2014

# Highway Infrastructure Data and Information Integration & Assessment Framework: A Data-Driven Decision-Making Approach

Asregedew Kassa Woldesenbet
*Iowa State University*

**Highway infrastructure data and information integration & assessment framework: A data-driven decision-making approach**

by

**Asregedew Kassa Woldesenbet**

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Civil Engineering (Construction Engineering and Management)

Program of Study Committee:
H. David Jeong, Major Professor
Douglas Gransberg
Charles Jahren
Jennifer Shane
Douglas Gemmill

Iowa State University

Ames, Iowa

2014

**DEDICATION**

To my mom, Hirut Dejene Yirdaw

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

# NOMENCLATURE

| | |
|---|---|
| AADT | Annual Average Daily Traffic |
| AASHTO | American Association of State Highway Transportation Organization |
| CI | cracking Index |
| DOT | Department of Transportation |
| ESAL | Equivalent Single Axle Loading |
| FWD | Falling Weight Deflectometer |
| FHWA | Federal Highway Administration |
| FTA | Fault Tree Analysis |
| *HIDI* | Highway Infrastructure Data Integration |
| HPMS | Highway Performance Monitoring System Reporting |
| MEPDG | Mechanistic–Empirical Pavement Design Guide calibration |
| PCI | Pavement Condition Index |
| SHA | State Highway Agency |
| SI | Structural Index |
| STIP | Statewide Transportation Improvement Program |
| LRTP | Long-Range Transportation Plan |

# ACKNOWLEDGEMENTS

First and foremost, I would like to pass my sincere appreciation and gratitude to Dr. H. David Jeong, my mentor and advisor for his continuous support, guidance, wisdom, and encouragement throughout my graduate study at both Oklahoma State University and Iowa State University. Thank you for all the experience and knowledge you shared with me. The patience you instilled in me, the positive and constructive discussions you provided me and valuable advices even in life made me have you as my second family.

I would also like to express my gratitude for Prof. Douglas Gransberg and Prof. Charles Jahren for their support, feedback, and relentless assistance and encouragement throughout the course of my dissertation and enhancing my academic career.

In addition, I would like to thank my committee members, Dr. Jennifer Shane and Prof. Douglas Gemmill for their guidance, insightful comments and support.

I would like to thank Iowa DOT Pavement Management Division, especially Francis Todey for his valuable inputs and feedback on my research. I would also like to extend my acknowledgement to Matthew Haubrich, John Selmer, Chris Brakee, Ben Behnami, Dr. Scott Schram, Thomas Tymkowicz, David Little, and Tony Gustafson for sharing their experience and serving as expert panels in my research. I would also like to acknowledge Mr. Brad Hoeffer and Mr. Mark Swenson for their inputs and support in extracting engineering contract data.

Furthermore, I would also like to thank my friends, colleagues, the department faculty (Dr. Yelda Turkan) and staff (Mary Muth, Therena Smith, Nancy Quavle, Marva Banks, Kathy Peterson, Denise Wood) for making my time at Oklahoma State University and Iowa State University a wonderful experience. I want to also offer my appreciation to those who

were willing to participate in my surveys, without whom, this dissertation would not have been possible.

Last but not least, I would like to praise God for blessing me with this opportunity and love my mom, Hirut Dejene. I wish you were here to see my final work, but I know you are looking up on me from heaven and you are always in my heart. A special love and appreciation goes to my dad, Kassa Woldesenbet who stood by my side through my journey and for his continuous love, support and understanding; without him I would not be standing where I am today. I also would like to extend my sincere gratitude to my sister Mahlet Kassa and brother Woudneh Kassa, many family and friends which I cannot even finish citing their names, who have supported and kept me in their prayers throughout the entire graduate program. As always, God lead me your ways through what the future will bring.

# ABSTRACT

State highway agencies invest a large amount of resources in collecting, storing and managing various types of data ranging from roadway inventory to pavement condition data during the life cycle of a highway infrastructure project. Despite this huge investment, the current level of data use is limited and is raising concerns whether the growing amount of data adds value to users and offers meaningful return on data collection efforts. This study presents a holistic approach that can systematically integrate and bridge data and information with decisions through incorporation of a unique and proactive performance assessment technique to improve the utilization of a growing amount of data in transportation agencies. With a focus on enhancing the active utilization of data and measuring level of data use, this research delivers i) Integrated Data Quality Assessment Framework, ii) Three-tiered Hierarchical Data-Information-Decision-making Framework and iii) Highway Infrastructure Data Integration (*HIDI*) index, new data and information performance assessment tool.

The study presents an integrated requirement analysis to identify the satisfaction level of various highway decision-makers in current data use and determine the quality requirements of highway data in an integrated and objective manner through the application of fault tree analysis. A three-tiered hierarchical framework is presented to understand the relationship between data and information and identify their use in supporting highway infrastructure decision-making processes. As part of this framework, key players in decision-making processes are identified and quantified through the application of a social network theory. A new index called, *HIDI* is also developed to evaluate the status of data utilization that may serve as *Highway Infrastructure Data Report Card* and help justify the return on investment on the continuous and growing data collection efforts.

This research study will allow agencies to interlink data, information and decisions and to develop active utilization plans of currently existing databases to place the right information in the hands of decision-makers. It will enhance the development of new data collection scheme to support key decisions that, historically, were not well-supported with information and data. The study uses pavement management data as a primary data set to illustrate the application of the framework along with preconstruction service data as a case study and validation data set. This new framework may be used as a benchmarking example for SHAs to make effective and reliable decisions through data-driven insights.

# Chapter 1

# INTRODUCTION

## 1.1 Research Background and Motivation

Today, business and government organizations are overwhelmed with a large amounts of data collected throughout their business cycles. There is a growing data torrent such that managers and potential users are "drowning in data while thirsting for information" (Hermann, 2001). The volume of data collected by organizations is increasing at an alarming rate stretching from kilobyte and megabyte to petabyte and zettabyte; the variety is alternating from structured or table format (numerical and categorical) to unstructured (text, figures and and video) format whereby the velocity at which they are collected is ranging from a batch of data acquired at specific time to real time (data acquired un a continuous manner). Han and Kamber (2006) estimate the amount of data stored in the world database doubles every 20 months, while the international data corporation estimates that approximately 7.9 zettabytes of data will be produced and replicated by fifteen of the seventeen US industry sectors in 2015. Presently, the term *Big Data* is used by these organizations to illustrate the size and complexity of data.

This rapid increase of data generation is due to the fact of recognizing the use and importance such that potential users and analysts claim data as the new science that goes beyond traditional statistics and holds the answers for businesses and services while others are considering data as the new oil of our era (Cleveland, 2001, Gelsinger, 2012 and Gerhardt et al. 2012). Basically, data are becoming a fundamental asset to organizations businesses. Davenport and Harris (2007) argue that the frontier of decision-making 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 quantitative analysis to improve the use of information available to managers (Kennerley and Mason, 2008).

Currently, many strategic business decisions are now supported by statistically reliable information and knowledge drawn from consistently collected data. These businesses utilize data to plan their program, design their activities, set their priorities and measure their performances with the aim of improving service, satisfying customers and increasing profit. For example, the credit card industry analyzes credit card holders spending behavior and demographic statistics to adjust customer's interest rate and identify any fraudulent activity. The retail industry deploys customers purchase habits to design coupons, plan store layout and attract new customers. The medical industry is actively utilizing patient's health care records and clinical data acquired during patient care to obtain optimal health, drive new medical discoveries and preventive measures. Even, basketball statistics are analyzed to identify key matchups in upcoming games (Bhandari et al. 1995).

These industries have kept pace with the complexity of data by investing in the growing capability of the digital infrastructure to address the importance of data with regards to:

- Data collection methods: automated systems and developments such as smartphones, camera, sensor, bar code, radio frequency identification, voice recognition and satellite navigation
- Data storage mechanisms: electronic and digital systems such as database, data warehouses, ontology frameworks and non-relational databases

- Data analytical tools: Data mining, knowledge discovery in database, machine learning and business intelligence tools, knowledge bases, and expert systems.
- Management approaches: Enterprise resource planning, total quality management, cloud computing, lean manufacturing, and business process management throughput their business cycles.

However, the construction industry is relatively behind other industry sectors in taking advantage of this valuable asset to generate reliable information and support decision-making processes. A study by US Bureau of Labor Statistics rates the construction industry's data usage opportunity or value as the lowest (with a negative productivity growth) when compared with seventeen other US industry sectors (Manyika et al. 2011). Nevertheless, it is important to note that the advancement in digital infrastructure and information technology is playing a key role in the data management of the construction industry and is heading in the right direction. Development of tablet-based computers, geographical positioning system (GPS), geographical information system (GIS) and advanced database management systems are good examples assisting the industry in better communication, improved productivity and facilitated management. Paper-based documents such as blue prints and two-dimensional drawings are being transformed into virtual and augmented environments where three and four-dimensional visualizations are becoming a reality. Growth of applications/programs such as building information modeling (BIM), construction operation building information exchange (COBIE), and automated machine guidance and are proof of this advancement.

As a result, this advancement in digital infrastructure is driving government agencies in the construction industry to invest their money and time in the collection and storage of

significant amount of data. For example, state highway agencies now collect highway condition data every two years. Local governments such as municipalities invest in geographical information systems to manage infrastructure asset data. Although the advancement in digital infrastructure is enhancing these agencies in data collection efforts, the current usage is very limited and minimal to support reliable and make informed decisions. In addition, the availability of data in various forms, speed and size are raising concerns if they are effectively utilized and communicated to support the decision-making processes. Thus, there is a need to investigate the current level of data use to improve the construction industry's data utilization and maximize the return on investment in the data collection efforts.

### 1.2 Problem Statement

Currently, highway agencies are one of the organizations in the construction industry that are overwhelmed with a large amounts of highway infrastructure data. These data range from roadway inventory to pavement condition data collected during the life cycle of a highway infrastructure project and stored in various database systems. For instance, the Iowa Department of Transportation (IaDOT) collects distress data along with video logs for right-of-way and pavement images for roadways by developing GIS databases throughout the state using data collection vendors as part of its pavement management program. These data collection costs $75 per mile for city streets and $60 per mile for county roads. In addition, the right of view and pavement images each cost $12 with an approximate collection of 300 images per mile (Iowa DOT, 2009). With the federal eligible highway network of Iowa covering around 27,000 miles, excluding the video log, images and pavement management software the cost of collecting only the distress data is more than two million dollars.

Similarly, the pavement management division of the Oklahoma DOT alone has approximately 1.5 million pavement condition records as part of their asset management program which covers 8000 miles of roadway consisting of 65 data fields every 0.01 miles. This means 800,000 records are collected annually which is approximately 52 million pieces of data (Calvarese, 2007). Despite this huge investment, the amount of information and knowledge generated from these data to support decision-making processes are questionable as compared to the amount produced. This has created a a big concern among various stakeholders whether the data currently being collected by these agencies adds value to the user and offer any meaningful return on the investment. Some of the questions associated with this limited data use are:

- Do currently available data include relevant and necessary data to support the decision-making process? Or does the collected data meet decision-makers requirements?
- Who needs these pieces of data and information in the organization?
- Which data are important and critical in decision-making process?

The effective use and challenges of data in supporting decisions have also been discussed by many industry sectors. Lee and Strong (2003) argue that the purpose of data production process is to produce data for users and should fit the users need. Schoefer et al. (2006) emphasized the value of data as a transportation asset where decisions are considered final products. Schoefer (2007) also stressed the importance of data in decision-making in the way that data meet users' needs and enhance data programs in support of performance measurements, while Harrison (2011) noted a lack of control of information and low responsiveness to the needs of decision-makers as two of the challenges associated with meeting data needs to provide timely decisions. A construction survey conducted in a Fiatech Conference in 2012 shows that more than 60% of participants agree that large amounts of data are impacting the way projects are

managed and is impeding collaboration among project team members with more 75% believing that their biggest challenge as collecting and finding the right information when needed for making decisions. Some of the reasons that may account for the limited usage of data and information might be due to:

1. Insufficient or missing data to perform meaningful analysis;
2. Nonstandard and non-digital data format (linguistic in nature and recorded differently);
3. Poorly defined procedures and mechanisms (user requirement not well defined) to extract, process, and analyze the data in generating usable information and knowledge to assist highway infrastructure decision-makers.
4. Minimal recognition or interest in using these data in the context of supporting various decision-making processes during the life cycle of a highway infrastructure management.

In the life cycle of a highway infrastructure management, data should be collected, stored and managed in a proficient manner to increase the value of data, facilitate transparent information and empower strategic decision-making by satisfying potential users' requirement. In other words, data should be a key basis of competition by utilizing them as information and knowledge generators, effective communication medium, and decision-making input resource. In current practice, there is not a well-structured system or procedure to effectively use and measure the importance of collected highway infrastructure data and integrate it with the decision-making process. On one end, large amounts of data collected across a highway infrastructure lifecycle are stored in different database systems which are managed by different highway divisions. On the other end, various types of decisions are made across a highway infrastructure life-cycle at different decision-making levels (layers) and dimensions. In the middle, data are replicated by different divisions to suit their demand where information are scattered everywhere and often

identified to be missing or not known. Sometimes, potential users of these data and information such as highway project engineers, estimators, and managers do not even know what type of data is available or how to access these data and information to support their decisions (Hummer et al., 1999). Therefore, a new procedure that can structure and address the questions and concerns of this explosion and chaos of data should be determined to improve the use in supporting highway infrastructure decisions.

### 1.3 Research Objectives

The primary goal of this research is to design and develop a holistic approach that can systematically integrate and bridge data with information and decisions through incorporation of a unique and proactive performance assessment technique which can ultimately revolutionize the way data are collected, utilized and managed in enhancing highway industry's decision-making process by meeting users requirement and organizations goal. In order to address this goal and answer the questions associated with limited data use identified in the problem statement, three objectives are set in this study:

1. Determine an integrated approach to identify highway infrastructure decision-makers data satisfaction and quality requirement.
2. Develop an innovative highway infrastructure data and information integration and assessment framework to understand the relationship and correlation of data, information and decisions in enhancing highway infrastructure decision-making processes and investigate the performance of highway agencies' data management.
3. Validate the developed integration and assessment framework through application of a different set of highway infrastructure data.

## 1.4 Research Methodology

Figure 3-1 illustrates the methodology flow chart.

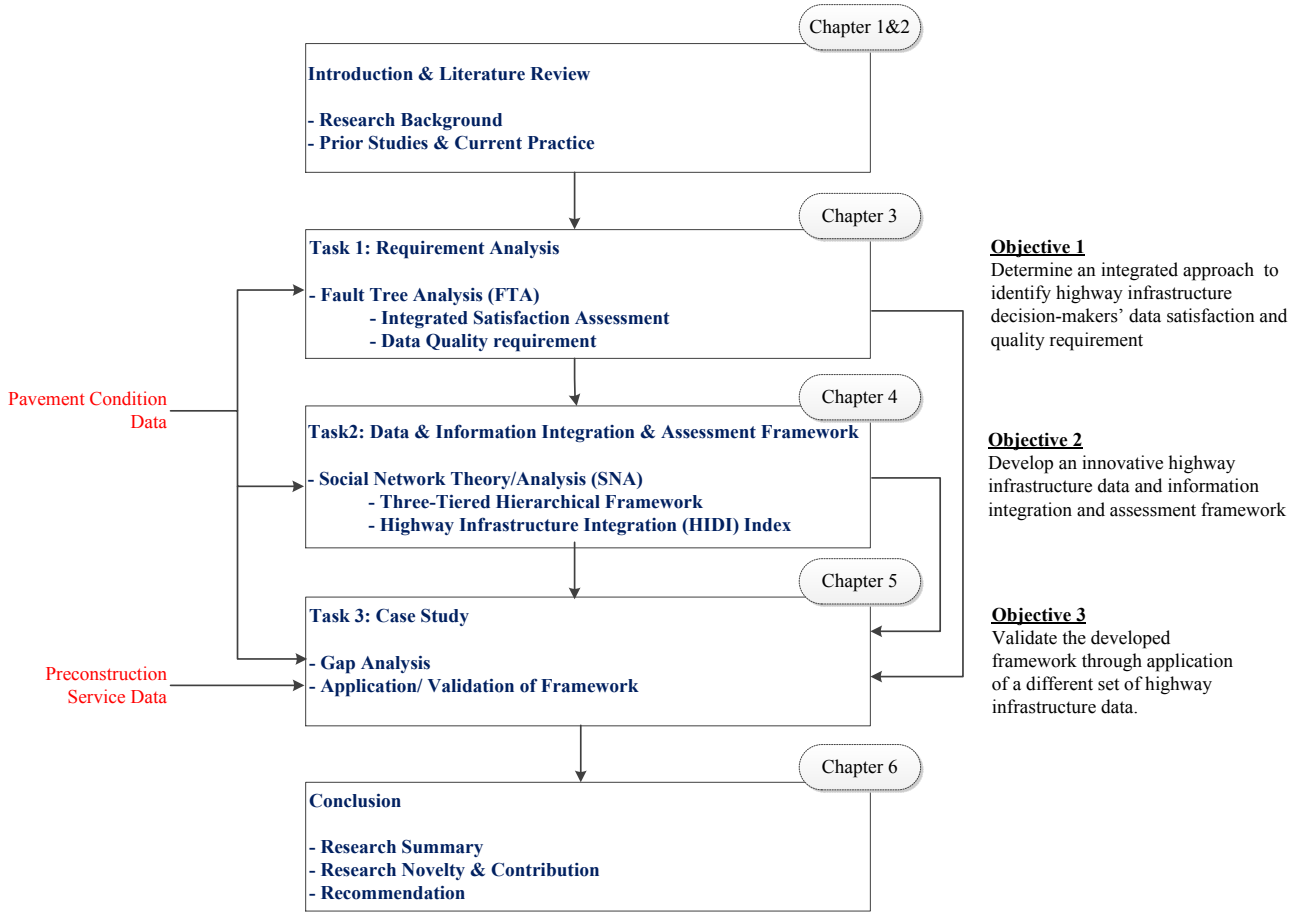



**Figure 1-1 Research Methodology**

Primarily, prior studies conducted in data and information integration, data utilization and current practices by various industries are summarized under a review of literature. Highway infrastructure data, information and decisions utilized in the construction industry are examined through published reports, manuals and interviews and meetings with highway agencies (chapter 2). This review will allow to understand the actual process of decision-making process, data requirement, and information extraction and utilization. In addition, review of prior studies on

specific methodologies adopted in this study are incorporated in each chapter. Then, the study implements three major tasks to fulfil the objectives of this research; requirement analysis, integration and assessment framework (network analysis), and case study.

### 1.4.1 Requirement Analysis

The first task implemented in designing and developing a holistic approach that can systematically integrate and bridge data with information and decisions is a requirement analysis. Typically, a requirement analysis is utilized to identify the features of system functions to fulfil the purpose by balancing customers' needs. This study utilizes a requirement analysis as a top-down approach to identify decision-makers data needs by assessing their values and determining how well current data are functioning in supporting highway infrastructure decisions. This requirement analysis is conducted from two perspectives:

i. Investigating the satisfaction level of decision-makers overall data use in current highway infrastructure decision-making process (at various decision-making levels).

ii. Determines the requirements of data through identification of the root causes behind the minimal usage from data quality perspective.

This requirement analysis is an integrated approach designed to systematically identify decision-makers data requirements from the standpoint of highway agency infrastructure management team through application of a fault tree analysis (FTA). The task answers questions with regards to identifying if current system include relevant and necessary data to support decision-making processes, meet decision-makers requirements, determine potential users and identify the reasons behind minimal data usage through an integrated satisfaction assessment. Chapter 3 discusses this task.

### 1.4.2 Integration and Assessment Framework

Once decision-makers' requirements are identified, the second task involves examining the highway infrastructure data correlations with information and decisions. These correlations can be determined through an integration and mapping of three entities; data, information and decisions. As part of this mapping, the study develops a three-tiered hierarchical framework in facilitating and enhancing integration. Then, the concept of social network theory is utilized to uncover patterns in data, information and decision-making relationships and investigate their correlations within their independent framework and overall framework. Social network theory or analysis (SNA) is a systematic approach to identify, examine and support processes of knowledge sharing (Muller-Prothman, 2006). This approach will allow determining important data, information and decisions or identify key players in decision-making process, examine their interactions and relationships, and assess the performance of highway infrastructure data management. As a result of this analysis, a new performance measure called, highway infrastructure data integration (*HIDI*) index is developed. This section is Chapter 4 of the study.

### 1.4.3 Case Study

The third task implemented in this study is a case study. The case study includes a gap analysis and validation of the integration and assessment framework developed in Task 2. A gap analysis is used to determine the difference between current data utilization and ideal data and information integration, while an external validation is used to evaluate the three-tiered hierarchical framework and implement the Highway Infrastructure Data Integration Index (*HIDI*) by utilizing a different set of data. Pavement condition data are used as primary data for developing the framework (conducting the requirement analysis and social network analysis),

while preconstruction service data are utilized for validating the developed framework. Chapter 5 covers the application and validation of the framework through a case study. Finally, in Chapter 6, the research findings, conclusion, contribution and recommendations for future study are presented.

### 1.5 Research Scope

State highway agencies store a large amount of highway project data throughout the life cycle of highway projects. Some examples of these databases include roadway inventory data, traffic inventory, contract data, daily work reports, and pavement condition assessment data. The scope of this study is limited to two primary data sets: preconstruction service data and pavement condition data. Agencies collect activities and tasks accomplished before the commencement of construction works to track engineering hours and works performed. These activities and tasks referred as preconstruction service data consists of data such as timesheet, contractor's expense, and travel cost. In post-construction, condition of pavements are recorded to assess the level of service of highway projects. The pavement condition data includes network level data such as rutting, roughness, ride quality, cracking, patching collected every other year. These data set are selected due to the fact that they represent two different phases (planning and design and operation and maintenance phases) of a highway infrastructure life-cycle.

It is important to note that in this study, decision-making refers to the process of collecting data at different stages in the life-cycle of a highway infrastructure and generating information/knowledge by applying appropriate data analysis methods that will be interpreted and communicated to support the decision-making process. *Data* refers to raw data collected during the life-cycle of a highway infrastructure and stored in data repository or databases

whereas *information* refers to data that are processed, structured and organized to make it meaningful where they are represented by key performance indicators or measures and outputs resulting from analysis of the raw data. *Decisions* refers to the selection or execution process from a set of available alternatives by utilizing collected *data* and generated *information*. Data and information usage are considered from highway agencies and department of transportation decision-makers' perspective.

# Chapter 2

# REVIEW OF LITERATURE

## 2.1 Introduction

The advancement in data collection methods, digital data storage technologies and database management systems is allowing industries to effectively collect, and store data. This chapter reviews current practices and trends of data usage in US industry sectors' business decision-making processes. Prior studies and lessons learned in construction industry's data utilization efforts are briefly discussed from data storage, interoperability and standardization, management and classification, and data analysis perspective. Potential methodologies that may be used to generate information using these data are also incorporated. In addition, the chapter summarizes state highway agencies (SHA) data management systems, potential information and decisions that may be supported by data across the life-cycle of a highway infrastructure management.

## 2.2 Information Hierarchy

Prior to addressing data utilization efforts, the usage and context of data, information, and decision adopted in this study is discussed as these terms sometimes overlap and are used interchangeably. Information hierarchy also known as data-information-knowledge-wisdom (DIKW) pyramid or knowledge pyramid is one of the fundamental models found in the information and knowledge management literature to illustrate the structural and functional relationship of data, information, knowledge and wisdom with the basic assumption of data are used to generate information; information are used to generate knowledge; and knowledge are in

turn used to generate wisdom (Rowley, 2007). A typical information hierarchy based on maturity level is shown in Figure 2-1.

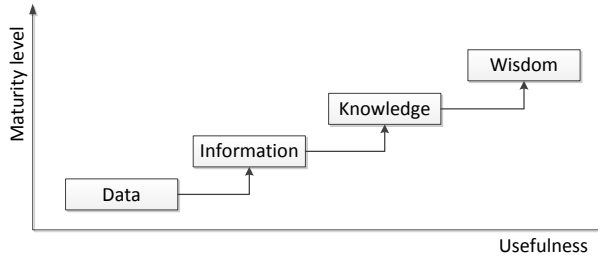



**Figure 2-1 Information Hierarchy (Ackoff, 1989)**

**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, images or audio. At this stage, the value of data is negligible unless it is converted to a usable form. In this study, data refers to raw data collected from highway infrastructure projects and stored in data repository or databases.

**Information:**

Information may 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 or a combination of one or more data that are processed, structured and manipulated to increase users' understanding and make it meaningful. Information 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 key performance indicators/ measures and outputs resulting from analysis of raw data.

**Knowledge:**

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

**Wisdom:**

The highest level of knowledge management hierarchy is wisdom. Wisdom is the ability to utilize knowledge through thorough realization, deep understanding and experience of terms and events that is 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:**

Decision may be represented by the application of knowledge and wisdom to promote business judgment, gain competitive advantage and visualize long-term goals. 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 or selection of choices based on a series of iterative process to

reach at a more reliable and justifiable result. In this study, decision is drawn by managers to execute highway infrastructure projects based on the knowledge acquired from data and information through a critical thinking process.

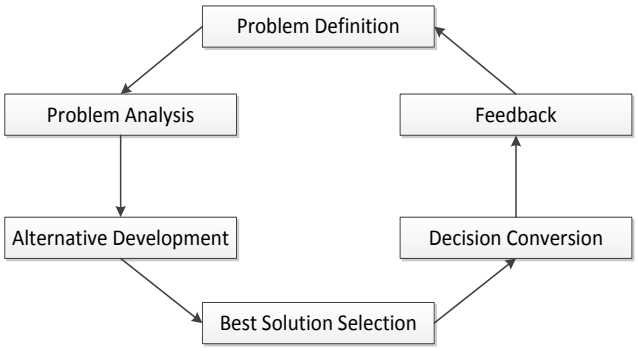



**Figure 2-2 Decision-Making Process (Drucker, 1955)**

**Decision-Making:**

Decision-making is the process or act of making final judgments or selection based on available alternatives to attain certain level of goals/objectives. It is a reasoning process that may range from making rational and formal decision to irrational and informal decision based on explicit or implicit knowledge. Although there might be differences in conducting a decision-making process, Drucker (1955) classified six key elements that should be incorporated in making scientific decisions from management perspective: a) define/identify managerial problem, b) analyze the problem, c) develop alternative solutions, d) select best solution out of alternatives, e) convert the decision into action, and f) ensure feedback for follow-ups (Figure 2.-2).

The European commission of project management guide map also recommend a six stage process as a general principle for making transportation decisions (May, 2003). These include problem definition, option generation, option assessment, decision-making, implementation,

monitoring and evaluation. Neely and Jarrar (2004) proposed a performance-planning-value-chain (PPVC) process to improve decision-making from data utilization perspective; collect, analyze, interpret; communicate; and make informed decisions. In this study, decision-making refers to the process of collecting raw data at different stages in the life-cycle of a highway infrastructure management and generating information and knowledge by applying appropriate data analysis methods that tends toward supporting the selection and execution process. Examples of highway infrastructure decision-makings' include pavement treatment selection, project selection, and contract time determination. Figure 2-3 shows decision-making process from data usage perspective modified from Neely and Jarrar (2004).

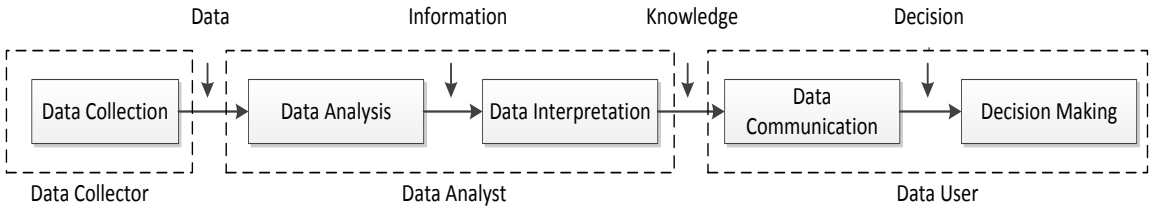



**Figure 2-3 Highway Infrastructure Decision-Making Process**

In this decision-making process, three major players are involved; data collector, data analyst and data user. Data collectors are responsible for collecting and storing data, while data analysts are in charge of analyzing data and generating information and knowledge. Data users are the final consumers that utilize the information and knowledge acquired from data analysis to make final judgments. In a highway agency, data may be collected by in-house or outsourced to data collectors. Data analysts refer to in-house experts, statisticians, data scientists, consulting firms or academic researchers, while data users refer to decision-makers ranging from program managers and administrators to project managers and division engineers. However, it is important to note that this linear decision-making process can be repetitive, parallel, or cyclical

that might trigger a second or iterative process depending on the type of decision. 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 all approaches to discuss prior studies and current practice of data utilization in highway infrastructure decision-making processes.

## 2.3 Trends in US Industry Sectors

Active utilization of data is greatly enhancing various industry sectors ranging from health care and energy to manufacturing and agricultural sectors in making reliable business decisions and generating significant financial values. Today, the US retail industry is estimated to have a 60% increase in net margin with 0.5 to 1.0 % of annual productivity growth, while the manufacturing industry will experience up to 50% decrease in their product development and assembly costs with up to 7% of reduction in working capital through active utilization of data (Manyika et al. 2011). A study by US Bureau of Labor Statistics shows that the information, computer products, manufacturing, wholesale, finance and insurance and government sectors are taking advantage in capturing values from data, while management, education service and construction have the lowest value with negative productivity growth (Figure 2-4).

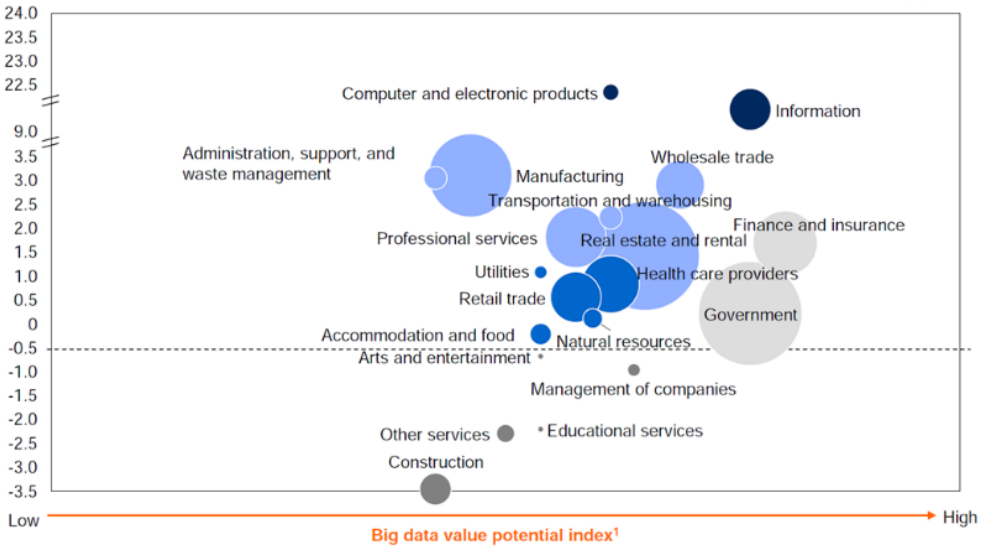



**Figure 2-4 Data utilization In US Industry Sectors (Manyika et al., 2011)**

Currently, the US federal government is even taking the initiative in addressing various issues through active utilization of data (OSTP, 2012). For instance, the department of defense and homeland security are working towards developing computational tools for use in targeted defense applications, establish alerts for activities to predict manmade and natural disasters. The department of health and human services and food and drug administration are working towards data-centric approach to public promote public health, improve clinical treatment control and prevent diseases. This section briefly summarizes the advancement in data utilization in four major sectors; retail, healthcare, manufacturing and construction industry from data storage, interoperability, standardization, management and classification, and data analysis perspective.

