Data analytics and visualization for enhanced highway construction cost indexes and as-built schedules

by

Krishna Prasad Shrestha

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Major: Civil Engineering (Construction Engineering and Management)

Program of Study Committee:
H. David Jeong, Major Professor
Douglas D. Gransberg
Charles T. Jahren
Jing Dong
Zhu Zhang

Iowa State University

Ames, Iowa

2016

DEDICATION

To my sister, brother, and mother, thank you all for your love and support in every step of my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my Advisor Dr. H. David Jeong for guiding me throughout the course of this research. It is my great privilege to have such a helpful and understanding Professor as my advisor. He always encouraged, motivated, and pushed me to the limit to get the best out of me.

Many thanks to the program committee members, Dr. Douglas D. Gransberg, Dr. Charles Jahren, Dr. Jing Dong, and Dr. Zhu Zhang for their support and feedback for the dissertation. I appreciate their constructive feedback, which gave me opportunities to improve the research.

I am grateful to the Mid-American Transportation Center (MATC) and Montana Department of Transportation for funding this research. The MDT played a very important role for this study by providing the historical bid data. I also wish to thank another state department of transportation that provided the daily work report data.

I would also like to express my gratitude to Dr. Pramen P. Shrestha, my master’s thesis advisor who introduced me to the wonderful world of cutting edge research. In addition, I would like to thank my friends, colleagues, the department faculty, and staff for making my time at Iowa State University a wonderful experience. I would especially like to acknowledge the help and support from Ahmed Adbelaty and Jorge Rueda throughout the three years of my doctoral study.

Finally, none of this would have been possible without my family, especially my sister Ms. Buddha Laxmi Shrestha. She always encouraged me to believe in myself and guided me in every step of my life. Her unconditional love and support is the first reason for my achievements and success. I cannot forget my brother, Mr. Babu Krishna Shrestha, for doing everything he can for my better future. Many thanks go to my mother, Ms. Ganga Laxmi Shrestha, and my father, Mr. Hari Krishna Shrestha, for their never-ending love.

ABSTRACT

A considerable amount of digital data is being collected by State Highway Agencies (SHAs) to aid project-planning activities, support various project level decision-making processes, and effectively maintain and operate constructed highway assets. However, the highway construction industry has been significantly lagging behind utilizing the growing digital data to support business decisions compared to other industry sectors such as health care and energy. The significant lack of understanding on the linkage between raw data collected and various decisions, proper computational methodologies, and effective guidance is considered as major barriers to the full utilization of the digital data.

This study uses digital datasets that are now commonly available in SHAs, to demonstrate the smart utilization of existing digital data to support and enhance decision-making processes using data analytics and visualization methods. This study will a) develop an advanced computational methodology to generate multidimensional highway construction cost indexes (HCCIs) using two new concepts of i) dynamic item basket and ii) multidimensional HCCI, b) develop an enhanced framework for collection and utilization of digital Daily work Report (DWR) data, c) develop an automated methodology to generate as-built schedules using data collected from existing DWR systems, and d) analyze as-built schedules to develop a knowledge base of frequent precedence relationships of activities. The study achieves those objectives by utilizing three digital datasets: bid data, DWR data, and project characteristics data. Further, two standalone prototype systems, namely, Dyna-Mu-HCCI and ABSS are developed to automate

computational methodologies for multidimensional HCCI calculation and as-built schedule development respectively.

This study will aid SHAs to utilize currently unused datasets for informed budgeting and project control decisions. It demonstrates the importance of data analytics and visualization to obtain more value from the investment made in collecting construction data. Overall, this study serves as a step in making a transition from experience driven to data driven decision making in the construction industry.

CHAPTER 1

INTRODUCTION

Background and Motivation

The size of digital universe is estimated to increase by 10 times – from 4.4 trillion gigabytes in 2013 to 44 trillion in 2020 – which has resulted in a phenomenon called "big data" (Turner et al. 2014). The big data is a "paradigm shift from hypothesis-driven to data-driven discovery" that allows to automatically extract "new knowledge about the physical, biological, and cyber world" (Wactlar 2012). Many other industries such as health care, energy, and agricultural sectors have utilized their digital data to make reliable business decisions and generate significant financial values (Manyika et al. 2011; McKinsey Center for Business Technology 2012).

The construction industry is known for collection, processing, and exchange of a large amount of data among project stakeholders (Cox et al. 2002; Hendrickson and Au 2008). Traditionally, most of the data collections and exchanges are paper-based and require manual effort for analysis. Project owners such as state Departments of Transportation (DOTs) in the highway industry have started to develop and implement digital systems to ease and streamline data collection, storage, and analysis. However, the construction industry is still lagging behind compared to other industries on utilizing digital data (Manyika et al. 2011; McKinsey Center for Business Technology 2012; Woldesenbet et al. 2015).

The data collected from one stage of a project can be useful resources for making decisions in the other stages of the same project as well as for life cycle decision-making for future projects. Despite the potential of using the data to generate meaningful, actionable, and hidden insights to support various decisions, state DOTs collect most of the data to meet federal and state requirements rather than for their actual analysis and utilization. Possible reasons for underutilization of data in state DOTs may include: a) lack of data attributes necessary for analysis; b) lack of methodologies to extract, clean, transform, and analyze the data; c) lack of resources to analyze the data; d) lack of automation for the analysis; and e) lack of visualization techniques to present the insights obtained from the analysis to support various decisions (Woldesenbet et al. 2014).

The concept of big data analytics and visualization can be applied to various datasets that are also growing rapidly with the introduction of digital project delivery and various digital data collection systems in the highway industry. However, currently, most of the decisions in the DOTs are still heavily dependent on engineers' experience and judgements that can be easily biased. This research is to study and demonstrate how emerging big data analytics and visualization techniques can be effectively used to generate actionable insights and to support and improve the major decision making process of state DOTs.

Problem Statement

Construction cost and schedule certainties are vital for successful planning and execution of construction projects. State DOTs develop cost estimates and schedules throughout the project life cycle. Although the cost estimates developed at various stages of a project life cycle are important, the planning level construction estimates are particularly important. The estimates become the budget for a project which are presented to the public and various agencies at local, regional, state, and federal levels for comments and reviews (Wilmot and Cheng 2003). If there are any changes in the costs of a project after its approval, state DOT officials usually have to defend the situation publicly or in the state legislature.

To keep the construction costs low by obtaining lower bids, realistic and optimal schedule development and contract time determination are essential (Iowa Department of Transportation (IADOT) 2012a). Proper schedules are also required to quantity the construction and inspection resources; and reduce road users' inconvenience, likelihood of crashes, and operating & maintenance costs of vehicles (Anastasopoulos et al. 2008; Zaniewski et al. 1982). During the construction, as-built schedules need to be developed by to document actual construction sequences and durations. This as-built schedule can be compared with the original schedule to ensure that construction projects are progressing at the desired pace. If not, the information can be used to detect any deviations, identify their causes, quantity its impact on overall schedule, identify corrective measures to get the schedule back on track, and resolve delay related claims filed by contractors (Alavi and Tavares 2009; Joint Federal Government/Industry Cost Predictability Taskforce 2012).

Despite the importance of cost estimation and scheduling, the cost overruns and delays are prevalent in the construction industry. Almost half of the large transportation projects in the U.S. overrun their initial budgets and experience delays (Bordat et al. 2004; Crossett and Hines 2007; Shane et al. 2009). Lack of a proper cost estimation procedure and, in particular, the underestimation of inflation rate is the most important factor resulting in inaccurate cost estimates and hence cost overruns overrun (Alavi and Tavares 2009). A proper Highway Construction Cost Index (HCCI) should be developed and used to account for the inflation. However, many state DOTs do not have a reliable methodology to calculate HCCIs and are looking forward to update their HCCI calculation methodologies (Walters and Yeh 2012).

Although studies have found that factors such as project location and item quantities affect the unit costs of construction items, those factors are currently neglected when calculating HCCIs (Jain et al. 2015; Rueda 2013). Thus, the estimates developed using such HCCIs are likely to be highly inaccurate and far off from the actual project costs. As such, the high level budget allocation decision driven based on such HCCIs can be significantly misleading and financial obligations expected by state DOTs can be severely different than the actual financial obligations.

