficulty Level	D56*					X					X
	Requirement of Special Skill	D57*										X
	Amount of Inspection Required	D58*										X
	Degree of Labor Intensity	D59*										X
	Political reasons	D60*										X
	Quality of Service	D61*										X

Figure 4-13 Ideal Data, Information, Decision-Making Input /Output Matrix II

Based on the framework, cost data (D_1 –D_9), contract data (D_{10} –D_{17}), functional data (D_{18} –D_{25}) and construction data (D_{49} –D_{53}) are the most probable data that can be used to generate information. Roadway (D_{26} –D_{36}), bridge (D_{37} –D_{44}), environmental (D_{45} –D_{46}) and right-of-way

(D_{47} –D_{48}) data can fairly be utilized in design analysis (I_1 –D_3) and estimating engineering hours, number of sheets and cost (I_6 –D_8). Almost all data attributes except outsourcing data can be utilized to develop prediction models for engineering hours, number of plan sheets and various cost models such as conceptual, preconstruction and construction cost estimating models. This aforementioned mapping is considered as inactive path due to the fact that data are stored in DOT systems, but they are not used in generating information and knowledge to support highway decisions. However, there is no existing path with regard to outsourcing data (D_{54} - D_{61}) such as time constraint, work volume and requirement of special skill, etc. that has the potential to develop project outsourcing analysis along with cost data (D_1 – D_9) and contract data (D_{10} - D_{17}) such as available fund, project cost, project type, etc. A data collection system could be developed to collect project outsourcing data using a structured domain along with its data attributes based on previously completed highway projects. It is important to note that the current utilization of data discussed in Figure 4-12 and 4-13 is considered as active paths as they are used to generate information and support decisions.

At present, there is no defined path to link the information with decision-making although some of the information is used to support decisions in informal manner. Based on the analysis, engineering hours (I_6) and number of plan sheets (I_7) can be utilized to support majority of the decision- making listed in this study. Multiple highway divisions can use these pieces of information to allocate their resources in terms of budget and personnel for each project, track the cost, control quality and decide as to perform the project in-house or outsource it to consulting firms (DM_1 and DM_3 - DM_{10}). Roadway analysis (I_1), bridge analysis (I_2), cost reporting (I_4), scheduling (I_5) and cost prediction models (I_8) also support various decision-making. For instance, cost prediction models can be used to support resource allocation (DM_1), outsourcing decision (DM_2), conceptual estimation (DM_2) during the planning phase, preconstruction or preliminary engineering cost estimation (DM_2) during design phase, construction cost estimation (DM_1) and analyzing bid (DM_1) during bidding phase. Overall, resource allocation, various cost estimation and outsourcing decisions have the most probable potential to utilize preconstruction information.

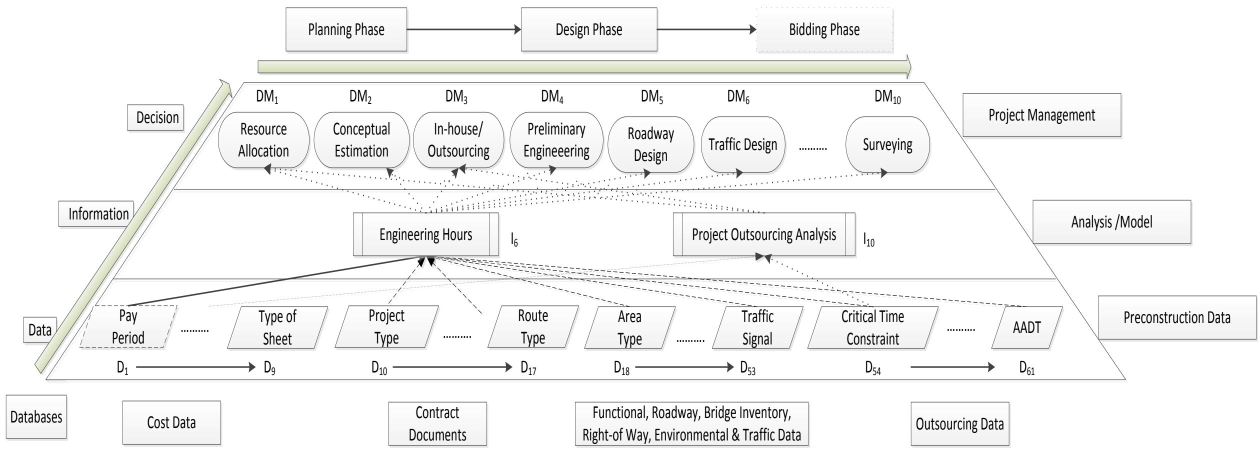


Figure 4-14 Three-Tiered Preconstruction Decision-Making Framework

As it is difficult to list all data, information and decision-making in this process, the application of the three-tiered framework discussed in the above paragraphs is shown as an example in Figure 4-11. It should be noted that solid lines represent the active paths, while the broken lines represent the inactive paths and the dotted lines show non-existing paths. In addition, these data, information and decision-making are not a comprehensive list, but rather illustrate the potential integration of these three entities through examples. It is also important to note that some of the data can be used as information without being processed. Furthermore, proper data analysis method and/or information generation technique and requirement analysis should be applied to address users' need, develop standard performance analysis or prediction models to support highway decisions. Previous chapter addressed potential qualitative and quantitative methodologies such as ranking and weighted score methods, statistical methods (regression models), artificial intelligence tools (neural networks), simulation models to develop preconstruction or preliminary engineering cost prediction models, select in-house/outsource design decisions and obtain optimized design decisions

4.2.5 Gap Analysis

Active utilization of preconstruction data can help DOTs reduce the risk of going under or over budget, selecting the right design alternative, constructing and completing projects on time with higher quality and minimum errors. It allows DOTs implement and analyze various scenarios to allocate resources efficiently and manage key decisions. This efficient management and reliable decisions translate into higher accountability of project developments and the overall highway program in terms of reducing project delays and inconvenience to the public by meeting public needs and agency's goal. However, current preconstruction data usage is minimal with respect to information and knowledge conversion and supporting key decisions. Preconstruction data are stored partially as individual projects in excel sheet spread across highway divisions and mostly in paper formats as part of engineering contract documents.

These contract documents consist of potential data in terms of scope definition, project attributes, cost data and work tasks that should be converted from linguistic format into digital and structured format. In addition, data such as NEPA documents should also be converted to structured format. Outsourcing data such as time constraint, difficulty level of project and quality of service, etc. are missing which can be utilized in the process of selecting a design firm. Contract data such contract let year, division or location of a project, etc. should be utilized in information generation. Important information such as estimation of engineering hours, number of plan sheets, various cost estimates should be generated using available preconstruction service data and application of the right type of tool to support decisions like resource allocation, various design decisions and bid analysis over the life-cycle of a highway project. Figure 4-12 shows a gap analysis between current and ideal preconstruction service data and information utilization in supporting highway decisions.

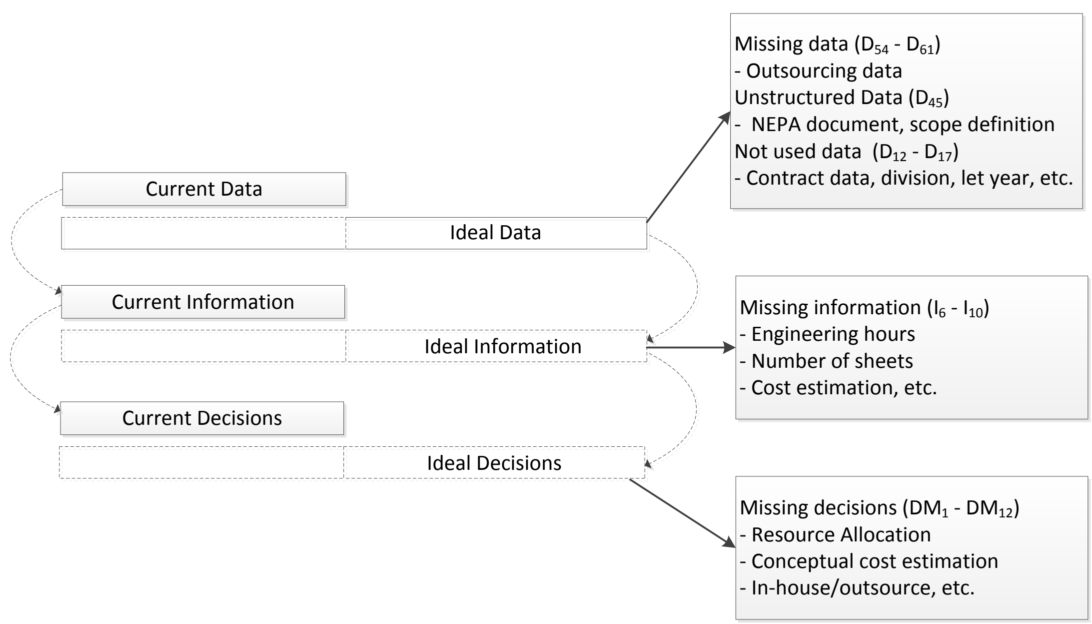


Figure 4-15 Gap Analysis

The previous OkTC project, OTCREOS10.1-19 entitled "*Procedures and Models for Estimating Preliminary Engineering Costs of Highway Projects",* showed the use of project attributes in developing quantitative prediction models of preliminary engineering (PE) costs for Oklahoma roadway projects. The study used data mining techniques to develop and compare three models as information generation methods, regression models, decision-tree model and neural network model. The study developed three types of models that consisted of a) engineering hours required per number of sheets, b) number of sheets, and c) cost per engineering hours for selected major plan development task outputs. Based on the decision tree models, the project length has a dominant effect on the majority of the PE cost components serving as a root node and decision node in the models. Route type has a fairly significant effect on decision making as compared to project type, fund type, consulting firm, and contract let year. Project type affects development of traffic sheets, cross-section sheet, summarizing sheets, and detail sheets. The contact let year has an effect on developing plan and profile, summarizing, construction sequence, and typical section sheets work effort. The type of consulting firm is important to typical section, drainage sheets, and summarizing sheets; while fund type is important to construction sequence, mass diagram, and summarizing sheets.