### 2.3.1 Retail Industry

The aggregation of big data is driving the retail and marketing industry to change its decision-making paradigm towards data-driven mindsets. Today, retail industries are able to acquire and analyze data ranging from demographics to call records to learn more about their customer. Retail giants such as Amazon, Netflix, and EBay are analyzing behavioral data to

design advertisements, identify movie tastes and predict future stock market purchases (Anderson, 2012). Currently, one of the trends that the retail industry is heading towards is next-best-offer (NBO). Next-best-offer or next-best-action is a customer-centric marketing strategy to deliver the 'best" one be either an offer, proposition or service to the customer (Sanjiv, 2007). The strategy is derived from a combination of customer's interest, need and organization's business policies and objectives. The science behind NBO is complex mathematical models and predictive analytics to process real time events such as birthdays, pregnancies, and accidents to do adaptive learning, behavior models and sentiment analysis integrated with business rules and customer management tools (Sanjiv, 2012).

This personalization technique will revitalize the traditional product-oriented marketing and indicates where the industry is heading. In near future, retail companies will be able to predict a customer's need before the customer realizes it. In this predictive analytics technique, Amazon is has paved the road in generating prompts like "you may also want this" when one visits a webpage or buys a product using a collaborative filtering mechanism (Kalakota, 2102). A similar approach used in the retail industry is a "taste graph". A taste graph is a recommendation engine developed based on a data structure designed to make decisions by connecting the user on the web to what they like (Hunch, 2013). This approach is not limited to prediction of single user behavior, but also other users' behavior that fall in the same group which is more like advanced clustering analysis. Other techniques used by the industry include use of developments like MapReduce to make use of data by running large distributed computations. MapReduce is a programing model or data architecture developed by Google to extract and process big data which uses parallel algorithm to filter, sort and search patterns of data stored in different systems where Handoop is its open source application (Lammel, 2008).

### 2.3.2 Biomedical Industry

The biomedical industry is another industry sector that is greatly excelling in data utilization to support decisions. Typically, the industry makes use of data captured from patients to measure patient's optimal health, drive new medical discoveries and preventive measures. Some of the data captured in this sector include clinical data; activity, claims and cost; pharmaceutical and medical products; and data about patient behavior. One of the interdisciplinary fields in the healthcare industry that deals with storing, organizing, analyzing and interpreting useful biological data using information technology is called bioinformatics. Bioinformatics is enhancing the industry through determination of DNA sequence analysis (genetics), alignment, prediction and interactions of protein structure (structural biology), and drug design and discovery (Hofacker et al. 1996). Bioinformatics uses data analysis techniques such as machine learning and pattern recognition algorithms, data mining, image processing and visualization based on theories and applications such as statistics, information theory, system theory and control theory.

In this sector, the concept of evidence-based management (EBM) has been used by clinicians, managers and policy makers to effectively utilize data to make reliable decisions by organizing, structuring, delivering and financing health service. Evidence–based health care is the provision of care based on data and information from well-conducted research into the effectiveness of health care interventions (Walshe and Rundall, 2001). The approach has influenced the decision-making of health care systems in the UK and US in terms of creating national database of health service research projects, reforming commission process through developmental strategy and national tracking methods and dissemination of findings (Black 1997 and Adelman et al. 2000). EBM has also been implemented in other fields including social care,

criminal justice and education to improve efficiency of services and reduce cost (Boruch et al. 1999 and Davies et al. 1999). Another approach utilized in the health care sector is conjoint analysis. Although conjoint analysis is widely used in marketing and operation research, it has been actively used in the healthcare to elicit views of patients and community's preference on the quality of health services. It also has been implemented in setting priority; developing outcome measures; determining optimal treatments; evaluating alternatives within randomized controlled trials; and establishing patient preferences in doctor-patient relationship (Ryan and Farrar, 2000).

### 2.3.3 Manufacturing Industry

The manufacturing or product development industry is also greatly advancing in the utilization of data to support decision-makings. Global competition, wide diversity of supply chain, and tight environmental regulations are some of the reasons for the industry to head into a data-driven insights. With the advancement in digital technology, the industry is able to acquire various data such as products, supply chain, service, employees and customers. To actively use these data, the industry is shifting to a data analytics where manufacturers may view upstream into a global and integrated supply chain and downstream into consumer base to understand current conditions and predict the future of a product life cycle (Dittmar, 2012). Manufacturers use data obtained from sensors implanted in products to track their usage, create proactive maintenance, avoid failures and improve design of new products. Others use data from a production line to develop self-optimization mechanism to avoid waste, reduce cost and improve product output (Mcguire et al., 2012). For instance, car producers are implementing an integrated approach of telecommunication and informatics (telematics) to collect real time data such as

distance, speed, and duration of travel, and location (highway, city, and street) of a driver to assess driving patterns and analyze accidents.

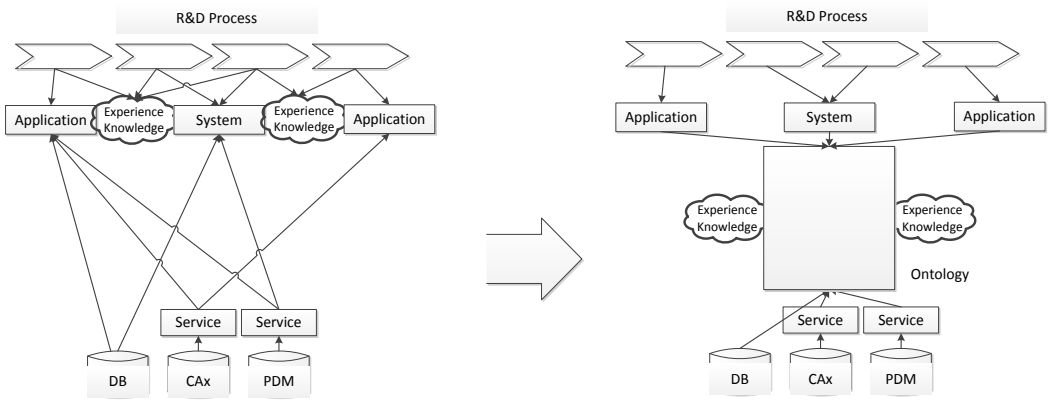



**Figure 2-5 Ontology Framework (Maier et al. 2003)**

Optimizing the developmental process from the planning of technical and organizational aspects to physical tests including the assembly processes is an essential strategic task in order to cope with data needs (Bullinger et al. 2003). One of the prominent methodologies used by the manufacturing industry in actively utilizing data is the use of semantic technologies or ontology-based approach. Ontology is defined as "an explicit specification of conceptualization" which may be considered as a more advanced knowledge representation model (Gruber, 1995). It is used as mediating medium between data that is stored in different departments' or data sources such as computer aided design (CAD), enterprise resource planning (ERP) and databases and applications. For example, Maier et al. (2003) illustrated the capabilities of ontology-based approach by integrating Audi car product components data and its permitted configurations in the product life-cycle management. Taisch et al. (2012) also showed a data life-cycle management semantic model which tends toward managing data across a product life-cycle to predict truck maintenance.

## 2.4 Construction Industry

### 2.4.1 Prior Studies

Efficient management and business decisions are supported through well documented data and better data management system. One of the noticeable approaches associated with data usage in the construction industry is the development and application of digital data storage. Prior studies have addressed the use of advanced database systems to enhance construction industry decisions. Noticeable works include the development of a prototype model-based system for bridge design processes using ISO STEP standards by Halfway et al. in 2005. 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. Yu et al. (1999) developed a computer-integrated facilities management (CIFM) framework that is supported by the facilities management classes (FMC) and the industry foundation class (IFC).

These studies were able to demonstrate how data should be properly stored in a structured data model and centralized project data repository to support decisions. They illustrate how systems such as computer aided design and drafting (CADD) can be integrated with various applications such as estimating, plan-generation and scheduling through a shared object-oriented database or how the use of project data in a design and construction phase could be used at later phases to support facilities management. However, some studies argue that these data models and integration approaches have limited semantic representation that might be difficult to make changes at later stages and the representations do not support multiple views from multiple domains due to predefined schema (O'Brien et al. 2000 and Rivard and Fenves 2000). These

studies have tried to address the data heterogeneity problems and limited semantic representation encountered in construction projects through the development of ontology-based frameworks (Shen, 2004, Wang et al. 2011). The studies showed how to manage context-sensitive construction data and integrate design knowledge with cost and schedule data. However, it is important to note that ontology-based frameworks practiced by the manufacturing and construction industry are in developmental/ conceptual stages and actual application on real projects is on progress.

Another potential approach associated with active data usage is classification and organization of project data to effectively generate information. Studies by Caldas et al. (2002) showed how machine learning, such as support vector machines (SVM) algorithm, can be used 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 support in integrating project documents or unstructured data types such as contract documents, filed reports and change orders in the architecture, engineering and construction (AEC) industry. Soibelman and Kim's work (2002) presented data preparation process to generate construction knowledge through knowledge discovery in databases (KDD). The study showed potential use of KDD to identify causes of construction activity delays using construction databases. 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) also presented the principles and methods of analyzing data using a decision support system architecture based on data warehousing through structured query language (SQL).

These studies have major contributions in the advancement of active utilization of data through application of database management system, development of standard models and data analysis techniques to improve the construction industry's decision-making process. However, prior studies do not show the interaction and relationship between specific data, information and decisions and the performances of these data if they meet decision-makers' requirement. Most of the efforts focus on vertical or building construction as compared to horizontal or highway construction. Although it is difficult to address every decision made over the life–cycle of a highway infrastructure management, this study focuses on integrating data, information and decisions through a three-tiered mapping and investigate their relationship based on decisions that may be supported by daily work report, pavement condition and preconstruction data.

### 2.4.2 Highway Infrastructure Data Integration

In the highway infrastructure management sector, the Federal highway Administration (FHWA) and American Association of State Highway and Transportation Officials (AASHTO) play a huge role in acquiring and managing data to support decisions by conducting studies and developing programs. For instance, one of the prominent studies conducted by FHWA's asset management program is "Data Integration Primer" (FHWA, 2010). The primer outlines a set of guidelines or key activities that should be incorporated in collecting, storing and managing data to support highway decisions. The primer summarizes five primary tasks as part of a data integration process from asset management perspective (Figure 2-6).

The first step in this process is to conduct a requirement analysis. A requirement analysis is concerned with analyzing user requirements, understanding the business process, and identifying the characteristics of existing systems. The second step involves recognizing the

relationship and mapping of data and process flows by identifying the inputs and outputs required in modeling the process. This step leads to identification of data storage or management systems by defining alternatives, evaluating and selecting database architecture, identifying the risks and the level of efforts and time required for development. The final step in the data integration process is the development and implementation of the chosen strategy in terms of computer programming, software/hardware setup and testing. This study utilizes FHWA's integration first two steps as a fundamental guide in analyzing the pieces of data and information to address users' requirement.

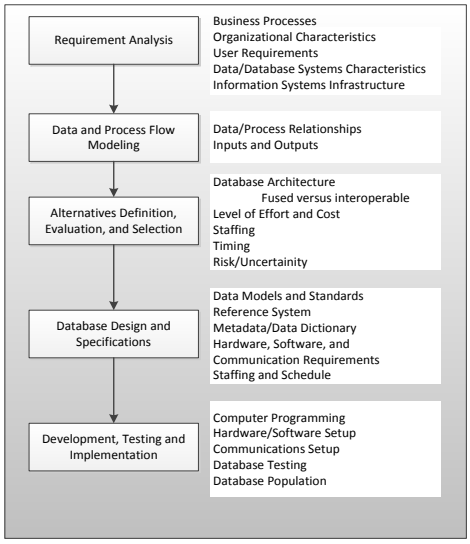



**Figure 2-6 Data Integration Process (FHWA, 2010)**

Another important study that is conducted by FHWA is the development of programs under AASHTO's transportation software management solution (Trns.port) program. The program aims at developing modules to convert data into information in supporting business decisions across the life-cycle of a highway infrastructure management. Some of these programs include *CES* (Cost Estimation System), *Trns.port Tracer*, *Trns.port Preconstruction System* and *Trns.port Estimator* which are designed to support project cost estimation during the

preconstruction phase. For instance, *CES* is used in preparation of parametric, cost-based and historical bid-based cost estimates, while *Trns.port Estimator* is used in the preparation of detailed cost estimates, supporting cost-based and historical bid-based estimation. *Trns.port Tracer* serves as a parametric 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 managing projects in terms of creating proposals, schedule letting and contract awarding.

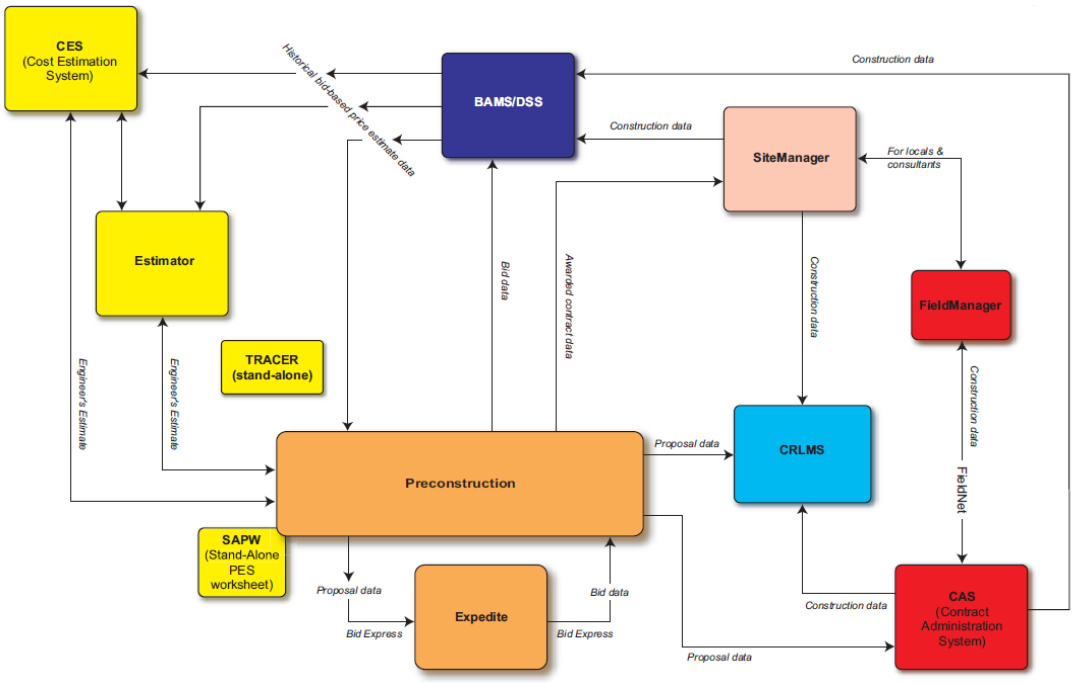



**Figure 2-7 AASHTO Trns.port Programs (AASHTO, 2009)**

Other modules that are currently in service/or under development include *BAMS/DSS* (Bid Analysis Management System/Decision Support System), *PES* (Proposal and Estimates System) and *LAS* (Letting and Awards System). *BAM/DSS* is a relational open architecture historical database and bid analysis software 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*. *PES* is used in the preletting phase of a bid, while *LAS* is used to help highway agencies 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 are also part of the program (AASHTO, 2009). Figure 2-8 shows AASHTO Trns.port modules.

These modules are specifically designed for highway agencies and their design consultants to be utilized as business intelligence tools and enterprise-wide data management systems to support agency's decisions at various levels. Although some of the programs are personal computer-based or standalone programs, they are capable of exchanging information and are interlinked with agency's websites and other databases. However, there are overlaps between some of the programs that might have the opportunity of duplicating data and their compatibility is limited.

## 2.5 Highway Infrastructure Databases

The development of the aforementioned programs and studies by FHWA and AASHTO along with in-house expertise, academia and consulting firms should allow department of transportations (DOTs) to store and maintain highway data in a more proficient manner. In a typical DOT, highway data management systems range from in-house spreadsheets to

commercially available programs acquired through manual and automated data collection across a highway infrastructure management. Some of these systems include *GripLite* or highway inventory, Contract Fee Proposal Spreadsheet, *Proposal and Estimates System* (PES), *Letting and Awards System* (LAS), *SiteManager*, *FieldManager*, *FieldBook*, and *Pavement Management Information System* (PMIS). A brief summary of major databases managed by DOTs is shown in Table 2-1.

**Table 2-1 Major DOT Databases**

| Division | Database | Category | Type of Data | 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. | |
| Pre-Construction | In-house Spreadsheets/ | Preliminary Engineering Cost Data | Work efforts (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 | 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 |
| | | Structural Data (Distress Data) | Longitudinal Cracking, Transverse Cracking, Patching, Spalling, Fatigue, etc. | Consultant |
| | | Functional Data | Average Roughness, Ride, Rut etc. | In-house |
| | | Other (structural) | Friction, Deflectometer (FWD), ESAL | In-house |

i. Highway Inventory

Typically, an enterprise-wide GIS database system is used to collect and manage highway inventory data utilized by DOT's planning and research division. This GIS system (e.g. Oklahoma DOT 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 includes data such as number of lanes, width, functional class, right of way,

route classification, railroad crossing, control type, and terrain area type for 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 may be accessed through its website, http://192.149.244.31/griplite/index.htm. Typically, most of these pieces of data are collected manually by specific divisions while others such as annual average daily traffic (AADT) may be collected in a semi-automated manner.

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, Oklahoma DOT 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. This spreadsheet is developed based on estimated work efforts (engineering man hours and hours per mile) along with associated costs to prepare a set of plans from the preliminary stage to the final plan preparation. The spreadsheet consists of a cross tab of major 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 work efforts required by engineers based on the amount of sheets required for each task and project length by comparing it with previous highway infrastructure projects. It is important to note that these pieces of data are partially stored electronically on the division engineer's computer, while the majority are stored in hard copies or 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). Highway infrastructure data such as bridge, 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 the data collected and stored through these databases are managed by the asset management program.

iv. Construction Database

DOTs' store construction project data through a combination of AASHTO's Trns.port construction administration programs such as *SiteManager*, *FieldManager* and/or *FieledBook*. For example, Oklahoma DOT utilizes *SiteManager* as its primary construction database. *SiteManager* is a multi-tier architecture construction management tool used for data entry, tracking, reporting, and analysis of contract data during the construction phase of a highway infrastructure project (FHWA, 2013). *SiteManager* consists of six basic functions to view and store highway construction project data; contract administration, daily work reports, contractor payments, change order, civil rights and material management systems. These functions allow data acquisitions such as materials and equipment, 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 a construction phase (FHWA, 2013). ODOT's *SiteManager* contains data sets of more than 1,500 previously completed and ongoing construction projects since 2002.

v. Pavement Database

Various DOTs have developed a pavement management system (PMS) to manage network level pavement condition data. Typically, PMS consists of three divisions, pavement history, structural and functional data. Pavement history is used to understand previous treatment applications in terms of pavement surface type, thickness, composition, and treatment cost. Pavement condition data takes functional and structural aspects, where the functional data considers pavement rutting, roughness, ride quality, etc., while the structural aspect considers pavement distress data and stiffness such as longitudinal cracking, transverse cracking, patching, bleeding, fatigue, etc. In addition, non-destructive evaluation test data such as friction, falling weight deflectometer (FWD) and equivalent single-axle load (ESAL) are recorded for checking structural adequacy. The pavement condition data mainly contain one record per $100^{th}$ mile of structural layer and surface condition of roadway collected in annual or bi-annual basis depending on the agency. For instance, ODOT collects 8000 miles of data every $100^{th}$ mile or 800,000 records annually that consists of 65 data fields which is 52 million pieces of data (Calvarese, 2007).

These data are collected using various methods ranging from manual surveys to semi-automated and automated data collection. The manual surveys involve collecting of surface distress by walking or travelling at low speed while semi-automated collection uses lasers and high speed cameras to capture digital images and usually a trained personnel rates visible distresses. Automated data collection utilizes the data collected using laser and cameras to classify pavement distresses in real time. For instance, rut depth can be estimated either by taking a manual spot measurement or sensor data in semi or fully automated manner.

It is shown that various types of data are collected in state highway agencies by taking Oklahoma DOT as an example. However, it is important to note that various DOTs collect different pieces of data and use diverse data management systems. For example, contrary to Oklahoma DOT, Iowa DOT uses a combination of *FieldManager* and *FieldBook* as its construction database and uses an accounting system along with a project scheduling system (PSS) to collect timesheet, contractor's expense, and travel cost as its preconstruction database as opposed to *SiteManager* and in-house spreadsheets respectively. Iowa DOT collects pavement condition data such as falling weight deflectometer, while Oklahoma DOT collects raveling and bleeding. In addition, it should be noted that DOTs utilize various methods ranging from manual surveys to semi-automated and automated systems to collect their data. Some may use manual surveys to collect surface distress by walking or travelling at low speed or automated mechanisms such as sensors, lasers and high speed cameras to capture digital images and classify pavement distresses in real time (Pierce et al. 2013). Furthermore, it is important to note that data from one database are utilized in conjunction with other databases to support various decisions. For example, pavement condition data in the maintenance and operation phase are utilized along with highway inventory and traffic data in the planning phase to support decisions such as treatment selection and project prioritization.

## 2.6 Highway Infrastructure Information

Once data are collected and stored, they should be properly analyzed and managed to meet their intended purpose. The purpose of highway data is to generate information and support decisions across the life-cycle of highway infrastructure management. This purpose may be enhanced through the use of relevant and appropriate data analysis methods. Prior to addressing

data analysis methods, it is important to note that in this study, highway infrastructure information is referred as raw data that are collected by an agency or a combination of one or more data that are processed and structured, represented by performance indicators/measures and outputs resulting from analysis of raw data. Highway information may include traffic analysis, economic analysis, cost estimation, crash analysis, bid analysis, production rate, and pavement condition indices. Table 2-2 shows examples of information that may be generated using highway infrastructure data in different phases.

During the planning phase information such as capacity analysis, traffic analysis, environmental assessment, needs study may be generated by utilizing traffic, cost, roadway inventory and pavement condition data. Similarly, various types of design analysis and cost estimation for bridge, roadway, environmental and right-of-way may be generated in the design phase by implementing roadway and bridge inventory, cost and pavement condition data. Vendor analysis, contract analysis, item price estimation, bid analysis and evaluation are examples of information that may be generated using bid data in the letting phase. Production rate estimation (quantity of work installed per day), cost tracking (actual cost per planned cost), schedule tracking (percentage completion), safety analysis (number of accidents per project), cost (cost per mile, percentage of construction cost, etc.) are information that may be generated in the construction phase by utilizing project cost, schedule and site condition data. Information in the operation and maintenance phase includes pavement condition indices, life –cycle and cost benefit analysis that may be generated from pavement management data along with traffic, roadway inventory and cost data.

**Table 2-2 Example of Highway Infrastructure Information**

| Phase | Information |
|---|---|
| Planning | Capacity analysis, economic analysis, traffic analysis, environmental assessment, needs study, sufficiency rating, etc. |
| Design | Design analysis, safety analysis, cost estimation, etc. |
| Bid | Vendor analysis, contract analysis, item price estimation, bid analysis, etc. |
| Construction | Production rate determination, project progress (percent completion), earned value management (cost and schedule tracking), etc. |
| Operation & Maintenance | Pavement condition indices, life-cycle cost analysis, cost/benefit analysis, etc. |

However, it is important to note that new information may always be generated by identifying new patterns and correlations through data analysis. Data analysis is a process of applying logical, statistical 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 common platform, making strategic and optimal business decisions. 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). Multiple-disciplines ranging from social science to information technology have developed various tools, techniques and applications to analyze, interpret and visualize data, extract patterns and knowledge from a vast amount of data. Some of these tools include knowledge management (KM) tools, and knowledge discovery in database (KDD) and/or data mining (DM) techniques, decision support systems, artificial intelligence, machine learning and business intelligence tools, and knowledge bases (KB).

Broadly, data analysis can be classified into explanatory and inferential analysis based on type of data usage from statistical 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 scatter plot, box-plot, and cross-tabulation. 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, 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, while Abdollahipour (2012) used association rules to categorize pavement treatment types in rehabilitation projects.

Quantitative analysis can further be divided into predictive modeling, and artificial intelligence and optimization techniques. Predictive modeling primarily includes parametric approaches 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 2-8 shows classification of data analysis techniques.

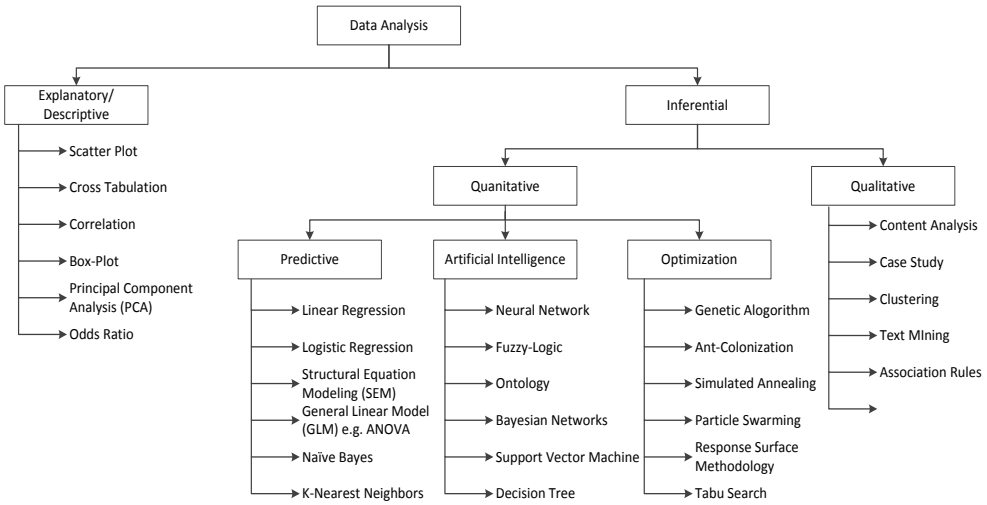



**Figure 2-8 Data Analysis Methods**

Examples of quantitative analysis from a cost estimating perspective by utilizing highway agency data include work done by Woldesenbet and Jeong (2012) who 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 10-year 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. Weisbrod and Backwith (1992) developed an economic simulation model called REMI (Regional Economic Model Inc.) to evaluate the development impacts of highway investment using 200 mile four-lane highway project Wisconsin DOT data as a case study. The study showed how economic benefits of highway projects can be estimated and used for cost-benefit analysis to support policy decision-making. 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. It is important to note these methodologies and studies are examples of potential data analysis or information generation methods.

## 2.7 Highway Infrastructure Decisions

Once information is generated using highway infrastructure data, the next step is to utilize the information along with other project data in supporting decisions. The basics of highway decision-making process commences with identification of opportunities to improve the transportation system for the user through transportation planning (FHWA, 2012). Highway planning starts with setting goals and visions based on critical factors such as population growth economic changes, transportation needs, public input, etc. These visions and goals are translated to 20 year or long-range transportation plan (LRTP) based on available alternatives. State department of transportation (DOT) and metropolitan planning organization (MPO) develop a short-range (4-year) improvement plan under Transportation Improvement Program (TIP) and Statewide Transportation Improvement Program (STIP) respectively. In this process, states and MPOs make various decisions ranging from estimation of project costs to evaluation and prioritization of treatment strategies along a highway infrastructure decision-making hierarchy.

Figure 2-9 illustrates examples of decisions made over the life-cycle of highway infrastructure management. Some decisions made in the planning phase include identification of projects, establishment of program objectives, evaluation of potential projects and allocation of budget. These planning decisions usually are part of a program level and network level decisions. Selection of design alternatives (pavement type, thickness, and bridge span), determination of contract time, selection of construction methods, right-of way, traffic control and allocation of

cost are some of the decisions made in the design phase, while contractor screening, contract awarding, procurement strategy selection are made in the bidding phase. Similarly, some of the decisions made in the construction phase include contractor payment approval, quality control, and cost estimation, while selection of treatment strategies, identification of maintenance needs, and prioritization of projects may fall under maintenance and operation decisions.

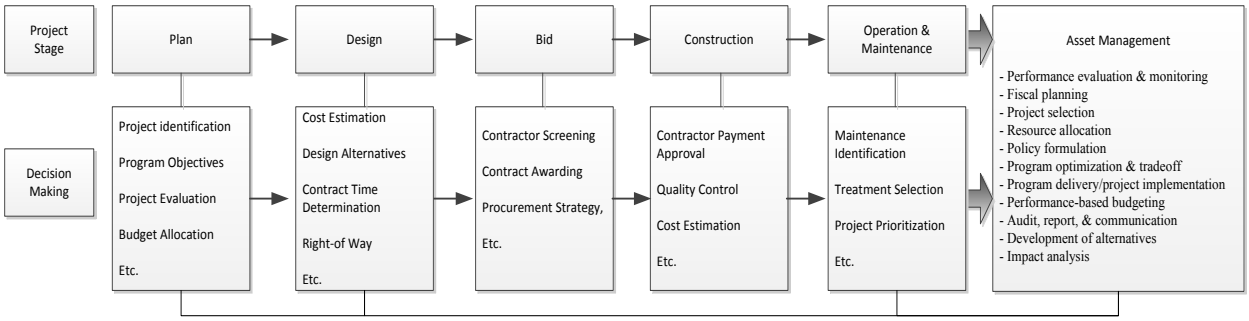



**Figure 2-9 Highway Infrastructure Decisions**

It is important to note that these decisions' require various pieces of information and data as input to support the decision-making processes. However, the current practice does not fully illustrate the types of data and information that is required by users' to support decisions. The relationship between these decisions, information and data are not clearly recognized from decision-makers' perspective. In addition, it should be noted that most of the planning level and design level decisions are heading towards the asset management program. Therefore, there is a need to differentiate the data and information requirements of decisions, study the relationship between them to integrate and improve data utilization in making reliable decisions. Specific decisions considered in this study are discussed in next chapters.

# Chapter 3

# REQUIREMENT ANALYSIS: ASSIMILATED DATA QUALITY ASSESSMENT

## 3.1 Introduction

Currently, effective data utilization in highway agencies is a critical issue with respect to the amount of data produced and replicated. Lee and Strong (2003) argue that the purpose of data production process is to produce data for data consumers and measure the value of data as 'data that are fit for use" by data consumers. Failure to address poor data usage can lead to wasting resources, failure to provide quality services and unsuccessful overall policy and management, while good data allows to achieve better reliability within an organization, improve decision support for managers and comply with external data requirements (Shekharan et al., 2006 and Audit Commission, 2007). Thus, the use of data collected by agencies should be investigated if they are useful and meet decision-makers' requirements.

Some fundamental questions that must be answered include: Do currently collected data quality meet decision-makers requirements? Are data interpreted in the same manner; Do data reflect the details of original observation; or Are relevant data collected to support decision-makers' requirements? Requirement analysis is a method to determine the needs of various stakeholders by minimizing ambiguity, analyzing, and managing the requirements in developing or improving a product (Kotonya and Sommerville, 1998). Requirement analysis helps in answering these types of questions by assessing the quality and understanding users' expectations. Requirement analysis sometimes referred as requirement engineering allows

detecting problems behind the minimal usage of data and help improve data performance to effectively generate and place the right information and knowledge in the hands of decision-makers.

The international council of systems engineering, INCOSE defines requirement analysis as "a process to review, assess, prioritize and balance all stakeholder requirements including constraints and to transform those requirements into a functional and technical view of a system capable of meeting the stakeholders' needs" where stakeholders may range from designers and managers to customers and the public in which their goals may vary with respect to the tasks they accomplish (INCOSE, 2001). Basically, requirement analysis involves understanding a problem (eliciting), modeling and analyzing the problem, attaining agreement on the nature of the problem, communicating the problem and managing the changes as the problem evolves (Nuseibeh and Easterbrook, 2001). These activities are performed to gather requirements, assess whether the requirements are clear, complete and unambiguous, and record them to meet stakeholders' needs such that the analysis may take functional, non-functional, technical, architectural, structural, or design aspects depending on the type of the product being developed or improved which might involve systems, machines, software, hardware, or databases.