A proper schedule development is another vital for successful construction project management and execution. Schedule development is a complex process that requires knowledge of construction methods, materials, and labor productivity (Bruce et al. 2012). Developing a realistic schedule is challenging for inexperienced as well as experienced schedulers (Fischer and Aalami 1996; Jeong et al. 2009). Current schedule development methods are manual and heavily dependent on schedulers' experience. Similarly, as-built schedules are also developed manually at the end of the project based on the outdated information (Hegazy et al. 2005; Kahler 2012). Daily Work Report (DWR) data contain valuable information such as activities conducted by date and resources utilized which can be used to generate as-built schedules and aid in developing schedules for future projects. However, state DOTs lack methodologies for developing the as-built schedules based on DWR data which can further be used to aid in developing the original planned schedule.

Thus, there is a need to develop methodologies to automate as-built and original schedule development.

Beside the schedule development, there are many other potential applications of DWR data such as progress monitoring, production rate estimation, contract time determination, contractor payment, dispute resolution, and risk identification. However, current systems are not necessarily designed with due consideration for those applications. As such, there is a need to develop a framework that can be used to collected DWR data properly for their improved utilization for making construction management decisions. This will also aid in moving the construction industry forward in terms of the data use and data analytics to make informed decisions. Some of the questions this study will aim to answer are:

- How can state DOTs improve current HCCIs to overcome the early cost estimation and budgeting issues?
- How can state DOTs improve the existing DWR data collection and utilization framework?
- How can state DOTs utilize existing DWR data to develop as-built and original schedules?

Research Objectives

The primary goal of this research is to develop and illustrate the methodologies to improve the collection of important data attributes, extract relevant data attributes for various analysis, transform and analyze the datasets using various data mining techniques, and visualize the results using advanced visualization techniques to aid in making decisions related to construction costs and schedules. The specific objectives of this study are to:

- Develop a methodology to calculate an advanced Highway Construction Cost Index (HCCI) using historical bid data.
- Develop a framework for better collection and utilization of Daily Work Report (DWR) data,
- Develop a framework to automate the as-built schedule development to improve project schedule control and settle claims;
- Develop a framework to discover precedence relationships of activities to aid in schedule development for future project using as-built schedules generated from DWR data;

Research Scope

The scope of this study is limited to the three major datasets: project information, Daily Work Report (DWR), and project bid data. Project information and bid data will be used to overcome early cost estimation issues by developing an advanced Highway Construction Cost Index (HCCI). A project information dataset includes data attributes such as project location, type, size, length, and total construction costs. A DWR dataset includes data attributes about ongoing work activities, equipment usage, labor hours, weather, and site conditions. A bid dataset contains bid items, their descriptions, quantities, unit costs, and corresponding project IDs.

The first paper in this study investigates the current practices of HCCI and develops a new methodology to calculate multidimensional HCCI using dynamic item basket that takes into account the important project characteristics such as project location, size, and type. The latter three papers deal with the development of a framework for improved DWR data collection and utilization; automation of as-built schedule development that can be used for progress monitoring and claim settlement; and development of precedence relationships of activities based on the as-built data using Sequential Pattern Mining (SPM) to aid in schedule development.

Methodology

The overall methodology for this study is presented in Figure 1. To develop a multidimensional HCCI, first, literature review and nationwide questionnaire survey is conducted to identify the current practices and processes. A sample bid and project information datasets are obtained from Montana Department of Transportation (MDT). The data is analyzed to quantify the effect of location, project size, and project type by developing a multidimensional Highway Construction Cost Index (HCCI).

On DWR side, first, existing DWR systems are reviewed. After that, two nationwide questionnaire surveys are conducted to understand current practices of collecting and utilizing DWR data in detail. Possible benefits, benefiting teams, and importance of various DWR data attributes are identified. The data attributes required to obtain those benefits by the teams are analyzed to develop to develop an enhanced framework that is later validated by DWR experts from the U.S. and sample DWR database.

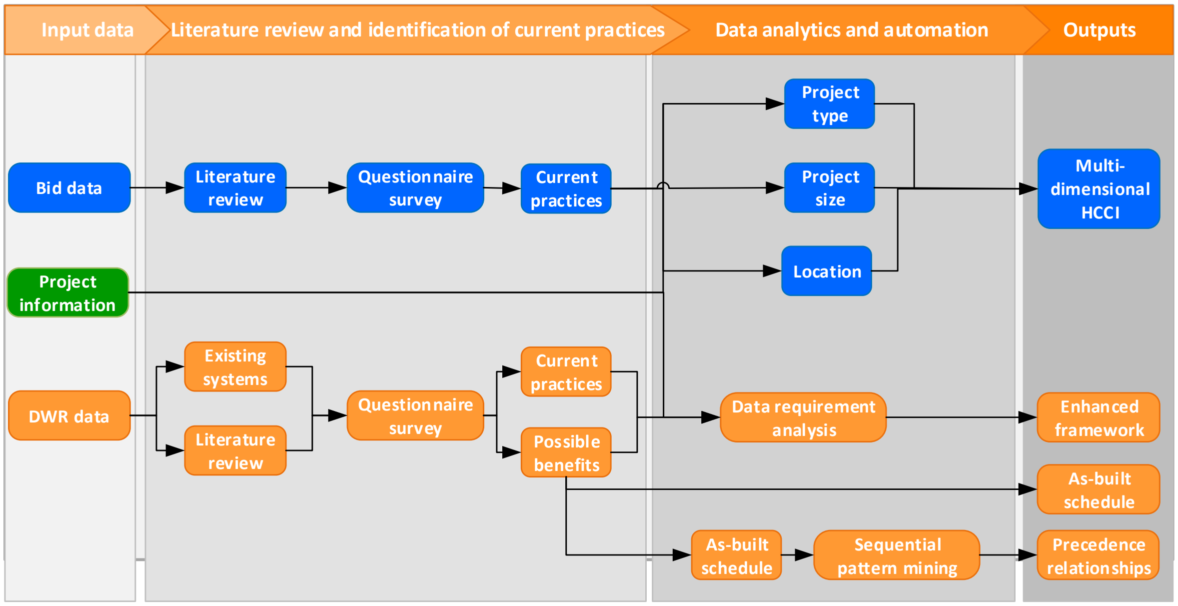

Figure 1 Overall research plan

An automated as-built schedule development methodology is developed and illustrated with an example based on a sample DWR and project information dataset. The methodology utilized Structure Query Language (SQL) to extract and Visual C#.NET frontend to visualize as-built schedules. The as-builts are further transformed and analyzed using SPM to obtain frequent construction activity sequences. The sequences are visualized in a precedence diagram, which can be used to aid as-planned schedule development for future projects.

Expected Contribution

This study utilizes the concept of big data analytics and visualization to the underutilized construction datasets, which will aid state DOTs in making data-driven decisions that are more reliable, accountable, defensible, and transparent. Specifically, the advanced multidimensional HCCI developed based on an extensive analysis of bid datasets will properly reflect the market conditions and aid state DOTs in making more reliable cost estimates for budgeting purposes. An enhanced DWR system based on the framework will aid state DOTs to improve DWR data collection and their utilization for making construction management and project control decisions. The automated method developed to prepare as-built schedules will help state DOTs in making project control decisions in real-time. The as-planned schedule development method based on an advanced SPM will enable state DOTs to develop reliable as-planned schedules efficiently with confidence and with less effort. The DWR framework developed in this study will be vital in developing a new DWR system or improving existing ones. Overall, this study will help in transforming the construction industry from current experience-based decision-making to the data-driven decision making.

CHAPTER 2

MULTIDIMENSIONAL HIGHWAY CONSTRUCTION COST INDEXES USING DYNAMIC ITEM BASKET

K. Joseph Shrestha[1], H. David Jeong[2], and Doug D. Gransberg[3]

Abstract

A Highway Construction Cost Index (HCCI) is an indicator of the purchasing power of a highway agency. Thus, it must reflect the actual construction market conditions. However, current methods used by most state departments of transportation are not robust enough to meet this primary goal due to a) a significantly insufficient sample size of bid items used in HCCI calculation and b) inability to address the need to track highway construction market conditions in specific sub-market segments in terms of project type, size, and location. This study proposes an advanced methodology to overcome these apparent limitations using two new concepts: a) dynamic item basket and b) multidimensional HCCIs. The dynamic item basket process identifies and utilizes an optimum number of bid item data to calculate HCCIs in order to minimize the potential error due to a small sample size, which leads to a better reflection of the current market conditions. Multidimensional HCCIs dissect the state highway construction market into distinctively smaller sectors of interest and thus,

[1] PhD Candidate; Dept. of Civil, Construction & Environmental Engineering, Iowa State University, Ames, IA 50011; email: shrestha@iastate.edu
[2] Associate Professor; Dept. of Civil, Construction & Environmental Engineering, Iowa State University, Ames, IA 50011; Phone: (515) 294-7271; email: djeong@iastate.edu
[3] Professor; Dept. of Civil, Construction & Environmental Engineering, Iowa State University, Ames, IA 50011; Phone: (515) 294-4148; dgran@iastate.edu

allow state departments of transportation to understand the market conditions with much higher granularity. A framework is developed to integrate these two concepts and a standalone prototype system, namely, Dyna-Mu-HCCI System is developed to automate the data processing part of the framework.