Based on the regression models developed, logarithmic and polynomial functions better fit the training data. The regression models showed that there is a steep increase of PE cost followed by a gradual slope with an increase in project length with regard to route type and project type. Although there is an overlap of PE costs between the project types for the first one mile of a project with widen/ reconstruct/ interchange type of projects having higher PE costs, there is an increase of PE cost in the order of bridges and approaches, widen/reconstruct/ interchange, and grade, drain, bridge, and surface type projects. According to ODOT's geographical classification, Divisions 8, 3, 4 and 1 have higher PE costs as compared to Divisions 2, 5, 7, and 6. In addition, PE costs tend to be higher for projects that encounter interstate highways as compared to US highways and state highways. Based on the neural network models, project length affects most of the plan development tasks, while route type has a fairly significant importance to most tasks. Project type affects traffic sheets, cross-section sheet, summarizing sheets, and detail sheets. The contact let year has an effect on developing plan and profile, summarizing, construction sequence, and typical section sheets work effort. The type of consulting firm is important to typical section, drainage sheets, and summarizing sheets; while fund type is important to construction sequence, mass diagram, and summarizing sheets.

Overall, project length is identified as the most significant factor affecting PE cost components using all three models. In addition, project type, fund type, and route type greatly affect PE plan development task outputs. Based on the decision tree and neural network models, representative models could not be developed for pay-item quantities and storm control plan development task outputs. A comparison of the developed models resulted in neural network models to outperform regression models and decision tree models based on the validation average squared error (ASE). Although the decision tree models and regression models outperform neural networks for some plan development task outputs, the variation of errors is not significant. The developed models help engineers determine a reliable number of sheets, work effort (engineering hours) required per number of sheets, cost per engineering hour, and total cost for the selected plan development task output with respect to project length, location, route type, and project type. In addition, this system not only allows engineers to easily manipulate the requirements for a specific task, but also helps them configure contingencies, as to whether any of the entities are either under- or over-estimated. It also helps to identify if a misallocation of resources (engineering hours assigned to the respective skilled manpower or number of sheets assigned to each task) exists at a specific level especially when negotiating PE costs with consulting firms. However, it should be noted that the study used 7 boxes of engineering contract documents to collect preliminary engineering cost data was time-consuming and rigorous process. A continuous use of the current system may result in data inconsistencies, missing important cost data and difficulty in retrieval and utilization of data to support preliminary design cost decisions. Therefore, an improved data collection system and better information generation analysis methods should be applied to make use of the collected data for better decision–making and move towards an integrated data and information era.

4.3 Case Study 3 – Pavement Management

4.3.1 Decision Making Process

One of the key components in a highway project life cycle is pavement management performed during the operation and maintenance phase. The aim of pavement management is to preserve, rehabilitate, reconstruct, and/or maintain the highway network to increase road users' safety and meet transportation goals. In this phase, highway agencies and department of transportations' (DOTs) perform multiple decisions at various levels of their program in order to sustain the network. Some of the major key decisions range from evaluation of system performance to setting policies and objectives by commissioners and governors at strategic level, 3R (Rehabilitation, Restoration and Resurfacing) fund distribution to allocation of budgets, forecasting short-term and long-range plans at network level, evaluation and prioritization of pavements/projects to treatment selection and reporting at project level by design office and district engineers.

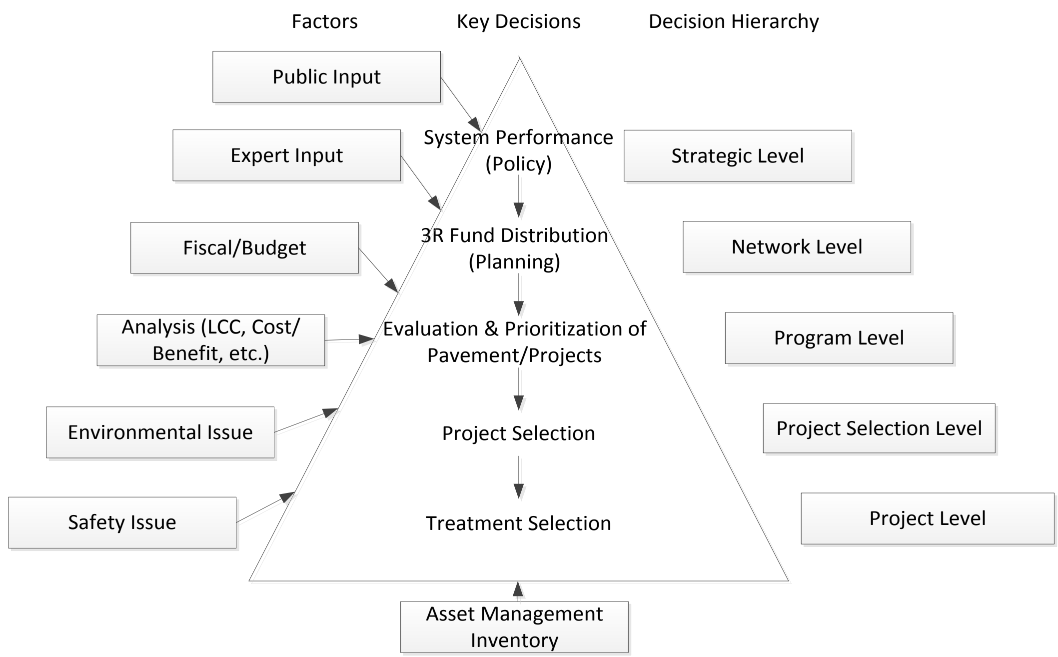


Figure 4-16 Pavement Management Decision Hierarchy

These decisions are supported through various data, information and knowledge that incorporate asset management inventory (condition data, structural history of pavement, traffic data and roadway data), external information (such as public input, expert opinion, safety data and environmental data) and performance analysis (life cycle cost analysis, cost/benefit analysis, deterioration curves, etc.) to address the socio-economic-environmental-political matters. Figure 4-13 shows a pavement management decision hierarchy based on top-down approach. A typical pavement management decision making process commences with identification of potential highway projects based on district staff and the public input. A list of Rehabilitation, Restoration and Resurfacing (3R) and maintenance candidate projects (MP) are reviewed by the districts and central office to identify the rehabilitation or maintenance need of projects. For projects that require maintenance, the districts match project funds for the first year and set the remaining candidate projects for future maintenance. Accordingly, design and schedule of the projects will pursue for selected first year projects. For 3R projects, with the support of pavement condition data, pavement history, treatment cost and traffic data, etc. projects are prioritized for first and second year along with statewide 3R projects under the 4-year program. In this stage, data analysis is performed to prioritize projects and select the best treatment option. The program allows to rank potential projects along with treatment options for four years. An example of a pavement management decision-making process adopted from Iowa DOT is shown in Figure 4-14.