Requirement analysis is widely utilized in systems engineering and software development as a primary step to determine the needs of the owner in developing software products or system designs. Previously, it has been used in designing aircrafts, computer chips, integrating software products (Kotonya and Sommerville, 1998). In the manufacturing industry, requirement analysis has been applied from a customer perspective by incorporating 'quality' in the development of automotive manufacturing and improving process design (Lochner and Matar, 1990, Klippel, 1998). In the construction industry, it has been employed to determine construction delays,

develop intelligent transportation systems, and identify high performance buildings (Yang et al. 2006, Ryen, 2008, Ye et al. 2009 and InfoTech 2009).

The goal of this chapter is to assess highway infrastructure decision-makers' data quality requirements as a primary step in determining a framework for integrating and assessing highway infrastructure data and information use. The chapter determines decision-makers' requirements by identifying and estimating the satisfaction level quantitatively to detect problems behind the minimal usage of data. This improves data performance at various decision-making hierarchy in a more integrated manner through the application of a fault tree analysis.

## 3.2 Methodology

Figure 3-1 shows the methodology used in this chapter which can be divided into three major phases. Primarily, review of literature is conducted to summarize prior studies, identify stakeholders involved in highway infrastructure management, review and categorize major data satisfaction attributes in performing a requirement analysis as phase I. As a result of this, decision-makers and a data satisfaction module called SESP are defined for determining data quality of highway infrastructure management. The importance of satisfaction attributes and overall data use are then evaluated based on decision-makers' requirement through project team or decision-makers defined in Phase I using a questionnaire survey as part of data generation (Phase II).

In Phase III, a relationship is mapped to integrate the decision-makers with the attributes through mathematical set theory and Boolean algebra. The level of data requirements is determined at each individual level of the decision-making hierarchy through a multi-attribute approach of probability laws and concepts (fault tree analysis). Based on the satisfaction

attributes, data quality requirements are assessed at different levels of the decision-making hierarchy. This analysis helps in identifying at which level of decision-making hierarchy, data are well-utilized and communicated and determine the data quality improvements required by decision-makers' to effectively meet their needs.

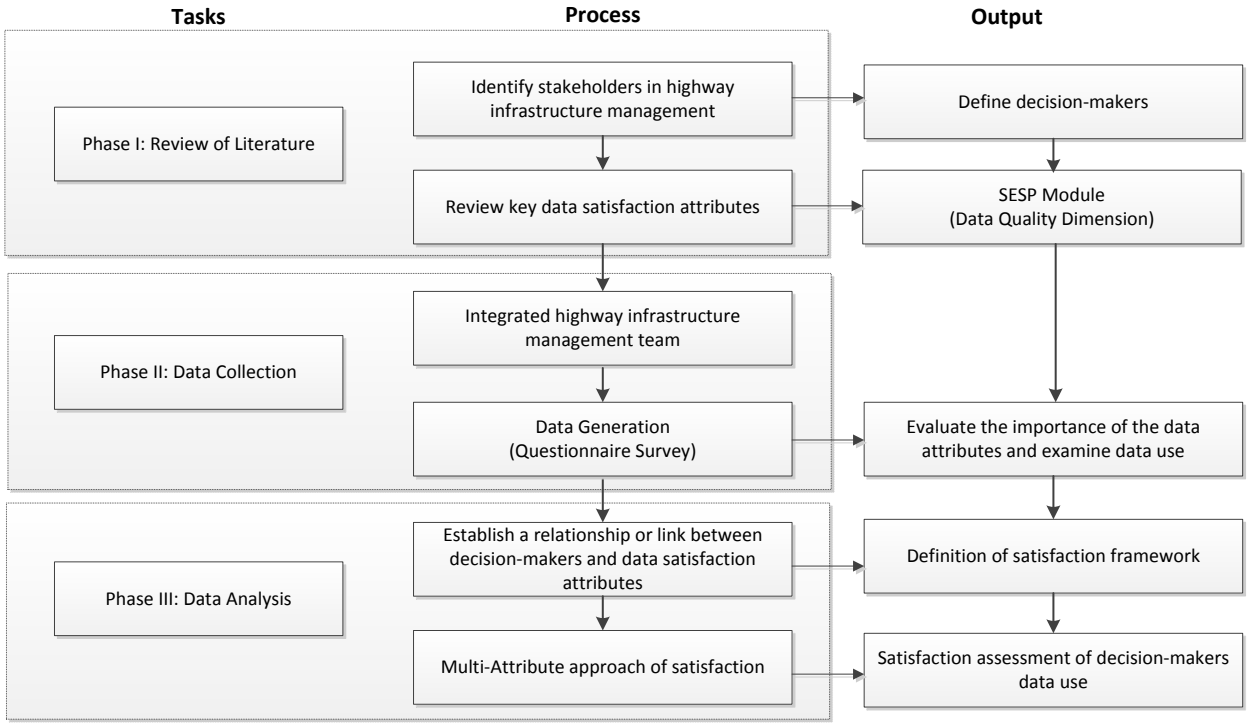



**Figure 3-1 Research Methodology for Data Quality Assessment**

## 3.3 Literature Review

Identifying the satisfaction of project participants is vital in determining the requirements and meeting the expectations of decision-makers and potential stakeholders. Te-King et al. (2003) argue that the success of a business is defined by how well it recognizes and satisfies its users and customers. Satisfaction can be said to have reached its goal if the desires and needs of stakeholders for a particular project, product or service are met. Although the study of stakeholder satisfaction originated in the marketing and psychology research, prior studies have

shown the satisfaction of project participants as an important measure and indicator for success in the construction sector as well (Cheung et al. 2000, Chan and Chan, 2004 and Nzekwe-Excel, 2009). Kamara and Anumba (2001) developed a client requirement processing model (CRPM) for processing client requirements in construction processes, while Mbachu and Nkado (2006) established a framework for examining client satisfaction during the design, management and construction services in the building process.

One of the issues that is associated with stakeholder satisfaction is the matter of identifying the underlying causes for dissatisfaction and preparing the appropriate remedy at a respective stage or level. The studies mentioned earlier address how satisfaction is an integral part or measure of a project success in terms of determining the satisfaction and sources of different stakeholders, project participants or each team's requirement at various stages in a more integrated manner. Vesely et al. (2002) argues that the ability to identify the cause of a particular event or predict the likelihood of occurrence of events is a critical element in managing risks and better planning. Studies by Wilemon and Baker (1983) and Ahmed and Kangari (1995) showed how cost, time, quality, customer orientation, communication skills and response to complaints as major components in measuring the satisfaction of stakeholders in meeting their project needs and requirements. According to Jang et al. (2003) material flow, schedule adherence, organizational structure and information flow are some of the major factors needed in creating satisfaction for the construction sector, while Tang et al. (2003) emphasized professionalism, competiveness, timeliness of service, quality of design, degree of innovation, completeness of factors considered, availability of support and supervision at implementation as key factors in evaluating construction sector's stakeholder satisfaction.

In identifying mechanisms to determine stakeholders' satisfaction and detect failure modes by incorporating multiple factors, various methods can been applied and implemented. Some of the potential methodologies available in evaluating stakeholders' satisfaction and detecting root cases of problems at the same time include failure mode and effect analysis (FMEA), fault tree analysis (FTA), fishbone diagram, failure modes, effects and criticality analysis (FMECA), quality function deployment (QFD) and event tree analysis. These methodologies have been previously utilized in analyzing failure, assessing satisfaction, risk identification and management (Vesely et al. 2002 and Nzekwe-Excel, 2008). Table 3-1 shows a brief summary of the methodologies.

**Table 3-1 Comparison of Potential Methodologies**

| Criteria | FTA | QFD | FMEA | Fishbone Diagram |
|---|---|---|---|---|
| Approach | Formal Deductive | Formal Deductive | Inductive | Informal Deductive |
| Link between customer need & organization goal | Yes | Yes | No | No |
| Corrective Action | No | No | Yes | No |
| Integration with other tools | Yes | Yes | Yes | Yes |
| Enables ranking of attributes | Yes | Yes | Yes | Yes |

Failure mode and effect analysis (FMEA) and event tree analysis are one of the most widely used techniques in identifying satisfaction and assessing failures. FMEA also known as Failure Analysis is an inductive approaches which helps in determining the resulting consequences of a failure and evaluating the risks associated with the failure. FMEA has the potential to prioritize attributes, integrate with other tools and assist in making corrective actions. FMEA has been widely utilized in aerospace, manufacturing, service sector and even construction in terms of determining the safety, reliability and repair cost of a system (O'Connor, 1995, Prince and Taylor, 2002, Anker, 2002). Although FMEA has a potential use in assessing

stakeholder satisfaction, it does not relatively help create a link between stakeholders' needs and organizations' requirements which is a key element of this study. Fishbone diagram is another potential methodology that can be used in identifying the satisfaction of stakeholders. Fishbone diagram also referred as cause and effect diagram or Ishakawa diagram is a method used for categorizing the cause of a problem to identify the root causes (Kenett, 2007). It is an informal deductive method used to identify potential causes of a failure in a product design and quality defect prevention. Although fishbone diagram has the ability to identify the cause and rank attributes, it is loosely-structured and only lists causes of a failure.

Quality function deployment (QFD) is another deductive approach which is widely used to translate customers' requirements into technical specification and design. QFD is a methodology practiced in the manufacturing industry to improve the quality of products to satisfy customers need and expectations (Lochner and Matar 1990). Akao (1990) defines QFD as "a method for developing a design quality aimed at satisfying the customer and then translating the customer's demands into design targets by utilizing quality assurance points throughout a production phase". The QFD is considered as a method to translate 'voice of the customer' to the 'voice of the designer' (Hauser 1993). QFD has been applied in construction industry in design-build contracts, conceptual design and renovation of housing projects (Abdul-Rahman et al., 1999, Pheng and Yeap, 2001 and Dikmen et al. 2005). QFD can be a potential methodology, but it is does not enable the integration of different attributes at various levels and lacks the possibility of identifying the critical elements in a system.

It is important to note that the aim of this study is to identify the satisfaction and causes for poor data usage to meet the requirements of decision-makers and organizational goals in which the fault tree analysis (FTA) fit appropriately for this study. FTA has the capacity to

identify the satisfaction and root cause of poor data usage by creating link between decision-makers need and organization goal which involves multiple stakeholders and functional divisions. This study implements a FTA to identify the satisfaction of current level of data use and determine the requirements of specific data quality in state DOTs in generating information and supporting decisions across the life-cycle of highway infrastructure management.

### 3.4 Fault Tree Analysis

Fault tree analysis (FTA) is a systematic deductive approach that follows a structured logical event based on a mathematical theory to identify the root causes or failures of an event. Vesely et al. (2002) describes FTA as "an analytical technique whereby an undesired state of the system is specified, usually from safety and reliability aspect where the system is analyzed in the context of its environment and operation to find all realistic ways in which an undesired event can occur". FTA is a qualitative and quantitative approach in such a way that it provides information on the cause of an event while predicting the probability of an event occurring and the importance of its causes (Vesely et al. 2002). It is one of the basic components in performing probabilistic risk assessment.

FTA was first developed by H. A. Watson of Bell Telephone Laboratories to evaluate missile launch control system for the United State Air force (Dhillon and Singh 1981). Previously, FTA has been widely utilized in the aerospace, nuclear, pharmaceutical, and chemical industries which involve high safety precaution to identify equipment/system failures, assess proposed design, identify effects of human errors, model risk assessments, and optimize tests and maintenances. Today, it is implemented in various industries including the medical industry and automobile industry to reduce patient healthcare and customer safety risks

respectively (Dhillon, 2003, Jetter et al. 2001). In the construction industry, FTA has been utilized to assess the satisfaction level of customers in public transport and project management team and estimate risk factors in project life-cycle (Strelcova 2007, Karaulova et al. 2008 and Nzekwe-Excel, 2008).

### 3.4.1 Fault Tree Structure

Fault tree analysis (FTA) is a top-down approach in which a top undesired event is broken down into components or a sequence of events until it reaches the initial cause (Vesely et al. 2002). Primarily, FTA consists of three entities; top event, intermediate event and basic events. These events are considered as the building blocks of a fault tree that are interlinked through logical gates such as "AND" and "OR" to determine the relationship between input and output events that is between the top, intermediate and basic events. These events and logical gates are represented by distinctive symbols. Figure 3-2 shows a simple structure of a fault tree.

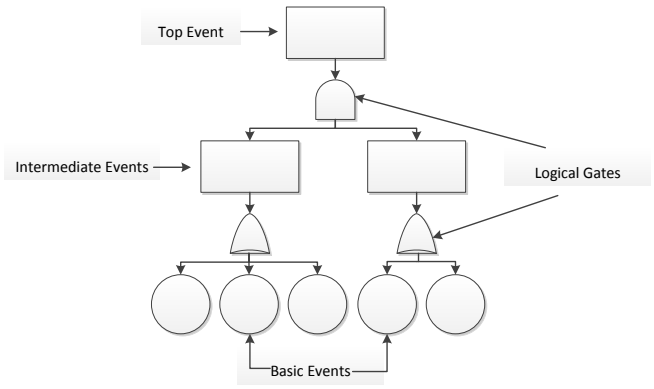



**Figure 3-2 Fault Tree Structure**

Primary Event Symbols:

i. Top /Intermediate Event: is the highest entity in the hierarchy which may consist of a series of intermediate, basic and undeveloped events whereas an intermediate event

may consist of basic and undeveloped events. It is represented by a rectangle shape (  ).

ii. Basic Event: is the final event or lowest entity of a cause. It is represented by a circular shape (  ).

iii. Undeveloped Event: is an event that is not developed due to insufficient consequence or unavailable information. It is represented by a diamond shape (  ).

Primary Gate Symbols:

iv. AND gate – results in an output if all the inputs are responsible for the fault to occur. This gate is represented by a bell shape (  ).

v. OR gate – results in an output if one or more of the inputs is responsible for the fault to occur. This gate is represented by inverted "V-shape" (  ).

Typically, the primary event symbols and gate symbols are the basic types of symbols in building a fault tree. However, there exists other event symbols (external event, conditioning event), gate symbols (exclusive OR gate, Priority AND gate, inhibit gate) and transfer symbols (in and out) used in special cases. For this study, the primary symbols are utilized to represent decision-makers' satisfaction of data usage to support highway infrastructure decisions.

## 3.5 Stakeholders in Highway Infrastructure Decision-Making Process

The first step in a requirement analysis is eliciting the requirements (Nuseibeh and Easterbrook, 2001). This is a process where stakeholders' requirements are gathered through business process documentation, interviews and/or surveys. Prior to gathering stakeholders' requirements, the stakeholders involved in highway infrastructure management and key data satisfaction attributes are determined based on interviews with DOT managers and review of

literature. In a typical highway infrastructure management, various stakeholders are involved in the planning, design, construction and maintenance phases of a project to enhance infrastructure life, improve quality, reduce cost, increase public-safety and meet transportation goals (Table 3-2). Stakeholders range from the public to federal transportation authorities. Stakeholders use data in some form knowingly or unknowingly during a highway infrastructure life-cycle. These stakeholders' data use can better be explained through the decisions they make.

**Table 3-2 Highway Infrastructure Stakeholders**

| Decision Level | Stakeholders | Decision-Makers | Decisions Use |
|---|---|---|---|
| Strategic Level | Commissioner, regulators, partners, etc. | Transportation Board or Committee, US DOT, FHWA, AASHTO, TRB, FTA, NHTSA, RITA, etc. | Set policy, develop guidelines (performance measure, assessment tools), allocate funds, develop program, decision support tools, best practice, etc. |
| Network Level | Program Managers (program level) | Capital Improvement, Local Governments, MPO, RPO, STIP, etc. | Administer transportation funds, identify project objective (long and short-term plan), determine priorities, develop program (3R and 4R), etc. |
| Program Level | Project Managers (project selection level) | Pavement Manager, Right-of-Way, Environmental, Bridge Manager, etc. | Project selection, safety improvement, traffic control, environmental studies, etc. |
| Project Level | Division Engineers | Scheduler, Designer, Superintendent, Maintenance Engineers, etc. | treatment selection, selection of design alternative (pavement type, thickness, bridge span), cost estimation, identify contract time, etc. |

Flintsch and Bryant (2006) classified highway infrastructure decisions into strategic, network, and project level from asset management perspective. Strategic level decisions deal with decisions made by higher level officials such as commissioners and directors along with regulators such as FHWA and AASHTO in setting system performance policies, developing guidelines and allocating funds. Partners such as Transportation Research Board (TRB) and Research and Innovative Technology Administration (RITA) utilize data to enhance program development and promote best practices. Network level decisions incorporate stakeholders such

as administrators and program managers that are responsible in developing long-term and short-term plans (set capital improvement plan), 3R (restoration, rehabilitation, and resurfacing) fund distribution, determining scope of agency, and transportation planning.

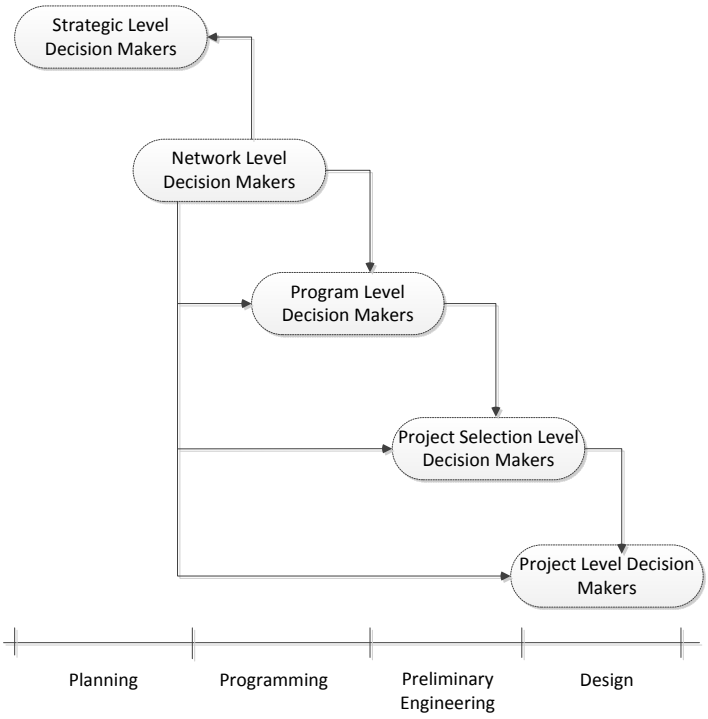



**Figure 3-3 Pavement Management Decision-Making Hierarchy (Flintsch and Bryant 2006)**

The network level is further broken-down into program level and project-selection level. The program level deals with evaluation and prioritization of projects, and administering of program while project-selection level deals with project selection, safety improvement and environmental studies at district level. Project-level decisions involve schedulers, designers, maintenance engineers responsible in the design and management of specific projects. Decisions can range from selection of design alternatives (treatment type, pavement type, thickness, bridge span length) to cost estimating, and contract time determination. In this study, pavement management decision-makers' at various decision-making hierarchy are considered as

stakeholders (data users or customers) in identifying the level of pavement condition data use by a state DOT (Figure 3-3).

## 3.6 SESP Module: Data Quality Satisfaction Attributes

One of the potential reasons that may be associated with the low use of data is quality (Kerr et al., 2007). Recently, the importance of data quality evaluation is receiving wide attention in meeting customers' requirements and meeting organizations goals. As part of this quality evaluation, various studies have addressed different measures and set of factors to improve the use of data. The United Kingdom Audit Commission (2007) identified six key characteristics to represent data quality; accuracy, validity, reliability, timeliness, relevance and completeness, while Loshin (2006) suggest eight dimensions: uniqueness, accuracy, consistency, completeness, timeliness, currency, conformance, and referential integrity. Meeting these quality dimensions help address potential users need and improve data use in making reliable decisions where good quality data portrays the generation of information and knowledge and data-driven insights.

In this study, a data quality satisfaction module called SESP (syntactic, empirics, semantics, and pragmatic) has been defined from a semiotics level to evaluate data quality in highway infrastructure management (Tejay et al. 2006). In a simple term, semiotics is the study of signs and symbols. Typically, signs consist of words, numbers, figures, and sentences that are used to mean, refer or portray messages across various users. Peirce (1998) defined sign as "something which stands to somebody for something in some respect or capacity". Similarly, data are represented by figures, characters, words, and numbers that are used to mean something as a means of communication with data users. According to Falkenberg et al. (1998) data is considered as meaningful symbolic constructs, which are a finite arrangement of signs and

symbols taken from an alphabet. The study of signs helps in understanding how data should be created, processed and used. As signs and symbols should be clear, significant, and well-defined that can be interpreted in the same manner, data should also be unique, consistent and have integrity. Thus, in this study, the concept of semiotics is used as a basis for defining data quality dimensions. The SESP module consists of eleven data satisfaction attributes or quality dimensions to determine the level of highway infrastructure data use from semiotics perspective. Table 3-3 lists the SESP module categorized under the four groups.

**Table 3-3 SESP Module: Key Data Quality Dimensions**

| Category | Attribute | Symbol | Description | Reference |
|---|---|---|---|---|
| Syntactic | Accurate | $SY_1$ | Data are precise and accurate | Delone and Mclean 1992, Wang and Strong, 1996 |
| | Consistency | $SY_2$ | Data are recorded in a consistent manner | Fox et al. 1994, Caby et al. 1995 and Wang and Strong, 1996 |
| | Completeness | $SY_3$ | Data is not missing and has sufficient depth and breadth | Fox et al, 1994, Gaby et al. 1995, Miller, 1996, Wang & Strong 1996 |
| | Structure | $SY_4$ | Data are in the right format and structure | Tejay et al. 2006 |
| | Integrity | $SY_5$ | Data reflect the full details of original observation | Brodie, 1980 |
| Empirics | Accessibility | $EM_1$ | Data are available and can easily be retrieved | Delone and Mclean 1992, Wang and Strong, 1996 |
| | Timeliness | $EM_2$ | Data are up-to-date | Fox et al. 1994 and Caby et al. 1995 |
| Semantics | Definition | $SE_1$ | Data are clearly defined in terms of its content | Caby et al. 1995 and Wang and Strong 1996 |
| | Ambiguity | $SE_2$ | Data are easily comprehended and interpreted in the same manner | Wand and Wang 1996 |
| Pragmatic | Relevant | $PR_1$ | Data are appropriate and applicable to support decisions | Delone & McLean 1992, Miller, 1996 and Wang & Strong 1996 |
| | Value | $PR_2$ | Data are beneficial and is useful | Wang & Strong 1996 |

1. Syntactic Dimension: deals with the structure and physical form of data rather than the content (Tejay et al. 2006). In this study, the syntactic data dimensions are characterized by five attributes; accuracy, consistency, completeness, structure and integrity.

a. *Accuracy:* deals with the conformity of data with the actual value that is collected either on the site or in the office. Accurate data represents correct, flawless, precise, reliable and

free of error (Delone and Mclean 1992 and Wang and Strong, 1996). For instance, the ODOT pavement management team uses severity levels (high, low and medium severity) to assess the condition of pavement and accuracy requirements (rutting depth should be within ±0.008 inches compared to manual survey with a resolution of 0.01 inches and minimum repeatability of ±0.008 inches for three repeat runs) to make it a representative data.

b. *Consistency*: Data should be continuously represented in the same format, compatible with previous data, succinct and compact in a continuous manner (Fox et al. 1994, Caby et al. 1995, and Wang and Strong, 1996). For instance, one percent difference in the areas of low-severity fatigue cracking can make 12 point difference in the 100-scale pavement condition index (PCI) based on ASTM D6433, Standard Practice for Roads and Parking Lots PCI Surveys (Pierce, 2013). Thus, data collected through different cycles should be consistent and make sense to perform reliable analysis.

c. *Completeness:* Data should include all relevant and necessary details required by the decision-maker to support his/her decisions. Data must be sufficient in breadth, depth and scope for the task at hand (Tejay et al. 2006). Proper care must be taken to monitor any missing, incomplete or duplicate data that may create problem in analyzing data and meet user requirement.

d. *Structure*: Data should be well-documented and recorded in the right format to perform a meaningful analysis. Data may be stored in paper format or pdf files which makes it difficult to extract information or perform analysis or in case of digital format, field names, value of data and the number of decimal places should be in the right format for a user to query data easily.

e. *Integrity*: Data should be sound and reflect the full details of an original observation such that data users such as analysts and managers feel confident in the data to make reliable decisions. Integrity is a measure of correctness of the implementation details (Brodie, 1980).

2. Empiric Dimension: deals with data that are used in repetitive manner to establish means of communication and data handling (Tejay et al. 2006). Data should be easily available at the right time for the user to make the right decision. In this study, empiric dimensions are represented by accessibility and timeliness.

f. *Accessibility:* Data should be available, accessible and easily retrieved to facilitate potential users' data need (Delone and Mclean 1992 and Wang and Strong, 1996).

g. *Timeliness*: Data should be recorded up-to-date and be available in a timely manner for the user to generate reliable information and make efficient decisions (Fox et al. 1994 and Caby et al. 1995).

3. Semantic Dimension: deals with the meaning of data in a certain context. In this study, semantic dimensions are represented by definition and ambiguity.

h. *Definition:* deals with clearly defining and representing data (Caby et al. 1995 and Wang and Strong 1996). Data should be well-defined with a common understanding and meaning in terms of its content. For instance, the level of pavement condition data collected by different agencies varies. Some agencies might collect the length of a longitudinal crack at centerline, while others collect the type, severity and extent of the crack at edges and centerline (Pierce et. al 2013). This level of variation would pose problems when using distress data. Data should be defined through the use of standards such as AASHTO and American Standard for Testing and Materials (ASTM).

i. *Ambiguity:* deals with data that may have more than one interpretation due to improper representation (Wand and Wang, 1996). Data should be easily comprehended and interpreted in the same manner. For instance, a classification of a transverse cracking into severity levels might create ambiguity as whether to classify the crack as transverse cracking or alligator cracking.

4. Pragmatic Dimension: deals with the implication of data from its usage perspective. In this study, pragmatic dimensions include relevancy and value.

j. *Relevance*: Data should be appropriate and applicable to meet its intended purpose or use in terms of generating information and supporting decisions (Delone & McLean 1992, Miller, 1996 and Wang & Strong 1996). Data that is of no use, should not be collected. For instance, use of the international roughness index (IRI) is becoming a huge concern among pavement management decision-makers in terms of whether it represents a reliable measure of pavement condition.

k. *Value*: Data should be worthwhile to the user. Wang and Strong (1996) suggested that the value of data should address the benefit and advantage of using data.

## 3.7 Data Collection

In this study, questionnaire surveys and interviews are used as the primary data collection method. The survey targeted highway agency decision-makers who utilize pavement condition data in their decision-making process which incorporate design engineers, project managers, and program managers. A total of eight experts participated in the survey. They include two pavement management decision-makers who have significant experience and deal with day-to-day pavement management decisions from Iowa DOT representing strategic level, network level

decisions, program level, and project level respectively. These experts determine the relative importance and level of influence by allocating weight to data satisfaction attributes. The level of importance helps create a stepwise relationship between data quality attributes and decision-making levels whereby priority is placed to meet the type of data quality dimension stakeholders require for their specific decision-making process. A data usage index is then defined to improve the data usage based on their satisfaction.

A five point Likert scale of 1-5 is used to note down decision-makers' level of importance for each attribute, where 1 refers to "no importance", 2 refers to "low importance", 3 refers to "somehow important", 4 refers to "important", and 5 refers to "high importance". In addition, a nine point Likert scale of 1-9, where 9 represents "strongly agree" and 1 refers to "strongly disagree" is used to evaluate the current level of satisfaction and quality of data use. This measurement determines the satisfaction requirement of the decision-makers and identifies the root cause for the minimal use of data in highway infrastructure decision-making processes. Appendix A illustrates the data collection sheet used in this study.

### 3.8 Data Analysis

Once data collection is completed, a multi-attribute approach is utilized to analyze the level of data usage among pavement management decision-makers by aggregating the SESP module (syntactic, empirics, semantics, pragmatics and external) or data quality dimensions. This assessment or multi-attribute approach is used to evaluate the decision-makers' current data satisfaction and allocate the weights to the data attributes in an integrated manner. It provides the option to incorporate all the attributes defined under the SESP module. Primarily, the relative importance of data attributes that may influence data usage is determined by the decision-maker.

Then, the current level of data usage is assessed for each data attribute. Finally, the overall data usage assessment is computed for the respective data quality dimension category defined under the SESP module. The sum of scores for all dimensions is considered as the satisfaction measurement or data usage assessment whereby a multi-attribute estimation is given by:

$$DU = DU_{sy} + DU_{em} + DU_{se} + DU_{pr} \quad \text{.....(3.1)}$$

Such that Eq. 3.1 can be further broken-down into;

$$DU = SY\sum_{i}^{N_{sy}} sy_i \;\; DU_{sy_i} + EM\sum_{i}^{N_{sy}} em_i \;\; DU_{em_i} + SE\sum_{i}^{N_{sy}} se_i \;\; DU_{se_i} + PR\sum_{i}^{N_{sy}} pr_i \;\; DU_{pr_i} \quad \text{.....(3.2)}$$

Where,

$DU$ = overall data usage

$DU_{\text{sy}}$ = Level of data usage with respect to syntactic quality

$DU_{\text{em}}$ = Level of data usage with respect to empirics quality

$DU_{se}$ = Level of data usage with respect to semantics quality

$DU_{pr}$ = Level of data usage with respect to pragmatic quality

$DU_{syi}$ = Syntactic assessment, whereas $sy_i$ is number of syntactic attributes ranging from 1 to N

$DU_{emi}$ = Empirics assessment, whereas $em_i$ is number of empirics attributes ranging from 1 to N

$DU_{sei}$ = Semantics assessment, whereas $se_i$ is number of semantic attributes ranging from 1 to N

$DU_{pri}$ = Pragmatic assessment, whereas $pr_i$ is number of pragmatic attributes ranging from 1 to N

*SY, EM, SE*, and *PR* represent relative importance indices.

Based on the multi-attribute approach, a data usage (*DU*) index for the data quality dimension is defined as summation of the product of decision-makers' importance rating ($DM_{IR}$)

and satisfaction percentage rating of a data usage attribute ($PR_{DU}\%$) out of the total number of responses involved in rating the data quality dimension. It is mathematically expressed as:

$$DU_i = \sum_{1}^{5} DM_{IR} \; X \; PR_{DU}\% \quad \text{.....(3.3)}$$

Where,

$DU_i$ = data usage index

$DM_{IR}$ = the decision-maker's importance rating based on a Likert scale of 1-5, where $1 < R < 5$

$PR_{DU}$ = the satisfaction percentage rating out based on a Likert scale of 1-9, where $1 < R < 9$ out of the total number of responses

### 3.8.1 Assimilated Linkage of Decision-Making Hierarchy

Having defined the satisfaction assessment through the SESP module of data quality dimensions, a technique should be established to assimilate the linkage between decision-makers at different levels. The linkage or relationship between decisions ensure integration, facilitates project delivery and enhances the efficiency of individual decision-making team whereby the success and failure of a project or an entire program depends on their contribution. In addition, this assimilated linkage provides an understanding of multiple stakeholders' view in meeting their requirements at different levels. This relationship can be systematically created in a more proactive manner through a construction of a fault tree.

Based on the pavement management decision-making hierarchy defined in Figure 3-3, a fault tree is constructed to illustrate the relationship of various decision-makers along with SESP data quality dimensions (Figure 3-4). Typically, agencies' decision-making process involves a team of project participants and progresses through a number of decision-making levels ranging from strategic level to project level. Each decision-making level are heterogeneous and the

requirements of the data quality needs of users are different. Though the level of decisions made at each level are different, identifying the decision-makers involved in these levels and their data quality satisfaction helps improve the requirements of the data users.

In this study, the linkage between these decision-making levels are integrated assuming that all decision-makers and potential data users have the same opportunity to contribute to the success of the overall agency. Integration of these users starts through communication and effective use of data their decision-making process whereby the decision of one level will influence the decision of another level. This can be explained by the fact that projects are initiated by project engineers at a project level and those projects undergo through various studies to identify maintenance requirements at program level. Once these requirements are identified, then selection of alternative methods and treatment selections are made through optimization at network level and finally budget allocation type of decisions are made at strategic level. This interdependency of decisions at various levels can be explained by a fault tree through the incorporation of data quality attributes.

