Highway Data Quality Report Card (HDQ$_{RC}$): A Conceptual Framework of Pavement Management Data

Asregedew Woldesenbet$^1$, H. David Jeong$^2$

$^1$ Doctoral Student, Dept. of Civil, Construction & Environmental Engineering, Iowa State University, Ames, IA 50011, asre@iastate.edu (405)-762-3261

$^2$ Associate Professor, Dept. of Civil, Construction & Environmental Engineering, Iowa State University, Ames, IA 50011, djeong@iastate.edu (515)-294-7271

ABSTRACT

Today, state highway agencies (SHA) invest a large amount of money in collecting, storing and managing various types of data ranging from roadway inventory to pavement condition data during the life cycle of highway projects. Despite this huge investment, the use of data in satisfying users’ need to extract information and knowledge and support decisions has become questionable. This paper presents a conceptual highway data quality report card (HDQ$_{RC}$) which can indicate the level of data quality by identifying decision-makers’ requirements through the application of a quality function deployment (QFD). The report card utilizes ten data quality dimensions identified from review of literature and interviews with department of transportation (DOT) managers to analyze the data needs of pavement management decision-makers’ at a project level. A multi-attribute approach is utilized to analyze the level of data quality with respect to active utilization. The output of this paper may be used as a tool to evaluate the status of various highway project data by linking decision-makers’ requirement with data quality dimensions. This type of report card will also help justify the return on investment on the continuous and growing data collection efforts.

Key Words: quality function deployment, pavement management data, pavement condition data, highway decision-making, requirement analysis, report card
Introduction

The advancement in data collection methods, digital data storage technologies and database management systems has enhanced the way highway agencies collect, store and manage data. For instance, the Oklahoma Department of Transportation (ODOT) pavement management division alone has approximately 1.5 million pavement condition records in its pavement management system (PMS). ODOT estimates 8000 miles of data are collected every 100th mile or 800,000 records annually that consist of 65 data fields that are approximately 52 million pieces of data as part of the asset management program (Calvarese, 2007). However, the benefits of using this valuable asset in generating information and knowledge needed to support highway decisions is questionable compared to the amount invested. ODOT spends at least $600,000 annually to collect statewide pavement condition data. Sometimes, potential users of these data and information such as highway project schedulers, estimators, and managers do not even know what type of data are available or how to access these data to support their decision-makings (Hummer et al., 1999).

Some of the reasons that may account for the poor usage of highway data might be 1) minimal recognition or interest in existing data; 2) lack of in-house resources and capabilities to analyze data; 3) insufficient data to perform any meaningful analysis, 4) nonstandard and non-digital data format; and/or e) poorly defined procedures and mechanisms used to extract, process, and analyze the data and generate usable information and knowledge to assist highway project decision-makers'. The purpose of this paper is to identify users’ requirement and assess the current level of data quality from decision-makers’ perspective by developing a conceptual highway data quality report card (HDQRC) through the application of quality function deployment (QFD).

Background

Currently, the importance of data quality is becoming a critical issue to highway agencies with respect to the amount of data produced and the information generated to make reliable decisions. Lee and Strong (2003) argue that the purpose of data production process is to produce data for users and define quality data as “data that are fit for use” by users. Quality data allows to achieve better credibility within an organization, integrate with other agency data, improve decision support for managers, increase accuracy in reporting existing condition and comply with external data requirements (Shekharan et al., 2006). One of the key tasks in addressing quality of data is to identify and understand decision-makers’ requirement and evaluate existing data to meet their needs: what type of data do designers, project managers, and schedulers require to support their decisions? Are currently available data interpreted in the same manner; reflect the details of the original observation; or include relevant data to support decision-makers’ requirements?