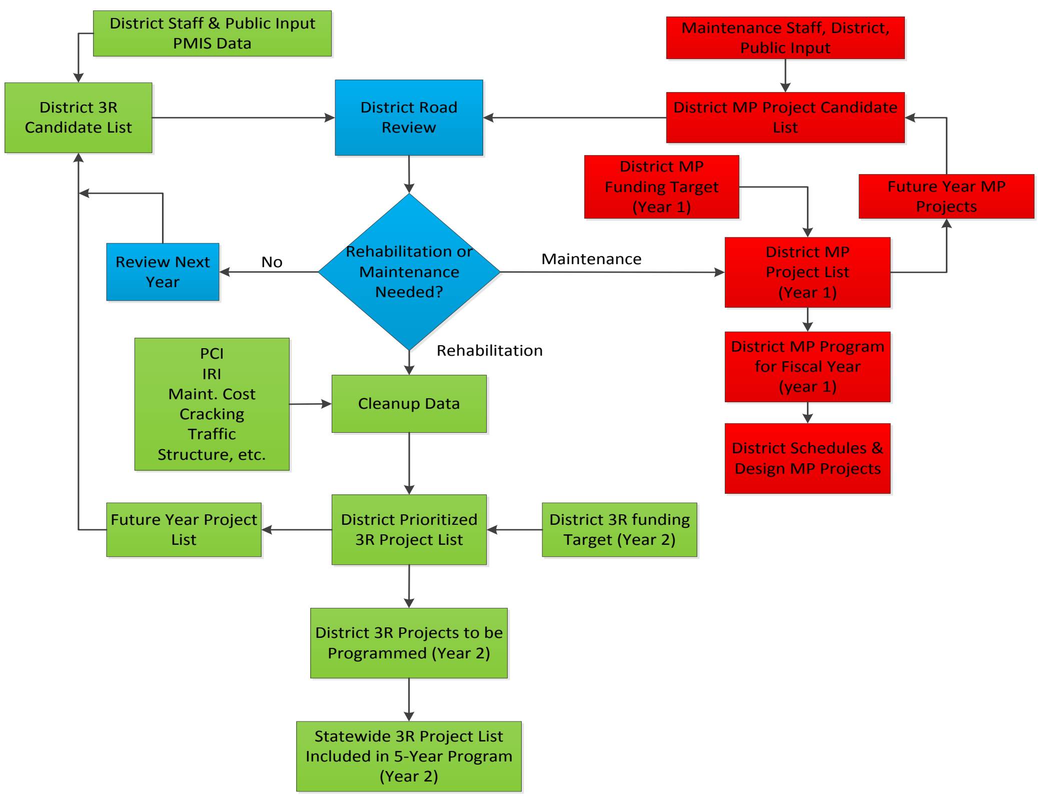


Figure 4-17 Example of Pavement Management Decision-Making Process

4.3.2 Database

Currently, the pavement management program along with other divisions collect and store a large amount of technical data and information as part of the asset management program. For instance, there are approximately 1.5 million pavement condition records in the Oklahoma Department of Transportation (ODOT) database. The program uses an automated data collection system to collect data on the condition of roads or distress data. It has established a contract with a consultant to collect network-level geometric and distress data. For example, the distress data collection is conducted on bi-annually basis for federal aid eligible roadways in the state. In the first year of cycle, data is collected for half of the state's distress data for the National Highway System (NHS) routes and approximately one-fourth of the non-NHS routes, while the remaining three-fourths of the non-NHS routes are collected in the second year of cycle.

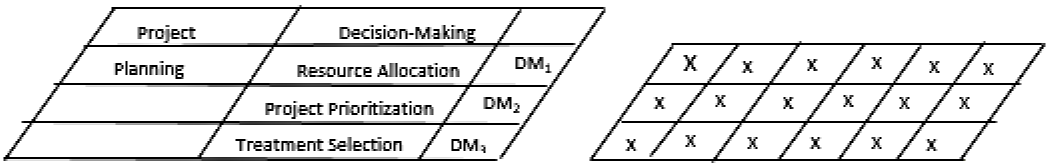
The pavement condition database is the primary source of data and/or information (key performance indicator) utilized in pavement operation and maintenance decision making such as treatment strategy selection. Pavement condition data takes functional and structural aspects of a pavement. The functional aspect considers pavement rutting, roughness, friction, ride quality, etc., while the structural aspect considers pavement distress data and stiffness such as longitudinal cracking, transverse cracking, fatigue (alligator cracking), etc. The functional data contains one record for each 100th mile of roadway surface condition. Structural data mainly contains one record per 100th mile of structural layer information while the analysis part contains condition indexes on structurally homogeneous segments or control sections of roadway.

In addition, pavement history, roadway inventory, and traffic data and/or information are used as supporting databases for pavement performance analysis. Pavement history database is used to learn and understand previous treatment applications in terms of pavement surface type, thickness, composition, and treatment cost. Roadway inventory database incorporates pavement classification, pavement type, section, length, width, etc., while traffic data incorporates the traffic profile or growth, annual average daily traffic (AADT), and traffic year to determine the structural capacity of existing and future pavement. Table 4-4 shows an example of pavement management database components. A detailed list of the major pavement distress data for various pavement types collected by Oklahoma DOT along with the data quality requirements are incorporated in Appendix C.

Table 4.4 Pavement Management Database

Division/ Source	Database	Sub-Elements	Collection Method
Pavement Management	Pavement History	Pavement surface type, thickness, composition, etc.	In-house - Automatic
	Distress Data	Longitudinal Cracking, Transverse Cracking, Patching, Spalling, Fatigue, etc.	Consultant
	Friction Data	Average Roughness, Ride, Average Rut etc.	Consultant
	Other (structural)	Deflectometer (FWD), ESAL	In-house
System Planning	Road Inventory, Traffic Data, 4-year program	Average Annual Daily Traffic (AADT), Functional Class, Right of Way, Route Classification, Terrain Area Type, etc.	Manual

Currently, it is important to note that most pavement management data are collected in digitalized format which can be transformed into different formats. Typically, the data are collected and stored in a relational database (Oracle) as a de-normalized data warehouse. The data can be exported or are compatible with various formats such as excel sheets (.xml), Arcview files (Shape), Google earth (.kml or .kmz formats), GIS and statistical software programs such as SAS. ODOT utilizes dTIMS™ pavement management software program to develop an inventory of physical assets and perform life cycle cost analysis to support their pavement management plan. dTIMS™ provides a generation of projects by year, recommended treatments by project and year, and overall summaries of condition, backlog, treatment cost, and treatment length. The program has the capability to do a network level analysis of the condition of the system and the funding levels needed to maintain status quo or improve the system. However, the program has been used in a limited extent for both of these purposes.



Type of Data	Data Attributes	Data	No Use	PCI	Condition Reporting	Needs Study	Sufficiency Rating	Cost/Benefit Analysis	Deterioration Curve
				Information					
				I1	I2	I3	I4	I5	I6
Distress	Transverse cracking	D1		X	X	X			X
	Fatigue cracking	D2		X	X	X			X
	Spalled joints	D3			X	X			X
	D-cracked joints	D4			X	X			X
	Corner breaks	D5		X	X	X			X
	AC patching	D6		X	X	X			X
	PC patching	D7		X	X	X			X
	Longitudnal cracking	D8		X	X	X			X
	Punchouts	D9			X	X			X
	Raveling	D10*	X						
	Bleeding	D11*	X						
	Surface Texture	D12*	X						
Structural Capacity	FWD	D13*	X						
Friction	Roughness (IRI)	D14			X	X			X
	Rutting	D15		X	X	X			X
	Faulting	D16			X	X			X
History	Age	D17			X		X	X	
	Grade	D18			X		X	X	
	Pavement Type	D19		X	X		X	X	
	Pavement Thickness	D20			X		X	X	
	Treatment Type	D21			X			X	
	Treatment Cost	D22			X			X	
	Shoulder Type	D23			X			X	
	Section	D24			X			X	
Roadway/Functional	Area Type	D25				X			
	Terrain Type	D26				X			
	Highway Type	D27				X			
	Highway Classification	D28				X			
	Number of Lanes	D29				X			
	Lane Width	D30				X			
	Shoulder Width	D31				X			
	Alignment	D32				X			
Traffic	AADT	D33				X	X	X	

Figure 4-18 Current Data Utilization

4.3.3 Current Data Utilization

The pavement condition data are relatively better collected, stored and utilized in generating information and supporting highway decisions as compared to construction daily work reports and preconstruction service data. Current pavement condition data are utilized in reporting conditions of pavements (I_2), assessing needs study (I_3), performing sufficiency rating (I_4), conducting cost/benefit analysis (I_5) and setting deterioration curves or developing models of pavement deterioration (I_6). This information is utilized in allocating budget (DM_1), prioritizing projects (DM_2) and selecting treatments (DM_3). Figure 4-18 shows current utilization of pavement management data. Based on the current trend, distress data (D_1 – D_9), friction data (D_{14} – D_{16}), history data (D_{17} –D_{20}) and traffic data (D_{33}) are the most utilized data in generating information. Potential distress data such as raveling, bleeding and surface texture (D_{10} – D_{13}) are missing.

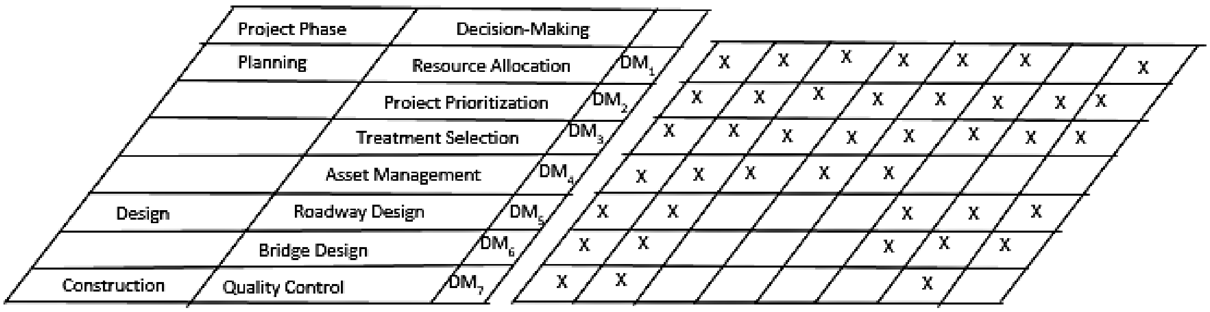
The analysis of current data utilization resulted in the majority of pavement condition data to be utilized in condition reporting (I_1) and needs study (I_1). A fair amount of the data is used in performing cost/benefit analysis, life-cycle cost analysis (I_5) and developing deterioration curves (I_6). Almost all the information generated are used in allocating resources (DM_1), prioritizing projects (DM_2), and selecting optimum treatments (DM_3). It is important to note that pavement condition index (PCI) is considered as information as it may incorporate various data to measure the performance of a pavement. The generation of information to support highway decisions, however, is far from complete.