The strategic level decision-making is defined at the top event. This is due to the fact that the strategic level decision-making is the backbone of the overall transportation program which deals with higher level decisions ranging from allocation of transportation funds to setting policies in meeting transportation goals. Network level, program level project selection level and project level are considered as intermediate events in a sequential manner (represented by a rectangle shape). The various data quality dimensions are taken as basic events for assessing the satisfaction (represented by circular shape). In this study, primary gate symbol “OR” is used to create the relationship between these decision-making levels or in other words connect the primary event symbols.

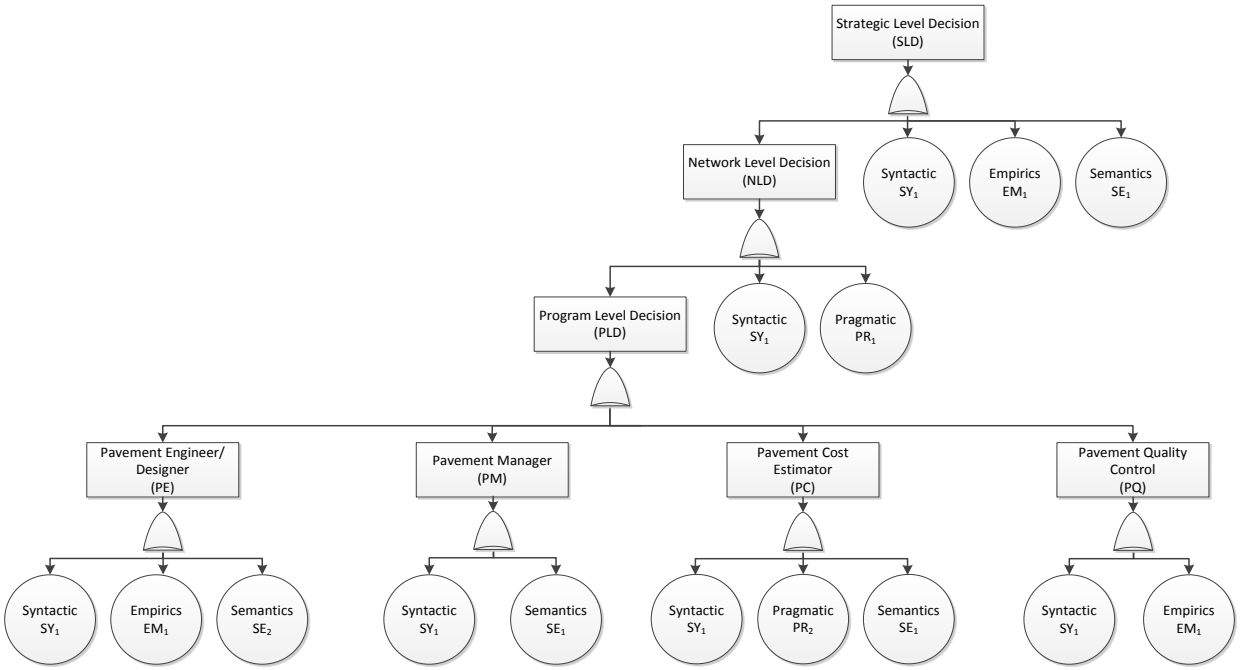



**Figure 3-4 Integrated Highway Decision-Making Hierarchy: Fault Tree**

Based on Figure 3-4, strategic level decision (*SLD*) being the top event is shown to be dependent on network level decision and three data quality dimensions of syntactic, empirics, and semantics which are represented by $SY_1$, $EM_1$ and $SE_1$ respectively. In other words, the success or failure of strategic level decision relies on the satisfaction of network level decision. In turn, the network level decision-makers (*NLD*) rely on program level decisions (*PLD*) and two data quality dimensions ($SY_1$, and $PR_1$) to meet the strategic level decision-makers (*SLD*) requirement. Similarly, program level decisions (*PLD*) depend on the outputs from project level decisions, pavement engineer (*PE*), pavement manager (*PM*), pavement cost estimator (*PC*) and pavement quality controller (*PQ*). The same procedure applies for *PE, PM, PC* and *PQ* to meet *PLD* requirements by meeting the respective data quality requirements.

Thus, the satisfaction of current data usage can be explained at the respective decision-making level and organizational level through this integrated system of fault tree construction. It

should be noted that in this study the success or failure of an organization is measured through the meeting of various stakeholders' requirements based on the SESP module to generate information and support decision-making processes through data-driven insights. It is also important to note that the use of hierarchical dependency on decision-making processes is assumed to be linear whereby upper level decisions are influenced by the inputs and analysis of lower level processes. In addition, the data quality dimensions shown in Figure 3-4 are used as an example for constructing a fault tree and a full analysis is shown on the results and discussion section later in the chapter.

The satisfaction assessment of this hierarchical and integrated approach is computed by combining three techniques; a mathematical set theory, Boolean algebra reductions and probability laws in an objective manner. The mathematical set theory is selected due to the fact that the output from the various decision-making levels developed in the fault tree can be ordered and structured in quantifying the probabilities of data quality satisfaction. The Boolean algebra is used to interpret the set operations used to interlink the decision-making levels, while probability laws are used to evaluate the data quality satisfaction of various levels assuming each data quality dimension requirements are distinct.

### 3.8.2 Mathematical Set Theory

The interdependency relationship between decision-makers is defined using mathematical set operations. These operations allow explain the interlinkage between the various decision-making levels in a stepwise manner. Basic logic gates are introduced whereby the operator union (∪) is used to represent "OR" gate and intersection (∩) is used to represent 'AND" gate. The operator '∪' is used to represent all decisions-makers and data quality dimensions required to assess the satisfaction of decision-makers, while the operator '∩' is used to represent the

decisions-makers and/or data quality dimensions that are mutually required to assess the satisfaction of decision-makers that is linking them. In this study, only the "OR" gate is used to represent all decision-makers' and data quality dimensions that are connected to them. For instance, based on the fault tree developed in Figure 3-4, a data usage relationship can be established mathematically as:

$$SLD = NLD\ U\ SY1\ U\ EM1\ U\ SE1 \quad \text{.....(3.4)}$$

$$NLD = PLD\ U\ SY1\ U\ PR \quad \text{.....(3.5)}$$

$$PE = SY1\ U\ EM1\ U\ SE1 \quad \text{.....(3.6)}$$

$$PM = SY1\ U\ SE1 \quad \text{.....(3.7)}$$

$$PC = SY1\ U\ PR2\ U\ SE1 \quad \text{.....(3.8)}$$

$$PQ = SY1\ U\ EM1 \quad \text{.....(3.9)}$$

Eq. 3.4 - 3.9 shows the satisfaction relationship developed based on the fault tree developed in Figure 4-4. For instance, in Eq. 3.4, the strategic level decision (*SLD*) data usage is met through the data quality attributes $SY_1$, $EM_1$ and $SE_1$ and network level decision (*NLD*). Likewise, the *NLD* data usage satisfaction is met through the integration of program level decision and $SY_1$, $EM_1$ and $SE_1$. The tree structure follows this trend until it meets basic event or data quality attributes.

### 3.8.3 Boolean Algebra

Based on the mathematical set operation, a Boolean reduction is used for union operator "+", and intersection operator ".". Applying Boolean algebra reduction to the mathematical set operators described in Eq. 3.4 to 3.9 can be expressed using Eq. 3.10 to 3.15 respectively:

$$SLD = NLD + SY1 + EM1 + SE1 \quad (3.10)$$

$$NLD = PLD + SY1 + PR \quad (3.11)$$

$$PE = SY1 + EM1 + SE1 \quad (3.12)$$

$$PM = SY1 + SE1 \quad (3.13)$$

$$PC = SY1 + PR2 + SE1 \quad (3.14)$$

$$PQ = SY1 + EM1 \quad (3.15)$$

### 3.8.4 Probability Laws

Then, a set of probability laws is applied for Boolean operators "+" and "." to determine the probability of data quality requirement satisfaction. The probability law is defined as:

$$P^{\cdot} = \prod_{i=1}^{n} P_i \quad (3.16)$$

$$P^{+\cdot} = 1 - \prod_{i=1}^{n} \{1 - P_i\} \quad (3.17)$$

Where,

$P^{\cdot}$ = Intersection output element

$P^{+}$ = Union output element

n = Total number of input elements

$P_i$ = Probability of data quality requirement

Applying probability law to the Boolean algebra of union output element results in the expression of:

$$P(SLD) = P(NLD + SY1 + EM1 + SE1) = 1 - \prod_{i=1}^{n} \{1 - P_i\} \quad (3.18)$$

$$P(NLD) = P(PLD + SY1 + PR) = 1 - \prod_{i=1}^{n} \{1 - P_i\} \quad (3.19)$$

$$P(PE) = P(SY1 + EM1 + SE1) = 1 - \prod_{i=1}^{n} \{1 - P_i\} \quad (3.20)$$

$$P(PM) = P(SY1 + SE1) = 1 - \prod_{i=1}^{n} \{1 - P_i\} \quad (3.21)$$

$$P(PC) = P(SY1 \ + \ PR2 \ + \ SE1) = 1 - \prod_{i=1}^{n} \{1 - P_i\} \quad \text{.....(3.22)}$$

$$P(PQ) = P(SY1 \ + \ EM1) = 1 - \prod_{i=1}^{n} \{1 - P_i\} \quad \text{.....(3.23)}$$

In these equations, it is important to note that probability law of independent events apply to the output of the satisfaction assessment. This implies that the satisfaction requirement of one data quality dimension does not affect the output of another data quality dimension requirement to signify the discrete nature of the dimensions defined under the SESP module.

### 3.8.5 Results and Discussions

In this section, the data collected from the experts is analyzed to quantify the satisfaction of data usage through the application of fault tree analysis. Table 3-4 shows a description of assessment measures and a survey result on the level of importance and the satisfaction level for SESP module data attributes obtained from highway decision-makers respectively.

**Table 3-4 Survey Result**

| Data Attribute | Strategic (SLD) | | | | Network (NLD) | | | | Program (PLD) | | Project Sel (PSLD) | | Project (PrLD) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | IL | DS | IL | DS | IL | D S | IL | DS | IL | DS | IL | DS | IL | DS |
| $SY_1$ | 4 | 7 | 5 | 7 | 3 | 7 | 5 | 7 | 4 | 7 | 4 | 6 | 5 | 4 |
| $SY_2$ | 5 | 7 | 5 | 8 | 5 | 8 | 5 | 7 | 5 | 7 | 4 | 6 | 4 | 3 |
| $SY_3$ | 4 | 8 | 5 | 5 | 5 | 7 | 4 | 8 | 4 | 7 | 3 | 6 | 4 | 3 |
| $SY_4$ | 5 | 7 | 5 | 7 | 5 | 8 | 4 | 6 | 4 | 6 | 3 | 4 | 3 | 4 |
| $SY_5$ | 5 | 4 | 5 | 7 | 4 | 8 | 5 | 8 | 4 | 7 | 4 | 3 | 4 | 3 |
| $EM_1$ | 5 | 2 | 5 | 5 | 5 | 6 | 5 | 6 | 5 | 4 | 5 | 3 | 5 | 5 |
| $EM_2$ | 4 | 2 | 5 | 6 | 4 | 8 | 5 | 8 | 4 | 4 | 4 | 8 | 3 | 3 |
| $SE_1$ | 5 | 3 | 5 | 7 | 5 | 9 | 4 | 6 | 5 | 5 | 5 | 3 | 4 | 4 |
| $SE_2$ | 4 | 4 | 5 | 7 | 4 | 8 | 4 | 6 | 4 | 6 | 4 | 3 | 4 | 5 |
| $PR_1$ | 5 | 6 | 5 | 8 | 5 | 9 | 5 | 7 | 5 | 5 | 5 | 5 | 4 | 3 |
| $PR_2$ | 5 | 7 | 5 | 8 | 4 | 8 | 5 | 7 | 4 | 6 | 5 | 7 | 5 | 5 |

IL – Importance Level DS - Degree of Satisfaction

Based on the level of importance, decision-makers believe that all data quality dimensions defined under the SESP module are considered to be critical in the use of data for generating information at all highway infrastructure decision-making hierarchy with a rating of 4 out of 5 (Table 3-4). However, the structure of data relatively does not have significant effect on decision-making processes at a program, project selection and project level. This may be due to the fact that at these levels there may not be a system that analyzes the collected data where data could be in text, image, or numerical format such that decision-makers use these data as inputs in their decision-making process based on their engineers judgment. In addition, completeness and timeliness of data are impartially important at project selection and project level or design of pavement decisions respectively. This may be a result of either pavement condition data are not significantly used at these levels or decision-making processes still persist through limited project scope in the early phases of a project planning and design.

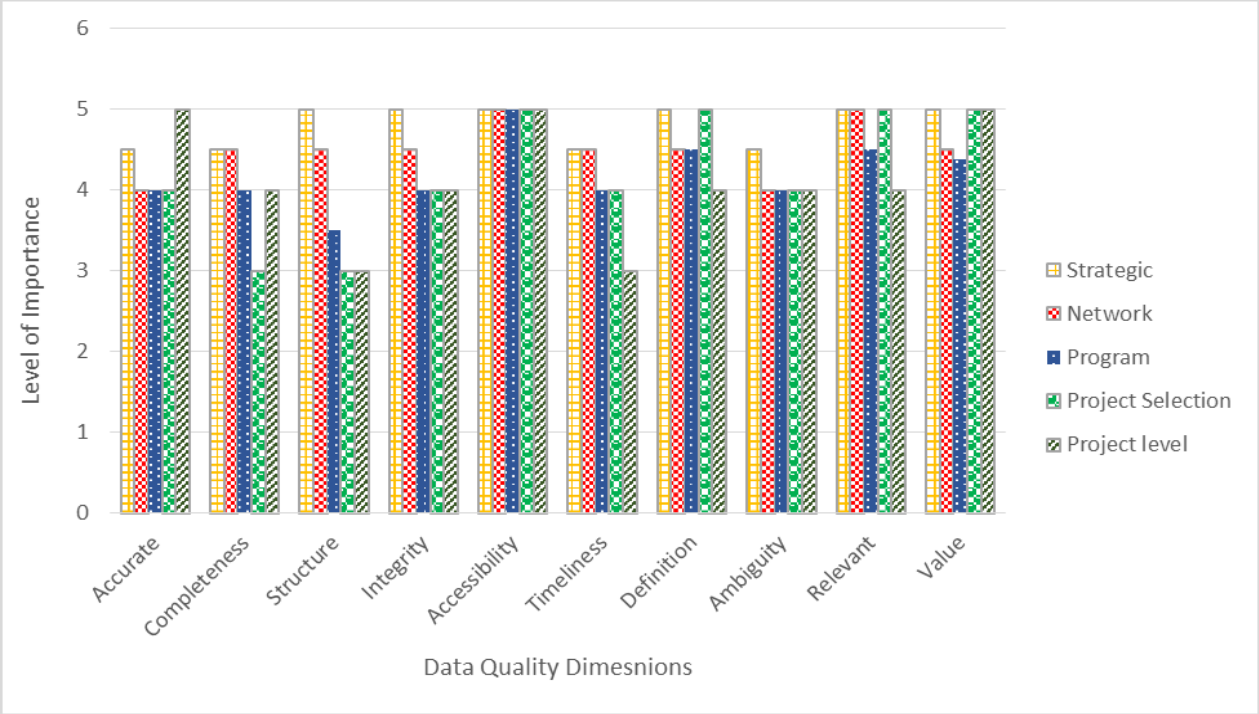



**Figure 3-5 Data Quality Dimensions Level of Importance**

Decsion-makers at all levels believe that data accesibility is one of the important data attributes in their decision-making processes. However, strategic, program, and project selection level decision-makers are not satisfied with the current level of data accessibility. This may be due to the lack of an integrated database system where data are collected, stored and managed by various divisions in decentralized manner across the project lifecycle. This creates difficulties for users to know what data are available in what division such that easy access, retireval, exchange and delivery of data is a challenge. In addition, obtaining up-to-date data at strategic, program and project levels are questionable as an integrated system is not currently available to facilitate data maintenance and update.

Project-selection level decision-makers are not satisfied with the structure, integrity, accessibility, definition and ambiguity of data, while decision-makers at design or project level are dissatisfied with regards to consistency, completeness, timiliness, relevancy and integity of data. However, decision-makers at startaegic, network and program level are relatively satisfied with respect to the consistency, completnesss, structure and integrity of data, while project selection level decision-makers requirements are well fulfilled with the timiliness of data. These results are due to the fact that pavement condition data are typically collected for federal aid eligible roadways at network level aiming to support strategic, network level and program level decisions. In addition, the Iowa DOT is comprised of reginal planning affiliations (RPAs) and metroplitan planning organizations (MPOs) such that the decisions to participate in the pavement condition data collection efforts are left to individual RPAs and MPOs mainly due to limited budget distribution and revenue generation differences. Furthemore, some of these agencies at project selection-level and project level use the data from network and program level as information only and use their own decision-making system due to social, environmental and

economical reasons such as detour cost, transportaion fees, and time factor assoicated in agricultural regions like the state of Iowa. Figures 3-6 and 3-7 show the satisfaction asssement results of current data usage.

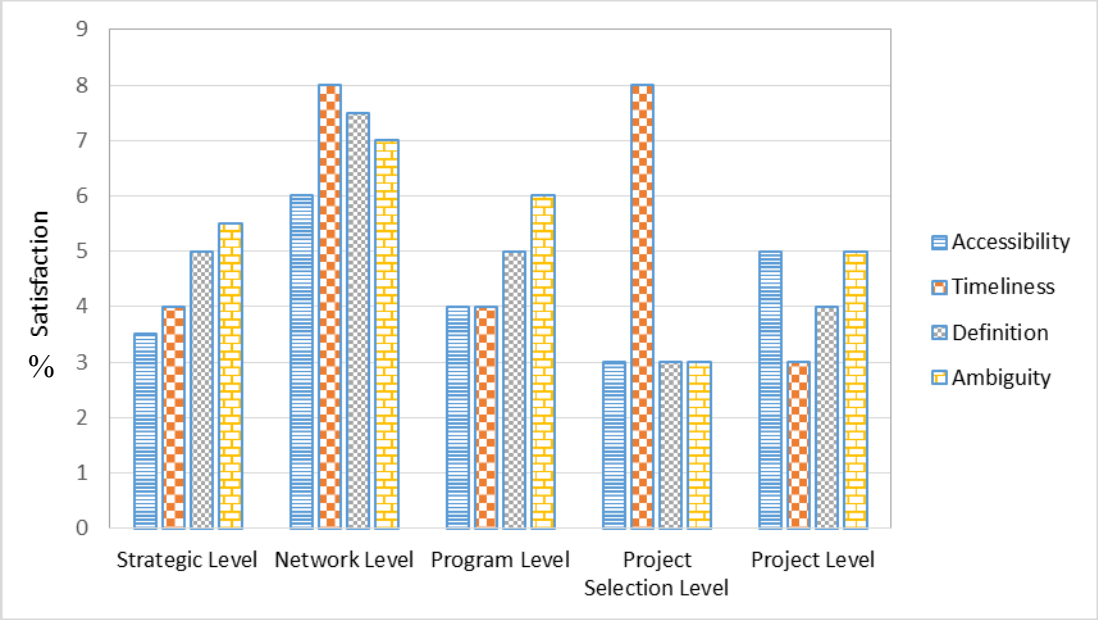



**Figure 3-6 Data Satisfaction I (Accessibility. Timeliness, Definition, Ambiguity)**

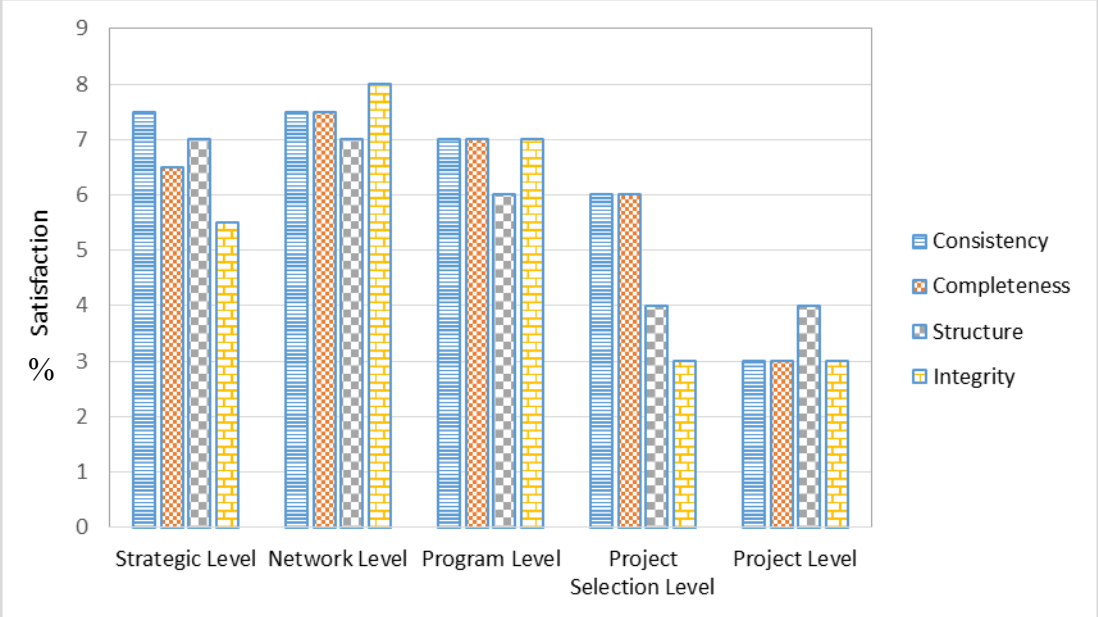



**Figure 3-7 Data Satisfaction II (Consistency, Completeness, Structure, Integrity)**

In order to quantify the data satisfaction requirement of highway agency decision-makers, a data usage (*DU*) index is developed using a multi-attribute approach by integrating the decision-making hierarchy with the level of importance and satisfaction. The data usage index is defines as the sum of the product of decision-makers' importance rating ($DM_{IR}$) and satisfaction percentage rating of a data usage attribute ($PR_{DU}$ *%*) out of the total number of responses involved in rating the data quality dimension defined in Eq. 3.3. Through the data usage index, the probability of satisfaction is estimated at a specific level of decision-making hierarchy and at an agency level. A proposition is first established based on the current practice of data usage from a highway agency decision-makers' perspective:

*Proposition 1: In highway infrastructure management, decision-makers' data quality requirement are well met in terms of generating information and supporting decisions.*

Table 3-6 shows the computed data usage index based on the results of the survey in Table 3-5. For instance, consider the ratings of a strategic level decision ($SLD_1$). The decision-maker might value the level of importance for accuracy ($SY_1$) of highway infrastructure data a rating of 4 based on a 5 point-scale and a satisfaction rating of 7 out of 9 for the current data usage. In this study, 11 attributes are considered with a satisfaction rating of a 9-scale measurement which results in a total of 99 points with the assumption that decision-makers' data requirements are fully met. Based on Eqn. 3.3, the satisfaction percentage rating point of $PR_{DU}$ for $SY_1$ is calculated as:

$$PR_{DU}Q_1 = \frac{7}{99}100\%$$

$$= 7.07$$

The data usage index $DU_i$ is calculated as:

$$DU_i = 4X7.07\%$$

$$= 0.283$$

Similarly, the data usage index for all data attributes is calculated and summarized in Table 3-5.

**Table 3-5 Data Usage Index**

| Data Attribute | $SLD_1$ | $SLD_2$ | $NLD_1$ | $NLD_2$ | PRD | PSLD | PLD |
|---|---|---|---|---|---|---|---|
| $SY_1$ | 0.283 | 0.354 | 0.212 | 0.354 | 0.283 | 0.242 | 0.202 |
| $SY_2$ | 0.354 | 0.404 | 0.404 | 0.354 | 0.327 | 0.242 | 0.121 |
| $SY_3$ | 0.323 | 0.253 | 0.354 | 0.323 | 0.283 | 0.182 | 0.121 |
| $SY_4$ | 0.354 | 0.354 | 0.404 | 0.242 | 0.212 | 0.121 | 0.121 |
| $SY_5$ | 0.202 | 0.354 | 0.323 | 0.404 | 0.283 | 0.121 | 0.121 |
| $EM_1$ | 0.101 | 0.253 | 0.303 | 0.303 | 0.202 | 0.152 | 0.253 |
| $EM_2$ | 0.081 | 0.303 | 0.323 | 0.404 | 0.162 | 0.323 | 0.091 |
| $SE_1$ | 0.152 | 0.354 | 0.455 | 0.242 | 0.227 | 0.152 | 0.162 |
| $SE_2$ | 0.162 | 0.354 | 0.323 | 0.242 | 0.242 | 0.121 | 0.202 |
| $PR_1$ | 0.303 | 0.404 | 0.455 | 0.354 | 0.227 | 0.253 | 0.121 |
| $PR_2$ | 0.354 | 0.404 | 0.323 | 0.354 | 0.265 | 0.354 | 0.253 |

Based on the mathematical Eq. 3.6 – 3.8, Boolean algebra Eq. 3.12 - 3.14 developed in the previous section, the probability of data quality needs required by a highway agency is estimated. The probability of project level or design engineer's satisfaction is calculated as:

$$P(PL) = P(SY1 + SY2 + SY3 + SY4 + SY5 + EM1 + EM2 + SE1 + SE2 + PR1 + PR2$$

$$= 1 - \prod_{i=1}^{n} \{1 - P_i\}$$

$$= 1 - \prod_{i=1}^{n} \{ [1 - P(SY_1)] [1 - P(SY_2)] [1 - P(SY_3)] [1 - P(SY_4)] [1 - P(SY_5)] [1 - P(EM_1)]$$

$$[1 - P(EM_2)][1 - P(SE_1)][1 - P(SE_2)][1 - P(PR_1)][1 - P(PR_2)] \}$$

$$= 1 - \{ [1 - 0.202] [1 - 0.121] [1 - 0.121] [1 - 0.121] [1 - 0.121] [1 - 0.202] [1 - 0.253] \}$$

$$[1-0.091]\,[1-0.162]\,[1-0.202]\,[1-0.121]\,[1-0.253]\,\}$$

$$= 0.8578$$

Similarly, the probability of program level is calculated as

$$P(PLD) = P(PE + SY1 + PR1\ )$$

$$= 1 - \prod_{i=1}^{n} \{1 - P_i\}$$

$$= 1 - \prod_{i=1}^{n} \{\ [1-P(SY_1)]\,[1-P(SY_1)]\,[1-P(SY_2)]\,[1-P(SY_3)]\,[1-P(SY_4)]\,[1-P(SY_5)]\}$$

$$[1-P(EM_1)]\,[1-P(EM_2)]\,[1-P(SE_1)]\,[1-P(SE_2)]\,[1-P(PR_1)]\,[1-P(PR_2)]\,\}$$

$$= 1 - \{\ [1-0.8578]\,[1-0.202]\,[1-0.121]\,[1-0.121]\,[1-0.121]\,[1-0.121]\,[1-0.253]\,\}$$

$$[1-0.091]\,[1-0.162]\,[1-0.202]\,[1-0.121]\,[1-0.253]\,\}$$

$$= 0.9249$$

**Table 3-6 Probability of Data Quality Satisfaction**

| Decision-Making Level | Probability |
|---|---|
| Strategic Level | 97.19% |
| Network Level | 98.96% |
| Program Level | 95.65% |
| Project Selection Level | 92.49% |
| Project Level | 85.78% |

Based on the probability estimate, strategic level decision (SLD) resulted in 0.9896. This output indicates that there is a 97% probability that the current data quality meets or satisfies decision-makers' requirements in terms of generating information and supporting decisions. At network level, this probability increases to 98.9% while, program level and project selection level decision-makers' requirements are met with a probability of 95.6% and 92.49% respectively and project level decision requirements are the lowest with a probability of 85.78%.

Although the probability difference between different levels in the decision-making hierarchy is not substantial, the result indicates that the pavement management data quality requirements are well met at network level, while improvement should be made at project level to satisfy decision-makers' requirements and improved use of data.

**Figure 3-8 Probability of Satisfaction for Individual Data Quality Dimensions -I**

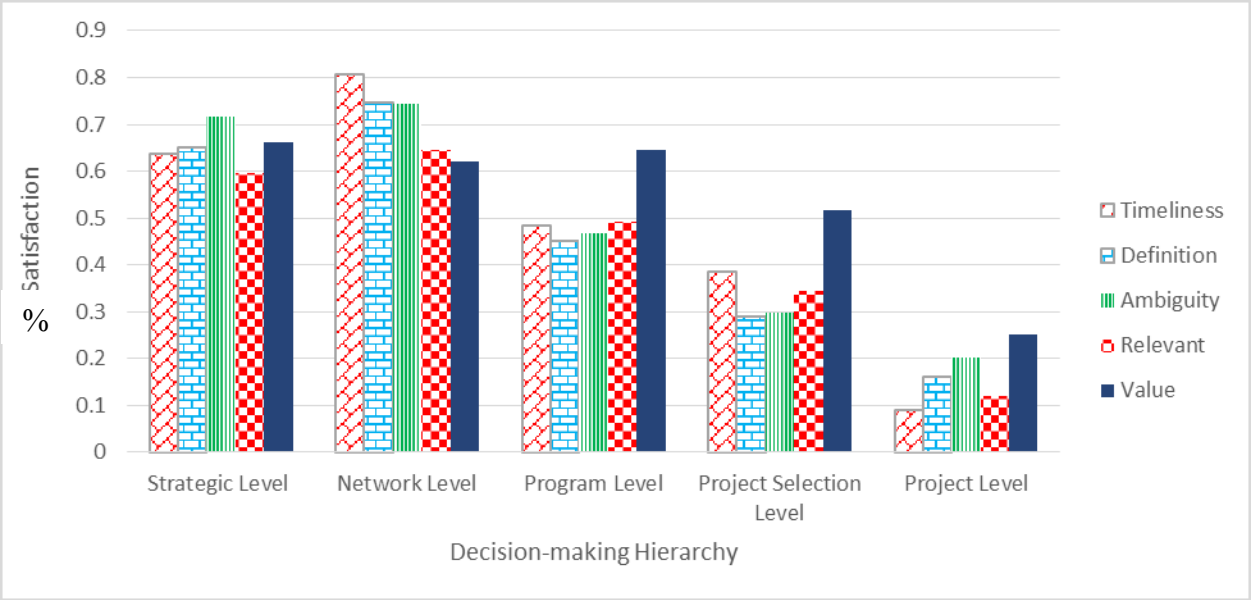



**Figure 3-9 Probability of Satisfaction for Individual Data Quality Dimensions -II**

The probability of level of satisfaction also shows that network level decisions requirement are well fulfilled with respect to individual data quality dimensions. Network level decisions are highly satisfied with the accessibility, timeliness, definition and ambiguity with a probability of 82.3%, 80.1%, 74.7% and 74.2% respectively. Based on the analysis, there is a decreasing trend of satisfaction when going from a strategic level decision to project level decision which indicates that most of the pavement management data are well utilized and understood at network and strategic level decisions as compared to project level decisions. This is due to the same reason of pavement condition data collection efforts decisions being left for

individual RPOs and MPOs as a result of budget constraints and are mainly collected for federally eligible roadways at network level. In turn this creates a gap in meeting decision-makers' requirements at other levels. In addition, the analysis shows that the completeness, structure and integrity of pavement management data quality may need improvement to meet users' need. Figures 3-8 and 3-9 illustrate the probabilities of satisfaction for individual data quality dimensions.

It is important to note that this study utilizes a semiotics-based data quality attributes to assess the satisfaction of data quality at different decision-making levels from the data users' perspective. This proactive satisfaction assessment of pavement management decision-making hierarchy allows data collectors to determine the level of data quality requirements of highway infrastructure managers and potential decision-makers in a more integrated manner. It will allow agencies' data management team to identify the causes behind the minimal usage of data to improve the quality in generating information and supporting decisions. In addition, the fault tree construction shows the interdependency of various decisions in the final output of a project/program and address the needs of potential data users. This approach can be applied to various data sets collected by various highway agency's divisions to meet the end users' need and help improve the overall data management.