Identifying and understanding data users’ requirement allows detecting problems behind the minimal usage and improving the quality of data to effectively generate and place the right information and knowledge in the hands of decision-makers’. Requirement analysis may be defined as 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). Thus, this paper proposes a requirement analysis through the application of a quality function deployment (QFD) to identify decision-makers’ needs in a more integrated manner by linking it with data quality dimensions. The approach allows to evaluate the level of
data quality through a multi-attribute approach in developing a conceptual highway data quality report card (HDQRC). This paper utilizes pavement condition data as an example to illustrate the application of QFD.

**Pavement Condition Data Collection & Use**

Pavement condition data are one of the critical data collected by highway agencies during the operation and maintenance phase as part of the asset management program. These data are usually stored as part of a pavement management system (PMS). 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. Roadway inventory, traffic and cost data are additional components included in a PMS to support various decisions. Roadway inventory incorporates pavement classification, pavement type, section, length, width, etc., while traffic data incorporates the traffic profile or growth, annual average daily traffic (AADT), and traffic year to determine structural capacity of existing and future pavements. Table 1 shows examples of types of data collected for an asphalt pavement in a typical highway agency PMS.

<table>
<thead>
<tr>
<th>Division/ Database</th>
<th>Type of Data</th>
<th>Sub-Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pavement Management System (PMS)</td>
<td>Pavement History</td>
<td>Pavement surface type, thickness, composition, etc.</td>
</tr>
<tr>
<td>Structural Data (Distress Data)</td>
<td>Transverse Cracking, Patching, Bleeding, Raveling, Fatigue, Polishing, shoving, etc.</td>
<td></td>
</tr>
<tr>
<td>Functional Data</td>
<td>Average Roughness, Ride, Average Rut depth, etc.</td>
<td></td>
</tr>
<tr>
<td>Roadway Inventory</td>
<td>Pavement type, section, length, width, etc.</td>
<td></td>
</tr>
<tr>
<td>Traffic Data</td>
<td>Traffic profile, AADT, etc.</td>
<td></td>
</tr>
<tr>
<td>Other (structural)</td>
<td>Friction, Deflectometer (FWD), ESAL</td>
<td></td>
</tr>
</tbody>
</table>

The pavement condition data mainly contain one record per 100th mile of structural layer and surface condition of roadway collected in annual or bi-annual basis depending on the agency. 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.
In a typical highway project, these pavement condition data are used as information and supporting various decision-making processes at various levels across the life cycle of a project. Some of these information use could be characterization of current pavement condition, development of pavement deterioration curves, life-cycle cost analysis and projecting future conditions, while decisions include treatment selection, prioritizing projects (short and long-term), selection of pavement designs (pavement thickness, type of pavement, etc.) and allocation of budgets to districts or regions. For instance, Oklahoma DOT uses pavement condition data such as ride, rut depth, structural and functional data to develop a pavement quality index (PQI) in describing current condition. Other uses of pavement condition data include highway performance monitoring system (HPMS) reporting, asset management, and calibration of the mechanistic–empirical pavement design guide calibration, MEPDG (Pierce et al. 2013). Overall, pavement condition data are essential in the planning, budgeting and programming of pavement preservation, rehabilitation and reconstruction of highway projects. It is important to note that data requirement for each type of decision varies from one to another. In this paper, a pavement treatment selection is used as an example decision to assess the level of pavement condition data quality by applying quality function deployment (QFD) from pavement managers’ perspective.

Quality Function Deployment (QFD)

Quality Function Deployment (QFD) is one of the requirement analysis methods 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”. In QFD, design quality of a product is achieved by identifying the relationship between the demands and quality characteristics using charts and matrices to trace customers need, evaluation of quantitative importance, and multi-discipline team in stepwise manner using top-down approach (Kamara et al. 1999). 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 the 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).