4.3.4 Ideal Data-Information-Decision Making Integration Framework

Based on the developed three-tiered framework, an ideal data and information integration was applied to pavement condition data. Although pavement condition data are well utilized in supporting highway decisions, important patterns and discoveries can be extracted from this big amount of data. Potential analysis and use of pavement condition data include asset management, Highway Performance Monitoring System (HPMS) and calibration of Mechanistic-Empirical Pavement Design Guide (MEPDG) (Linda et. al, 2013). Some of the basic decisions made include:

i. Resource Allocation (DM_1) - deals with allocating resources between regions, districts and/or assets based on treatment scenarios and available budget.
ii. Project Prioritization (DM_2) – includes preparing and ranking projects to identify candidate projects an annual and multi-year basis in program development as part of a short term and long term planning process respectively.
iii. Treatment Selection (DM_3) – refers to decisions related to choosing treatment strategy, the timing and cost of the strategy for various segments of a highway.
iv. Asset Management (DM_4) – refers to managing and preserving highway agencies' assets in terms of safety and policy making considering high level decision.
v. Roadway design (DM_5) – refers to design decisions associated with selecting pavement type, shoulder type and thickness of pavement based on the condition and history of pavements.
vi. Bridge design (DM_5) – Similar to roadway design decisions associated with selecting bridge span, pavement type, and type of bridge structure the condition and history of bridges.
vii. Quality control (DM_7) - deals with measuring the performance of work tasks according to contracts and specifications.

Figure 4-19 shows an ideal framework using an input/output matrix analysis. Based on the analysis, more data may be utilized to generate information to support various decisions across a highway project life-cycle. For instance, the current data utilization of pavement condition index, PCI (I_1) as a product of functional aspects of the pavement is enhanced through the inclusion of structural capacity (D_{13}), skid resistance, roughness (D_{14} -D_{16}), and history such as the age and grade of a pavement (D_{17} – D_{19}). A well-developed PCI may in turn influence various decisions ranging from planning phase to construction and operation & maintenance. Distress data (D_1 – D_{13}), friction data (D_{14} – D_{16}), history data (D_{17} –D_{20}) and traffic data (D_{33}) may be used in performing material analysis (I_7) and developing treatment trigger mechanisms (I_8). PCI (I_1), condition reporting (I_2), deterioration curve (I_6), material analysis (I_7) and trigger mechanisms (I_8) have the most potential information utilization in supporting decisions. In addition to budget allocation (DM_1), project prioritization (DM_2) and treatment selection (DM_3), decisions such as asset management, roadway and bridge design and quality control (DM_4- DM_7) may make use of the information generated from pavement condition data.



Type of Data	Data Attributes	Data	PCI	Condition Reporting	Needs Study	Sufficiency Rating	Cost/Benefit Analysis	Deterioration Curve	Material Analysis	Trigger Mechanism
			I1	I2	I3	I4	I5	I6	I7*	I8*
Distress	Transverse cracking	D1	X	X	X			X	X	X
	Fatigue cracking	D2	X	X	X			X	X	X
	Spalled joints	D3	X	X	X			X	X	X
	D-cracked joints	D4	X	X	X			X	X	X
	Corner breaks	D5	X	X	X			X	X	X
	AC patching	D6	X	X	X			X	X	X
	PC patching	D7	X	X	X			X	X	X
	Longitudinal cracking	D8	X	X	X			X	X	X
	Punchouts	D9	X	X	X			X	X	X
	Raveling	D10*	X						X	X
	Bleeding	D11*	X						X	X
	Surface Texture	D12*	X						X	X
	Structural Capacity	D13*	X						X	X
Friction	Roughness	D14	X	X	X			X	X	X
	Rutting	D15	X	X	X			X	X	X
	Faulting	D16	X	X	X			X	X	X
History	Age	D17	X	X		X	X		X	X
	Grade	D18	X	X		X	X		X	X
	Pavement Type	D19	X	X		X	X		X	X
	Pavement Thickness	D20		X		X	X		X	X
	Treatment Type	D21		X			X		X	X
	Treatment Cost	D22		X			X		X	X
	Shoulder Type	D23		X			X		X	X
	Section	D24		X			X			
Roadway/Functional	Area Type	D25			X					
	Terrain Type	D26			X					
	Highway Type	D27			X					
	Highway Classification	D28			X					
	Number of Lanes	D29			X					
	Lane Width	D30			X					
	Shoulder Width	D31			X					
	Alignment	D32			X					
Traffic	AADT	D33			X	X	X		X	X

Figure 4-19 Ideal Data, Information and Decision-Making Input /Output Matrix

The three paths developed in the framework are shown using an example extracted from the input/output matrix (Figure 4-15). As stated earlier, pavement condition data incorporate functional data (rutting, roughness, etc.), structural data (cracking, patching, etc.), traffic data (AADT), pavement history, and cost data (treatment cost, budget, etc.). Pavement condition indexes or performance measure such as PCI (I_1), condition reporting (I_2) and various analyses such as cost/benefit analysis (I_3) and material analysis (I_7) are considered as information in this study. Major decisions made at current stage incorporate fund distribution or budget allocation (DM_1), project prioritization (DM_2), and treatment selection or improvement strategy (DM_3). The current data utilization explained in Table 4-7 are active paths. Data that are inactive or under-utilized incorporate distress data such as raveling, bleeding and structural data in developing a PCI. Information that are inactive includes Material analysis that can be performed based on type of distress data along with the type, age and grade of a pavement. Non-existing paths include either data that are missing to generate information or unknown information that can be utilized to support a specific decision. Some examples include the development of a material analysis to control quality of pavements in the construction phase or use it in selecting type or thickness of pavement in the design phase of a highway project.

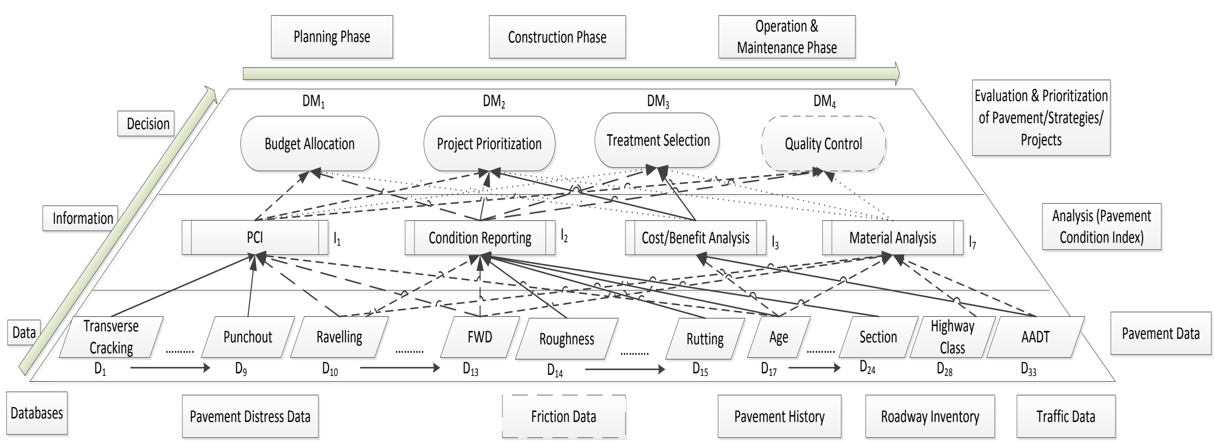


Figure 4-20 Three-Tiered Pavement Management Decision-Making Framework

4.3.5 Gap Analysis

Currently, ODOT utilizes the pavement management data in project selection, prioritization and treatment selection that ranges from pavement replacement to rehabilitation and maintenance projects. Information is well-generated as compared to daily work report (DWR) and preconstruction service data. In many cases, the pavement condition index (PCI) and the international roughness index (IRI) are used for comparing the performance of the various sections of roadways which in turn help prioritize sections of roadway and select an optimum type of treatments. There is however, still a gap in effectively utilizing data and converting them into information and knowledge to support highway decision-making compared to the amount of data collected every year. One of the reasons for this gap is the lack of integration between data, information and decision-making across various processes in project life-cycle which creates a problem in utilization of upstream data and information to be input for downstream decision-making and vice versa. A particular data that is neither converted to information and knowledge nor supports decisions in some way or another may be considered as a waste of resources.