# Chapter 4

# DATA AND INFORMATION INTEGRATION AND ASSESSMENT FRAMEWORK

## 4.1 Introduction

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. In highway infrastructure management, a number of decisions are made across a project life cycle, from the planning to the operation and maintenance phase and from the project level to strategic and network level. These decisions require different types of data and pieces of information collected, stored and managed in various database systems. This phenomenon results in an overlap and multiple use/dependency of data, information and decisions, which generate many-to-many relationships between these three entities across highway divisions, decision-making hierarchy and project phases. Raw data collected or information analyzed in one phase can be utilized as data and/or information in another phase, or information generated in one division can be utilized by another division or decision-making hierarchy to manage active projects and/or plan future projects. This creates difficulty in quantifying the efficiency, value and integration of data and information that are critical in supporting highway management decisions.

Previous chapter illustrated the use of a top-down approach to determine decision-makers' data requirements. The goal of this chapter is to present a new framework to systematically integrate and assess data use and enhance the efficiency in highway infrastructure

management from a decision-making standpoint. The proposed framework incorporates both top-down and bottom up approach to to attain three major objectives:

a. Interlink data with information and decisions
b. Identify key players in decision-making processes and
c. Determine the overall performance of highway infrastructure data management

A new performance measure called the Highway Infrastructure Data Integration (*HIDI*) index has been developed as part of this new framework in which a social network theory is utilized as a principal component in identifying new patterns between data, information and decisions within highway infrastructure and assessing their performance in lieu of facilitating integration by studying their correlation. The study uses pavement condition data, information generated from these data and decisions made in pavement management as an application of this new framework. Primarily, the chapter outlines the definition, evolution and vision for highway agencies' data and information management and prior studies conducted in data and information integration and assessment and application of social network theory.

## 4.2 Definition

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 the various data and pieces of information available from different sources and b) preparing the data and information in a usable and accurate form to make it available to the various end users. Integration will allow organizations and agencies 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. In addition, integration will act as a knowledge base. In this study, integration is viewed from data preparation in generating information and supporting highway decision-makings whereby

i. **Data** is defined as raw data collected during the life cycle of a highway infrastructure project and stored in data repositories or databases.

ii. **Information** refers to data that are processed, structured and generated through proper data analysis methods. Information is represented by key performance indicators or measures and outputs resulting from analysis of raw data.

iii. **Decision** refers to the selection or judgment process from a set of available alternatives based on data-driven insights by utilizing the collected raw data and information generated from the raw data.

## 4.3 Evolution of Data and Information Integration

The evolution of data and information management in the highway industry can be characterized by three categories in terms of project data collection, data storage (system) and usage (approach) utilized to generate information and support business decisions. Figure 4-1 conceptually illustrates three generations in data and information generation and utilization in the highway industry by using these three perspectives.

**1st Generation:**

The process of collecting raw data manually using paper based documentation and the utilization of judgment in making decisions is considered as the First Generation in data and

information management. The traditional process of collecting raw data has been in practice for long time for the purposes of keeping records, communicating and sharing data, reporting and dispute resolution. Typically, data were collected manually in a paper-based format, in which a project participant completes a report based on their activities and observations. These data are usually kept in a file cabinet or a storage room. In some cases, data are stored partially in a digital database system such as personal computers that are utilized by the user (project manager, engineer or designer) in some phases of a project. At this stage, users or decision-makers utilize judgment to make decisions. The user may look up the data collected to support the judgment, but does not use a structured approach to extract information and knowledge from the raw data.

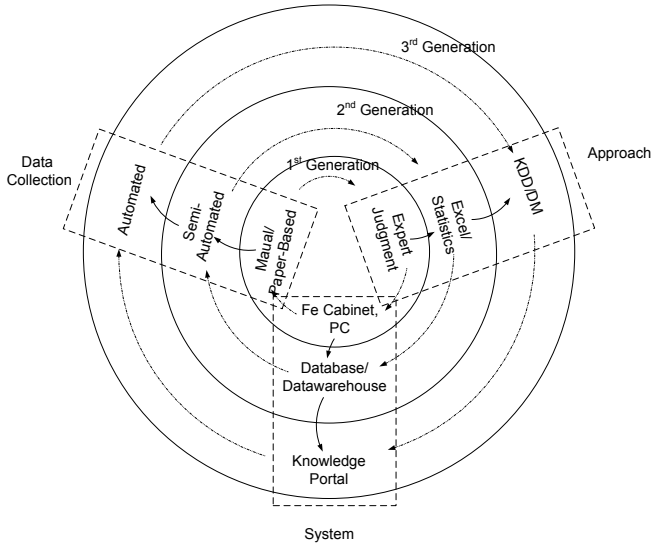



**Figure 4-1 Three Generation of Data and Information Management for Highway Infrastructure**

**2nd Generation:**

Use of technological measures and analytical techniques for improved extraction of information and knowledge can be considered as the Second Generation of data and information management. It is characterized by the transition and interpretation of raw data into valuable

information for use in the highway industry. This generation features the transition of paper-based data collection to semi-automated data, through the implementation of computer applications such as spreadsheets, and database systems. The use of manual data collection still persists, but automated technology systems such as sensors, smartphones, cameras, tablets, and geographical position system (GPS) have greatly improved data collection in terms of data quantity and quality. The advancement in database administration programs and data warehouses has also improved the way data are stored and managed, which provides easier access and retrieval by the user. In addition, the use of analytical techniques such as statistical methods, artificial intelligence and decision-support tools has greatly improved the ability to extract information and knowledge for specific decision support activities during the second generation.

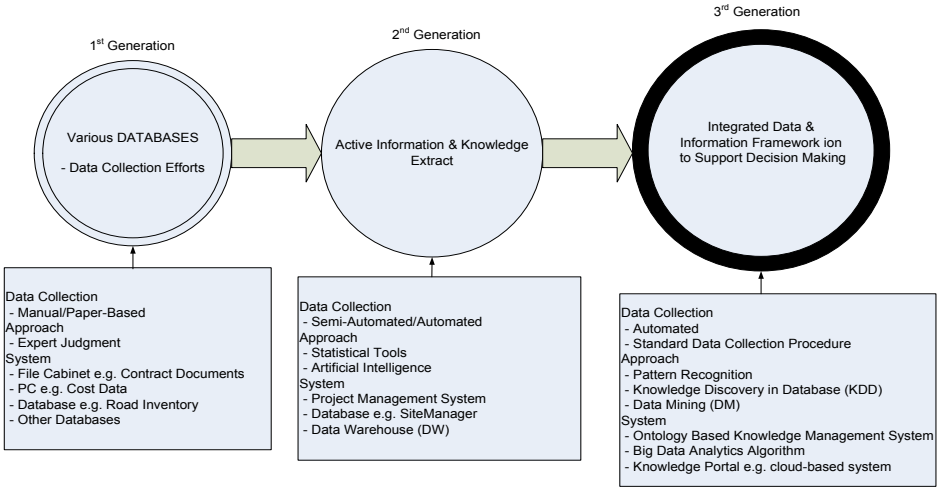



**Figure 4-2 Evolutions of Data and Information Integration for Highway Agencies**

**3rd Generation:**

The Third Generation is characterized by the emergence and applications of advanced knowledge management (KM) tools, big data analytics algorithms and knowledge discovery in database (KDD) approaches. It includes data mining (DM) techniques, machine learning and business intelligence tools, knowledge bases (KB), expert systems, cloud computing and

ontology-based frameworks. 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 generation, data are collected through automated systems and are actively utilized in generating information and knowledge to support various decisions. The Third Generation is greatly affecting the decision-making process and information system management in terms of capturing, storing, organizing, sharing, integrating, and communicating data. This generation has also improved pattern extraction as well as information dissemination at a high level and in a more efficient manner.

This evolution of data and information integration of state DOTs is summarized and illustrated in Figure 4-2. Many DOTs are considered to be in the transitional stage between the First and Second Generations. Although manual and paper-based types of data collection and the use of file cabinets and personal computers for data storage is shifting towards semi-automated/automated data collection systems and use of advanced database systems, a limited amount of effort has been put into extracting information and knowledge and actively utilizing the data. A limited number of decision-making processes are supported through the use of statistical tools and spreadsheets. The use of preliminary engineering (PE) contract fee proposal spreadsheet used by design divisions to negotiate PE costs with consulting firms and the collection of semi-automated/automated pavement condition data by asset management team for use in performance modeling, pavement treatment selection and prioritization of highway projects are good examples.

## 4.4 Data and Information Integration and Assessment Framework

Various concepts have been developed to address the efficiency and use of data in organizations. Conceptual data quality models and frameworks have been proposed to assess information systems through the incorporation of different characteristics and data/information quality dimensions (Khan et al. 2002; Shanks and Corbitt 1999 and Wang and Strong, 1996). Data and information management system measurements – which consist of system quality, information quality, service quality, and information economics – have been established (DeLone and McLean, 1992 and Knight and Burn, 2005). However, most of these studies focused on quality improvements of information systems and/or data quality from a data collection process perspective, and little effort has been put in to determining data efficiency from the potential users' or decision-making position.

In current practices, on one end, various data are collected across infrastructure life cycles and are stored in different database systems and are managed by multiple highway divisions. On the other end, various types of decisions are made across different decision-making hierarchy and different project phases. In the middle, data are being produced and replicated by different divisions or users to suit individual division or department needs, which causes data and information to be scattered everywhere and often identified to be missing or unknown. In addition, previous data studies focused on learning about a certain datum by associating the observation or case with its attribute where the datum is considered an independent occurrence (Scott and Carrington, 2011). For example, a specific datum can be associated with its location, type and size; one can describe the datum through its attributes. However, in this study, the datum is described through its relations: how the datum relates with another datum, how it is influenced/interacts with information and decisions in a system from a network approach rather

than an individual/mutually exclusive group where data are no longer considered independent. These data relations can be learned and measured through the integration and mapping of data with information and decision-making processes.

A three-tiered hierarchical framework is developed to explain the concept of data integration by matching and converting data into meaningful information and knowledge, which in turn support decisions in a set of graduated steps. The framework consists of raw data ($D_{ij}$) as Tier I, information ($I_i$) as Tier II and decisions ($DM_i$) as Tier III. These three tiers are integrated and mapped using hierarchical dependency and inclusive relationships based on decision-makers' requirements (Figure 4-3). Mapping these three-tiers or entities determines three types of paths. The first path is an active path that indicates active use of data currently being employed by highway agencies to generate information and support decisions. The second path is an inactive path, meaning that data are currently available (collected by agencies) but are not utilized in decision-making processes. As data required by decision-makers are available, transforming data into information would make an inactive path an active path. The third path is a non-existing path indicating that either data are not available for decision-makers to generate required information or information extraction method is not known to support decisions.

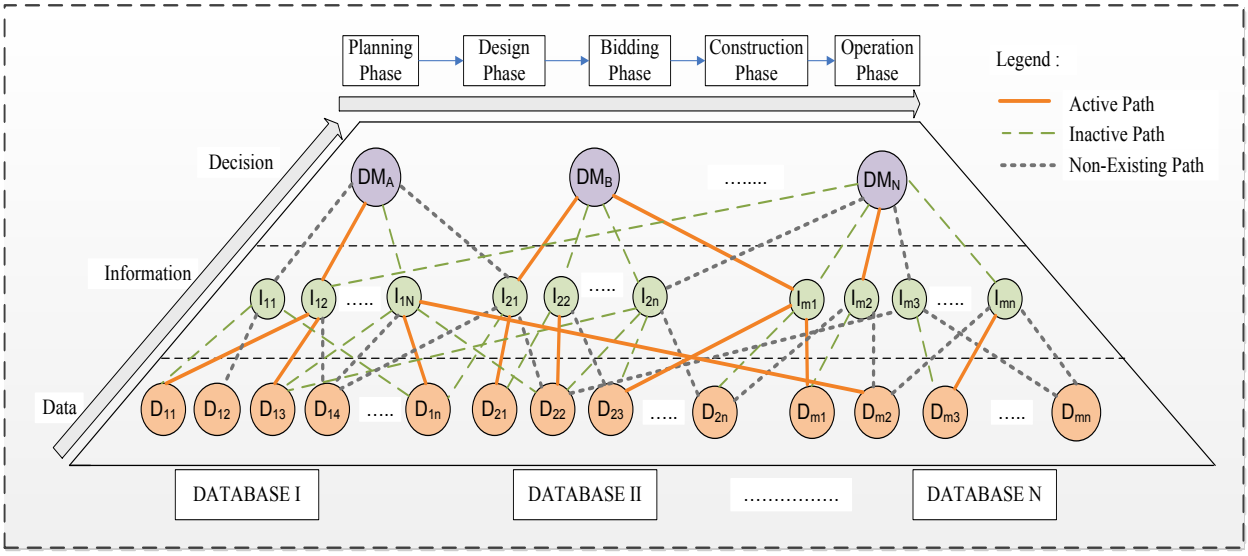



**Figure 4-3 Three-Tiered Hierarchical Framework**

These three types of paths are demonstrated in Figure 4-3 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. Once these mappings are figured, the framework (the three identified paths) is used as a basis for measuring the performance of state highway agencies data utilization and analyzing the gap between current status of data management and ideal data collection and information creation using four major components from a bottom-up and top-down approach. A bottom-up approach is used as an internal measure to provide insights on how individual data, pieces of information and decisions are embedded in the three-tiered framework. In other words, it determines the key players in highway infrastructure management. This bottom-up approach addresses the needs of decision-makers at a project level or network level by identifying the specific use and importance of data and information in supporting their day-to-day decision-making processes. In this study, three individual components are used as internal measures to

determine important data at Tier I, information at Tier II and decisions at Tier III in infrastructure decision-making processes respectively from a bottom-up approach.

The top-down approach is used as an external measure or fourth component to understanding the whole structure or data-information-decision integration framework that may be predictive of the data and information path dynamics. The approach is utilized as a principal instrument in identifying and assessing the overall performance of data, information and decision framework by understanding the interrelation or correlation of the paths as one system. Based on this fourth component and the use of the three paths identified from data, information and decision mappings or integration efforts, a new performance measure, the Highway Infrastructure Data Integration (*HIDI*) index, has been developed. This index helps higher-level decision-makers and the public evaluate the performance of data usage and visualize the overall framework. Figure 4-4 shows the framework for highway infrastructure data and information integration and assessment.

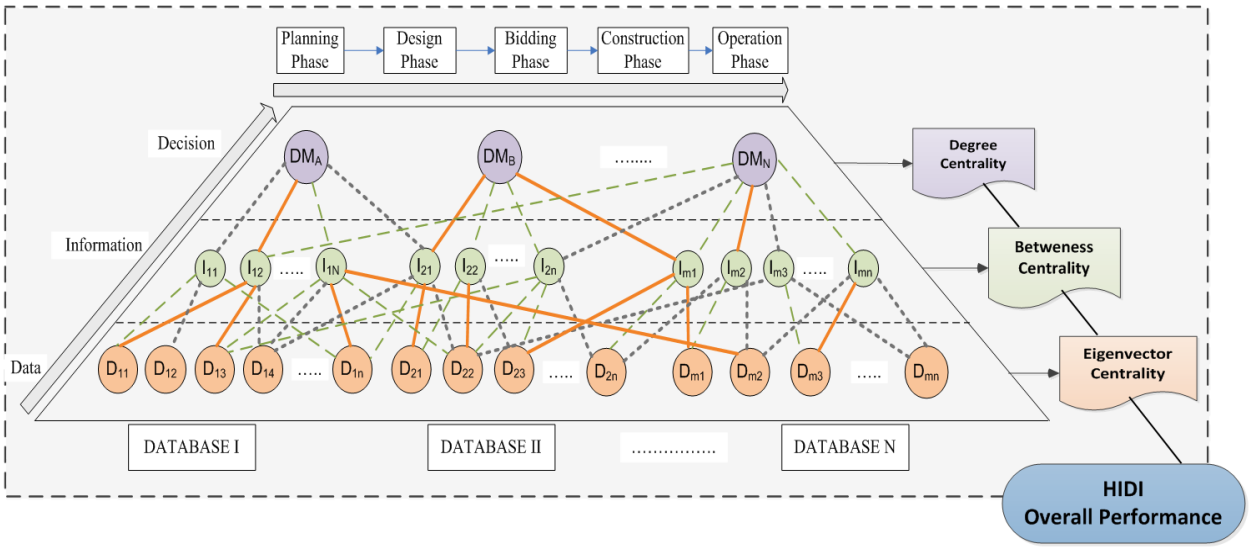



**Figure 4-4 Data and Information Integration and Assessment Framework**

In this study, social network theory or social network indicators of centrality measure degree centrality, betweenness centrality and eigenvector centrality. These are utilized as the three bottom-up approach components in determining the importance of data, information and decisions respectively. A cohesive measure or density indicator is used as the fourth component in assessing the overall integration framework from a top-down approach. This measure explicitly analyzes the relationship between highway infrastructure management data, information and decisions in meeting various SHA stakeholders' requirements and making reliable decisions through data-driven insights.

## 4.5 Social Network Theory

In a conventional social science, the focus of data study is to learn about a certain observation or case by associating it with its attribute where the observations are considered as independent occurrences. For instance, a data user can be associated with its demographic attributes such as age, sex or height of the subject where one can describe the user through these attributes. However, in social network theory, the user is described through its relation. It investigates how the user relates with another user, how it is influenced, and interacts within the network or a social institution. Social network theory studies this relation from a network approach rather than individual/mutually exclusive group. In this study, this concept of social network theory is utilized to identify patterns and determine interactions between data, information and decisions within the highway infrastructure network in lieu of facilitating integration, increased data use and improving decision-making efficiency where data are no longer considered as independent. In other words, the social network theory will be applied to

the concept of three data-information-decision paths as a means to quantitatively measure the interrelation or cohesion and strengths of the three paths.

Social network theory, or social network analysis (SNA), is a methodology used to identify the relationship among social entities by uncovering patterns and analyzing the interactions between them (Wasserman and Faust, 1997). Leinhadt (1977), Scott and Carrington (2011) argue that SNA is more of a paradigm for conceptualizing and analyzing social life that guides the selection of social behavior data, influences the way data are organized for analysis and specifies the questions addressed. However, Muller-Prothman and Finke (2004) define social network analysis as a systematic approach to identify, examine and support processes of knowledge sharing. In short, SNA can be considered as an approach or method for describing and explaining the social structure of a certain environment through a relational measure from a network approach rather than a separate entity.

The domain of SNA originated from sociology and social phycology to study the relations of social units originating through society. Moreno (1934) introduced the sociogram, a graphical demonstration of relationships as a formal representation of patterns to examine interpersonal relationship within groups. Later, anthropologists used this concept to understand the similarities and differences occurring in primitive cultures and tribal societies. This is where the term “social network” emerged (Barnes 1954; Chinowsky and Taylor 2012 and Mitchell 1969). However, Mitchell’s work (1969) shifted the approach through the introduction of graph theory, in which social interactions were mathematically expressed, which transformed SNA from a qualitative approach to a quantitative approach. This evolution has led to the development of standard measurements, mathematical analysis, modeling and algorithms (Freeman, 1977 and Granovetter, 1973).

**Table 4-1 SNA Application in Construction Industry (Timurcan and Dogan, 2013)**

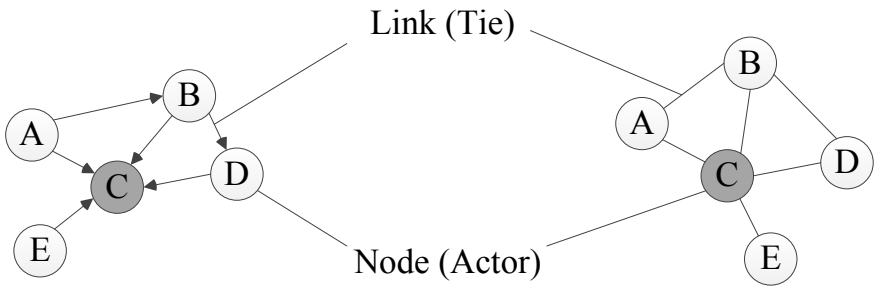
| Author | Social Network Study/Application | Contribution/Data Use | Network Metrics |
|---|---|---|---|
| Pryke, S. (2004) | Construction project coalition | Applied SNA as a new quantitative measure in UK construction procurement | Point centrality, degree and closeness centrality |
| Hossain, L. (2009) | Communication and coordination in construction projects | Explored the association between network centrality and coordination of construction project | Network centrality |
| Chinowsky et al. (2010) | Project organization | Utilized four engineering companies to introduce an approach to enhance trust and communication | Density, centrality, betweenness and power |
| Park et al. (2011) | Collaborative ventures in oversees construction projects | Employed 389 construction projects in Korean firms to provide collaborative strategies and measure level of performance | Degree, density, betweenness, closeness centrality and triad |
| Alsamadani et al. (2012) | Modeling and measuring safety communication in small work crews | Utilized nine construction firms in Denver metropolitan region to measure and model safety communication patterns | Density, centrality, betweenness |

Over the last three decades, SNA has been utilized in a wide range of applications, from computer and life sciences to law enforcement agencies that study internet traffic and webpages, explore food chains and identify criminal and terrorist networks (Scott and Carrington, 2011 and Wasserman and Faust, 1997). Even social media sites and search engines such as Facebook and Google utilize basic elements of SNA to recommend potential friends and rank web pages. In the construction industry, SNA has also been utilized recently to explore the coordination of projects, project organization, safety communication and the procurement and performance of construction firms (Alsamadani et al. 2012; Chinowsky et al. 2010; Hossain, 2009; Park et al. 2011 and Pryke 2004). Table 4-1 summarizes some of these studies in the construction industry. This trend reveals the potential use and application of SNA in the construction industry over the

last decade. However, most of these studies focused in communication, collaboration and the procurement and performance measurement of construction projects, and less attention has been given to the utilization of data integration in efficient decision-making for construction projects.

## 4.6 Social Network and Network Indicators

A social network is a set of socially relevant nodes connected by one or more links (Scott and Carrington, 2011). A node, or actor, may represent a subject, case or observation within a network, while a link or tie may represent interactions, flows, similarities or social relations between actors. Figure 4-5 shows an example of nodes and links in a social network. In a typical social network, the relation between actors can be represented using two types of ties: a directed or undirected network. In a directed network, one actor initiates while another receives; it investigates the relation between an active and passive actor, which is asymmetric. In an undirected network, a relation between two actors is mutual, or symmetric. A directed approach answers the question of who contacted whom, whereas an undirected approach answers the question of who knows whom. A directed network is represented by an arc (one-directional arrow) which can take an inward (input) relation or an outward (output) relation depending on the direction of the arrow, while an undirected network is represented by a line without arrowheads. For example, Figure 4-5a shows a directed network in which actor B has one inward tie and two outward ties with its neighboring actors, while Figure 4-5b shows an undirected network in which actor B is connected with three ties. In this study, an undirected network is utilized to investigate the mutual relationship between data, information and decisions with each entity being represented by a node or actor.




A. Directed Network B. Undirected Network

**Figure 4-5 Social Network**

Social network theory investigates the relationships and characteristics of a network based on various properties. Three types of properties are mainly used to assess a network: tie strength, key players and cohesion. Tie strength identifies the strength of connection, meaning whether a relationship has strong or weak ties in the network. Key players identify central nodes that play major roles in spreading information or influencing others in a network. Cohesion represents the measurement of the overall performance of a network structure. These properties have various indicators that help determine the characteristic of the network. For example, centrality indicators such as degree centrality, geodesic distance, betweenness centrality, closeness centrality and eigenvector centrality help identify key players by measuring how well nodes are connected in the network. Degree of reciprocity, density and clustering measure overall performance by evaluating the whole network (Wasserman and Faust, 1997, Scott and Carrington, 2011). For example, in Figure 4-5, actor C may be considered the critical or base actor (key player) as it is directly connected to all four neighboring actors (A, B, D and E), and a failure of actor C will break the relationship that actors A, B, and D have with actor E. In this study, four network indicators are utilized to develop a data and information integration and assessment framework.

***Degree Centrality*** is defined as the number of ties incident upon a node (Freeman, 1979 and Borgatti, 2005). It is one of the centrality measures used to describe the power and influence of a node based on its connection with other nodes in the network (Park et al. 2011). A degree centrality is the number of links that lead in or out of the node or the number of ties that a node has (Eq. 4.1). It is important to note that for an undirected network, the degree of centrality is identical. Degree centrality can be mathematically expressed as

$$DC_i = \frac{\sum_{j=1}^{n}(Z_{ij} + Z_{ji})}{\sum_{i=1}^{n}\sum_{j=1}^{n} Z_{ij}} \quad \text{.....4.1}$$

Where,

$DC_i$ *is the* deg*ree centrality such that* $0 \le DC_i \le 1$

$Z_{ij}$ *is the number of* deg*ree that a node i receives from a node j*

$Z_{ji}$ *is the number of* deg*ree that a node j receives from a node i*

$n$ *is the number of available nodes*

***Betweenness Centrality*** is defined as the share or number of times that a node 'i' needs a node 's' in order to reach to a node 't' via the shortest path (Freeman, 1979 and Borgatti, 2005). It is considered the mediator or brokerage which signifies the extent to which a node lies between other pairs of nodes (Kim, 2007 and Park et al. 2011). It is represented by a proportion of the number of shortest or geodesic paths that pass through a node to all the geodesic paths in the network (Eq. 4.2). It is mathematically expressed as:

$$BC_i = \sum_{s \ne t \ne i}^{n} \frac{\sigma_i(s,t)}{\sigma(s,t)} \quad \text{.....(4.2)}$$

Where,

$BC_i$ *is the betweenness centrality*

$\sigma_i(s,t)$ *is the number of shortest paths from node 's' to node 't' through node i*

$\sigma(s,t)$ *is the total number of shortest paths from node 's' to node 't'*

***Eigenvector Centrality*** is defined as the principal eigenvector of an adjacency matrix defining the network (Bonacich, 1972 and Borgatti, 2005). A node's eigenvector centrality is proportional to the sum of the eigenvector centralities of all nodes directly linked to it (Knoke and Yang, 2008). In other words, if a node influences another node and that node subsequently influences other nodes, the first node is influential in the network. Mathematically, it can be expressed as:

$$\lambda v = Av \quad \text{.....(4.3)}$$

Where,

$v$ *is the eigenvector*

$\lambda$ *is a constant* (*eigenvalue*)

$A$ *is the adjacency matrix of the network*

***Density*** is defined as the sum of ties divided by the number of possible ties (Scott and Carrington, 2011). Density is the most widely utilized measure and indicates how well a network is interconnected. It is mathematically expressed as the ratio of the number of existing relations or links to the maximum possible number of relations (Eq. 4.4):

$$D = . \frac{l}{n*(n-1)/2} Av \quad \text{.....4.4}$$

Where,

$l$ *is the number of existing links or ties*

$n$ *is the number of existing nodes*

## 4.7 Highway Infrastructure Data Integration (*HIDI*) Index

In the framework of integrating data-information-decision, the degree centrality indicator will identify important data attributes that are responsible for generating information and influencing decision-making processes, while the betwenness centrality indicator will help identify a key piece of information that is responsible for keeping a network intact. In determining decisions that are highly supported by data, information and/or other decisions, eigenvector centrality is employed. Thus, a data attribute participating in the integration framework that exhibits higher degree centrality is considered important data in highway infrastructure management, while a piece of information participating in the integration framework with higher betweenness centrality is considered critical in making better, more reliable decisions by employing data (acts as a communicator between data and decision). A decision participating in the integration framework that exhibits higher eigenvector centrality is assumed to reach a reliable conclusion through the support of data and information. In other words, data with the most number of connections with other data, pieces of information and decisions are considered key players in highway planning and management decisions, whereas pieces of information providing the most number of paths (creating a bridge) between data and decisions are considered critical. Decisions that create the most number of paths requiring information inputs and that in turn demand data inputs are considered data-driven decisions.

Based on Figure 4-3, the three types of paths are demonstrated using a solid line to represent active data usage, a dotted line to represent available but under-utilized data usage and a broken line for non-existing paths. In order to assess the overall performance of SHAs' data, information and decision integration framework (data and information utilization in supporting highway infrastructure decisions), the densities of these three paths are analyzed. The measure of

cohesiveness is used as an indicator whereby an efficient highway infrastructure data management is assumed to have a higher density measure that utilizes available data to generate reliable information and knowledge, which in turn support decisions. The density measure expressed in Eq. 4.4 is modified to Eq. 4.5 to represent the three paths which are expressed as the percentage ratio of the number of paths (either active, inactive or non-existing) to the total number of possible paths in measuring the overall data integration framework. This new criteria or measure is called the Highway Infrastructure Data Integration (*HIDI*) index.

$$HIDI_{i=1}^{3} = \frac{i}{l+m+n} \quad \text{.....4.5}$$

Where,

$l$ *is the number of existing or active paths*

$m$ *is the number of inactive paths*

$n$ *is the number of non – existing paths*

$i$ *is* $l, m, or\ n$

This *HIDI* index allows for an assessment of the relative proportions of active, inactive and non-existing paths in the data framework. For example, if the *HIDI* indexes for active path, inactive path and non-existing path result in 20%, 30% and 50% respectively, then a large amount of data are either missing, under-utilized, not in a usable format or do not meet decision-makers' requirements, as the percentage ratio of non-existing paths is higher than inactive and active paths. The assessment using *HIDI* indexes will allow agencies to evaluate their current data usage in terms of the data collection, the analysis and the management required to generate information and knowledge in support of decision-making processes through data-driven insights. Lee and Strong (2003) argue that the purpose of data production process is to produce data for data users and that it should measure the value of data as "data that are fit for use" by

data users. An agency can utilize this new index to reduce the cost and time associated with the data collection effort if some data attributes are identified as having no value to the decision-makers.

**Table 4-2 *HIDI* Grading System**

| Grade | Criteria | Grade Breakdown | Definition |
|---|---|---|---|
| A | $80\% \le HIDI_l$ and $HIDI_m, HIDI_n \le 20\%$ | $A^+$ if $HIDI_m > HIDI_n$ | *Highway infrastructure management are well supported through active utilization and integration of data and information and decision* |
| | | A if $HIDI_m = HIDI_n$ | |
| | | $A^-$ if $HIDI_m < HIDI_n$ | |
| B | $60\% \le HIDI_l \le 80\%$ and $HIDI_m, HIDI_n \le 40\%$ | $B^+$ if $HIDI_m > HIDI_n$ | *Highway infrastructure management collects data well, but requires active utilization* |
| | | B if $HIDI_m = HIDI_n$ | |
| | | $B^-$ if $HIDI_m < HIDI_n$ | |
| C | $40\% \le HIDI_l \le 60\%$ and $HIDI_m, HIDI_n \le 50\%$ | $C^+$ if $HIDI_m > HIDI_n$ | *Highway infrastructure management does not actively utilize data and needs major changes in terms of developing well defined method to generate information and support decisions* |
| | | C if $HIDI_m = HIDI_n$ | |
| | | $C^-$ if $HIDI_m < HIDI_n$ | |
| D | $20\% \le HIDI_l \le 40\%$ and $HIDI_m, HIDI_n \le 60\%$ | $D^+$ if $HIDI_m > HIDI_n$ | *Highway infrastructure management's current data and information use are questionable if they meet the standards or decision-makers' requirement* |
| | | D if $HIDI_m = HIDI_n$ | |
| | | $D^-$ if $HIDI_m < HIDI_n$ | |
| F | $HIDI_l \le 20\%$ and $HIDI_m, HIDI_n \le 80\%$ | $F^+$ if $HIDI_m > HIDI_n$ | *Highway infrastructure management needs new data collection and information generation plan* |
| | | F if $HIDI_m = HIDI_n$ | |
| | | $F^-$ if $HIDI_m < HIDI_n$ | |

Based on the results of the *HIDI*, a grading system is developed to easily understand the performance and to provide a holistic assessment of infrastructure data utilization by state highway agencies (Table 4-2). The grading system converts the quantitative assessment data into

different grades by assigning letter grades ranging from A to F depending on the percentage ratio of the active paths, inactive paths and non-existing paths to the overall framework. In the grading system, active path value is the determining factor, followed by the comparison of values of inactive and non-existing paths. For instance, grade "A" is assigned if the percentage of active paths to the overall paths (active, inactive and non-existing) accounts for greater or equal to 80%, and the percentage of inactive and non-existing paths are less than 20%, implying that current highway infrastructure management is well supported through the active utilization of data and generation of information (well integrated). In this grading system, "+" is used if the inactive path percentage ratio is greater than the non-existing path, and "-" is used if the non-existing path is greater than the inactive path. If this grading system were applied to the example in the previous paragraph of active paths (20%), inactive paths (30%) and non-existing paths (50%), the result will be the grade of "$D^{-}$" where the active path is between 20% and 40% such that the non-existing path (n) representing 50% is greater than the inactive path (m), which is 30% of the integration framework.