The historical bid data of the Montana Department of Transportation is used to evaluate the performance of the Dyna-Mu-HCCI System and measure the effects of the DIB and multidimensional HCCIs. The results show an eight-fold increase in terms of the number of bid items used in calculating HCCIs and at least 20% increase in terms of the total cost of bid items used. In addition, the multidimensional HCCIs reveal different cost change patterns from different highway sectors. For example, the bridge construction market historically shows a very different trend compared with the overall highway construction market.

The new methodology is expected to aid state departments of transportation in making more reliable decisions on preparing business plans and budgets with more accurate and detailed information about the construction market conditions. Further, the prototype, Dyna-Mu-HCCI System is expected to significantly facilitate the HCCI calculation process and rapidly implement this new system.

Keywords: highway-construction-cost-index (HCCI), inflation, dynamic-construction-item-basket, multidimensional-HCCI, construction-market-basket, construction-market-conditions, planning-and-budgeting, big-data, data-analytics, visualization, automation.

Introduction and Background

A Highway Construction Cost Index (HCCI) is an indicator of the purchasing power of a highway agency (Guerrero 2003; Strickland and Beasley 2007; White and Erickson 2011). It is calculated to shows highway construction cost changes over time as a function of unit costs and quantities of various bid items used in highway construction.

State departments of transportation (DOTs) use it to track changes in highway construction costs over time and reasonably estimate future highway funding needs (Erickson and White 2011; Guerrero 2003). An HCCI is also used by some DOTs as an inflation factor for preliminary and detailed cost estimates and life cycle cost analysis (LCCA) of their highway projects (Gransberg and Diekmann 2004; Iowa Department of Transportation (IADOT) 2012b; Mack 2012; Slone 2009; Wilmot 1999). HCCIs are also recommended as a factor to determine the gas tax rate to generate revenue necessary to properly maintain the existing highway infrastructure system (Arkansas Highway and Transportation Department (AHTD) 2013; Dodier 2014; Institute on Taxation and Economic Policy 2013). Thus, it is very important that HCCIs accurately reflect the actual construction market conditions.

The Federal Highway Administration (FHWA) pioneered the concept of HCCI in the U.S. highway construction industry in 1933 by introducing Bid Price Index (BPI) (White and Erickson 2011). Subsequently, some DOTs have adopted FHWA's methodology to develop their state level HCCIs (Luo 2009; Wilmot 1999). In 2011, FHWA introduced an updated National HCCI (NHCCI) as the replacement of the BPI (Erickson and White 2011). HCCI experts consider this change the most significant update in the national HCCI methodology. Among many notable changes such as a wider coverage of projects and electronic bid data collection processes, the switch to an enhanced indexing formula (Fisher index) is considered the major change. Currently, at least 21 DOTs compute their state level HCCIs, but most of them have not yet updated their methodologies to reflect the changes in the NHCCI methodology primarily due to lack of appropriate guidance (Shrestha et al. 2016; Walters and Yeh 2012).

In addition, current HCCI calculation methods adopted by most DOTs are not sophisticated enough to assure that an HCCI can be used as a reliable indicator of the changing market conditions. One of the reasons is the use of a significantly insufficient sample size of bid items in HCCI calculation. Since an HCCI is calculated using the cost information of bid items, ideally, the entire bid dataset should be used to truly reflect actual market conditions (International Monetary Fund (IMF) 2010). Currently, the coverage of bid items ranges from as little as 14% to not more than 50% of the total construction costs (Nebraska Department of Roads (NDOR) 2015; West Virginia Division of Highways (WVDOH) 2015; Wilmot 1999).

Another area for improvement in DOT's HCCI calculation methodology is in the current method's inability to address the need to track highway construction market conditions with higher granularity. Current methodologies typically produce only one overall HCCI as a representative index to indicate the entire state's highway construction market condition. However, highway construction costs are heavily affected by availability of local materials, equipment, and even specialty contractors. In addition, the project size and quantity of work significantly affect construction methods and their productivities which are directly associated with project costs. Moreover, many DOTs are forced to shift their highway project portfolio from new construction to maintenance and rehabilitation projects due to aging roadway systems. These unique characteristics of highway construction and changing business environments require DOTs to have customized HCCIs designed to better understand specific market conditions and trends based on local regions, project sizes and project types. The current system fails to address this issue.

The goal of this study aims at addressing the two specific issues described above by developing an advanced HCCI methodology with new concepts of dynamic item basket and multi-dimensional HCCIs. Specifically, this study will: a) develop a methodology to generate a *Dynamic Item Basket* (DIB) with a higher coverage of bid items, b) develop multidimensional HCCIs that can show construction market conditions with a higher granularity, c) automate the process to reduce efforts required to compute multi-dimensional HCCIs, and d) evaluate the performance of the new HCCI methodology.

Theory of Cost Index

The calculation of any type of cost index starts with the identification of product items that are relevant to and representative of the specific industry sector of interest. The collection of those items is called 'market basket' or '*item basket* (IB).' An IB with 'n' items has two important properties: a cost vector (p) = [p_1, p_2, p_3, ..., p_n] and a quantity vector (q) = [q_1, q_2, q_3, …, q_n] that represent the cost and quantity of each item in the IB. The subscript in each element of cost and quantity vectors represents a specific item. Theoretically, a cost index measures the movement of the cost vector from one period to another. Oftentimes, the quantity vector is used to indicate the importance of items in the IB. Generally, the cost movement in the *current period* (t) is measured relative to the *base period* (t=0). The cost index for the base period is typically set to 1.00 or 100. Thus, cost and quantity vectors from the current period (p^t, q^t) and base period (p^0, q^0) must be available to compute a cost index at a minimum.

In the highway construction industry, Laspeyres, Paasche and Fisher indexing methods are three most popular formulas among DOTs to compute HCCIs (Shrestha et al. 2016). Their formulas are presented in equations *(1), (2)*, and *(3)* respectively as functions of cost and quantity vectors from the base period to the current period.

$$\text{Laspeyres index, } L_{t,0}\ (p^0, p^t, q^0, q^t) = \frac{\sum_{i=1}^{n} p_i^t q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0} \tag{1}$$

$$\text{Paasche index, } P_{t,0}\ (p^0, p^t, q^0, q^t) = \frac{\sum_{i=1}^{n} p_i^t q_i^t}{\sum_{i=1}^{n} p_i^0 q_i^t} \tag{2}$$

$$\text{Fisher index, } F_{t,0}\ (p^0, p^t, q^0, q^t) = \sqrt{L_{t,0} X P_{t,0}} = \sqrt{\frac{\sum_{i=1}^{n} p_i^t q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0} X \frac{\sum_{i=1}^{n} p_i^t q_i^t}{\sum_{i=1}^{n} p_i^0 q_i^t}} \tag{3}$$

Laspeyres index is the ratio of the total expenditure in the current period to the total expenditure in the base period assuming that the same quantities of items are purchased in the current period as in the base period. Paasche, on the other hand, utilizes the quantity vector for the current period and assumes it to be the same for the base period. Because those two formulas consider the quantity vector from only one period, Laspeyres overestimates the impact of cost increases while Paasche underestimates it. Fisher index is calculated as a geometric average of the Laspeyres and Paasche indexes which can theoretically cancel out those two biases, (International Labour Organization (ILO) et al. 2004)

Over time, not only the quantities, but also the IB itself might be outdated because of changes in the market resulting in the addition, removal, and substitution of items. This results in a *sampling error*. Thus, the base year and IB are recommended to be updated periodically (i.e., every five or ten years). However, it is very possible that the IB and the quantity vectors might get outdated before the base year is changed. Thus, a chained cost index is recommended to overcome this error by calculating a cost index between two consecutive periods. In a chained cost indexing process, the *net cost index* between two periods [say current period (t) and some arbitrary base period (t=0)] is calculated by multiplying all consecutive cost indexes ($I_{k,k-1}$) between the two periods (equation (4)).

$$\text{Chained index}, CI_{t,0} = \prod_{k=1}^{t} I_{k,k-1} \tag{4}$$

Thus, the chained Fisher index formula is considered the most ideal method for calculating a cost index. This formula is used by FHWA for its NHCCI computation and is recommended for DOTs' HCCI calculation (Erickson and White 2011).