The QFD House of Quality is one of the tools that is used to transform customer’s voice to the designer. The term “house” acquired its name from the fact that the tool has a shape of a house.
The house consists of six basic functions to interlink customers need with designers’ goal (Figure 1). These include:

1. Voice of Customer: addresses the customer’s requirement presented as a list of needs where customer rate the importance of each need. It is referred as the “What” which is located at the right part of the house.
2. Voice of organization: also referred as the “How” is used to put the characteristics required by a customer within a product. It is located at the top of the house.
3. Relationship Matrix: is the middle of the house that is used to show the relationship between the voice of the customer, the “what” and voice of the organization, the “how”.
4. Correlation Matrix: also called the roof of the house is used to show the relationship or between voices of the organization or the “Hows”.
5. Competitive Analytics: is used as comparison of current system with other competitors to show the relative performance of a system. It is also referred as marketing information and customer perception of other competing products. It is located on the left part of the house.
6. Design Target: is the bottom part of the house where technical assessment of a system or product is conducted.

Application of QFD

The application of QFD house of quality is illustrated hypothetically through the use of a pavement condition data in selecting pavement treatments by taking a particular highway agency, Oklahoma DOT (ODOT). The goal of this paper is to assess the current level of data use and indicate if data quality needs improvements to meet decision-makers’ requirement in an agency. The first step in QFD house of quality is to determine the “what” or the voice of the customer and “how” or the voice of the organization. In this paper, pavement management decision-makers’ involved in the decision-making of treatment selection are considered as the customers or data users, where specific asphalt pavement condition data that may attribute to the selection of treatments are considered as their requirement or “what” or voice of the customer. Examples of structural, functional condition data and pavement history along with roadway inventory, traffic and cost data are listed to left of the house (Figure 2). In order to meet decision-makers’ requirement and improve the quality of data (how), this paper uses ten key data dimensions or attributes categorized into two categories as the voice of the organization based on review of literature and interviews with highway decision-makers’.

1. Quality: One of the decision-makers’ requirement in generating reliable information and knowledge is acquisition of quality data. The United Kingdom Audit Commission (2007) has set six characteristics to represent data quality: accuracy, validity, reliability, timeliness, relevance and completeness, while Loshin (2006) suggests eight quality dimensions: uniqueness, accuracy, consistency, completeness, timeliness, currency, conformance, and referential integrity. In this paper, quality is characterized by seven attributes:
   a. Definition: deals with proper and common understanding of a data being collected. Various agencies define pavement condition data differently. Some agencies might collect the length of a crack, while others collect the type, severity and extent of the crack (Pierce et. al 2013). Thus, data should be well-defined in terms of its content that can be easily communicated throughout users without any confusion. Data should have the same meaning across users through the use of standards such as American Association of State
Highways and Transportation Officials (AASHTO) and American Standard for Testing and Materials (ASTM).

b. Precision: deals with the accuracy of data collected either on the site or in the office. The output of accurate data is represented through its use in generating information and developing reliable prediction models. For instance, the ODOT pavement management team has a set types of severity levels 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) in collecting pavement condition data.

c. Validity: It can also be referred as 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.

d. Relevant: Data should be specific and relevant to meet its intended purpose of generating information and supporting decisions. Data that is of no use, should not be collected. For instance, the use of the international roughness is index (IRI) is becoming a huge concern among pavement management decision-makers’ if it represents a reliable measure of pavement condition.

e. Consistency: The level of data consistency or data variability is another huge concern. For instance, a 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). Or wrong data might be recorded by mistake where a faulting may be reported in asphalt pavement. Thus, data collected through different cycles should be consistent and make sense to perform reliable analysis.

f. Completeness: Data should include all relevant and necessary data required by the decision-maker to support decisions. Proper care must be taken to monitor any missing, incomplete or duplicated data that may create problem in analyzing data to meet user requirement.

g. Structure: Data should be well-documented and recorded in the right format and structure to perform meaningful analysis. Data may be stored in paper format or pdf files which makes it difficult to extract information or perform analysis. 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.