### 4.8 Application of Framework

In this section, the application of the framework is illustrated by considering pavement condition data, information generated from these data and decisions made in pavement management as a representative infrastructure decision–making process. For the purpose of this study, Iowa DOT's pavement management system is used as a representative data management system in applying the developed framework. Twenty-eight pavement condition data attributes classified into six categories: structural, functional, history roadway inventory, traffic and cost data. Five pieces of information (cracking index, structural index, pavement condition index, cost

analysis and deterioration model) and five representative decisions (treatment selection, project selection, rehabilitation or maintenance selection, project prioritization and 3R budget allocation) are used as potential actors in the framework.

**Table 4-3 Pavement Management Data**

| Source | Type of Data | Sub-Elements |
|---|---|---|
| Pavement Management System | Pavement History | Pavement surface type, thickness, composition. |
| | Structural Data | Transverse Cracking, Patching, Bleeding, Raveling, Fatigue, Polishing, Shoving |
| | Functional Data | Average Roughness, Ride, Average Rut depth |
| | Roadway Inventory | Pavement type, section, length, width |
| | Traffic Data | Traffic Profile, AADT |
| | Other (Structural) | Friction, Deflectometer (FWD), ESAL, Cost |

One of the potential data collected by highway agencies during the operation and maintenance phase are pavement management data (Table 4-3). These data are usually collected and managed as part of the asset management program through the pavement management system. Basically, a pavement management system consists of pavement history and pavement condition data. Pavement history is used to understand previous treatment applications in terms of the surface type, thickness, composition and treatment type, while pavement condition data includes functional and structural aspects. The functional data includes pavement rutting, roughness, ride quality, etc., whereas the structural data are comprised of pavement distress and stiffness such as longitudinal cracking, transverse cracking, patching and fatigue. In addition, roadway inventory, traffic and cost data such as pavement classification, pavement section, length, width, annual average daily traffic (AADT), traffic year and cost are also incorporated as supplemental data for this study.

In a typical highway infrastructure management program, these pavement condition data are converted into information. Some of these pieces of information include the characterization of current pavement conditions, development of pavement deterioration curves, life cycle cost analysis and projection of future conditions. For instance, the Iowa DOT uses pavement condition data such as ride, rut depth, structural and functional data to develop a pavement condition index (PCI) – a key performance indicator. In turn, this information is utilized to support various decisions at different levels and phases across the life cycle of a highway infrastructure. Some of these decisions include treatment selection, project prioritization, pavement design selection and allocation of funds to districts.

Flintsch and Bryant (2006) classified highway infrastructure decisions into strategic, network, and project level from an asset management perspective. Strategic level decisions deal with decisions made by higher level officials, such as commissioners and directors, in setting system performance policies, developing guidelines and allocating funds, while the network level incorporates development of long-term and short-term plans or capital improvement plans, 3R (restoration, rehabilitation and resurfacing), determination of scope and transportation planning made by administrators and program managers. The network level is further broken down into the program and project-selection levels. The program level deals with the evaluation and prioritization of projects and the administration of programs. The project-selection level deals with safety improvement and environmental studies at district levels. Project-level decisions involve schedulers, designers and engineers who are responsible for the design and maintenance of specific projects where decisions can range from the selection of design alternatives (treatment type, pavement type, thickness, bridge span length) to the estimation of cost and determination of contract time.

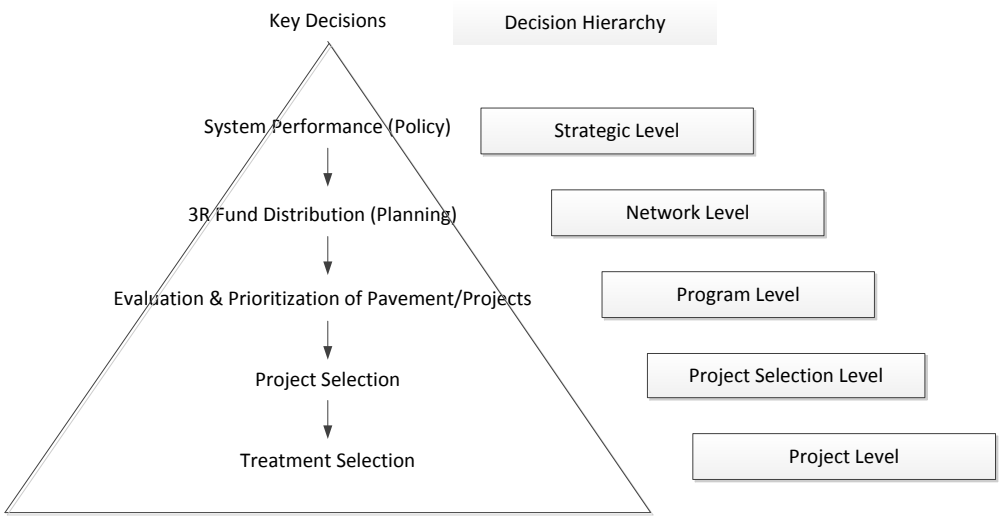



**Figure 4-6 Pavement Management Decisions at Different Levels**

This study selects five representative decisions based on Flintsch and Bryant's (2006) decision-making hierarchy (Figure 3), which are described as:

1. Treatment selection and timing – refers to determining an appropriate type of treatment need to be selected, as well as a proposed time frame for when the treatment will be placed.
2. Project selection – refers to projects selected for various pavement sections and their possible treatments.
3. Project prioritization – refers to the ranking of projects for scheduling purposes based on available funding to address the needs of the roadway system.
4. Rehabilitation or maintenance selection – refers to the selection of maintenance type based on the investment strategy and condition of the roadway system.
5. 3-R fund distribution – refers to the identification of funding levels needed by districts across a state.

### 4.8.1 Data Collection

Primarily, an expert panel consisting of five pavement management decision-makers from Iowa DOT representing different decision-making hierarchy listed in the previous section was formed. The expert panel included Matt Haubrich (strategic level), Chris Brakee and Francis Todey (network level), Thomas Tymkowicz (program level) and Ben Behnami (project level) who have significant knowledge and understanding in pavement management data, information and decisions. A pilot study was then conducted through a series of meetings, interviews and brain-storming sessions with the expert panel to acquire in-depth knowledge of current system, identify and define data, information and decisions used in pavement management system. Based on the pilot study, an adjacency matrix was designed to allow the experts interlink data with information and decisions and identify the three paths developed in the framework. As part of the matrix, a range of values were assigned to gather the information on the current status of data use in generating information and supporting pavement management decisions (1 representing active path, 2 representing inactive path and 3 representing non-existing paths). The adjacency matrix used in the data collection is attached as Appendix B.

### 4.8.2 Data Analysis

Based on the matrix, a graph network was used to represent and analyze the relations between these three entities. This analysis will help in identifying the current status of pavement management data-information-decision and ideal data management system. In addition, the analysis will allow determining important data, information and decisions that play key roles in the pavement management system. In order to perform this analysis, the study utilized Ucinet™, a SNA analysis tool to examine the data, information and decision-making integration framework because of its capability of developing network diagrams and analyzing network

indicators. Before analyzing and applying the data integration framework, it is important to note that the following assumptions were made:

i. Every datum, piece of information and decision is considered an individual node or actor.
ii. Data, information and decisions are considered one-level network.
iii. Undirected edges or lines without arrowheads are drawn between data, information and decisions to represent a mutual relationship or symmetric approach.
iv. A key data participating in the integration network which exhibits a higher degree centrality is critical in generating information and making reliable decisions.
v. A key information participating in the integration network which exhibits a higher betweenness centrality reaches at a reliable and better decision (interlinks data with decision well).
vi. A key decision participating in the integration network which exhibits a higher eigenvector centrality reaches at a reliable and better decision (well-supported with data and information).
vii. An efficient highway infrastructure data management is experienced by a higher density measure which utilizes data to generate reliable information and knowledge which in turn support decisions (data-driven insights).

#### 4.8.2.1 Primary Data

Based on the analysis, a data-information-decision integration framework or data network is developed (Figure 4-7). In the framework, data are represented by boxes (□), information are represented by circles (O) and decisions are represented by upper triangles (Δ). As stated in the

assumption, an undirected network is utilized to illustrate the mutual relationship between data, information and decisions.

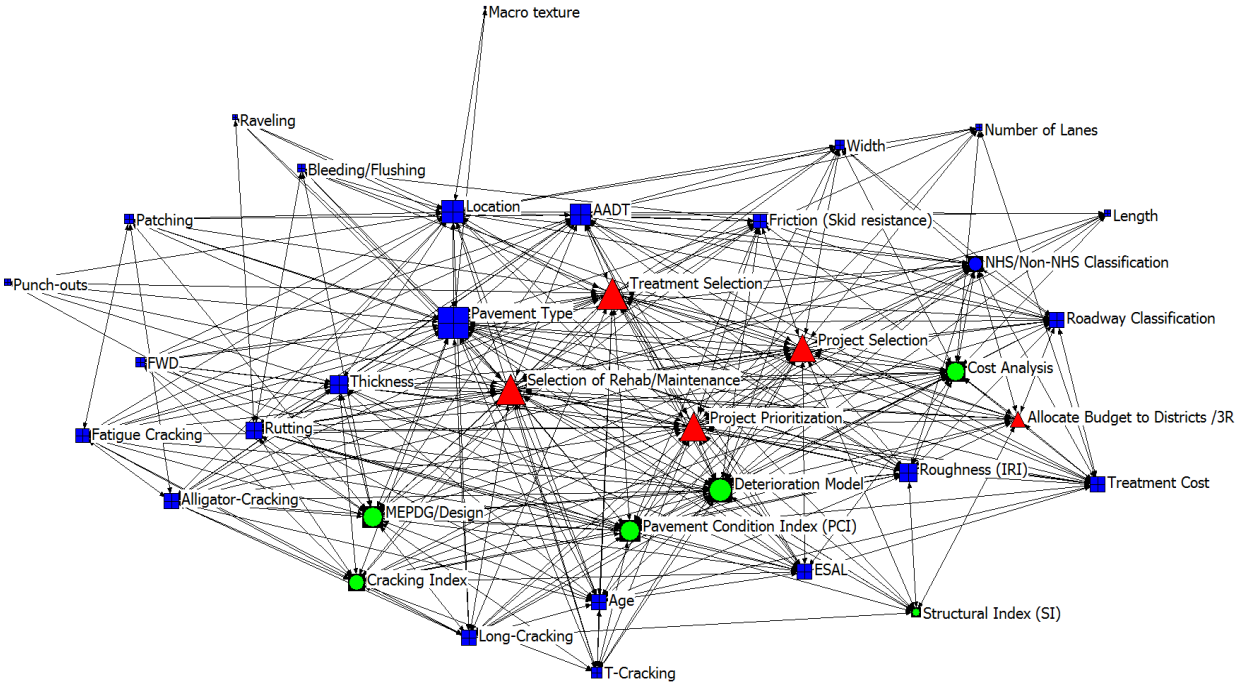



**Figure 4-7 Data-Information-Decision Integration Framework**

A degree centrality was used as a centrality indicator to identify important data that play a crucial role in generating information and supporting decisions. Based on the analysis, pavement type, location, annual average daily traffic (AADT), roughness, pavement thickness, rutting, alligator cracking, age, national highway classification (NHS) and treatment cost were determined to be critical (Figure 4-8). Although equivalent single axle loading (ESAL), falling weight deflectometer (FWD), patching, friction (skid resistance) and fatigue cracking are important data incorporated in the framework (are currently collected), their current usage is limited in generating information and supporting decisions indicating inactive data. Macro texture, number of lanes, transverse cracking, length and width of pavement are relatively not important data in pavement management decisions. The framework determined that

bleeding/flushing, raveling, and punch-outs are important highway data for the infrastructure pavement management decision-making process, which indicates the data are missing even though they are important indicators of pavement condition data.

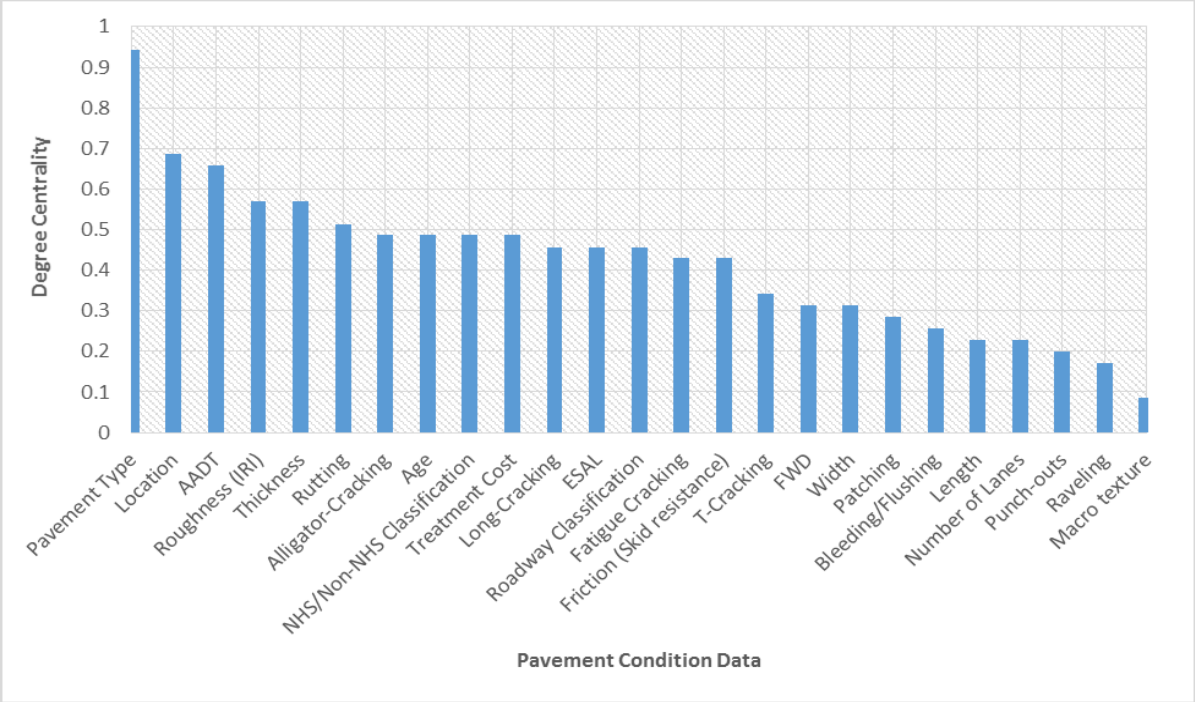



**Figure 4-8 Important Pavement Condition Data: Degree Centrality**

#### 4.8.2.2 Primary Information

Equal to the degree of centrality, a betweenness centrality is used as an indicator to identify key information that keeps the data and decisions bond tight. Based on the betweenness centrality measure, pavement condition index (PCI), cost analysis and mechanistic-empirical pavement design guide (MEPDG or design analysis) are crucial in integrating data with highway infrastructure decisions (Figure 4-9). In other words, these pieces of information utilize available data to generate insights for decision-making processes. As a piece of information has more data points connected through it to decisions in the shortest or most geodesic paths possible, that information is considered to be critical. However, it is important to note that a piece of

information that connects only one datum with one decision should also be taken into account which is one of the limitation of this analysis.

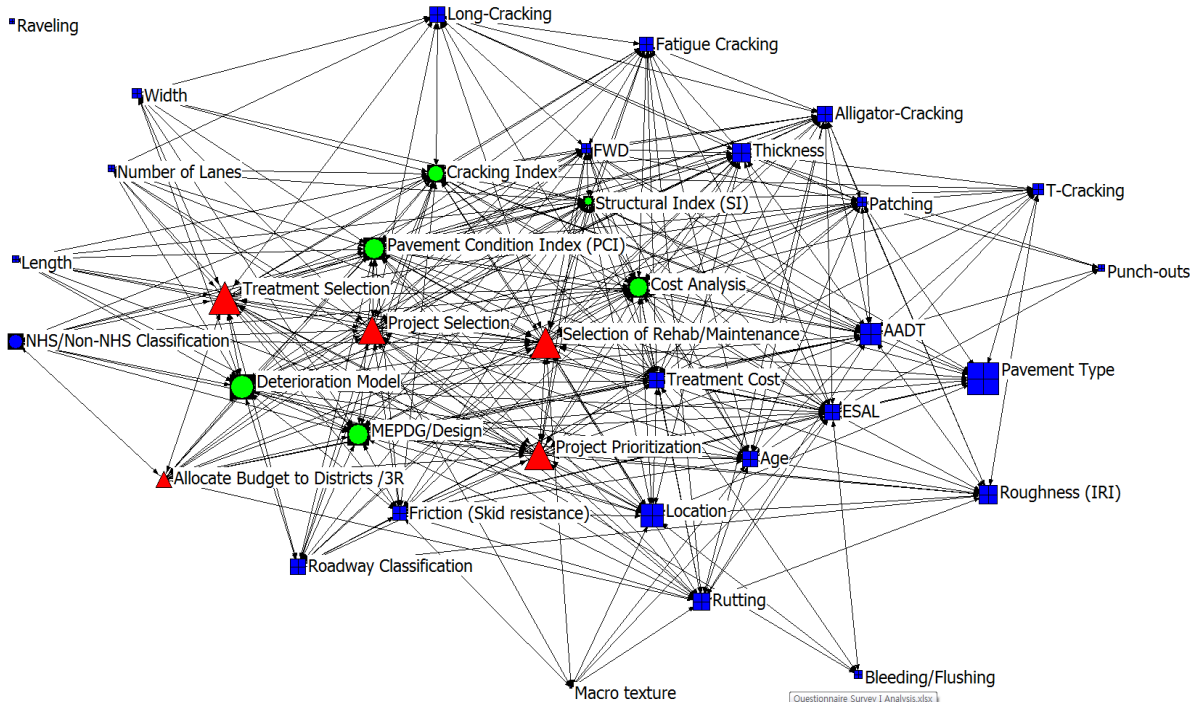



**Figure 4-9 Important Pieces of Information: Betweenness Centrality)**

Although, structural index, cracking index and deterioration model are critical pieces of information, where these pieces of information are either missing or their information extraction method is unknown. For instance, cracking index can be developed by integrating the four cracking data (longitudinal, alligator, fatigue and transverse cracking), and pavement type data along with deterioration model and pavement condition index information which in turn can be used in treatment selection and selection of projects. A detailed analysis of these information paths is discussed in the next chapter as part of a gap analysis.

#### 4.8.2.3 Primary Decisions

Another indicator that allows for the identification of key players in the decision-making integration framework is eigenvector centrality. This indicator is used to assess the efficiency of

pavement management decisions in terms of utilizing data and information. In this study, a decision with high eigenvector centrality is considered well-connected to other decisions or pieces of information with high eigenvector centrality, or those with the most connection. The concept behind eigenvector centrality is that if a node influences another node and that node subsequently influences other nodes, the first node is influential in the network. Thus a decision with higher eigenvector centrality is supported by other decisions or utilizes pieces of information, which in turn utilize data.

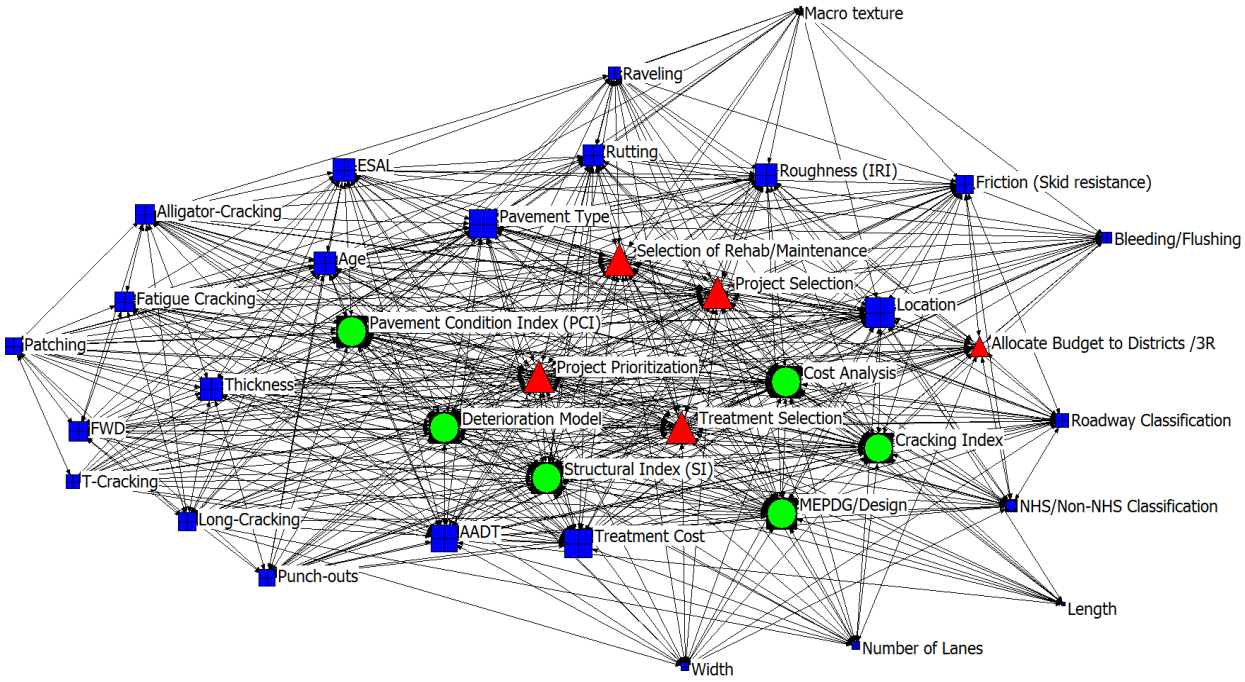



**Figure 4-10 Important Decisions: Eigen Vector Centrality**

Based on the network analysis, the selection of treatment (0.371), rehabilitation and maintenance selection (0.364), project prioritization (0.354), project selection (0.344) and allocation of budget (0218) are well interlinked with data, information and each other in the highway infrastructure decision-making process. This close value of decisions' eigenvector analysis can be explained by the fact that these decision-makings tend to operate in an iterative

process, in which multiple iterations are needed to come to the best solution. Treatment selection may change based on the timing of the project or the available funding. This change impacts the other decisions that will be made. Part of this new framework is a test to verify the impact the investment has on the entire highway infrastructure data management system. This integration is used to optimize the decision-making process to get the greatest increase in data and information use for the funding invested.

#### 4.8.2.4 *HIDI*

Once the internal measures are employed to determine important data, information and decisions, the next step is to assess the overall data integration framework based on the new performance measure, the Highway Infrastructure Data Integration (*HIDI*) index. Table 4-4 shows the density measures of the Highway Infrastructure Data Infrastructure index. The analysis resulted in an *HIDI* of 40.4% for active path, while inactive and non-existing path results account for 41.1% and 18.7% respectively. A perfectly integrated framework has a density of 1. It is important to note that the densities described in Table 4-4 are the results measured with respect to the maximum possible relation in the framework. However, a relative measure of densities based on Eq. 4.4 or the number of ties (paths) based on Eq. 4.5 shall result in the same *HIDI*. The low value of the densities estimated based on Eq. 4.4 is due to the fact that all data, information and decisions are not correlated to each other.

**Table 4-4 *HIDI***

| Path | Density | Number of Ties | *HIDI* |
|---|---|---|---|
| **Active (l)** | 0.471 | 594 | **0.402** |
| **Inactive (m)** | 0.481 | 606 | **0.411** |
| **Non-existing (n)** | 0.219 | 276 | **0.187** |
| **Total** | 1.17 | 1476 | 1 |

This *HIDI* index indicates that more than 80% of the data required by decision-makers to generate information and support pavement management decisions are available in the current data management system. However, applying the proposed grading system defined in Table 4-2, the overall highway infrastructure data integration framework or system will receive a grade of 'C+,' with the inactive path accounting for more than 41.1% of the integration framework whereas which is greater than active (40.2%) and non-existing (18.7%) paths. This implies that although the current infrastructure management is rich with data, the amount of information extracted out of these data to support decisions is minimal which requires major changes in terms of establishing a well-defined method to generate information and support decisions.

This grading system allows agencies to measure their data performances in terms of their use by integrating it with information and decisions. It will help identify new correlations between data, information and decisions. This new system improves the way data are collected, stored and managed in highway infrastructure management in which agencies might be able to identify the potential uses of data and information in supporting decision-making processes. This new framework can be applicable to other highway infrastructure data to improve utilization and help justify the return on investment in data collection efforts.

# Chapter 5

# CASE STUDIES

## 5.1 Introduction

This chapter presents two case studies: a gap analysis between current data utilization and an ideal data-and information integration to illustrate the ultimate benefit of the three tiered data-information-decision framework developed in the previous chapter and a validation of the framework and the highway infrastructure data integration (*HIDI*) index to evaluate the effectiveness of the integration and assessment framework by implementing it on a different data management system. The study uses pavement condition data to perform the gap analysis (case study I) and preconstruction service cost data management as part of the validation (case study-II) in implementing the framework to identify key decisions made, data and information utilized, map the data and information flow and evaluate the overall performance.

## 5.2 Case Study I - Gap Analysis

A gap analysis is a technique that is used to determine the steps required to meet a desired state by identifying the current state or present situation of system. In this study, a gap analysis is performed to identify the ideal pavement management data-information-decision integration system (desired state) based on its current state of data usage. In order to assess this gap, the new data and information integration and assessment framework developed in the previous chapter is used as main component for conducting a gap analysis. Based on the three paths identified (active, inactive and missing) in the framework and the social network analysis results of pavement management experts, the steps needed to achieve an ideal pavement

management data-information-decision is determined. Basically, network level decision-makers are used as representative example to identify the active, inactive and missing data, information and decision paths where integration of these paths result in an ideal data-information-decision system. This gap analysis helps to determine possible underutilized and missing data and identify unknown pieces of information that are not part of the current system. It will help evaluate where agencies are standing today and where they should head in the future with respect to actively utilizing the data they collect during the life cycle of highway infrastructure in making data-driven insights and reliable decisions.

Currently, agencies utilize pavement condition data in project selection, prioritization and treatment selection that ranges from pavement replacement to rehabilitation and maintenance projects. However, there is still a gap in effectively utilizing these data and converting them into information and knowledge to support highway infrastructure decisions compared to the amount collected. For example, information such as pavement condition index (PCI) is generated from available data to measure the performance of various sections of roadways which in turn help prioritize roadways projects and select optimum treatments. Typically, PCI is a subjective method of evaluating the condition of the surface of a road network based on manual inspection and visual observation of a road network. This index uses a numerical value between 0 and 100 where points are deducted from 100 based on the various distresses and severity combinations. Although the index can be a helpful assessment tool, the subjectivity nature creates difficulty in deciding whether a road network is in good or poor condition. This would result in either spending more money on maintaining a pavement before it really needs rehabilitation or delaying the maintenance period as a result of being indulgent about the condition.

While many agencies use an overall condition indicator, some agencies convert raw pavement condition data to a comprehensive information or index to fit their needs in pavement management decisions. For instance, Iowa DOT uses distress and severities data such as transverse cracking, roughness and pavement type to develop a PCI based on statistical models. But the question is if current PCI is a key performance indicator which is representative of pavement condition with the use of only surface distress data or whether additional data and information should be incorporated to calibrate or develop a new PCI index as it might have significant relationship with other pavement condition data. Thus, with PCI dealing only with surface distress condition, it may not indicate the underlying cause behind the deterioration of the pavement and other indicators should be considered to diagnose the condition.

Based on the gap analysis, structural (patching, alligator cracking, fatigue cracking, FWD) and functional (rutting) data, history (age, ESAL, roadway classification, AADT) and cost data are important data that should be incorporated in pavement analysis and improving the PCI index. Although these data are currently available, they are underutilized. In addition, data such as raveling, skid resistance, macro texture, thickness of pavement and location are potential data that are missing in the development of PCI. This ideal integration will strengthen the use of PCI through incorporation of structural, functional, history and cost data which can be a key performance indicator or information. Table 5-1 shows a gap analysis between current and ideal data, information and decisions in highway infrastructure pavement management system.

**Table 5-1 Gap Analysis of Pavement Data Management System at Network Level**

| Category | | Data | | | | | | | | | | | | | | | | | | | | | | | | | Information | | | | | | Decision | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Structural | | | | | | | | | Functional | | | | History | | | | | | | | | | | Cost | | | | | | | | | | | |
| Type | | Patching | Long-Cracking | Alligator-Cracking | Fatigue Cracking | T-Cracking | Bleeding/Flushing | Raveling | Punch-outs | FWD | Rutting | Roughness (IRI) | Friction (Skid resistance) | Macro texture | Pavement Type | Thickness | Age | Length | Width | ESAL | Roadway Classification | NHS/Non-NHS Classification | Number of Lanes | Location | AADT | Treatment Cost | Deterioration Model | Cost Analysis | Pavement Condition Index | Structural Index (SI) | Cracking Index | MEPDG/Design | Treatment Selection | Project Selection | Project Prioritization | Selection of | Allocate Budget to Districts |
| Information | Deterioration Model | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Cost Analysis | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Index (PCI) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Structural Index (SI) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Cracking Index | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | MEPDG/ Design | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Decision | Treatment Selection | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Project Selection | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Project Prioritization | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Rehab/Maintenance | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Allocate Budget | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Active Inactive Missing

Based on the gap analysis, cracking data (longitudinal, alligator and fatigue cracking), FWD, ESAL, project length, patching and pavement type are the most underutilized data with more than 50% not meeting their intended use, while AADT, rutting and treatment cost are impartially required (more than 40%), but not used to generate information and support decsions. The green cells represent currently active data, yellow cells represent inactive data and red cellss represent missing data or

information. More than 80% of punchout and location data are currently missing to generate information such as cost analysis and PCI and support decisions such as project selection and treatment selection. Based on the analysis, transverse cracking, bleeding, ravelling friction (skid resistance), pavement width, number of lanes, roadway and NHS classification are effectively utilized in meeting the ideal usage or fulfilling more than 65% of decision-makers requirements. In ideal situation, the incorportion of FWD, pavement type, AADT, treatment cost, and roughness data play a crtical role in the generation of mutiple information and supporting decisions, while inclusion of age, rutting, and pavement thickness also play a significant role if missing data are collected and available data are actively utilized. For instance, performing a pavement analysis by including potential factor such as falling weight deflectometer (FWD) will help quantify a pavement condition not only from the functional performance (surface distress and roughness), but also from structural strength aspect through structural capacity analysis (Denso, 2009 and Flora et al. 2010). The gap analysis of data usage is shown using an ideal progress percentage in Figure 5-1.

With regard to information, structural Index (SI), MEPDG and cracking index (CI) are identified to be key performance indicators that are missing in the current decsion-making processes, while PCI and deterioration model can be reinforced with additional data where only 14% and 20% of available data are actively utilized to develop the PCI index and deterioration models repectively. Currently, PCI being the only key performance indicator, it provides a broad measure of pavement condition to help prioritize projects, select treatments and allocate budgets. However, this indicator should be used in conjunction with other indictaors to evaluate the road network and support these decsions. With the current level of severity being clasisifed into low, medium and high severity, a high severity transverse or longitudnal cracking may fall into an

alligator cracking. Development of a new information or indicator such as cracking index can be a performance indicator that can support PCI. Jackson (2008) showed the singificance of a cracking index in measuring pavement distress by combining longitudnal, transvers and multiple cracking. In addition, generation of structural index can be another influential indicator of the structual condition of pavement where Flora et al. (2010) showed its importance as potential knowledge to support network and project level pavement and maintenance management decision-making processes. The study developed a structural strength index based on FWD, pavement type, design, traffic and segment for different pavement families through cumulative probabilistic functions. Thus, the development of these additional indicators is benefical to decision-makers to have more insights in anlyzing pavement conditions and create a comprehensive measure of a road network in making reliable decsions in a more objective manner.