Another reason is that current indexes do not use specific values that would lead to a “yes/no” decision on whether a treatment should be used. For instance, the accuracy of treatment selection criteria or treatment triggers developed based on expert judgment or visual measurement of performances cannot be justified. Do a PCI value greater than 50 and a pavement age greater than 10 years really indicate that a pavement needs to be replaced? Should additional data and/or information be incorporated for replacing a pavement? Does the PCI have any significant relationship with other pavement condition data or key performance indicators? At what level of rutting should a trigger be developed to recommend a treatment strategy? At what IRI level is it determined that the ride is poor, and rehabilitation is needed? Is there a level of cracking amount, severity, or a combination of both that would trigger different treatment strategies? These types of questions should be addressed and well defined to support decision making.

Detailed analysis of pavement data should be conducted to investigate the reactions and performances of different pavement types with respect to the design and use of materials. Different relationships and correlation of pavement condition data should be studied to characterize current conditions and predict future conditions. For instance, currently the pavement condition index (PCI) provides the condition of a pavement based on the collected distress data. However, the measure does not incorporate structural capacity and does not provide a measurement of skid resistance or roughness. This would enhance redundant data and decrease the correlations of various pavement condition data in predicting a reliable performance measure. In addition, the calibration of input models or treatment trigger mechanisms should be performed to find better performance models for each pavement family that will allow highway agencies to make more effective pavement management decisions.

The type of data needs by each decision-maker hierarchy should be well identified to meet data needs in generating information and knowledge. For instance, smoothness and distress data are collected more to support a network level decision-making process whereas structural capacity such as deflection and friction data are collected more at a project level. This will help in determining the type and quality of data requirements of the decision-maker. ODOT has an automated pavement data quality assurance system to evaluate the quality of data collected by consultants as part of its pavement management system. The program allows for accessing and changing the data in the database which helps to check the quality of data before making any decisions. Overall, the pavement management data are well-documented in terms of using databases and digital formats to extract information. But still, there is a need to map and extract patterns from the data and information to support specific decisions where most are inactive. Some of the data, information and decisions missing in the ideal framework are shown in Figure 4-21.

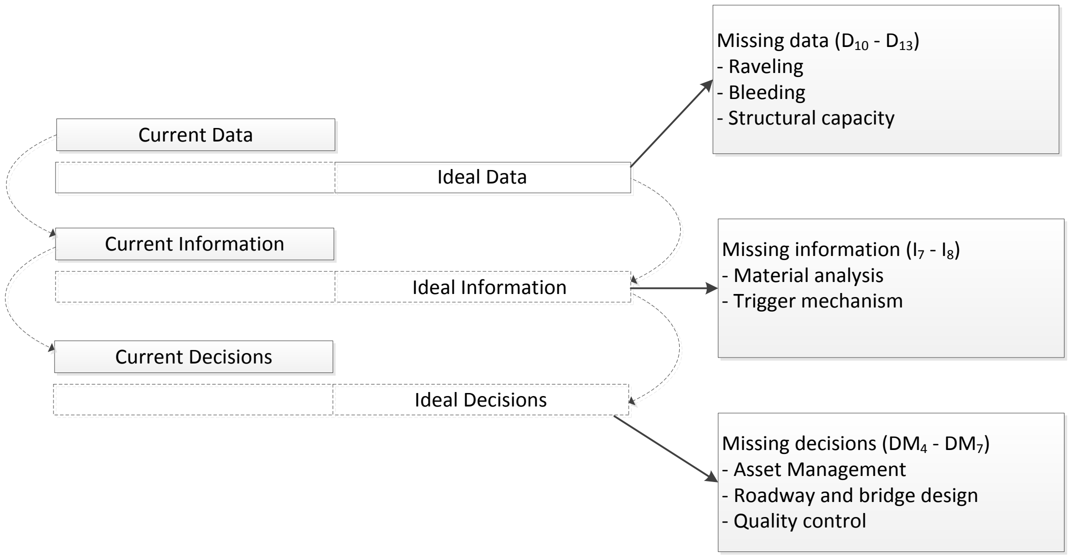


Figure 4-21 Gap Analysis

Therefore, based on the amount of data available in highway agencies, information generation analysis methods should be applied to identify the relationships and correlations between the various pavement condition data and determine appropriate selection criteria and treatment triggers to make more reliable decisions. Knowledge Discovery in Database (KDD) approach should be utilized to classify pavement families, assess the effectiveness of various treatment options, and develop deterioration curves. Better performance measure indicator models for PCI and IRI value can be developed based on statistical regression models or a clustering technique can be utilized to categorize similar condition data. A decision tree model can be developed using a data mining technique to determine at what age, PCI level, cracking level or roughness level a pavement needs replacement or rehabilitation. This allows highway agencies to calibrate controlling or better performing input models for a pavement management system. This helps in validating methods to generate information and knowledge from pavement condition assessment data and justify the benefits of pavement condition data collection efforts.

5. Conclusion

5.1 Summary

This study presented a three-tiered framework to integrate data, information and decision-making in highway projects. The Juran's triple role concept and context graph was used to illustrate the relationship between data, information and decision-making. The study also showed the complexity of data and information flow through a conceptual three dimensional data flow diagram with each axis representing a) level of decision-making, b) database utilization and c) project phase. The evolution of data and information integration in highway agencies was also outlined in the study. The study discussed potential methodologies in extracting information from raw qualitative and quantitative data. In addition, the study evaluated the framework through three case-studies, a) construction daily work reports, b) preconstruction cost estimation, and c) pavement management. In each case study, a gap analysis was performed by applying the three tiered framework to identify the gap between the current and ideal levels of data collection efforts and key decisions.

A combination of a top-down and a bottom-up approach was utilized in this study to link data, information and decision-making. Four divisions of key decision-making levels made throughout a project life-cycle from an asset management perspective were adopted in this study. These include strategic level, network level, program level, project level with the degree of data granularity increasing and the degree of decision-making decreasing accordingly. Data attributes were classified into four basic classification based on basic project management principle; time, cost, quality and safety and environment. In this study, information was considered as performance measures and/or analysis outputs as a result of one or more processed data. These data attributes, information, key decisions made across the highway project life-cycle were presented as examples along with responsible stakeholders through the decision hierarchy. Statistical methods and knowledge discovery in databases are the most influential data analysis methods to extract information and knowledge from large data sets to support highway decision-making.

A three-tiered framework was used to map data attributes, information and key decision-making. The framework used three types of paths to indicate existing data utilization (active path), available but underutilized data (inactive path) and non-existing path in generating information and knowledge and supporting highway decisions. Although there are some existing data, information and decision-making paths, a well-documented path supported with data analysis methods was missing. The study identified potential non-existing paths of decision-making that are not well-supported with data and information. Based on the study, preconstruction cost estimation, project outsourcing and resource allocation are some of the prominent decisions made during the preconstruction phase. These decisions can be supported by information like number of sheets, engineering hours, and project outsourcing analysis. In turn, the information or analysis can be supported by cost data, project attributes, traffic data obtained from the various databases (contract documents, roadway inventory cost database and traffic inventory) through the utilization of regression analysis, decision tree models, neural network models and qualitative ranking methods.

Contract time determination, project control, future design and resource allocation are some of the key decisions made during various phases based on construction data. These decisions can be supported by production rate identification, as-built analysis, and material analysis. Daily work reports such as reported quantity, pay items, unit cost data, and crew and equipment data along with project attributes obtained from contract administration software programs such as SiteManager are good sources of information. Pavement treatment selections, project prioritization, and improvement strategies are currently supported by various information such as pavement condition indexes (PCI) and international roughness indexes (IRI). However, these performance measurements are not well-supported or related with other pavement data attributes or condition data or performance indicators. The indexes do not use specific values that would lead to a “yes/no” decision on whether a treatment should be used. Therefore, appropriate data analysis methods such as regression models or decision tree models should be applied to identify the relationship between these performance indicators and data attributes obtained from the pavement management system and make reliable decisions. If there are existing models, calibration of input models should be performed to update the model with new data and check the relationship between data attributes and performance indicators.

Existing databases managed by highway agencies and/or departments of transportation have been summarized to analyze the data collection effort and current usage. It was identified that pavement management is the most utilized database management system whereas construction database is the least utilized. Especially, pavement condition data are currently being utilized in project prioritization and treatment selection process. The preconstruction database has also a relatively well-developed system compared to the construction database. However, there is a gap in terms of converting the collected data to information and knowledge which requires application of information generation or analysis methods.

The developed framework analyzes gaps in the current level of data usage and ideal situation in making better use of data. It shows what types of data should be collected, what methods can be used to convert the data into meaningful information and knowledge to support various highway project decisions. It will guide DOTs on how to generate and place right information and knowledge in the hands of decision makers. The framework will empower engineers to make informed and justifiable decisions, and lead to the improved accountability of project development and management. In addition, it will allow active utilization plan of currently existing databases and justify the current rate of return in the continuous and growing data collection efforts by DOTs. Furthermore, it allows DOTs a chance to measure their current performance and develop a new or revised data collection and information/knowledge generation scheme to support key decisions which historically were not well supported with information and data.