2. Time: Another potential data dimension that may be attributed to address decision-makers’ requirement is time that is spent to acquire, process and analyze data. Data should be easily available at the right time for the user to make the right decision.

h. Accessibility: Data should be accessible and ease of use to query data must be facilitated for potential users across the department or division.

i. Timeliness: Data should be up-to-date and be available in a timely manner for the user to generate reliable information and make efficient decisions.

j. Constancy: The meaning of data should remain constant over a period of time. If there is a change in meaning or use over time, proper mechanism should be applied to address it.

These data quality dimensions or attributes are listed as the voice of the organization (top of the house or how decision-makers’ data requirements may be satisfied for improved use. Once the “what” and “how” are set, the second step is to assign weights to the respective pavement condition data to reflect the relative importance of data requirement to the decision-maker in
selecting treatments. The decision-maker may assign a value ranging from 0 to 1. For instance, in Figure 2, fatigue cracking and treatment cost have higher importance level to decision-makers’ with a value of 0.15 whereas age, pavement type and traffic profile have low importance with values of 0.01, 0.02 and 0.02 respectively. It should be noted that these values are assigned for demonstration purposes and might have equal importance in real conditions and the summation of relative importance should add up to 1. In addition, the pavement data are examples used for illustration purposes and additional data or missing data can be included based on the requirement analysis.

Third, the decision-maker evaluates how close each type of pavement management data is to the optimum quality to be utilized in generating information and knowledge in the selection of pavement treatment decision. The decision-maker can use the columns in the respective attribute to evaluate the quality of data listed in each raw. The decision-maker may allocate weights based on a Likert scale approach of 1 – 9 where 1 refers to low quality for use in generating information and supporting decision while a rating of 9 refers to excellent quality data. Basically, this represents the relationship matrix between specific pavement management data and quality attributes shown in the middle of the house (Figure 2). For instance, patching data performs well with most of the data dimensions which have a value of 9 except definition (A) and precision or accuracy of data (B) which have both a value of 1. This indicates that the quality of data with respect to clearly defining the data based on recognized standards and acquiring precise and accurate data needs improvement the use of data in supporting treatment selection decision.

![Figure 2 Pavement Condition Data Quality Report Card](image)
Fourth, the overall performance of the pavement management data is compared with other systems with respect to each data dimensions listed through A – J in the top of the house. The decision-makers’ perception may be used as competitor analytics to measure the current status of pavement management data (DB1) with other database systems (right of the house). The goal of this section of the house is to visually identify the relative performance of pavement management data with respect to other data systems in each data quality dimension. A rating based on a Likert scale of 1-5 may be used where 1 represents ‘very poor’ performance requiring new data collection scheme, 2 represents ‘poor’ performance which needs major adjustments, 3 represent ‘average’ use, 4 represent ‘good’ quality data that can be used for information generation with minor adjustments and 5 represents ‘excellent’ performance and does need any change. For this paper, existing database systems, roadway inventory (DB2) and traffic database (DB3) are used as an example. Based on the competitive analytics, pavement management data performs well with respect to precision, validity, and completeness, while it needs improvement in terms of definition, relevance, structure, accessibility and temporal reliability of data. For instance, pavement management could adopt lessons learned or best practices from traffic database with respect to definition and structure or format.

Finally, a multi-attribute approach is utilized to assess the level of data quality and develop a report card by aggregating the key quality attributes. Weighted score method is utilized to match the relative importance with the decision-makers’ evaluation (Eq. 1). For example, based on Eq. 1, the weighted score for definition data dimension (A) is 3.98. This weighted score is then summed and divided by 9 to normalize the score which resulted in 0.44. The normalized score is multiplied by 100 to obtain an aggregate grade of 44% as a percentage value (Eq. 2).

\[
WT_i = \sum_{j} R_i \times PR_j
\]  

\[
NS_i = \frac{WT_i}{9} \times 100
\]

Where, $WT_i$ = Weighted total for specific type of a pavement management data i  

$R_i$ = the decision-maker’s relative importance of data based on value of 0-1 where j represent the ten data quality attributes $1 < J < 16$  