Current Practices in HCCI Calculation

Despite the clear advantages of the chained Fisher index, only Colorado, Ohio, and South Dakota DOTs currently use the Fisher index and Wisconsin and North Dakota DOTs are updating their methodologies to use the chained Fisher index (Shrestha et al.2016).

Also, state level HCCIs are calculated using IBs with its cost coverage as low as 14% and as low as 7% in terms of its bid item coverage (Table 1). The highest IB coverage in terms of total costs is 60% for FHWA's NHCCI. The coverage of 271 bid items in Utah DOT may appear to be large, but considering that DOTs typically use more than 2,000 bid items, it is quite small. There are several possible reasons for using IBs with such small coverages.

Table 1 Item basket coverage comparison

DOT	**Item Basket (IB) coverage**	
	Number of bid items	**% of total construction costs**
West Virginia	7	14%
Wisconsin	91	-
Colorado	-	45%
Nebraska	101	46%
Ohio	-	48%
Mississippi	116	-
Iowa	190	-
Utah	271	--
FHWA	-	60%

First, lump sum items are typically removed from HCCI calculation, because these items are mostly unit-less and their costs do not have consistent relationships with their quantities, if there were quantities assigned. Removal of lump sum items such as mobilization is likely to reduce the IB coverage in terms of costs substantially due to the significant percentage of lump sum items in total project costs.

Second, DOTs generally remove data from smaller projects and item data with smaller quantities. For example, Minnesota, California, and Wisconsin DOTs remove data from projects smaller than $100,000 in value (Hanna et al. 2011; Lacho 2015; Minnesota

Department of Transportation (MnDOT) 2009). Similarly, Iowa DOT removes concrete items with quantities less than 125 cubic yards and Colorado DOT removes excavation items less than 1,000 cubic yards (Colorado Department of Transportation (CDOT) 2015; Iowa Department of Transportation (IADOT) 2013). They also utilize various outlier detection techniques to remove items whose unit costs appear to be different than most of the unit costs. However, removal of such data may create a *sampling error*, i.e. the HCCI becomes more representative of a specific segment of the market rather than the entire market (Hanna et al. 2011; Lacho 2015; Minnesota Department of Transportation (MnDOT) 2009).

Third, DOTs choose a few important bid items from various work categories such as asphalt, concrete, and earthwork with a rationale that those selected items can represent all items in the category (Hanna et al. 2011). In this process, most DOTs consider items with high unit costs and/or high frequency as the important items with reasonable rationale that non-frequent items should be excluded mathematically in HCCI calculation and higher cost items may have higher impact on project costs (Shrestha et al. 2016). Such sampling process is common in the general inflation calculation such as consumer price index as it requires a significant amount of effort to use a larger IB, and it is practically impossible to use an IB of the entire product items in general inflation calculation (Bureau of Labor Statistics (BLS) n.d.; International Monetary Fund (IMF) 2010). However, for HCCI calculation, the entire bid dataset is readily available in an electronic format which provides an opportunity to potentially eliminate any sampling error. Next section presents the concept of Dynamic IB (DIB) to address this issue by improving the coverage of IB. Then, the concept of multidimensional HCCI is also presented.

Concept of Dynamic Item basket

An IB should contain all items used in the market if the costs and quantities of the items are available for both base and current periods. If that is not possible, an IB should still be a good representor of actual items used in the market to ensure that the cost index is a good reflector of the current market conditions (Bureau of Labor Statistics (BLS) 2015; International Monetary Fund (IMF) 2010). Since highway project bid data are now available in a digital format in DOT's contracts office, it is practically possible to use the entire population of bid items for HCCI calculation.

In dynamic IB (DIB), the items in the IB, and corresponding cost and quantity vectors are updated automatically based on the current purchasing behavior of DOTs. The DIB generation process identifies the largest IB that can be generated from the bid data and hence increases the coverage of the IB to the maximum possible value. To explain the DIB generation process, consider a universal set U consisting of all standard bid items used by DOTs (Figure 2).

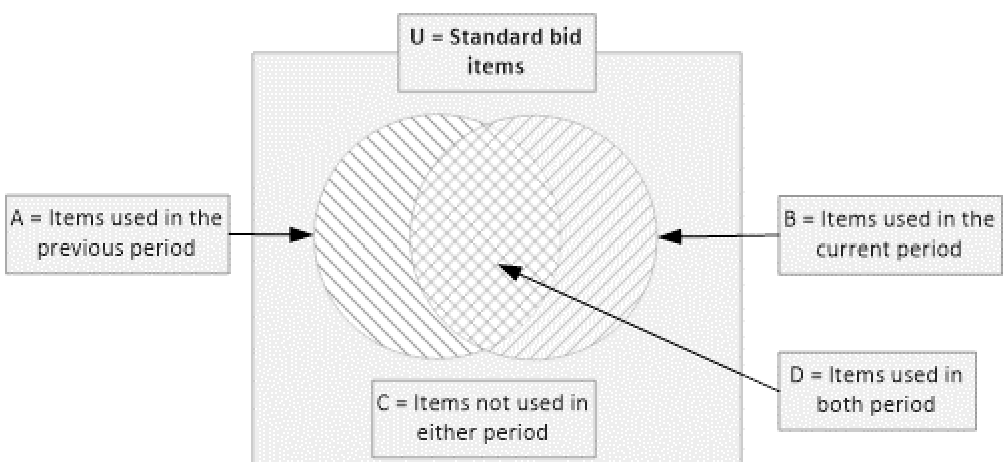

Figure 2 Dynamic item basket

Some of those items will be used in the current period (B), some in the previous period (A), and others will not be used in either period (C). The items that are not used in either period or the items used for only one of the two periods cannot be mathematically included in HCCI calculation. But, all items that were used in both periods (D) can be used in HCCI calculation and DIB consists of these items (D). Using this DIB with those items instead of a small-sampled IBs that are currently used by most DOTs, can significantly improve the HCCI calculation process with higher accuracy and reliability by removing the sampling error.

Concept of Multidimensional HCCIs

The concept of multidimensional HCCIs is to develop cost indexes for highway construction market sectors defined by project size, project type, and location. Thus, in addition to an overall HCCI that is used to indicate the state level market conditions, three-dimensional sub-HCCIs are developed: project size specific HCCIs (S-HCCI), project type specific HCCIs (T-HCCI), and location specific HCCIs (L-HCCI) which are visually depicted as HCCI cubes in Figure 3.

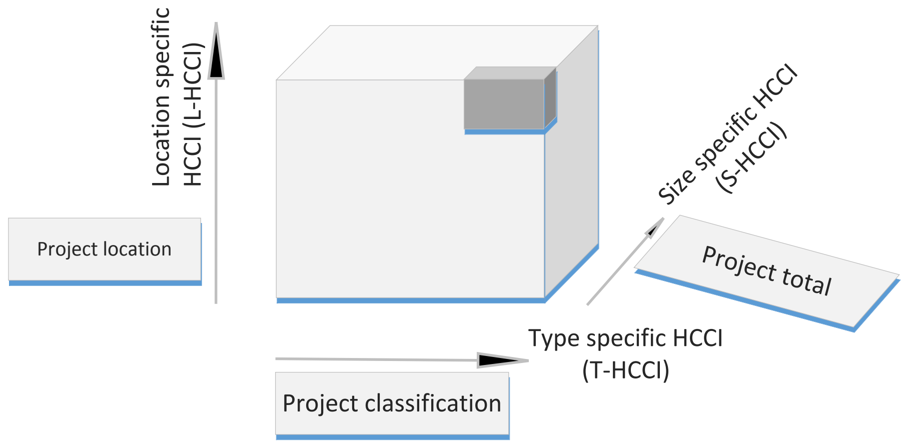

Figure 3 HCCI cubes

The size specific sub-HCCIs (S-HCCIs) are necessary because of the effect on costs by the economies of scale. The cost of an item is less when purchased in bulk. As such, larger projects that would contain larger quantities of items are likely to have a different market trend than that of smaller projects. Further, the level of competition for projects of different sizes also varies because contractors often need to be prequalified to perform larger projects. Similarly, contractors are often specialized to perform a certain type of projects. In addition, work items for different types of projects also vary. Those reasons necessitate a project type specific HCCI (T-HCCI) (Erickson and White 2011; Rueda and Gransberg 2015). One may argue that DOTs already calculate item category specific HCCIs (I-HCCIs) for different work categories such as structures, pavements, etc. However, a typical highway project consists of various work items from different item categories. Thus, T-HCCIs are different from I-HCCIs.