5.2 Recommendations

Currently, there is a huge amount of data collected and stored in state highway agencies. However, there is a limited and minimal use of these data and information in supporting highway decision-making. Based on the study, there is a huge gap in integrating data, information and decision-making. Sometimes, potential users of these data such as highway project schedulers, estimators, and managers do not even know what type of data is available or how to access these data to support their decisions. One of the primary reasons or difficulties in this integration effort is the lack of skilled data analysts and/or experts to analyze data and convert these data into information and knowledge. Today, various industries have data analysis departments to make use of available data and extract information to support their business decisions. Even though some data are well-documented in proper databases, but highway divisions are putting limited effort towards generating information from these data to support their daily decisions. State highway agencies may need to hire data experts and/or create an analyst team/department.

Another potential issue in integration efforts is the lack of database systems that can address potential users' need. Usually, data are collected not based solely on decision-makers' need and does not thoroughly consider the user's requirements. Standard data formats should be used that are compatible with various information generation tools utilized by different divisions or users to reduce data multiplication and to allow easier data and information exchange and extraction. Linguistic data types should be limited as these types will increase the inconsistencies and difficulties in the analysis stage. A panel of experts may be created to determine or set the criteria for collecting structured data attributes. A possible drop-down list of structured data attributes would not only save the agency's time, money and energy spent on data collection, but also make the extraction of information and knowledge easier for the user/decision-maker. Implementation of performance measurement tools would aid in evaluating the use of data and information utilization in terms of satisfying users' requirement. Therefore, a well-developed requirement analysis and comprehensive structured database system greatly influences information generation, knowledge retrieval.

In addition, highway agencies focus on development of specific or standalone programs for their specific division or business processes or project phase to efficiently manage and support their own division. However, it should be noted that these data and information can be utilized by the same or other divisions at a later stage during a project phase. For instance, a culvert as-built data during a construction phase would later be used during planning, design or maintenance phases by the design team, traffic and safety team, the bridge team and the maintenance team. These teams look for various data attributes as part of their data requirement. The design team needs the vertical and horizontal diameter of drainage structure, horizontal super elevation, horizontal and vertical curves, while the traffic and safety team need the location, pole mounting height and fixture type of a lighting or a signage. The bridge team needs the location of the culvert in terms of latitude, longitude, and mile post whereas the maintenance divisions look for the pavement type, shoulder type, width and number of lanes. Therefore, the requirements of these divisions should be identified, classified, and collected in terms of their data needs to minimize missing important data and satisfy all users need for future planning, designing, operation and maintenance decisions.

There is a lot of work to be done in developing an enterprise-based integration effort where potential users can benefit from data at any stage of project development. Future studies may need to be geared toward developing an enterprise wide ontology-based framework for the highway industry.

References

1. AASHTO (2009). “Estimation, AASHTO - Trns.port”, American Association of State Highway and Transportation Officials, Inc. and InfoTech.
2. AbouRizk, S., Knowles, P., and Hermann, U.R. (2001). “Estimating Labor Production Rates for industrial Construction Activities”, *Journal of Construction Engineering and Management*, 127 (6), 502-511.
3. Audit Commission (2007), “Improving Information to Support Decision Making: Standards for Better Quality Data”, www.audit-commision.gov.uk, accessed June 30, 2013.
4. Bao, Yang and Zhang, LuJing (2010). “Decision Support System Based on Data Warehouse.” *Journal of World Academy of Science, Engineering and Technology*, 71, 172-176.
5. Bhandari, I., Colet, E., Parker, J., Pines, Z, Pratap, R. and Ramanujam, K. (1997). “Advanced Scout: Data Mining and Knowledge Discovery in NBA Data”, *Data Mining and Knowledge Discovery*, Kluwer Academic Publishers, Boston, 121-125.
6. Björk, Bo-Christer (1999). ”Information Technology in Construction: Domain Definition and Research Issues”, *International Journal of Computer Integrated Design And Construction*, SETO, London, Volume 1, Issue 1, 1-16.
7. Boulicaut, J. F., and Jeudy, B. (2010). “Constraint-based data mining”, *Data Mining and Knowledge Discovery Handbook*, 339-354.
8. Cabena, P., Hadjnian, Stadler, Verhees, and Zanasi (1998). *Discovering Data Mining: From Concept to Implementation*, Int. Technical Support Organization (IBM), Prentice Hall, Upper Saddle River, N.J., 42–44.
9. Caldas, C., Soibelman, L., and Han, J. (2002). “Automated Classification of Construction Project Documents”, *Journal of Computing in Civil Engineering*, 16(4), 234-243.
10. Caldas, C., Soibelman, L., and Gasser, L. (2005). “Methodology for the Integration of Project Documents in Model-Based Information Systems”, *Journal of Computing in Civil Engineering*, 19(1), 25-33.
11. Chao, L, and Skibniewski, M.J. (1994). “Estimating Construction Productivity: Neural-Network-Based Approach”, *Journal of Computing in Civil Engineering*, 8(2), 234-251.
12. Chen, Z. (2001). *Data mining and Uncertain Reasoning: an Integrated Approach*, Wiley-Interscience, 1st Edition, New York, NY.

13. Chong, W. (2005). “Construction Production Rate Information System for Highway Projects”, Dissertation for Ph.D. at the University of Texas at Austin, May 2005.
14. Christian, J. and Hachey, D. (1995). “Effects of Delay Times on Production Rates in Construction”, *Journal of Construction Engineering and Management*, 121(1), 20-26.
15. Davenport T. H. and Harris, J. G. (2007). *Competing on Analytics: The New Sciences of Winning*, Harvard Business School Press, Boston, MA.
16. Deis, D., Watson, E., and Wilmot, C. (2002). “Designing a Comprehensive Model to Evaluate Outsourcing of Louisiana DOTD Functions and Activities,” Project Report prepared for Louisiana Department of Transportation and Development, Baton Rouge, LA.
17. Eastman, Chuck, Teicholz, Paul, Sacks, Rafael, and Liston, Kathleen (2011). *BIM Handbook, A Guide to Building Information Modeling for Owners, Managers, Engineers, and Contractors*, John Wiley and Sons, Inc., Hoboken, NJ.
18. Eger III, R., Knudson, D., Marlowe, J., and Ogard, L. (2002). “Report on Evaluation of Transportation Organization Outsourcing: Decision Making Criteria for Outsourcing Opportunities”, University of Wisconsin, Milwaukee. Available at http://www.wistrans.org/mrutc/research/completed-research/01-03/.
19. El-Rayes, K. and Moselhi, O. (2001). “Impact of Rainfall on the Productivity of Highway Construction”, *Journal of Construction Engineering and Management*, 127(2), 125-131.
20. El Wakil, E. and Zayed, T. (2012). “Data Management for Construction Processes Using Fuzzy Approach”, *Proceedings of the ASCE Construction Research Congress 2012: Construction Challenges in a Flat World*, American Society of Civil Engineers, West Lafayette, IN.
21. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). “From Data Mining to Knowledge Discovery in Databases,” *AI Magazine*, MIT Press, Cambridge, Mass 17(3), 37.
22. FHWA (2010). “Data Integration Primer,” U.S. Department of Transportation Federal Highway Administration, Office of Asset Management, Washington, DC.
23. Flintsch, G. and Bryant, J. (2006). “Asset Management Data Collection for Supporting Decision Processes”, U.S. Department of Transportation Federal Highway Administration, Office of Asset Management, Washington, DC.

24. Froese, T. (1992). “Integrated Computer-Aided Project Management Through Standard Object-Oriented Models”, Center for Integrated Facility Engineering, Technical Report No. 68A, Stanford University, Palo Alto, CA.
25. Halfawy, Hadipriono F., C., Duane, J. and Larew, R. (2005). “Development of Model-Based Systems for Integrated Design of Highway Bridges”, National Research Council, Canada.
26. Ittner, C.D. and Larker, D.F. (2006), “Moving from Strategic Measurement to Strategic Data Analysis” Chapman, C. S. (Ed.), Controlling Strategy, Oxford University Press.
27. Jiang, Y. and Wu, H. (2007). “Production Rates of Highway Construction Activities”, *International Journal of Construction Education and Research,* Vol. 3, 81-98.
28. Juran, J.M. (1988). *Juran on Planning for Quality,* Free Press, New York, NY.
29. Jang, H., Russell, J. and Yi, J. (2003). “A Project Manager’s Level of Satisfaction in Construction Logistics”, *Canadian Journal of Civil Engineering*, NRC, Canada, 1133-1142.
30. Kennerley, M. and Mason, S. (2008). “The Use of Information in Decision Making: Literature Review for the Audit Commission”, Centre for Business Performance, Cranfield School of Management.
31. Molenaar, K.R. (2005). “Programmatic Cost Risk Analysis for Highway Mega-Projects,” *Journal of Construction Engineering and Management*, American Society of Civil Engineers, 131(3), 343-353.
32. Nassar, K., Hegab, M., and Jack, N. (2005). “Design Cost Analysis of Transportation Projects” *Proceedings of the Construction Research Congress 2005: Broadening Perspectives*, American Society of Civil Engineers, San Diego, CA.
33. Neely, A. and Jarrar, Y. (2004). “Extracting Value from Data – The Performance Planning Value Chain’, *Business Process Management Journal*, 10(5), 506-509.
34. NIBS, http://www.buildingsmartalliance.org/index.php/projects/activeprojects/25, Accessed August 1, 2012.
35. Oberlender, Garold D. (2000). “Project Management for Engineering and Construction”, McGraw Hill, 2nd Edition, Boston, MA.
36. O’Brien, W., Issa, R., Hammer, J., M. Schmalz, Guenes, J. and Bai, S. (2000). “SEEK: Accomplishing Enterprise Integration across Heterogeneous Sources,” *ITCON-Electronic Journal of Information Technology in Construction: Special Edition on Knowledge Management*, 7(2): 101-124.