$PR_i$ = the performance of data based on a Likert scale of 1-9  

$NS_i$ = Normalized score as percent grade

A grading system based on 100 scale is used to convert the scores into letter grades. In this paper, definition of data will have a grade of ‘F’ which indicates a new data definition scheme should be developed to improve the quality except for pavement type, thickness and functional class data types. These steps are performed for all pavement condition data and an aggregate grade is used to measure the overall data performance of a certain database system by taking an average value. For this example, the pavement management data appears to have a grade of “C”. This design target is shown at the bottom of the house to interpret the technical evaluation of data. The grading system used to convert the scores into letter grades are:
A = 90-100% - Outstanding data for performing meaningful analysis and decision-making
B = 80-89% - Good data for use in analysis and decision-making
C = 70-79% - Acceptable data that can be used as information without any analysis
D = 51-69% - Poor data that is difficult to use
F = 50% or lower – Needs new data collection scheme

Discussion
The proposed approach fits appropriately in addressing the issue of data quality and assessing the current level of use by interlinking decision-makers’ data requirements and quality dimensions in a highway project decision-making process. QFD house of quality can be implemented in a stepwise manner at various levels of decision-making such as network level, program level, project-selection level and project level to identify decision-maker’s requirement at each stage of a highway project development (Figure 3). Or it can be applied at specific project level across a project phase starting from the design phase to construction, maintenance and operation phase to address various users requirement. This allows the flexibility in addressing data needs of various users either across a highway decision-making hierarchy or a project life-cycle. Once individual measurements at specific highway decision-making hierarchy are assessed, this process may be repeated to evaluate all data sets collected in an agency to develop a comprehensive report card.

This QFD house of quality approach can serve two main purposes: 1. Internal evaluators or knowledgeable decision-makers’ within a highway agency may assess their data quality needs to improve quality of data as part of a quality control or quality assurance (QC/QA) program, or 2. External evaluators may evaluate the status of data quality and usage that can serve as periodic data report card that may be adopted for state highway agencies similar to ASCE’s infrastructure report card. However, it is important to note that the maturity level of agencies varies from one to another with respect to the amount, type and level of detail of data collected along with the technologies and analysis utilized to support decisions. An agency may use fully automated data collection using laser scans and utilize econometric approaches to select treatments while another may use manual-based data collection and expert-based analysis or a simple subjective decision-
making process to select treatments. Thus, an internal assessment must not be used to evaluate the performance of one agency with another agency, an external evaluation based on benchmarks and required level of service should be made by higher level authorities like the Federal Highway Administration (FHWA).

Conclusions

This paper demonstrated a conceptual approach for measuring and tracking highway data quality through the application of quality function deployment (QFD). The QFD used a correlation matrix to increase the cross functional integration between decision-makers’ data requirements and specific technical data characteristics to convey the voice of the decision-maker to the quality management team. This approach is suitable to the highway industry to address various data users’ requirement that involve multiple highway divisions and decision-makers’ across a project lifecycle. By utilizing this approach, a highway data quality report card (HDQRC) was developed through the use of pavement condition data at a project level of decision-making.

The developed report card will have the ability to enhance the efficiency of highway data in generating information and making reliable decisions by addressing user requirements. It will set a benchmarking example to agencies in the area of data quality assessment from decision-maker perspective. This type of report card not only guide DOTs’ on how to effectively collect data and improve its quality, but also help to develop active utilization plan of currently existing databases and place the right information in the hands of decision-makers’. The report card may also be communicated among higher officials and the public to easily understand current status of data and help justify the return on investment on the continuous and growing data collection efforts. It is important to note that this paper implemented five of the six components of QFD house of quality with the exception of correlation matrix. The correlation matrix is in developmental stage of this study which aims at developing an integrated framework of highway data, information and decisions through the application of social network analysis (SNA).

Reference