Existing literature also recognizes the importance of developing location specific sub-HCCIs (L-HCCIs) (Anderson et al. 2007; Erickson and White 2011; Ghosh and Lynn 2014; Gransberg and Diekmann 2004; Shahandashti 2014). The rationale behind L-HCCI can be explained with the Tobler (1970)'s First Law of Geography which states that "everything is related to everything else, but near things are more related than distant things." Specifically, in highway construction, the availability of resources and their hauling distances to the jobsite such as qualified materials, equipment, and labor significantly affect the total construction cost and hence the market trend. Also, the market trend is likely to vary differently in mountainous areas and plain areas.

Framework for Multidimensional HCCI with DIB

The framework to integrate DIB into multidimensional HCCI calculation process is illustrated in Figure 4. The framework can be divided into four components: a) database development, b) project filtering, c) DIB generation, and d) multidimensional HCCI calculation. In the first component, data required for calculating multidimensional HCCIs with DIB are collected and systematically compiled in a structured database. *Project filtering* is a process to filter project data in three stages to obtain a list of projects relevant to a particular sub-HCCI. In *DIB generation*, two sets of cost and quantity vectors from previously selected projects are extracted. Finally, the Chained Fisher index formula is applied in the final component to generate sub-HCCIs. The project-filtering component and the following components are repeated to generate various sub-HCCIs (such as small, medium, and large sized S-HCCIs).

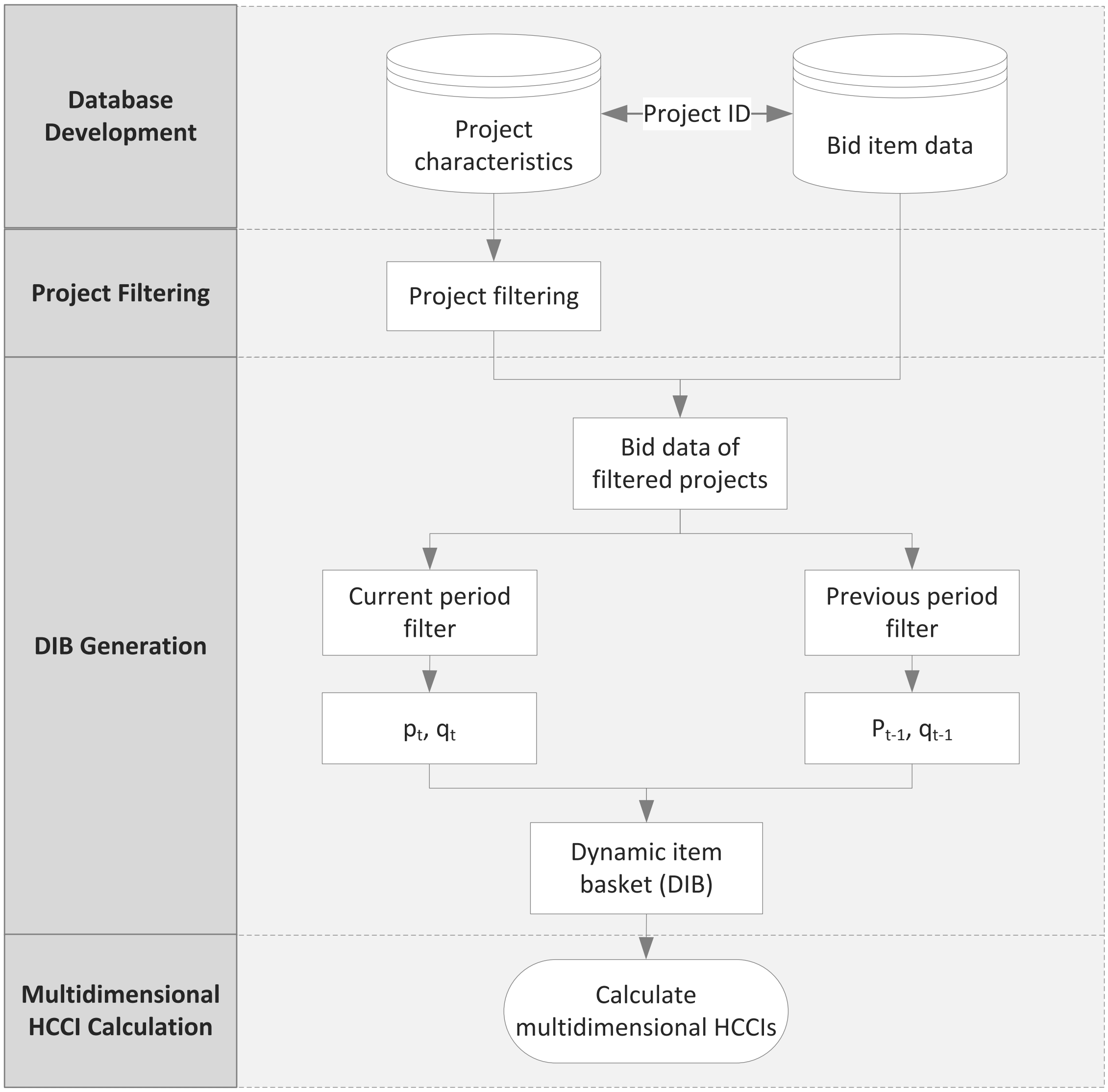

Figure 4 Framework for advanced multidimensional HCCI calculation using DIB

Database development

In this component, project characteristics and bid item data that are necessary for HCCI calculation are obtained from electronic bid letting systems and compiled into a new database for further processing. Currently, 41 DOTs use AASHTOWare Project Expedite

System that stores data in a structured database such as Oracle and Microsoft SQL (Structured Query Language) Server (American Association of State Highway and Transportation Officials (AASHTO) 2015, 2016). SQL queries can be executed in the database used by such systems to generate relevant data. Alternately, those databases can be used directly as the database for this framework. At minimum, the database should contain *project characteristics* and *bid item data*. Project characteristics should include project size, type, and location. Bid item data should include information such as the item number, quantity, and cost for each bid item. These two datasets need to be tied together by a unique project ID as shown in Figure 4 so that relevant bid items from a list of projects of our interest can be obtained by automated filtering process.

Project filtering

In this component, projects relevant to calculating sub-HCCIs are selected in three phases: a) removal of non-design-bid-build projects, b) selection of projects from the current and previous or base periods, and c) selection of projects of a particular category corresponding to the selected sub-HCCI. Figure 5 shows the detailed procedure for project filtering. The third phase (c) is required only to generate sub-HCCIs and is skipped for an overall HCCI calculation. For an overall HCCI calculation, data from all project sizes, types, and locations are used.