37. O'Connor, J. and Huh, Y. (2005). "Crew Production Rates for Contract Time Estimation: Bent Footing, Column, and Cap of Highway Bridges", *Journal of Construction Engineering and Management*, American Society of Civil Engineers, 131(9), 1013-1020.
38. Pan, N. (2005). "Assessment of Productivity and Duration of Highway Construction Activities subject to Impact of Rain", *Expert Systems with Applications*, Vol. 28, 331-326.
39. Rivard, H. and S.J. Fenves (2000). "A Representation for Conceptual Design of Buildings," *Journal of Computing in Civil Engineering*, American Society of Civil Engineers, 14(3), 151-159.
40. Rogge, D., Carbonell, T. and Hinrichsen, R. (2003). "Evaluation of Oregon Department of Transportation Project Delivery: Outsourcing Project Delivery in State Departments of Transportation" Report for SPR Project 351 for Oregon Department of Transportation and Federal Highway Administration, Corvallis, OR.
41. Saito, M., Sinha, K., and Anderson, V. (1991). "Statistical Models for the Estimation of Bridge Replacement Costs", *Transportation Research A*. Vol. 25A, Issue 6, 339-350.
42. Shen, Z., Issa, R., and O'Brien, W. (2004). "A Model for Integrating Construction Design and Schedule Data," *Proceedings of the Fourth Joint International Symposium on Information Technology in Civil Engineering*, 1-13.
43. Smith, D., and Tardif, M. (2009). *Building Information Modeling – A strategic Implementation Guide for Architects, Engineers, Constructors, and Real Estate Asset Managers*, John Wiley and Sons, Inc., Hoboken, NJ.
44. Smith, S., (1999). "Earthmoving Productivity Estimation Using Linear Regression Techniques", *Journal of Construction Engineering and Management*, American Society of Civil Engineers, 125 (3), 133-141.
45. Soibelman, L. and Kim, H. (2002). "Data Preparation Process for Construction Knowledge Generation through Knowledge Discovery in Databases", *Journal of Computing in Civil Engineering*, American Society of Civil Engineers, 16(1), 39-48.
46. Soibelman, L., Liu, L. Y., & Wu, J. (2004). "Data Fusion and Modeling for Construction Management Knowledge Discovery," International Conference on Computing in Civil and Building Engineering, Weimar, Germany.
47. Tribelsky, E. and Sacks, R. (2010). "Measuring Information Flow in the Detailed Design of Construction Projects", Springer-Verlag London Limited.

48. Williams, T., Parker, N., and Klotz, J. (2012). "Development of an Estimating Tool for Transportation Project Design Effort", Transportation Research Board, Washington, DC. In Press.

49. Woldesenbet, A. and Jeong, H.S. (2012). "Historical Data Driven and Component Based Prediction Models For Predicting Preliminary Engineering Costs of Roadway Projects." *Proceedings of the ASCE Construction Research Congress 2012: Construction Challenges in a Flat World*, American Society of Civil Engineers, West Lafayette, IN.

50. Yu, K., Froese, T., and Grobler, F. (2000). "A Development Framework for Data Models for Computer-Integrated Facilities Management," *Automation in Construction*, Vol. 9, 145-147.

Abbreviations

US DOT – United States Department of Transportation

FHWA – Federal highway Administration

AASHTO – American Association of State Highway and Transportation Officials

TRB – Transportation Research Board

FTA – Federal Transit Administration

NHTSA – National Highway Traffic Safety Administration

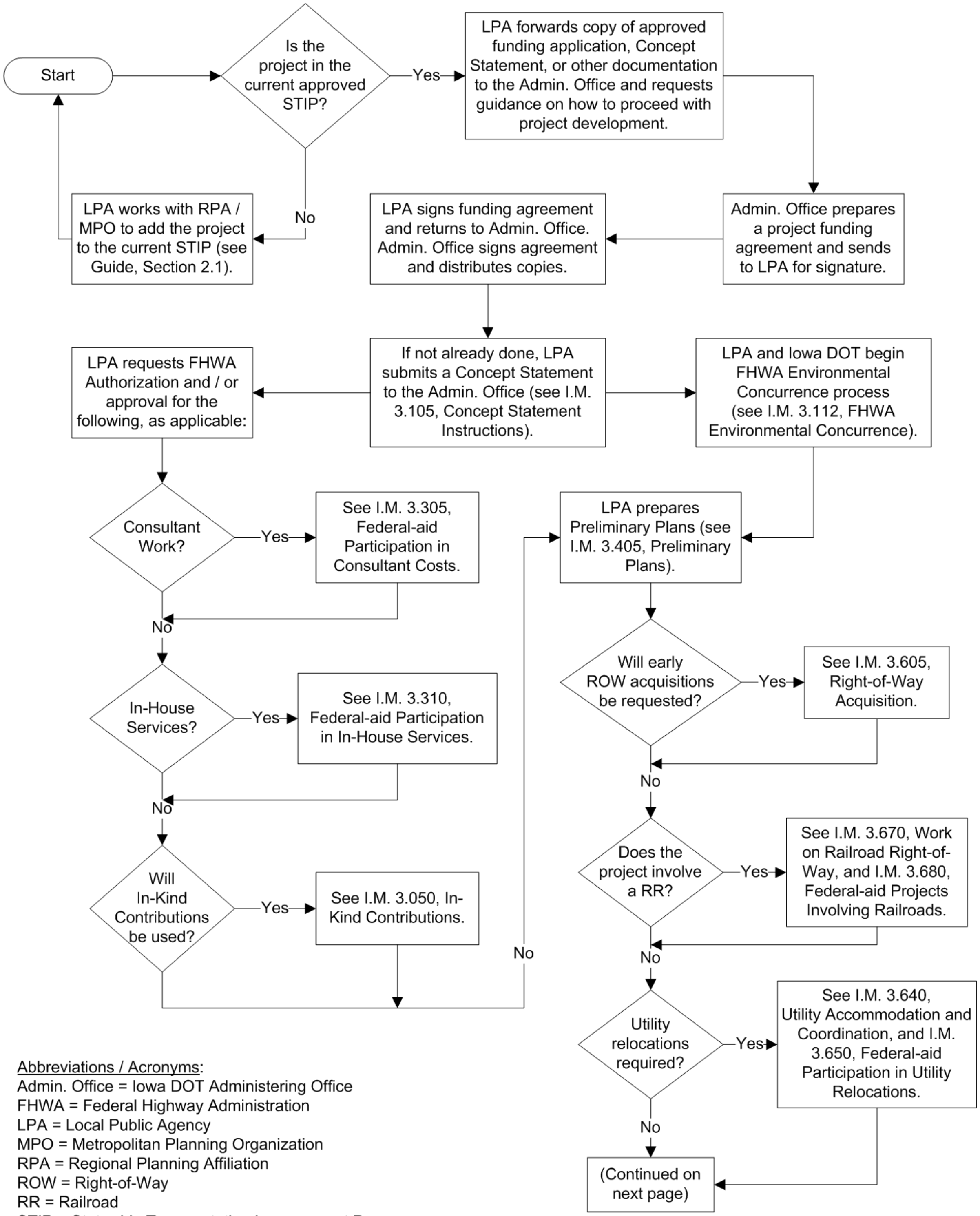
RITA – Research and innovative Technology Administration