Typically, a deterioration model is developed based on the type and age of the pavement through time. Based on the experts analysis, more than 60% of data which have the potential of generating reliable deterioration models are missing, while 20% which are currently available are not actively utilized. However, a well developed deteroration model that consists of structural conditions such as patching and FWD, functional data such as roughness through the inclusion of location of the pavement and traffic can be a major input in terms of selecting the right treatment, priortize projects and/or allocate budgets. Ozbay and Laub (2001) justified this by developing a neural newtork pavement derioration model which consists of roughness, age, traffic (ESAL). An ideal cost analysis also plays an important role in allocating budget, rehab and maintenance selection by incorporating treatment cost, location, AADT, FWD, rutting, roughness, skid

resistance, pavement type and thickness, where currently 38% of these data are inactive and 31% are still missing for conducting the analysis.

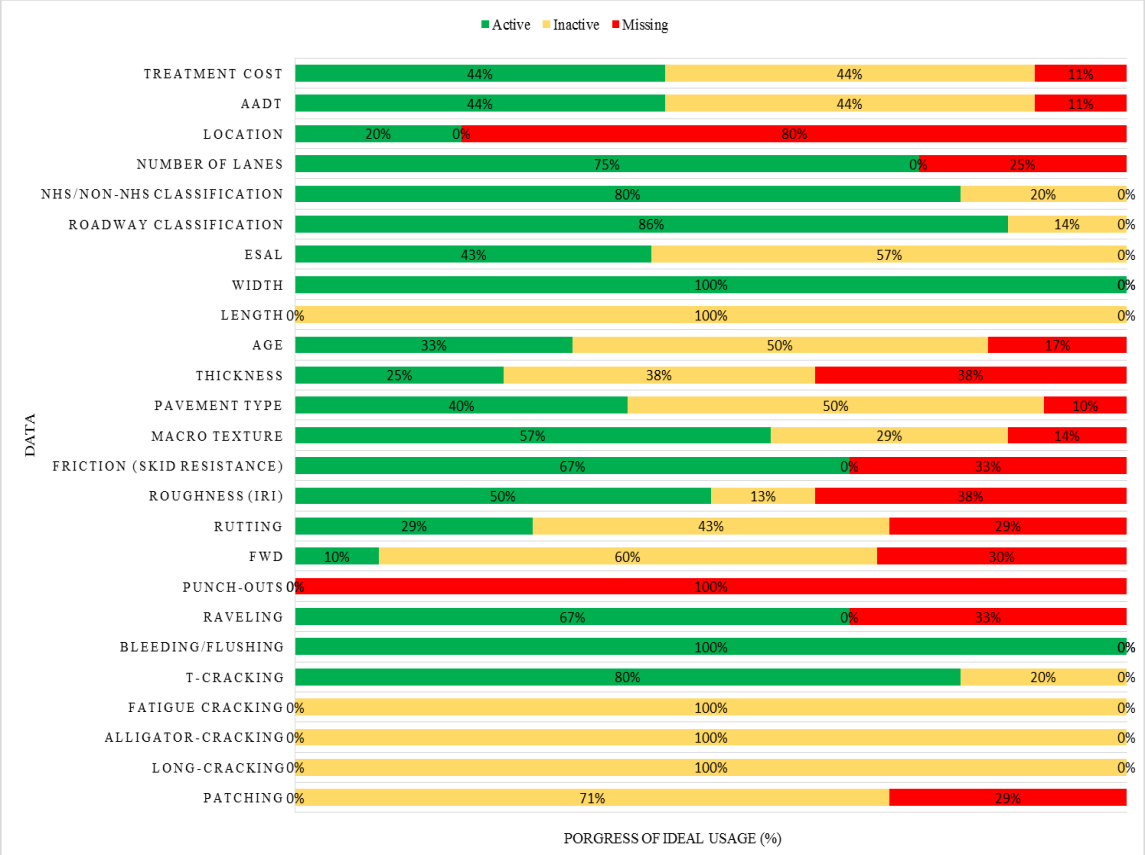



**Figure 5-1 Ideal Progress of Data Integration**

Various Studies conduced in multiple DOTs including California, Illinois, Indiana and Iowa also showed how these factors such as traffic, structural integrtity, skid and distress type should be incorporated as trigger values in generating information and supporting decsions such as conducting life-cycle cost analysis, developing deterioration models, performing treatment selection and rehabilition (Caltrans, 2007, Jackson 2009, Smadi 2010 and Wolters et al. 2011). The gap analysis on information and decisions is shown using ideal data usage percentage in Figure 5-2.

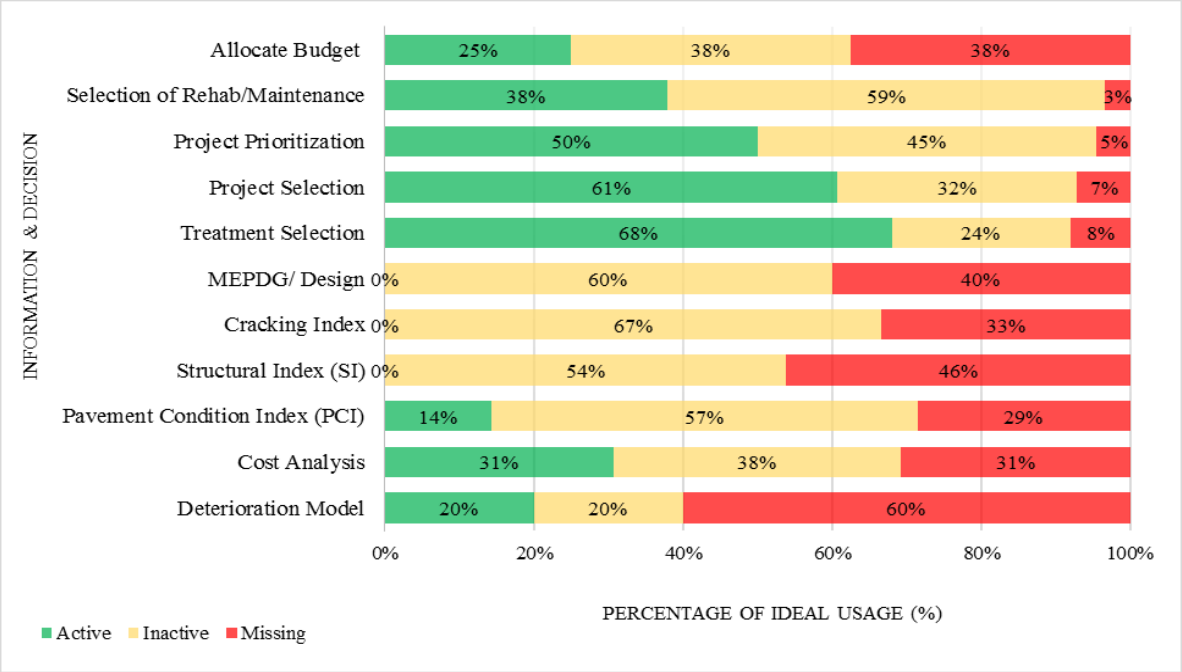



**Figure 5-2 Ideal Data Usage Percentage of Information and Decisions**

Based on Figure 5-2, project selection, treatment selection and project prioritization decsison-makers seem to actively utilize currently available data by generating information from the collected data in their decsion-making process as compared to allocating budgets. This might be due to the fact that allocation of budget can be influenced highly by political, economic and social factors as compared to needs and technical analysis outputs. For instance, current treatment selection utilizes more than 65% of the available data and information, while project selection uses 61%. However, the inclusion of cracking data, patching, and incoproration of pieces of information such as deterioiration model and pavement design analysis creates an ideal treatment selection. Figure 5-2 shows the progress of decisions towards an ideal integration.

## 5.3 Case Study II - Preconstruction Services

One of the data collected during the early stages of a highway project life cycle is preconstruction service data. In this section, preconstruction service data management are used

as part of a case study to evaluate the efficiency of the developed framework in chapter 4. This case study is used to identify key decisions made, data and information utilized, map the data and information flow and evaluate the overall performance. Primarily, data, information and decisions utilized in preconstructions services are determined to perform the case study. Then, an adjacency matrix is developed to map the data and information flow with decisions. Two preconstruction service experts at Iowa DOT are used to map this linkage. This helps identifying the active, inactive and missing paths of the current preconstruction service system by implementing the three-tiered hierarchical system. During this stage, a social network theory is used to identify important data, information and decisions by applying social network indicators discussed in the previous chapter. Finally, highway infrastructure integration index (HIDI) is applied to assess the overall performance of preconstruction service data management system.

During the early stage of planning and design phases, various decisions are made which can impact the overall cost, schedule, performance and quality of a highway project. These decisions range from determination of feasibility options and budget allocation at strategic level to selection of design alternatives, project control and review of construction documents at project level to meet the level of service. Various types of cost estimation (conceptual, preliminary engineering, construction or detailed cost estimation), identification of right-of-way and acquisition of permits such as NEPA, selection of performing design works by in-house or consultants, and analysis of bid in contractor selection are also some of the decisions made in preconstruction services. Typical preconstruction services decisions made at various levels are summarized in Table 5-2. For validation purposes, five decisions representing different level of decision-making hierarchy from cost estimation perspective were considered in this study (Figure 5-3).

**Table 5-2 Preconstruction Service Decisions**

| Decision Hierarchy | Decision | Definition |
|---|---|---|
| Strategic Level | Budget allocation | Refers to but not limited to distribution of transportation funds or budget across projects |
| Network Level | Conceptual cost estimating | Estimation of project cost based on limited scope in making a "go" or "no go" decisions as to perform the project or reject it |
| Program Level | Environmental approval | Refers to making "yes" or "no" decision in assessing environmental impacts to be considered significant or not |
| | Traffic and safety design decisions | Choosing the right type of light fixtures, identifying the number of traffic signs and posts required and type of guardrail needed |
| | Right-of-way approval | deals with selecting optimal alignment that may consist of relocating environments and utilities that is cost effective and convenient to public |
| Project Selection Level | In-house/outsourcing decision | decision made in the selection of design works to be performed by in-house or outsource it to consulting firms |
| Project Level | Survey Design/Estimate | refers to decisions to estimate cost of survey, selection of resources (mobilizing equipment) and setting alignment |
| | Construction cost estimation | Refers to detailed cost estimate based on final design data. It may be represented as bid estimate |
| | Analysis of Bid | Includes decisions as to the selection of contractors by setting criterion based on previous performance, cost data and scope of work. |
| | Roadway design decisions | 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.) |
| | Bridge design decisions | selection or identification of span length, width, and number of bridges required, etc. that may be optimized based on functional and design data |
| | Resource Allocation | Assignment and time allotment of skilled manpower and assignment of equipment in the planning of rehabilitation and new highway projects. |

In this case study, budget allocation, conceptual cost estimation, environmental decision, in-house/outsource projects and resource allocation were used as primary decisions representing different decision hierarchy levels. These decisions require various pieces of information as input to reach at reliable decisions. Some of the information that can be utilized during the preconstruction phase include roadway and bridge analysis, traffic analysis, needs study, estimation of number of sheet plans, work hours and project cost. This study takes into consideration estimation of cost and work hours or engineering hours as potential information to support the decision-making processes listed in Figure 5-3. The estimation of cost information

refers to the prediction of engineering cost associated with performing preconstruction services based on scope definition and project data acquired from preliminary survey and study, while the estimated work hours refer to the projection of amount of time engineers or skilled labors spend in developing design and preliminary engineering services.

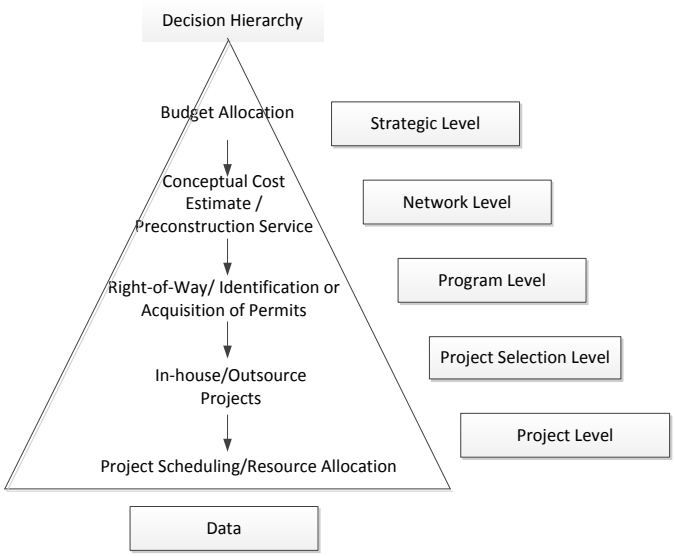



**Figure 5-3 Preconstruction Service Decision Hierarchy**

These pieces of information in turn require various data as inputs. In this study, existing preconstruction service data are classified into contract data, functional data, design data (roadway, bridge, environmental, right-of-way and traffic), cost data and outsourcing data. Most of these data attributes are structured data types with a combination of interval, nominal and ordinal variables. This study focuses on five potential data categories listed in Table 5-3 as boundary condition to meet the requirements of the five decisions listed in the previous paragraph and validate the three-tiered hierarchical integration framework and HIDI index developed in chapter 4. A complete list of preconstruction service data collected during the planning and design stages is provided in Appendix C.

**Table 5-3 Preconstruction Service Data and Information**

| Source | Type of Data | Sub-Elements |
|---|---|---|
| Preconstruction Service Data | Contract Data | Pavement surface type, type of work, location, project length |
| | Functional Data | Roadway classification, NHS/Non-NHS classification |
| | Roadway Data | ROW, grade, number of lanes, pavement type, length of project, lane width |
| | Traffic Data | Traffic Profile/AADT |
| | Cost Data | Labor hours, labor dollars, vehicle dollars, personal expenses |

## 5.4 Three-Tiered Hierarchical Integration and Assessment Framework

Once these data, information and decisions were identified, the next step was the application of the integration and assessment framework and *HIDI* developed in the previous chapter. An adjacency matrix was first designed to interlink data with information and decisions and identify the three paths. Then, two Iowa DOT experts were asked to evaluate the framework by filling out the matrix to gather the actualities on current data use in generating information and supporting preconstruction service decisions. Primarily, the bottom-up approach of identifying important data, information and decisions were determined using centrality indicators; degree centrality, betweenness centrality, eigenvector centrality as internal measures. Then, highway infrastructure integration index (*HIDI*) is applied to examine the overall performance of preconstruction services data management system using a cohesion or density measure through top-down approach. The analysis results are discussed in the following sections.

### 5.4.1 Primary Data

Based on social network analysis of degree centrality, project type, type of work, engineering hours, vehicle dollars, project length, and labor dollars are the critical data (degree centrality>10) that are used in generating preconstruction service cost information and supporting decisions, while pavement type, location, highway classification and personal

expenses are fairly important (5<degree centrality<10). The study identified that the importance of roadway classification and AADT data are not significant as compared to grade and lane width which do not have any importance in preconstruction service decisions (degree centrality < 5). Figure 5-4 shows the level of data importance based on degree centrality.

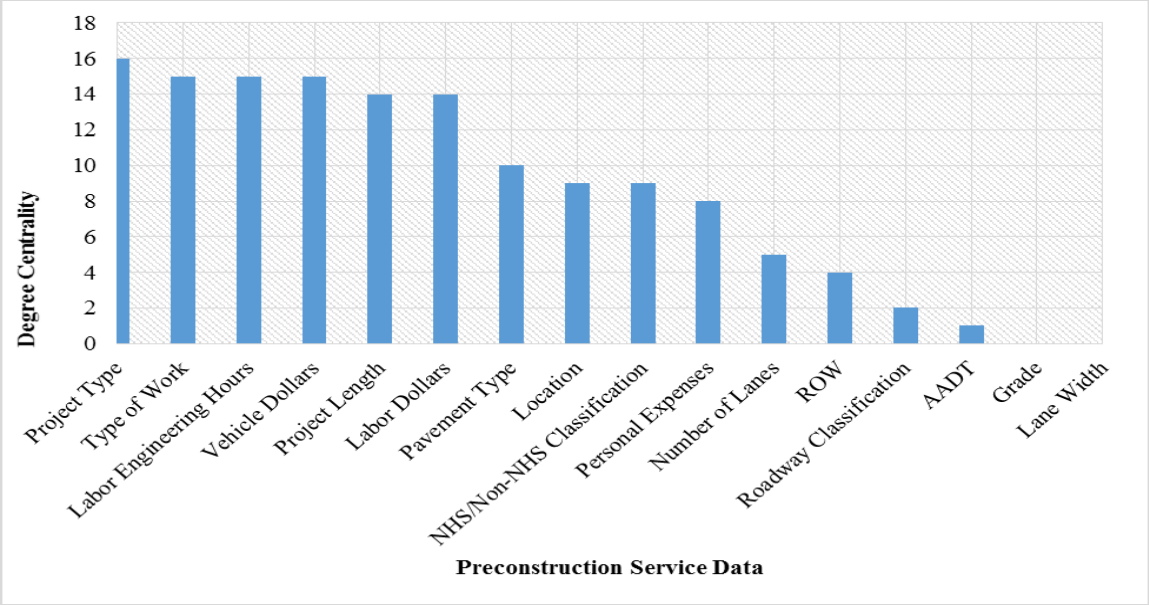



**Figure 5-4 Preconstruction Services Data Importance: Degree Centrality**

Figure 5-5 shows the mapping between data-information-decisions network where data nodes are represented by box symbol (blue), pieces of information are represented by circle (green), and decisions are represented by triangle (red). As shown in Figure 5-5, lane width and grade of pavement are outliers with no connection to neither data, information nor decisions. It is important to note that the size of a node also symbolizes the importance of data (where a higher degree centrality is shown as a larger symbol and vice versa).

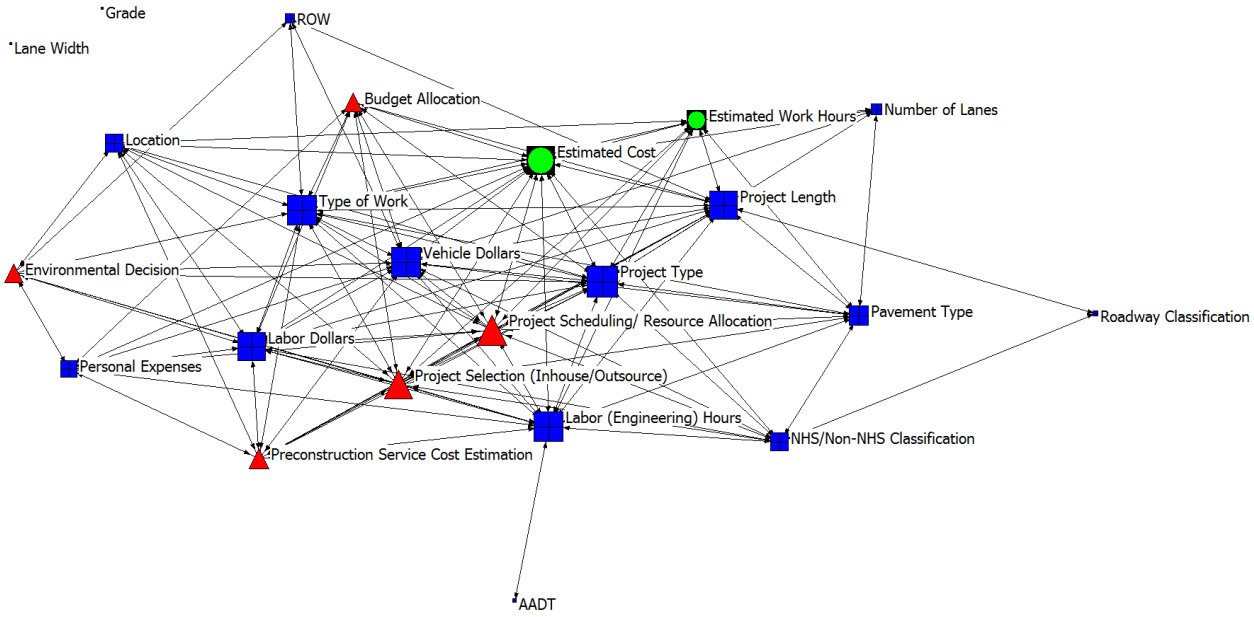



**Figure 5-5 Preconstruction Service Data-Information-Decision Network**

### 5.4.2 Primary Information

Based on the measure of betweenness centrality, the engineering labor hours is the important piece of information that is used in supporting preconstruction service decisions. Although two pieces of information (estimated cost and estimated work hours) are considered in the study, contract data and cost data play a major role in the decision-making process. It is important to note that based on discussions with the engineers/experts that were validating the framework, it was identified that these data are utilized directly as inputs making decisions as there are not any models currently available to predict estimated cost and work hours of preconstruction services. Table 5-4 shows the level of importance of data and pieces of information.

**Table 5-4 Important Pieces of Information: Betweenness Centrality**

| Attribute | Betweenness Centrality |
|---|---|
| Engineering Labor Hours | 23.367 |
| Project Length | 16.962 |
| Vehicle Dollars | 12.701 |
| Project Type | 8.745 |
| Type of Work | 7.726 |
| **Estimated Cost** | **6.96** |
| NHS/Non-NHS Classification | 6.471 |
| Project Scheduling/ Resource Allocation | 5.34 |
| Project Selection (In-house/Outsource) | 4.671 |
| Labor Dollars | 4.264 |
| Environmental Decision | 3.375 |
| **Estimated Work hours** | **2.419** |
| Pavement Type | 2.333 |
| Preconstruction Service Cost Estimation | 1.469 |
| Personal Expenses | 0.989 |
| Location | 0.943 |
| Budget Allocation | 0.702 |
| ROW | 0.226 |
| Number of Lanes | 0.225 |
| Roadway Classification | 0.111 |
| AADT | 0 |
| Grade | 0 |
| Lane Width | 0 |

### 5.4.3 Primary Decision

Another internal measure used to identify the importance level of decisions is eigenvector centrality. Based on this indicator, the results showed that all preconstruction service decisions almost equally require data and pieces of information with project scheduling/resource allocation taking the highest importance with 0.291 Eigen value and environmental decisions taking the lowest with 0.173 Eigen value. Table 5-5 shows the eigenvector centrality results of preconstruction service decisions.

**Table 5-5 Important Decisions: Eigenvector Centrality**

| Decision | Eigenvector Centrality |
|---|---|
| Project Scheduling/ Resource Allocation | 0.291 |
| Project Selection (In-house/Outsource) | 0.272 |
| Preconstruction Service Cost Estimation | 0.212 |
| Budget Allocation | 0.196 |
| Environmental Decision | 0.173 |

### 5.4.4 *HIDI*

The overall preconstruction service data utilization is evaluated using the *HIDI* index from a network approach. Based on the index, the preconstruction service framework has only 26.6% of data being collected are actively utilized in decision-making, while the majority of data and information are either inactive (29.4%) or missing with a density measure of 44% (Table 5-6). This indicates that currently more than 70% of data collected by the division are underutilized and potential information generation methods or procedures are not available to support decision-making processes.

**Table 5-6 *HIDI*: Preconstruction Service**

| **Path** | **Density** | **Number of Ties** | ***HIDI*** |
|---|---|---|---|
| **Active (l)** | 0.301 | 58 | **0.266** |
| **Inactive (m)** | 0.329 | 64 | **0.294** |
| **Non-existing (n)** | 0.456 | 96 | **0.440** |
| **Total** | 1.086 | 218 | 1 |

Based on the *HIDI* grading report card, the preconstruction service data management system receives a grade of "D-". This report card shows how the preconstruction service decisions need major adjustments and the data management system is questionable in meeting the standards and

decision-makers' requirements. Thus, a new data collection or information extraction procedure needs to be designed to effectively utilize data and generate information in supporting preconstruction service decisions. Active development of information extraction method (to develop engineering hour and estimated cost models) alone can increase the utilization of existing data (cost data such as engineering hours, labor dollars, vehicle dollars and personal expenses) whereby the report card can improve from grade $D^-$ to grade $C^+$.

# Chapter 6

# CONCLUSION

## 6.1 Conclusion

This study presented a new framework that can systematically integrate and bridge highway infrastructure data, information and decisions through the incorporation of a unique and proactive performance assessment technique. The approach uses a combination of top-down and bottom-up approaches by combining quantitative and qualitative methods to interlink and map data, information and decisions and assess the performance of data use in generating information and supporting decisions in highway infrastructure management. The study utilizes social network theory as a principal component for developing this new framework. The theory is employed to quantify key data, information and decisions by determining the correlation between these entities to enhance the level of data use. Pavement condition data, information and decisions available in pavement management systems and preconstruction service data are employed as representative data in highway infrastructure lifecycle to illustrate the application of the framework.

The developed framework would serve two main purposes:

1) Internal evaluators or knowledgeable decision-makers within a State Highway Agency (SHA) could determine the data requirements to improve their use and support their decisions and

2) External evaluators may be able to evaluate the status of data collection and utilization efforts of an agency by analyzing the gap between the current status of data management

and the ideal data utilization, which can serve as a periodic data integration report card just like the ASCE "Infrastructure Report Card."

This framework not only can be used as an infrastructure data report card to assess the current status of data management, but can also help to justify the return on the huge investments being made by SHAs in data collection efforts. It will allow for mapping highway infrastructure data, information and decisions using the concept of three tiered process, help agencies to develop an active utilization plan of currently existing databases and place the right information in the hands of decision-makers. In addition, it will enhance the development of new data collection scheme and information/knowledge generation plans to support key decisions that, historically, were not well-supported with information and data. Furthermore, SHAs can visually examine the interactions and relationships of data, information and decisions and identify their importance in decision-making processes.

Understanding the needs and requirements of stakeholders is a crucial component in meeting of the goals and objectives of an organization. As part of this data and information integration and assessment framework study, an assimilated requirement analysis is also presented to determine the satisfaction level of potential stakeholders or data users in current highway infrastructure data use in their decision-making processes. In the study, an integrated hierarchical system was developed to estimate the probability of satisfaction at various decision-making levels and agency level through the utilization of a fault tree analysis (FTA), where FTA was used as a technique to quantify data users satisfaction.

This FTA-based requirement analysis showed the interdependency of various decision-makers in the final output of a program and address the needs of potential data users at various levels. The approach has also allowed the identification of the root causes behind the limited and

minimal use of data in a more integrated and objective manner in addition to evaluating the level of satisfaction. A module was defined based on four dimensions, SESP (syntactic, empirics, semantics, and pragmatic) to improve data quality in generating information and supporting decisions. This analysis helps in identifying the level of decision-making process where data are well-utilized and communicated and determining the data quality improvements from users' perspective. The approach can be applied to various data sets collected by various highway agency divisions to meet the end users need and help improve the overall data management.

A particular datum that is neither converted to information and knowledge nor supports decisions in some way or another is considered as a waste of resource. Through a gap analysis, the difference between current data usage and ideal data and information utilization has been illustrated. This gap analysis enhances redundant data and increase the correlations of various data in generating reliable pieces of information and/or performance measure thereby creating an ultimate decision-making path. For instance, current pavement condition index (PCI) provides the condition of pavement based only on distress data. However, the measure neither incorporates structural capacity nor provide functional indicators such as FWD, skid resistance or roughness. This ideal condition allows calibration of input data in finding better information or performance models for various pavement families that will support highway agencies to make more effective pavement management decisions. Overall, this new framework will significantly improve the way data are collected, utilized and managed in enhancing the SHAs decision-making processes by meeting users requirements and organizations goals.

### 6.2 Research Contribution

The study develops a new framework that gains strategic advantage from highway infrastructure data by integrating with information and decisions through fault tree analysis (FTA) and social network analysis (SNA) theories. The framework provides a top-down and bottom-up approach to integrate data, information and decision and measure the performance of data use to generate information and support highway infrastructure decisions in state highway agencies. The developed framework will set a benchmarking example in the area of data and information integration to make effective and reliable highway infrastructure decisions. It will help guide DOTs develop an active utilization plan of currently existing databases and place the right information at the right time in the hands of decision-makers.

The framework will not only guide the effective use of existing data, but also create a paradigm shift in the collection of new types of data through a systematic design of data acquisition and identification of data analysis methods by keeping the end users in mind and/or recognizing highway agencies decisions-makers requirements. In addition, it will help agencies measure their data performances and justify the level of return on investment in the data collection efforts by minimizing unrequired and redundant data, incorporating missing data and information. The output of this study will tremendously improve the proficiency of the overall highway infrastructure decision-making processes by changing the culture of owner engineers and project managers view on data and information. It will influence how efficiently and economically highway infrastructure systems are planned, executed and managed through a data-driven insight. Through this research study:

1. An innovative three-tiered data and information integration framework is developed that can ultimately support highway infrastructure decisions and make the highway industry a data–driven industry.
2. A new report card, highway infrastructure data integration, (*HIDI*) index is developed to identify status of highway infrastructure data management and help justify the return on investment of highway agencies data collection efforts.
3. An assimilated approach that logically integrates highway decision-makers data requirements with their satisfaction attributes and quality needs is developed to recognize various decision-makers requirements and improve data quality.

### 6.3 Recommendations

This research study has shown that addressing users' requirement and evaluating the performance of data and information integration allows improved use of data and information in supporting decision-making processes and help justify data collection efforts by connecting the data collector with the data user. Although the study can be applied in various agencies and divisions, there are some improvements that can be addressed in future research.

*Maturity Model*

In utilizing *HIDI* report card, it is important to note that the maturity level of an agency's data management varies from one to another with respect to the amount, type and level of detail of data collected along with technologies and analysis methods implemented to support decisions. For example, an agency may use a fully automated data collection method using laser scans and utilize Markov chain deterioration models to select treatments, while another agency may use a manual-based data collection method, an expert-based analysis or a subjective

decision-making process to select treatments. Thus, the application of an external evaluation should be based on benchmarks developed to address the maturity level of organizations. In the future, a comprehensive data maturity model can be developed to set a standard measure to evaluate various agencies data integration performance.

*Key Performance Indicator Model Development*

Based on the amount of data available in highway agencies, information generation analysis methods should be applied to determine appropriate selection criteria and treatment triggers to make more reliable decisions. Data mining technique can be utilized to classify pavement families, assess the effectiveness of various treatment options and develop deterioration curves. Better performance measure indicator models for structural index and cracking index can be developed based on statistical regression models or a clustering technique to categorize pavement condition data. Decision tree models can be developed 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 pavement management system. It also helps in validating the gap analysis to generate information and knowledge from pavement condition assessment data and justify the benefits of pavement condition data collection effort.

*Return on Investment, ROI*

Due to the limited availability of cost associated with collecting, storing and managing data, the study measured the efficiency of data through its use in generating information and supporting decisions. In the future, the return on investment can be estimated through a cost benefit analysis by incorporating additional cost related data.