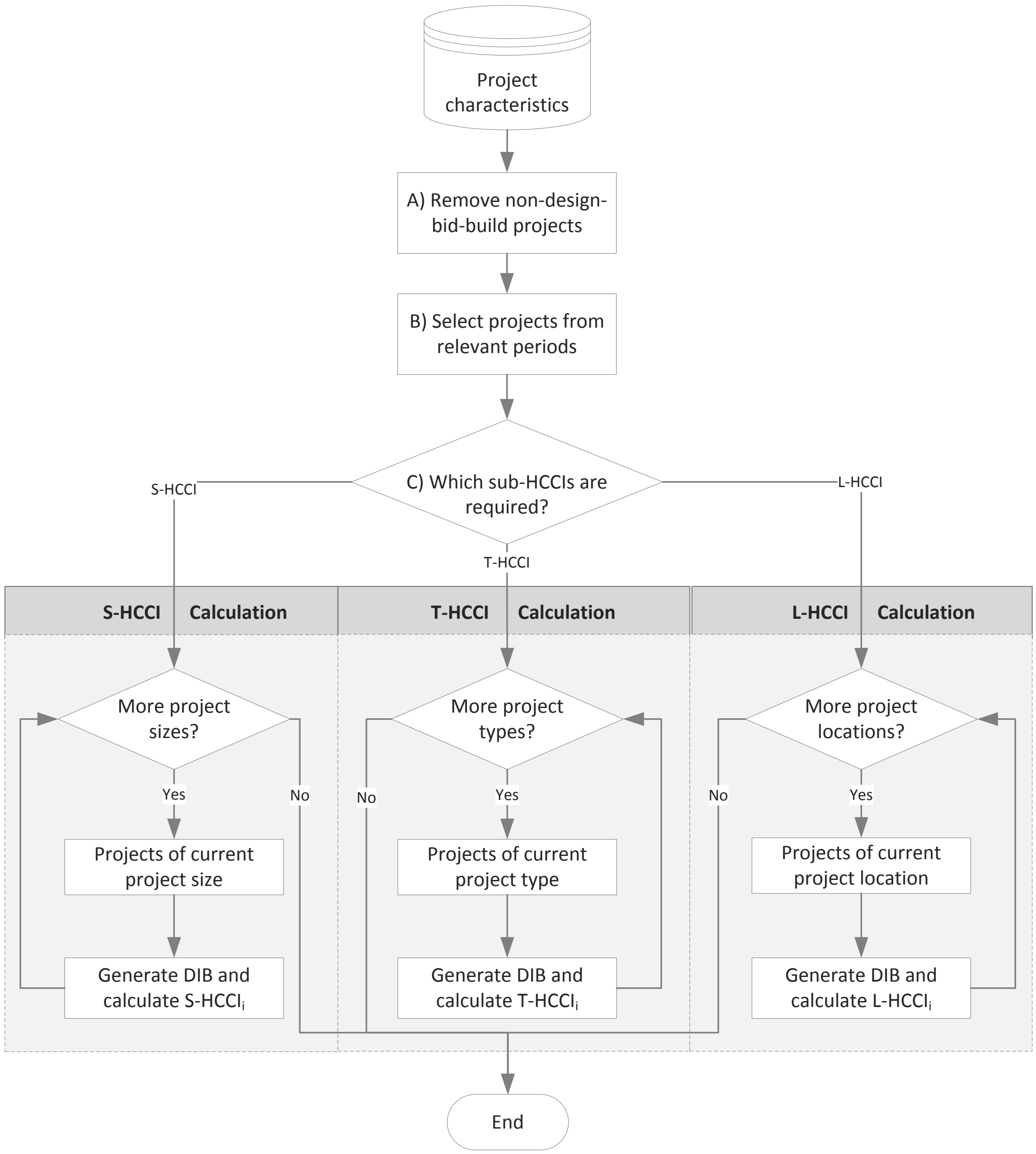

Figure 5 Project filtering component

In the first phase, projects that are procured through nonstandard design-bid-build procurement method are removed. For example, in ‘indefinite delivery infinite quantity’ contracts, a predetermined inflation rate is used (Rueda and Gransberg 2014) and in 'design-build' contracts, non-standard bid items are used. Thus, those projects need to be eliminated. In the second phase, projects let in the current period or previous period are selected. Finally, projects relevant to the specific sub-HCCI are shortlisted using one of the three subcomponents (S-HCCI Calculation, T-HCCI Calculation, and L-HCCI Calculation) shown in Figure 5 for generating DIBs for the sub-HCCI in the next phase. Further, each sub-HCCI consists of multiple sub-HCCI values (i.e., S-HCCIs for small sized projects, medium sized projects, and large sized projects). The list of projects for each of the sub-HCCI value calculation is filtered separately and each list is sent to the DIB generation component one at a time.

DIB generation

In this component, a DIB and corresponding cost and quantity vectors required to calculate sub-HCCIs are generated in three phases: a) extraction of relevant bid data, b) splitting the data into current and previous period data, c) generation of initial cost and quantity vectors, and d) removal of irrelevant items to generate the final cost and quantity vectors.

First, all bid data corresponding to the projects selected from the project filtering component is *extracted*. This can be achieved by SQL (Structured Query Language) command *Inner Join* (LeCorps 2001). The inner join can be considered as SQL equivalent of *intersection* in the set theory (Jech 1978). In this case, project ID is used for the intersection operation (equation (5)).

$$\text{Bid data of filtered projects} = (\text{Bid data and } \textit{Project IDs} \text{ of all projects}) \cap (\textit{Projects IDs} \text{ of selected projects}) \tag{5}$$

The resulting dataset is *split* into two groups: one for the current period and another for the previous period. As items for all projects are based on a finite list of standard bid items used by DOTs, same bid items appear in various construction projects. However, for HCCI calculation, data from each unique item needs to be converted into a single line of data to *generate initial cost and quantity vectors*. For that, quantities are generated as a sum of quantities of the same items from all the projects (equation (6)) while costs are generated as weighted averages of the costs (equation (7)).

$$\text{Total quantity of an item } (q_i) = \sum Quantity\ of\ the\ item \tag{6}$$

$$\text{Weighted average cost of an item } (p_i) = \frac{\sum(\text{Cost of the item X Quantity of the item})}{\sum Quantity\ of\ the\ item} \tag{7}$$

So far, the item lists (ILs) and corresponding cost and quantity vectors are obtained for both periods. These ILs are further processed to develop DIB using equation (8). The cost and quantity vectors corresponding to this DIB is the final vectors required for the next component. First, an item should coexist in both periods to use it for HCCI calculation. Thus, an *intersection* operation is performed between the two ILs.

$$DIB_t = IL_{t-1} \bigcap IL_t - IL_{irrelevant} \tag{8}$$

Then, this dataset obtained from the intersection operation is cleaned by *removing* all items that are not relevant to measuring the market conditions ($IL_{irrelevant}$). These items include lump sum items and items whose costs do not have a consistent relationship with their quantities. For example, costs for mobilization and utility relocation may vary widely despite its constant quantity (one unit). Some DOTs also remove seemingly outlier items based on cost fluctuation (Collins and Pritchard 2013; Federal Highway Administration (FHWA) 2014; Nassereddine et al. 2016). However, HCCIs are meant to measure the cost fluctuations and hence the removal of items with high cost fluctuations may not be the best approach. Thus, in this framework, those items are also included.

The items obtained using this process described above is the largest IB that can be generated from any given bid and project datasets. Further, the process updates IB dynamically based on the project characteristics and bid item data, current period selection, and sub-HCCI that is calculated. Thus, this IB can also be called an *optimum IB*. Unlike traditional methods where smaller and/or less frequent items are ignored and only larger and more frequent items are used, this method utilizes all items if they are purchased in both the current and previous period. This DIB and corresponding final cost and quantity vectors are transferred to the next component for multidimensional HCCI calculation.

Multidimensional HCCI calculation

In the final component, single staged chained Fisher index (equation (3)) based sub-HCCIs are calculated using the cost and quantity vectors generated from the previous component. In equation (3), instead of base period (t=0) cost and quantity vectors, previous period cost and quantity vectors (t-1) are used. Different chaining intervals can be used depending on the DOT's needs. In quarterly chained HCCIs, the chaining error can occur if both cost and quantity vectors of the IB oscillate over time (Nygaard 2010). In case of annual HCCIs, such oscillation is less likely to occur which reduces the chaining error. Finally, the sub-HCCI can be chained using equation (4). A base year can be selected arbitrarily, for which the cost index is set to 1.00 or 100. Generally, the base year is selected when the market is in a normal economic condition (e.g. not affected by heavy recession, etc.).

Prototype Development

A prototype, namely, Dynamic Multidimensional HCCI Calculation System (Dyna-Mu-HCCI-System) is developed with MS Access database (Figure 6) and Visual C#.NET frontend (Figure 7) to implement the framework. Seven data tables are created using Entity-Relation Model (ERM) to optimize the database (Stephens 2010). The 'm_project_characteristics' and 'm_bidtabs_winning' contain the required project characteristics and bid item data. The 'm_bid_item_specs' and 'm_item_type' contain additional information about the standard bid items.

The Graphical User Interface (GUI) has the menu items on top to calculate various sub-HCCIs and perform some additional bid data analysis. The prototype is capable of generating sub-HCCIs using the raw bid data in a single click. Users can select a year as the current year to calculate sub-HCCIs for that particular year. Figure 7 shows the item basket generated for T-HCCI on the left and six T-HCCI values on the right. Next section discusses the analysis of the results regarding the performance of this new methodology generated using this prototype.