MPO - Metropolitan Planning Organizations

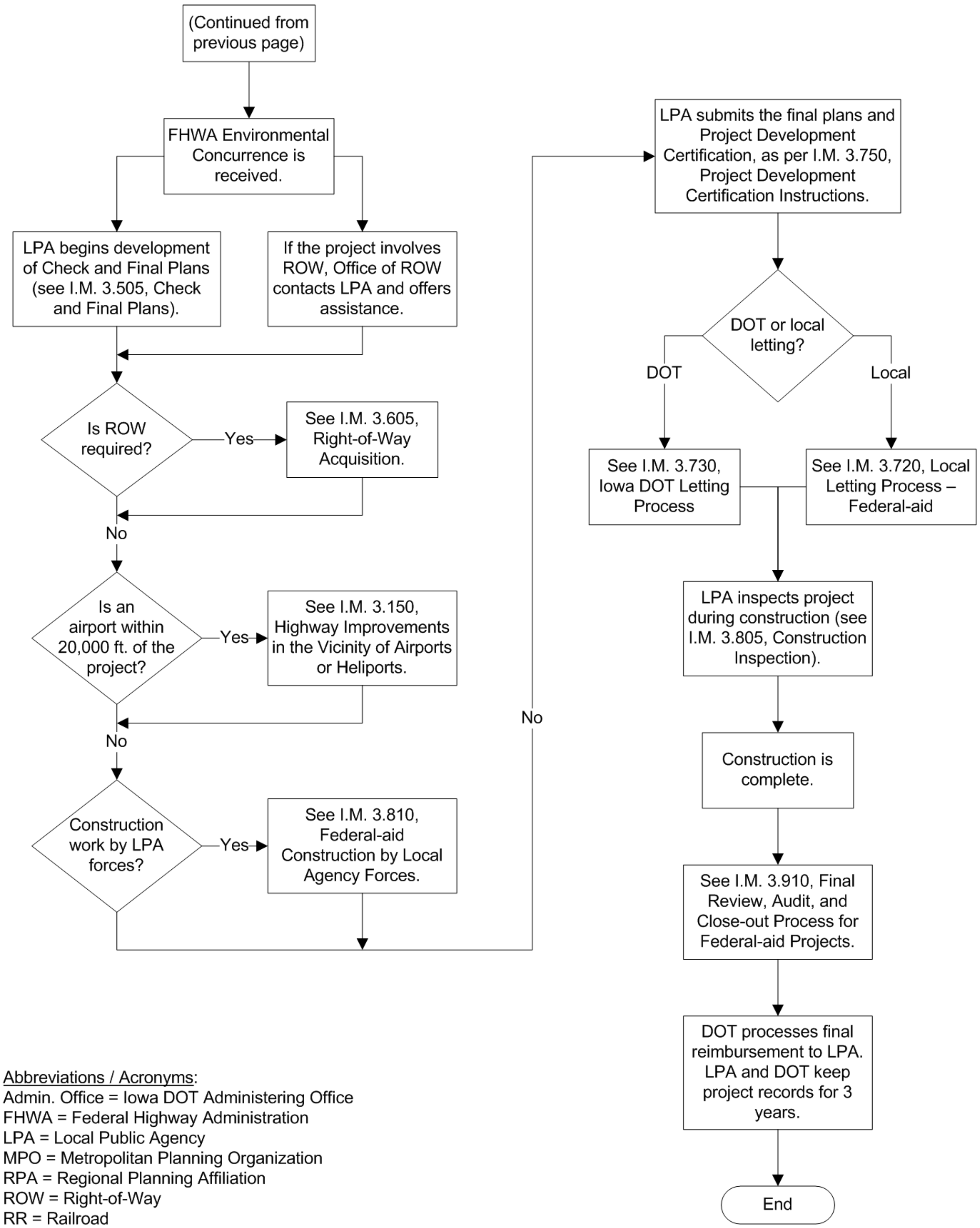
RPO - Rural Planning Agencies

STIP – Statewide Transportation Improvement Plan

Appendix A
Project Development Flow Chart (Iowa DOT, 2013)



Abbreviations / Acronyms:
Admin. Office = Iowa DOT Administering Office
FHWA = Federal Highway Administration
LPA = Local Public Agency
MPO = Metropolitan Planning Organization
RPA = Regional Planning Affiliation
ROW = Right-of-Way
RR = Railroad
STIP = Statewide Transportation Improvement Program
Note: Project submittals shall be as per I.M. 3.005, Project Development Submittal Dates and Information.



Abbreviations / Acronyms:
Admin. Office = Iowa DOT Administering Office
FHWA = Federal Highway Administration
LPA = Local Public Agency
MPO = Metropolitan Planning Organization
RPA = Regional Planning Affiliation
ROW = Right-of-Way
RR = Railroad
STIP = Statewide Transportation Improvement Program
Note: Project submittals shall be as per I.M. 3.005, Project Development Submittal Dates and Information.

Appendix B

Assessment of Data and Information Utilization in Highway Project Decision Making

Dear Survey Participant,

Oklahoma State University is working on a research project entitled "Data & Information Integration Framework for Highway Project Decision Making". This research is focused on developing a framework to map the data and information path utilized to support decision-making processes. Large amounts of data are collected, stored and managed daily to support highway decisions; however, there is a concern in the current level of usage of these data to support decision making. The purpose of this research is to improve the quality of data and information and identify those that are missing or are currently available but are not utilized to support the highway decision making process.

We would like you to participate in the survey and provide us with your valuable opinions for identifying the data and information that should be utilized to support highway projects. The time required to complete this form is approximately 15 minutes. You may return the completed survey form in the following ways. Please return the completed forms by **<u>March 5th 2013</u>**.

Electronic Copy	Mail Copy: Dr. Michael Phil Lewis, Assistant Professor
Please e-mail to: mkhaleg@ostatemail.okstate.edu	Oklahoma State University
Or fax to: 405-744-7554	Civil & Environmental Engineering
	207 Engineering South
	Stillwater, OK 74078

If you have any questions, please feel free to contact me, via phone or e-mail. All data provided for this survey will be considered confidential.

We appreciate your support.

Sincerely,

Phil Lewis, Ph.D., P.E.
Assistant Professor
207 Engineering South
School of Civil and Environmental Engineering
Oklahoma State University
Stillwater OK 74078-5033
Telephone: 405-744-5207/Fax: 405-744-7554
Email: phil.lewis@okstate.edu

1. Please provide:

Contact Person Name: Position:

Work Experience: Division:

Phone: Ext: Email Address:

2. What do you consider to be the major decisions that you make related to highway project lifecycle process?

3. Do you consider these decisions to be made at the network level, program level, or project level? (need to define each level):

4. Please provide examples of the type of data and information that you need to support each decision:

5. Based on your answer to #4, please list the type of data that are currently available and you use to support your decision:

6. Based on your answer to #4, please list the type of data that are currently available, but you do not use. For example, the data may not be used because it is formatted incorrectly, there may be difficulty in retrieving the data, or the data may not be considered reliable:

7. Based on your answer to #4, please list the type of data that are currently missing but you feel could be used to support your decision:

8. Does your division have its own database or information collection and storage system to support your decisions? If so, please provide a brief description of the database or system:

9. Please list the tools or technologies that you use to support your decisions For example, information management systems, software programs, spreadsheets, or subjective judgment by the decision-maker:

10. Please list databases or systems from other divisions that you may use in conjunction with your own division's databases or systems:

11. In your opinion, please rate the current level of data and information usage in terms of supporting your decisions:

Factors		1=Strongly Disagree, 3= Neutral 5=Strongly Agree
a. Definition	Clearly defined in terms of its content	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
b. Precision; Accuracy	Precise and accurate to support my decisions	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
c. Validity & Integrity	Reflect the full details of original observation	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
d. Meaningfulness	Useful and relevant to support my decisions	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
e. Recording	Recorded in a consistent way using standard definitions, such as AASHTO	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
f. Completeness	Includes all relevant and necessary data needed for my decisions	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
g. Subjectivity	Data is more qualitative than quantitative	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
h. Utilization	Used frequently in decision making	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
i. Accessibility	Available to support my decision	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
j. Timeliness	Readily available when needed without delay	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
k. Constant	Meaning or intent remain constant over time	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
l. Data Format	Is data mostly paper-based (3) or digital (5)	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
m. Data Type	Is data mostly numerical (3) or non-numerical (1)	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

12. Based on your experience, what are some of the current problems or needs in terms of utilizing existing data and information to support decision making?

13. Please provide additional comments regarding data and information usage in highway decision-making:

Appendix C

Pavement Data Collection

Asphalt concrete & composite pavements		Jointed concrete pavements		Continuously reinforced concrete pavements	
Transverse cracking	4 Severity Levels	Transverse cracked slabs	2 Severity Levels	Longitudinal cracking	2 Severity Levels
Fatigue cracking	3 Severity Levels	Spilled joints	2 Severity Levels	Punch-out's	3 Severity Levels
Miscellaneous cracking	3 Severity Levels	D-cracked joints	2 Severity Levels	AC patching	No level
AC patching	No level	Corner breaks	2 Severity Levels	PC patching	No level
		AC patching	No level		
		PC patching	No level		
Additional data :	Chain age, Direction, Surface type, International roughness index (IRI), Rutting, Faulting, Joints, Raveling, Macrostructure, Geometrics, GPS coordinates				

Accuracy Requirement

Data Element	Minimum Accuracy	Resolution	Minimum Repeatability
Rut Depth	± 0.008 in compared to manual survey	0.01 in	± 0.008 in run to run for three repeat runs
International Roughness Index	± 5% compared to Dipstick or Class I Profiler	1 in/mi	± 5% run to run for three repeat runs
Faulting	± 0.004 compared to manual survey	0.01 in	± 0.004 in run to run for three repeat runs
Distress Ratings	± 10% compared to ODOT rating	N/A	N/A
GPS Coordinates	± 0.0005 degrees as compared to ODOT provided coordinates	0.00000001 degree	N/A