# REFERENCES


1. Abdul-Rahman, Hamza, Kwan, C. L. and Woods, Peter, C. (1999), "Quality function Deployment in Construction Design: Application in Low Cost Housing Design", *International Journal of Quality and Reliability Mana*gement, Vol. 16 No. 6, pp. 591-605.
2. Akao Y. (1990), "*Quality Function Deployment: Integrating Customer Requirements into Product Design*", Cambridge: Productivity Press.
3. Alsamadani, R., M. Hallowell, et al. (2012). "Measuring and Modelling Safety Communication in Small Work Crews in the US Using Social Network Analysis." Construction Management and Economics, 1–12.
4. Barnes, J. A. (1954). "Class and Committees in a Norwegian Fishing Parish." Human Relations, 7(1), 39–58.
5. Bonacich, P. (1972). "Factoring and Weighting Approaches to Status Scores and Clique Identification." Journal of Mathematical Sociology, Vol. 2, pp. 113–120.
6. Borgatti, Stephen P. (2005). "Centrality and Network Flow." Social networks, Vol. 27, pp 55-71.
7. Caltrans (2007), " Framework for Treatment Selection", Caltrans Division of Maintenance, MTAG-Rigid Pavement Preservation, Vol. II, 2nd edition
8. Calvarese, Justin. (2007). "QA/QC Oklahoma DOT." National Conference on Pavement Management, Norfolk, Va.
9. Chinowsky, P. S., J. Diekmann, et al. (2010). "Project Organizations as Social Networks." Journal of Construction Engineering and Management.
10. Chinowsky, Paul & Taylor, John E. (2012). "Networks in Engineering: An Emerging Approach to Project Organization Studies." The Engineering Project Organization Journal, Vol. 2, pp. 15-26.
11. Cleveland, William S. (2001). Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics." Bell Laboratory.
12. Deng, Fei & Smyth, Hedley. (2013). "Nature of Firm Performance in Construction." Journal of Construction Engineering and Management,
13. Delone, W. H., & McLean, E. R. (1992). "Information Systems Success: The Quest for the Dependent Variable." Information Systems Research, 3(1): 60-95.

14. Dikmen I, Birgonul MT, and Kiziltas S. (2005), "Strategic Use of Quality Function Deployment in the Construction Industry", *Building and Environment*, Vol 40 pp. 245-255.
15. Farris Paul W., Bendle, Neil T., Pfeifer, Philip E. & Reibstein, David J. (2010). "Marketing Metrics: The Definitive Guide to Measuring Marketing Performance." Pearson Education.
16. Flintsch W., Gerardo & Bryant, W. J. (2006). "Asset Management Data Collection for Supporting Decision Processes." U.S. Department of Transportation Federal Highway Administration.
17. Flora, William F. and Ong, Ghim P. (2010), "Development of a Structural index as an Integral Part of Overall Pavement Quality in the INDOT PMS", Joint Transportation Research Program
18. Freeman, L.C. (1979). "Centrality in Networks: I. Conceptual clarification." Social Networks, 1, 215–239.
19. Gelsinger, Pat. (2012). "Data: The New Science." EMC Word and Data Science Summit.
20. Gerhardt, Bill, Griffin, Kate & Klemann, Roland. (2012). "Unlocking Value in the Fragmented World of Big Data Analytics." CISCO Internet Business Solution Group.
21. Granovetter, M.S. (1973). "The Strength of Weak Ties." The American Journal of Sociology, 78(6), 1360–80.
22. Harrison, Frances D. (2011). "Meeting Critical Data needs for Decision-Making in State and Metropolitan Transportation Agencies." Transportation Research Board, Conference Proceedings on the Web 9, Irvine, California, http://onlinepubs.trb.org/onlinepubs/conf /CPW9pdf.
23. Hauser, J. R., and Clausing, D. (1988). ''The house of quality.'', Harvard Business Rev., 66(3).
24. Hossain, L. (2009). "Communications and Coordination in Construction Projects." Construction Management and Economics, Vol. 27.
25. Iowa Department of Transportation (2009). "Iowa's Pavement management Program", Office of Local Systems, http://www.iowadot.gov/local_systems/mailing/2009/march/08_ipmp_update.pdf
26. Jackson N. and Nichols Consulting (2009), "Development of Revised Pavement Condition Indices for PCC Pavement for the WSDOT PMS", WA State Transportation Centre

27. Julnes, Patria De Lancer. (2006). "Performance Management: An Effective Tool for Government Accountability? The Debate Goes On." Evaluation, Vol. 12(2), pp. 219-235, Sage Publications
28. Khan, Bevely K., Strong, Diane M. & Wang, Richard Y. (2002). "Information Quality Benchmarks: Product and Service Performance." Communications of Association for Computing Machinery, 45 (4), 84–192
29. Kaplan, Robert S. & Morton, David P. (1996). "The Balanced Scorecard: Translating Strategy into Action." Harvard Business Press.
30. Kenett, R. S. (2008), "Cause and Effect Diagrams", Encyclopedia of Statistics in Quality and Reliability
31. Knight, Shirlee-ann & Burn, Janice. (2005). "Developing a Framework for Assessing Information Quality on the World Wide Web." Informing Science, Vol. 8.
32. Knoke, David & Yang, Song. (2008). "Social Network Analysis." Sage Publications, 2nd Ed.
33. Larson, C.D. & Skrypczuk, O. (2004). "Lessons of Asset Management Data Collection at Virginia DOT." Paper No. 471, 6th International Conference on Managing Pavements
34. Lee Yang, W. & Strong, Diane M. (2003). "Knowing-Why About Data processes and Data Quality." Journal of Management Information System, Vol. 20, No. 3 pp. 13-39
35. Leinhadt, S. (1977). "Social networks: A Developing Paradigm." New York Academic.
36. Lochner, R.H. and Matar, J.E. (1990), "*Designing for Quality*", Chapman & Hall, New York.
37. McCabe, B., Tran, V., & Ramani, J. (2005). "Construction Prequalification Using Data Envelopment Analysis." Canadian Journal of Civil Engineering, 32(1), 183–193.
38. Mitchell, J.C. (1969). "Social Networks in Urban Situations: Analyses of Personal Relationships in Central African Towns." Manchester University Press, Manchester, UK.
39. Moreno, J. L. (1934). "Who shall survive? A New Approach to the Problem of Human Interrelations." Washington, D. C, Nervous and Mental Disease Publishing.
40. Mueller-Prothman, T. & Finke, I. (2004). "SELaKT: Social Network Analysis as a Method for Expert Localization and Sustainable Knowledge Transfer." Journal of Universal Computer Science, 10 (6), 691–701.

41. Nabitz, U., Klazinga, N., & Walburg, J. (2000). "The EFQM excellence model: European and Dutch Experiences with the EFQM Approach in Health Care." International Journal of Quality Health Care, Vol. 12, pp. 191–201.
42. Neely, A.D., Gregory, M. J., & Platts, K. (1995). "Performance Measurement System Design: a Literature Review and Research Agenda." International Journal of Operations & Production Management, Vol. 15, No. 4, pp. 80-116.
43. Nuseibeh, Bashar & Easterbrook, Steve. (2001). "Requirement Engineering: a Roadmap." IEEE Software.
44. Ozbay, K. and Laub, R. (2001), "Models for Pavement Deterioration Using LTTP", Final Report, FHWA and New Jersey Department of Transportation
45. Ozbek, E. Mehmet, De La Garza, M., Jesus, & Triantis, Konstantinos. (2009). "Data Envelopment Analysis as Decision Making Tool for Transportation Professionals." Journal of Transportation Engineering, Vol. 135, No. 11.
46. Park, H., M. A. Seung H. Han, Eddy M. Rojas, JeongWook Son and Wooyong Jung (2011). "Social Network Analysis of Collaborative Ventures for Overseas Construction Projects." Journal of Construction Engineering and Management, vol. 137, pp. 344–55.
47. Pheng, Low Sui and Yeap, Larry (2001), "Quality Function Deployment in Design/build Projects", *Journal of Architectural Engineering*, Vol. 7, No. 2
48. Pierce M., Linda, McGovern, Ginger & Zimmerman, Kathryn A. (2013). "Practical Guide for Quality Management of Pavement Condition Data Collection." U.S. Department of Transportation, Federal Highway Administration.
49. Poister, T. H. (2004). "National Cooperative Highway Research Program Synthesis 326: Strategic Planning and Decision Making in State Departments of Transportation." NCHRP, Transportation Research Board, Washington, DC.
50. Pryke, Stephen D. (2004). "Applying Construction Project Coalitions: Exploring the Application of Social Network Analysis." Construction Management and Economics, Vol. 22, pp. 787-797.
51. Schofer, J. L., T. Lomax, T. Palmerlee, J. Zmud. (2006). "Transportation Research Circular E-C109: Transportation Information Assets and Impacts." TRB, National Research Council, Washington, D.C., December 2006. http://onlinepubs.trb.org/onlinepubs/circulars/ec109.pdf.

52. Schofer. (2007). "Information Assets to Support Transportation Decision Making Report of a Peer Exchange of State Transportation Organizations." Kansas City, KS, April 17-18, 2007, Transportation Research Board Electronic Circular E-C121, 23 pps. http://onlinepubs.trb.Org/ onlinepubs/circulars/ec121.pdf
53. Scott, John & Carrington, Peter J. (2011). "The Sage Handbook of Social Network Analysis." Sage Publications.
54. Shanks, G. & Corbitt, B. (1999). Understanding data quality: Social and cultural aspects. Proceedings of the 10th Australasian Conference on Information Systems; p. 785
55. Smadi, O. (2010), "Pavement Management and Decision-Making", Pavement Evaluation, Va
56. Timurcan, Pelin Ipek & Dogan, Sevgi Zeynep. (2013). "A Review of Research Trends in Social Network Analysis," Creative Construction Conference, Budapest, Hungary
57. Thomas, H. Randolph, Korte, Q. Coco, Sanvido, E. Victor, & Parfitt, M. Kevin. (1999). "Conceptual Model for Measuring Productivity of Design and Engineering." Journal of Architectural Engineering, ASCE Vol. 5 No. 1
58. Tribelsky, Effy & Sacks, Rafael. (2010). "Measuring Information Flow in the Detailed Design of Construction Projects." Springer-Verlag London Limited
59. Torbett, Richard, Salter J., Ammon, Gann M., David, & Hobday, Mike. (2001). "Design Performance Measurement in the Construction Sector: A Pilot Study." IEEE Transactions of Engineering Management, Science & Technology Policy Research.
60. Yin, Y., Qin, S., & Holland, R. (2008). "Development of Project Level Performance Measurement Model for Improving Collaborative Design Team Work." IEEE, pp. 135-140.
61. Wang, R.Y., & Strong, D. M. (1996). "Beyond Accuracy: What Data Quality Means to Data Consumers." Journal of Management Information Systems. 12(4): 5-34.
62. Wasserman, S. & Faust, K. (1997). "Social Network Analysis: Methods and Applications." Cambridge University Press, Cambridge.
63. Wolters, A., Zimmerman, K,, Schattler, K, and Rietgraf, A, (2011), "Implementing Pavement Management Systems for Local Agencies", Illinois Centre for Transportaion

# APPENDIX A

## QUESTIONNAIRE SURVEY I

## Assessment of Data Quality in Pavement Management Decision-Making Process

Iowa State University is working on a research topic, *Data & Information Integration Framework for Pavement Management Decision-Making Process*. Today, large amount of pavement condition data are collected, stored and managed under the pavement management information system. However, there are concerns if currently collected data are used to generate information and knowledge and support pavement management decision-making processes. Part of this research is used to assess the current use of pavement condition data by identifying the satisfaction level of potential data users and stakeholders from decision-maker standpoint. The study is set to determine if current data usage meets decision-makers requirements and identify the root causes for the minimal use of data.

We would like you to participate in the survey and provide us with your valuable opinion as a decision-maker in identifying the level of agreement of current data usage in generating information and supporting your decisions as a potential data user. Please use the value from "**1 to 5" to evaluate the importance of each attribute and use "1 to 9" to rate the current Pavement Management System (PMS)** of how well it currently meets the stated attribute. The time required to complete this form is approximately **5 minutes**. You can return the completed survey form in the following ways. Please return the completed forms by **March 20th 2014**.

| Electronic Copy | Mail Copy: Dr. David Jeong, Associate Professor |
|---|---|
| Please e-mail to: asre@iastate.edu | Iowa State University |
| Or fax to: 515-294-3845 | Dept. of Civil, Construction and Environmental Engineering |
| | 404 Town Engineering |
| | Ames, IA 50011 |

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,

David H. Jeong, Ph.D.
Associate Professor
404 Town Engineering
Dept. of Civil, Construction and Environmental Engineering
Iowa State University
Ames, IA 50011
Email: djeong@iastate.edu

**A. General information**

1) Pleas provide

Contact Person Name: ☐ District: ☐

Phone: ☐ Ext: ☐ Email Address: ☐

2) What is your responsibility or position in your project?

| | | | |
|---|---|---|---|
| Program Director | ☐ | Project Engineer | ☐ |
| Program Manager | ☐ | Field Engineer | ☐ |
| Project Manager | ☐ | Other (please specify):_____ | |

3) Explain your job description based on your answer to #2

| |
|---|
| |

4) For what level of decision-making are you using PMS pavement condition data for?

| | | | |
|---|---|---|---|
| Strategic Level Decision | ☐ | Project Selection Level Decision | ☐ |
| Network Level Decision | ☐ | Project Level Decision | ☐ |
| Program level Decision | ☐ | Other please specify):_____ | |

**B. Data Quality Dimensions in Pavement Management Decision-Making Process**

The following items list potential data quality attributes that may affect the use of pavement condition data in generating information and supporting pavement management decisions. Please indicate the **level of importance (1-5) with 1 being the lowest and 5 being the highest importance** and your **level of agreement (1-9) with 1 being the lowest and 9 being the highest rating** about the current status of pavement management system (PMS) of how well it currently meets the stated attribute as a decision-maker **based on the question you answered on #4.**

| Category | No | Attribute | Description | Level of Importance (1- 5) | Degree of Agreement | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Highly Disagree ← → Highly Agree | | | | | | | | | Don't Know/Not Important |
| | | | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| Syntactic (Structure & Form of Data) | 1. | Accurate | Data are precise and free of error for use as information and decision-making process | | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| | 2. | Consistency | Data are recorded in a consistent manner to generate information and support | | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| | 3. | Completeness | Data are not missing, has sufficient depth and breadth and include necessary details | | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| | 4. | Structure | Data are in the right format and structure | | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| | 5. | Integrity | Data reflect the full details of original observation and has not been manipulated which has no bias (representative) | | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Empirics (Means of Communication) | 6. | Accessibility | Data are readily available and can easily be retrieved when needed | | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| | 7. | Timeliness | Data are sufficiently up-to-date & current | | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Semantics (Data Meaning) | 8. | Definition | Data are clearly defined in terms of its content | | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| | 9. | Ambiguity | Data are easily comprehended and | | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

| Category | No | Attribute | Description | Level of | Degree of Agreement | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pragmatic (Data Usage) | 10. | Relevant | Data are appropriate and applicable to support | | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| | 11. | Value | Data are beneficial and adds value to the decision-maker | | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

## C. Additional Information

5) In your opinion, what other data quality attributes that were not listed above would make the use of data valuable and effective in decision-making processes?

# APPENDIX B

## QUESTIONNAIRE SURVEY II

## Assessment of Data and Information Utilization in Pavement Management Decision-Making Process

Iowa State University is working on a research topic, *Data & Information Integration Framework for Pavement Management Decision-Making Process*. Today, large amount of pavement condition data are collected, stored and managed under the pavement management information system. However, there are concerns if currently collected data are used to generate information and knowledge and support decision-making processes. The purpose of this research is to improve the use of pavement condition data and information by identifying three types of data paths; i) data that are actively utilized, ii) data that are currently available, but underutilized and c) missing data through the development of a framework that can integrate the data, information and decisions'.

We would like you to participate in the survey and provide us with your valuable opinion as a decision-maker in mapping this data, information and decisions' by identifying their relationship using attached Microsoft Excel Matrix Sheet. Please use the value **"1"** if there data is actively utilized in generating information and supporting decisions', **"2"** if data is currently available, but underutilized, **"3"** if data is missing and leave **blank** if there is no relationship. Please indicate also if the data is of no use or do not know about the data using the table next to the matrix. The time required to complete this form is approximately 20 minutes. You can return the completed survey form in the following ways. Please return the completed forms by **March 20th 2014**.

| Electronic Copy | Mail Copy: Dr. David Jeong, Associate Professor |
|---|---|
| Please e-mail to: asre@iastate.edu | Iowa State University |
| Or fax to: 515-294-3845 | Dept. of Civil, Construction and Environmental Engineering |
| | 404 Town Engineering |
| | Ames, IA 50011 |

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,

David H. Jeong, Ph.D.
Associate Professor
404 Town Engineering
Dept. of Civil, Construction and Environmental Engineering
Iowa State University
Ames, IA 50011

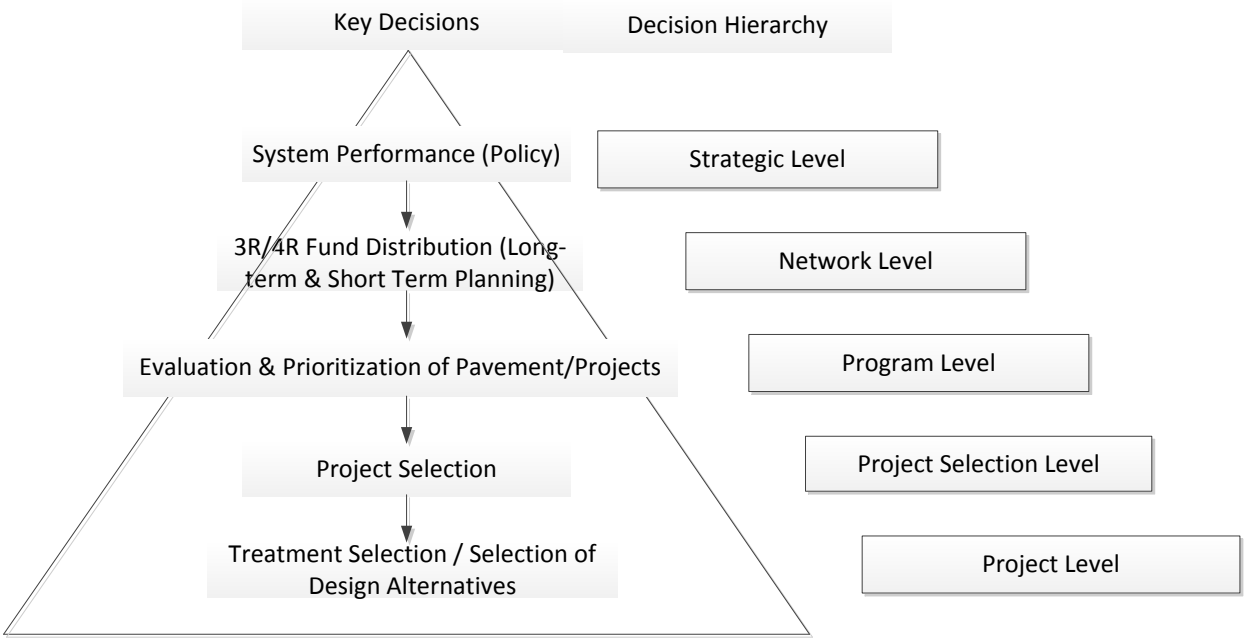
**Definitions:**

**Data** refers to raw data collected and stored in data repository or databases.

**Information** refers to data that are processed, structured and generated through proper data analysis method. It is represented by indicators or measures and outputs resulting from analysis of raw data.

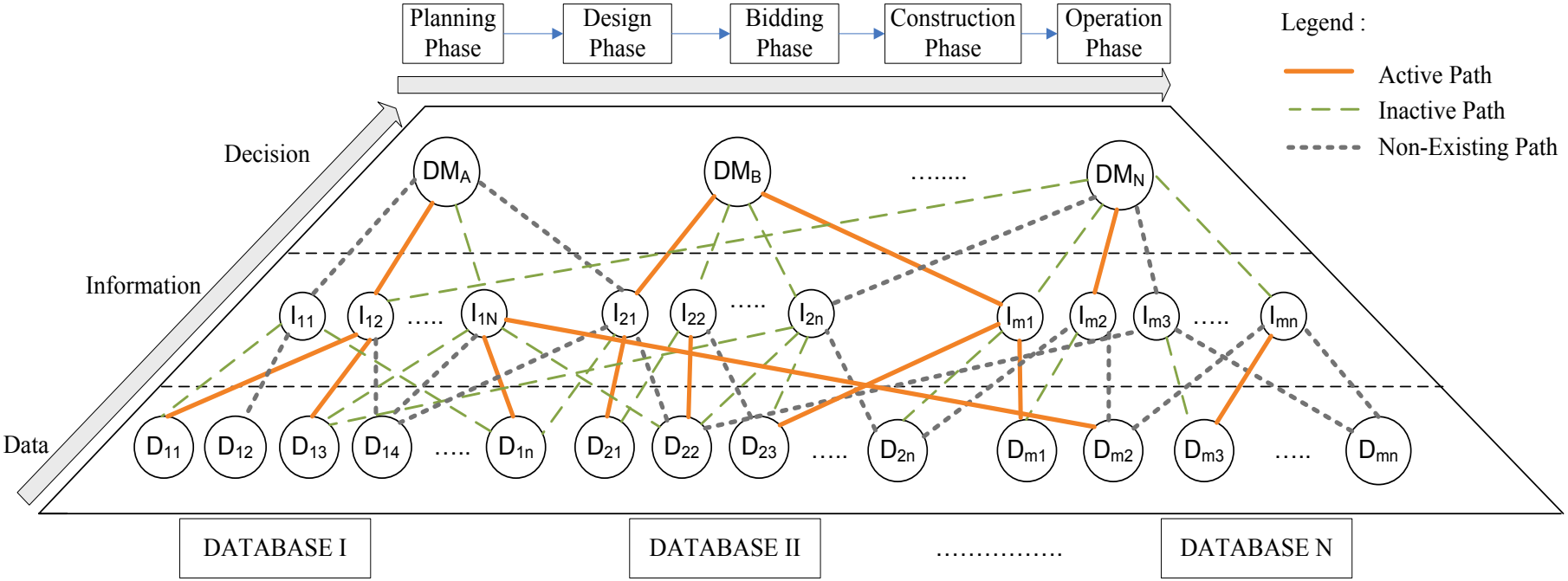
**Decisions** refers to the selection or judgment process from a set of available alternatives based on data-driven insights by utilizing the collected data and information generated.

1. Treatment selection and timing – refers to determining an appropriate type of treatment needs to be selected, as well as a proposed time frame when the treatment will be placed.
2. Project Selection – refers to possible treatments for pavement sections where projects will be selected for various pavement sections.
3. Project prioritization – refers to the ranking of projects for scheduling purposes based on available funding to address the needs of the roadway system
4. Rehabilitation or Maintenance Selection – refers to the selection of maintenance type based on the investment strategy and condition of the roadway system
5. 3-R Fund Distribution – refers to identification of funding levels needed by districts across state.



**Abbreviations:**

1. (HPMS) - Highway Performance Monitoring System Reporting
2. MEPDG - Mechanistic–Empirical Pavement Design Guide calibration
3. FWD - Falling Weight Deflectometer
4. ESAL – Equivalent Single Axle Loading
5. AADT – Annual Average Daily Traffic
6. Roadway Classification – refer to the classification of roadways into planning classes: Planning Class I (interstate system), Planning Class 2 (Commercial/Industrial System) and Planning Classes 3 and 4 (Lower level roads




Data and Information Integration Framework

| Category | | Type | | Patching | Long-Cracking | Alligator-Cracking | Fatigue Cracking | T-Cracking | Bleeding/Flus | Raveling | Punch-outs | FWD (falling | Rutting | Roughness (IRI) | Friction (Skid | Macro textur | Pavement Type | Thickness | Age | Length | Width | ESAL | Roadway Classification | NHS/Non-NHS | Number of La | Location | AADT | Treatment Cost | Deterioration Model | Cost Analysis | Pavement Condition | Structural Index (SI) | Cracking Index | MEPDG/ Design | Treatment Selection | Project Selection | Project Prioritization | Selection of Rehab/Maint | Allocate Budget to | No Use | Do not know |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | D10 | D11 | D12 | D13 | D14 | D15 | D16 | D17 | D18 | D19 | D20 | D21 | D22 | D23 | D24 | D25 | Info1 | Info2 | Info3 | Info4 | Info5 | Info6 | DM1 | DM2 | DM3 | DM4 | DM5 | | |
| Data | Structural | Patching | D1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Long-Cracking | D2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Alligator-Cracking | D3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Fatigue Cracking | D4 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | T-Cracking | D5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Bleeding/Flushing | D6 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Raveling | D7 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Punch-outs | D8 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | FWD | D9 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Functional | Rutting | D10 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Roughness (IRI) | D11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Friction (Skid resistance) | D12 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Macro texture | D13 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | History | Pavement Type | D14 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Thickness | D15 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Age | D16 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Length | D17 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Width | D18 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | ESAL | D19 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Roadway Classification | D20 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | NHS/Non-NHS Classification | D21 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Number of Lanes | D22 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Location | D23 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | AADT | D24 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Cost | Treatment Cost | D25 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Information | | Deterioration Model | Info1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Cost Analysis | Info2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Pavement Condition Index (PCI) | Info3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Structural Index (SI) | Info4 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Cracking Index | Info5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | MEPDG/Design | Info6 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Decision | | Treatment Selection | DM1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Project Selection | DM2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Project Prioritization | DM3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Selection of Rehab/Maintenance | DM4 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Allocate Budget to Districts /3R | DM5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Please assign the value as

1 If the datum is currently available and is actively used in generating information and supporting decisions

2 If the datum is currently available, but underutilized in generating information and supporting decisions

3 If the datum is missing

Leave blank if datum has no relationship or use

# APPENDIX C

## PRECONSTRUCTION SERVICE DATA

| Type of Data | Data Attributes | Description | Data Type |
|---|---|---|---|
| Cost Data | Pay period | xx/xx/xxxx date | Numeric : Ordinal |
| | Labor Hours | Number of Hours | Numeric : Interval |
| | Vehicle Miles | Mileage | Numeric : Interval |
| | Labor Dollars | Amount of Dollar | Numeric : Interval |
| | Vehicle Dollars | Amount of Dollar | Numeric : Interval |
| | Personal Expense | Amount of Dollar | Numeric : Interval |
| | Unit Price | Cost in Dollar | Numeric : Interval |
| | Plan Sheets | Number of Sheets/plans | Numeric : Interval |
| | Type of Sheet | Plan & Profile, drainage, etc. | Characters :Nominal |
| Contract Data | Project Type | Replacement; Interchange New Construction, etc. Construction; Reconstruction; Rehabilitation; Widen & Reconstruct | Characters :Nominal |
| | Type of Work | Bridges & Approaches; Grade & Drain, etc. | Characters :Nominal |
| | Let Year | Years xxxx - yyyy | Numeric : Interval |
| | Fund Type | SSP; STPY; BRFY, etc. | Characters :Nominal |
| | Project Length | z - miles | Characters :Interval |
| | Division | Geographical Division 1 , 2, 3, etc. | Numeric : Ordinal |
| | Consulting firm (CS) | $CS_1$; $CS_2$; $CS_3$; etc. | Characters :Nominal |
| | Route Type | SH; I; US: City Street | Characters :Nominal |
| Functional | Area Type | Rural; Urban; Suburban | Characters :Nominal |
| | Terrain Type | Rolling; Flat | Characters :Nominal |
| | Highway Type | Collector; Principal arterial; Freeway; Major Collector; Minor Arterial; | Characters :Nominal |
| | Highway Classification | NON-NHS; NHS | Characters :Nominal |
| | Access Control | Full, Partial, None | Characters :Nominal |
| | Vertical Alignment | "K" values | |
| | Horizontal Alignment | Degree of curve | Numeric : Interval |
| | Superelvation | X ft | Numeric : Interval |
| Roadway | Pavement Type | Asphalt; Concrete; Asphalt concrete; etc. | Characters :Nominal |
| | Shoulder Type | Sod; Asphalt; Concrete; Grass; Asphalt Concrete; None | Characters :Nominal |
| | Number of Lanes | 2; 3; 4; 6; 8 | Numeric : Interval |

| Type of Data | Data Attributes | Description | Data Type |
|---|---|---|---|
| | Lane Width | 10'; 11'; 12' | Numeric : Interval |
| | Shoulder Width | 2'; 4'; 5'; 6'; 8'; 10'; | Numeric : Interval |
| | Alignment | Existing, New located; Offset; Parallel Lanes, etc. | Characters :Nominal |
| | Section | 2D; 4D: 2L; 4L; 5O; 6L; 7L; 8L | Characters :Nominal |
| | Typical Section | Open section; Curb & Gutter; Combination | Characters :Nominal |
| | Storm Sewer | Yes (0); No (1) | Characters :Dummy |
| | Sidewalks | Yes (0); No (1) | Characters :Dummy |
| | Detour | Closed route; Closed signed route; phased; shoo-fly, etc. | Characters :Nominal |
| Bridge | Bridge Number | ID xxxxxxx | Numeric : Ordinal |
| | Span | X ft Y inch | Numeric : Ordinal |
| | Sufficiency Rating | 0- 100 | Numeric : Interval |
| | Construction Year | XXXX, built year | Numeric : Ordinal |
| | Bridge Width | X ft | Numeric : Interval |
| | Bridge Length | Y ft | Numeric : Interval |
| | Clearance | Z ft | Numeric : Interval |
| | Load | M ton | Numeric : Interval |
| Environment | NEPA Document | Linguistic | Text |
| | Permit Type | COE; OWRB; FAA; COE; OWRB; USACE; USCOE | Characters :Nominal |
| Right-of-Way | ROW Requirement | Yes (0); No (1) | Characters :Dummy |
| | Utility Conflicts | Yes (0); No (1) | Characters :Dummy |
| Traffic | ADT | Traffic Count | Numeric : Interval |
| | New Guardrail | Yes (0); No (1) | Characters :Dummy |
| | End Treatment | Yes (0); No (1) | Characters :Dummy |
| | Highway Lighting | Outside, median or no lighting | Characters :Dummy |
| | Traffic Signals | Yes (0); No (1) | Characters :Dummy |
| Outsource | Critical time Constraint | Number 1 - 10 | Numeric : Ordinal |
| | Work Volume | Number 1 - 10 | Numeric : Ordinal |
| | Planning Difficulty Level | Number 1 - 10 | Numeric : Ordinal |
| | Requirement of Special Skill | Number 1 - 10 | Numeric : Ordinal |
| | Amount of Inspection Required | Number 1 - 10 | Numeric : Ordinal |
| | Degree of Labor Intensity | Number 1 - 10 | Numeric : Ordinal |
| | Political reasons | Number 1 - 10 | Numeric : Ordinal |
| | Quality of Service | Number 1 - 11 | Numeric : Ordinal |

# APPENDIX D

## QUESTIONNAIRE SURVEY III

## Assessment of Data and Information Utilization in Preconstruction Service Decision-Making Process

| Category | | Type | | Project Type | Type of Work | Location | Project Length | Roadway Classification | NHS/Non-NHS Classification | ROW | Grade | Number of Lanes | Pavement Type | Lane Width | AADT | Labor (Engineering) | Labor Dollars | Vehicle Dollars | Personal Expenses | Estimated Work Hours | Estimated Cost | Budget Allocation | Preconstruction Service Cost | Project Selection | Environmental Decision | Project Scheduling | Resource Allocation | No Use | Do not know |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | D10 | D12 | D13 | D14 | D15 | D16 | D17 | Info1 | Info2 | DM1 | DM2 | DM3 | DM4 | DM5 | DM6 | | |
| Data | Roadway Inventory | Project Type | D1 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Type of Work | D2 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Location | D3 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Project Length | D4 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Roadway Classification | D5 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | NHS/Non-NHS Classification | D6 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | ROW | D7 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Grade | D8 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Number of Lanes | D9 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Pavement Type | D10 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Cost | Lane Width | D12 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | AADT | D13 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Labor (Engineering) Hours | D14 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Labor Dollars | D15 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Other | Vehicle Dollars | D16 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Personal Expenses | D17 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Information | | Estimated Work Hours | Info1 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Estimated Cost | Info2 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Decision | | Budget Allocation | DM1 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Preconstruction Service Cost Estimation | DM2 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Project Selection (Inhouse/Outsource) | DM3 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Environmental Decision | DM4 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Project Scheduling | DM5 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Resource Allocation | DM6 | | | | | | | | | | | | | | | | | | | | | | | | | | |

Please assign the value as

1 If the datum is currently available and is actively used in generating information and supporting decisions

2 If the datum is currently available, but underutilized in generating information and supporting decisions

3 If the datum is missing

Leave blank if datum has no relationship or use