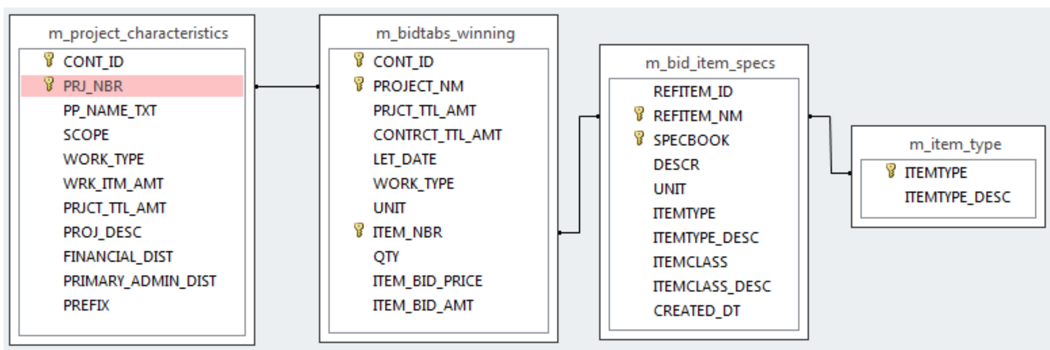

Figure 6 MS Access database of Dyna-Mu-HCCI-System

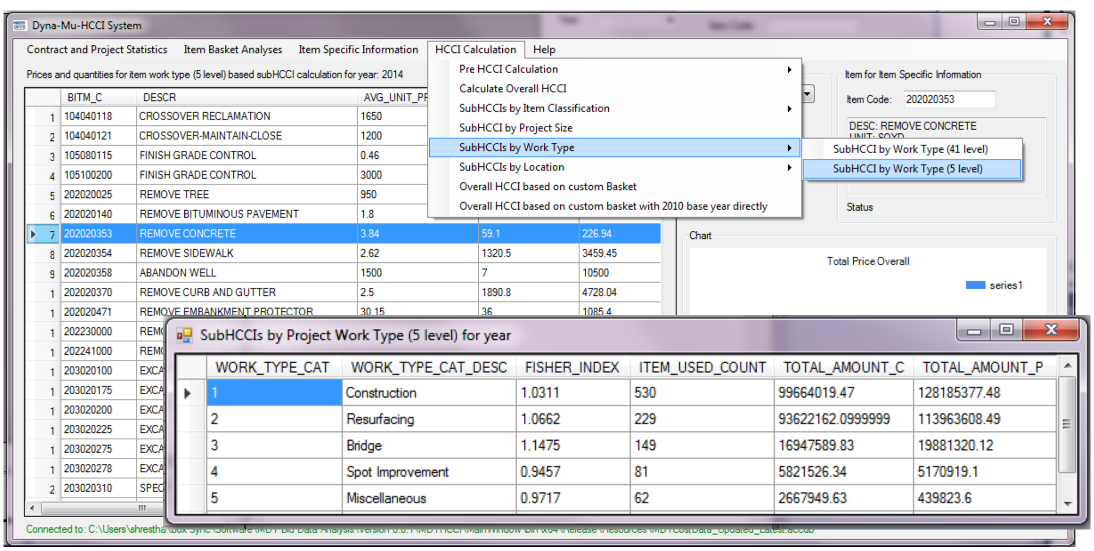

Figure 7 Visual C#.NET frontend of Dyna-Mu-HCCI-System

Performance Evaluation of Dyna-Mu-HCCI-System

Historical bid data from Montana Department of Transportation (MDT) are collected and analyzed to evaluate improvements in the IB coverage using the DIB. It further discusses the results on the fluctuation of specific segments of the highway construction market using the multidimensional HCCI approach by comparing the sub-HCCIs with the overall HCCIs.

Data collection

The researchers obtained the historical bid data from MDT in an excel format which was imported into the database. The database consists of bid data of 687 projects let from 2010 to 2014 that represent more than $1.8 billion of construction projects. The dataset consists of 33,975 lines of items based on 2,529 standard bid items from MDT's specification. MDT has developed a list of 5,645 unique bid items in its 2006 specification manual (Montana Department of Transportation (MDT) 2006). Each bid item represents a unique work item. For example, bid item "402020091" represents ASPHALT CEMENT PG 64-22. All bid items that begin with 402 represent bituminous materials and include the cost of "furnishing and applying bituminous materials, on bases and surfacing." The obtained bid data was imported into the Dyna-Mu-HCCI-System.

Improvements in IB using DIB

To evaluate the effect of the DIB, overall HCCIs are calculated using DIB ($HCCI_{DIB}$) and the current IB used by MDT ($HCCI_{current\ IB}$). MDT's current item basket includes 71 high cost items handpicked by MDT. In the DIB, items are selected automatically using the framework developed in this study. The number of items in the DIB ranges from 610 to 735 items in various years (2010 – 2014). This indicates that DIB consists of items more than eight times the number of items in the original IB. In terms of the cost coverage, the current MDT's item basket represents less than 50% of the total project costs. The DIB improves the cost coverage to over 70% of the total project costs indicating at least 20% increase in the coverage. The overall HCCI values calculated from year 2011 to 2014 are presented in Table 2. Year 2010 is assigned as the base year with the base cost index of 100. The difference in terms of percentage ranges from 2.34% up to 5.98%.

Table 2 Comparison of overall HCCI calculated using DIB and current IB

Current year	**$HCCI_{DIB}$**	**$HCCI_{current\ IB}$**	**% difference**
2010	100.00	100.00	0%
2011	110.46	114.37	-3.54%
2012	111.12	117.77	-5.98%
2013	113.06	115.70	-2.34%
2014	115.46	119.92	-3.86%

A correlation coefficient is calculated to compare the trend of the two series. The correlation coefficient (r) is a statistical factor used to access the linear relationship between two variables (say x and y) (Taylor 1990). Mathematically, the correlation coefficient can be calculated as:

$$Correlation\ coefficient\ (r) = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}. \tag{9}$$

The value of r can vary from -1 to +1. A positive value indicates that both variables have similar trends, i.e. increase in one variable is associated with the increase in another variable. The higher the value is, the stronger the correlation is. Negative values indicate that an increase in one variable is associated with a decrease in the other. The r-value calculated for these two HCCIs series is 0.98, which indicates a very similar trend between the two series.

Further, an overall error between the two HCCI series is calculated using Mean Absolute Percentage Error (MAPE) (equation (10*)*). The higher MAPE indicates more variation between the two series. Generally, one may expect to have a higher MAPE value associated with a lower r-value and vice-versa.

$$MAPE = \sum_{i=2011}^{2014} \frac{\frac{|HCCI_{DIB,i} - HCCI_{current\ IB,i}|}{HCCI_{DIB,i}} * 100}{n} \tag{10}$$

The results show a MAPE value of 3.93%. While 3.93% may seem to be a small error, this is a large error considering that an average inflation itself is recommended as 4% by the FHWA (Mack 2012). In addition, the absolute percentage difference between the two series is as high as 5.98% in 2012. This implies that the use of the current IB may result in an erroneous decision-making on highway construction market evaluation, preliminary transportation budgeting and planning, etc.

Fluctuations of multidimensional HCCIs

MDT uses several project characteristics to classify their highway projects (Table 3). It uses a six-level project type classification system, which is further sub-divided into 41 types. MDT also divides the state into five administrative and construction districts and five financial districts. These two types of districts overlap closely. MDT also uses three different bid item classification systems: division, class, and type. However, no project size classification is found in the current MDT business practices. For this study, MDT projects are classified using a clustering algorithm known as Simple Expectation Maximization that resulted into three clusters. Based on the clusters, project sizes are divided into three ranges representing small (0 - $3,500,000), medium ($3,500,000 - $10,500,000), large ($10,500,000 - $50,000,000).

With those classification systems, 107 series of chained sub-HCCIs can be calculated. For chained sub-HCCIs, their continuity over time is very important to utilize them. Sixty-eight sub-HCCIs have continuous values from 2010 to 2014. Continuous values for other sub-HCCIs are not available because of the lack of items in the DIB. Such scenarios can occur when projects of a particular category are not let frequently. For example, a type of project - ‘facilities’, is not very frequent in MDT and hence very limited data points are available. In addition, some item categories such as ‘unknown’ are used for lump sum items. Thus, it is not possible to calculate sub-HCCIs for such categories as the Dyna-Mu-HCCI-System removes all lump sum items. In addition, as the number of classification levels in a given category increases, the possibility of generating a non-empty DIB for that specific classification level decreases causing a discontinuity in sub-HCCIs. The extended project type T-HCCIs (41 levels) and item class IHCCI (31 level) have many non-continuous sub-HCCIs and are hence not included for further analysis.

Table 3 Sub-HCCI calculation parameters and number of sub-HCCIs

Sub-HCCI type			Number of sub-HCCIs	Sub-HCCIs	Number of continuous sub-HCCIs
Project characteristics based	Project Type	Project Type	6	Construction; Resurfacing; Bridge; Spot Improvement; Miscellaneous; Facilities	5
		Extended Project Type	41	New Construction; Reconstruction – with added capacity; Reconstruction – without added capacity; Resurfacing – Crack Sealing; New Bridge; Bridge Replacement with added capacity; etc.	13
	Project Location	Administrative and Construction District	5	Glendive; Billings; Great Falls; Missoula; Butte	5
		Financial District	5	Glendive; Billings; Great Falls; Missoula; Butte	5
	Project Size		3	Small (0 - $3,500,000) Medium ($3,500,000 - $10,500,000) Large ($10,500,000 - $50,000,000)	3
Item characteristics based		Item Division	6	General Provisions; Earthwork; Aggregate Surfacing and Base Courses; Bituminous Pavements; Rigid Pavement and Structures; Miscellaneous Construction	6
		Item Class	31	Liquid Asphalt; Base Course; Concrete Paving; Crushing; Drainage; Earthwork; Removals; Signing; Structures; Surface Treatment, etc.	24
		Item Type	10	Grading/ Drainage; Paving; Structures/ Buildings; Materials; Equipment; Traffic Control; Landscaping; Other, misc.; Trucking; Unknown	7

The values of overall HCCIs and all continuous sub-HCCIs are presented in Table 4. Correlation coefficients and MAPE values are calculated for the two series to quantify the similarities and differences between them. Most of the bituminous pavement and paving sub-HCCIs have a very high correlation (r = 0.94 and 0.96) with the overall HCCI. However, T-HCCI for bridges has r-value of -0.04 indicating slightly negative correlation. It might be because a large portion of bridge costs are associated with concrete and steel but the majority of construction projects are asphalt intensive roadway projects. Concrete and steel costs do not necessarily follow the cost movement of asphalt items. This weak relationship is also visible in structures/buildings HCCI (r = 0.10) and rigid pavement & structures HCCI (r=0.02). From L-HCCI perspective, Glendive district has the strongest correlation (r = 0.99 for both financial district and administrative & construction district) while others have lesser correlation but still strong correlation. In terms of project sizes, the overall HCCI was a better representative of small and large sized projects rather than medium sized projects.

MAPE confirms correlation analysis results and provides additional insights. For instance, in most cases such as T-HCCI for resurfacing projects and S-HCCI for large projects, MAPEs are less than 5%, which is in accordance with the strong correlations observed with higher r-values. The MAPE and r-value for the T-HCCI for spot improvement might seem contradictory at first sight. The T-HCCI has the highest MAPE value (68%) as well as a high r-value (0.94). This indicates that spot improvement projects do have a similar trend to an overall HCCI, but their rates of change (i.e. inflation rates) are very different. Specifically, while the overall HCCI increased from 100 in 2010 to only 115.46 in 2014, the spot improvement project T-HCCI increased to 207.12 during the same period.

Finally, project characteristics based sub-HCCIs provide more granular insights than the item based sub-HCCIs. For example, while paving HCCI has a strong correlation (r-value = 0.96) and small error (MAPE = 2%), construction and resurfacing T-HCCIs shows relatively weaker correlations (r-values = 0.91 and 0.89 respectively) and higher errors (MAPE = 3% each). Further, construction and resurfacing projects have varying sub-HCCIs:

while construction T-HCCI grew from 100 in 2010 to 112.64 in 2014, resurfacing T-HCCI grew only to 116.83 during the same period indicating 3.52% MAPE value between the two types of paving projects.

Overall, T-HCCIs have the highest deviations from the overall HCCIs while S-HCCIs have the lowest. However, S-HCCI might have varying deviations based on the different range of size categories developed.

Table 4 Overall HCCIs, sub-HCCIs and their correlation coefficients (r)

Sub-HCCI Type		**sub-HCCI**	**2010**	**2011**	**2012**	**2013**	**2014**	**r**	**MAPE**
Overall		Overall HCCI	100.00	110.46	111.12	113.06	115.46	-	-
Project characteristics based	Project Size	Small (0 - $3,500,000)	100.00	106.76	109.01	107.73	109.15	0.96	4%
		Medium ($3,500,000-$10,500,000)	100.00	107.73	115.50	117.29	112.81	0.86	3%
		Large ($10,500,000-$50,000,000)	100.00	114.15	113.50	116.87	115.55	0.97	2%
	Project Type	Construction	100.00	112.03	106.90	109.24	112.64	0.91	3%
		Resurfacing	100.00	106.83	114.12	109.57	116.83	0.89	3%
		Bridge	100.00	104.89	89.94	91.50	105.00	**-0.04**	13%
		Spot Improvement	100.00	169.33	162.56	219.01	207.12	**0.94**	**68%**
		Miscellaneous	100.00	91.68	42.72	72.10	70.06	-0.60	39%
	Financial District	**Glendive**	100.00	114.55	113.11	115.41	121.58	**0.99**	3%
		Billings	100.00	106.73	104.62	105.42	114.30	0.83	4%
		Great Falls	100.00	107.25	101.06	114.44	119.12	0.77	4%
		Missoula	100.00	118.07	125.21	123.67	113.64	0.76	8%
		Butte	100.00	102.94	117.79	110.81	128.74	0.76	7%
	Primary Administrative and Construction District	**Glendive**	100.00	114.98	113.11	116.75	119.19	**0.99**	3%
		Billings	100.00	106.85	104.20	106.67	112.95	0.88	4%
		Great Falls	100.00	107.61	103.11	122.66	121.28	0.77	6%
		Missoula	100.00	109.97	127.49	125.46	118.39	0.78	7%
		Butte	100.00	101.98	118.29	119.62	130.99	0.81	8%

Table 4 continued

Sub-HCCI Type		sub-HCCI	2010	2011	2012	2013	2014	r	MAPE
Item characteristics based	Item Division	General Provisions	100.00	154.91	95.02	144.04	131.14	0.51	24%
		Earthwork	100.00	124.64	106.80	115.40	116.02	0.68	5%
		Aggregate Surfacing and Base Courses	100.00	107.50	103.08	116.36	107.96	0.68	5%
		Bituminous Pavements	100.00	109.83	116.76	118.91	117.80	**0.94**	3%
		Rigid Pavement and Structures	100.00	109.51	110.47	90.39	103.06	**0.02**	8%
		Miscellaneous Construction	100.00	104.12	104.04	104.36	113.45	0.79	5%
	Item Type	Grading/ Drainage	100.00	117.93	100.23	108.21	111.98	0.54	6%
		Paving	100.00	109.69	113.96	116.59	115.62	**0.96**	2%
		Structures/ buildings	100.00	106.46	112.92	93.69	103.52	**0.10**	8%
		Materials	100.00	107.44	107.79	110.42	111.11	0.99	3%
		Traffic Control	100.00	117.94	122.83	121.00	119.56	0.92	7%
		Landscaping	100.00	93.25	91.93	106.51	124.36	0.45	12%
		Other, misc.	100.00	99.59	105.05	102.75	121.73	0.61	7%

Conclusions

This study identifies a gap in the knowledge on the current HCCI calculation methodology in DOTs and develops an advanced methodology to fill the gap. It develops a concept of Dynamic Item Basket (DIB) to improve the coverage of Item Basket (IB) used to calculate HCCIs. A concept of multidimensional HCCIs is also developed to enable more granular overview of the market conditions. A prototype system is developed to automate the framework. The automated system will facilitate the use of advanced concepts and reduce the time and effort required to compute HCCIs. The results of this study can serve as a guide to DOTs that desire to update their current methodology.

The study used bid data from Montana Department of Transportation (MDT) to validate the new methodology. The new DIB methodology improves the coverage of the bid items dramatically more than 8 times higher in terms of the number of bid items used and at least 20% higher in terms of the total project costs covered. Multidimensional HCCIs revealed high fluctuations in specific construction markets such as bridges compared to the overall market conditions. These granular and more accurate HCCIs are expected to aid DOTs to assess their market condition accurately and develop more customized business plans for different project types and sizes in different locations.

Acknowledgements

The researchers would like to thank Montana Department of Transportation (MDT) for funding this study. The study would not have been possible without help from MDT staff who provided valuable datasets and guidance on the current MDT practices. We would also like to thank DOT representatives who provided insights about their current practices of calculating HCCIs.

CHAPTER 3

ENHANCED AUTOMATION FRAMEWORK FOR COLLECTION AND UTILIZATION OF DAILY WORK REPORT DATA

K. Joseph Shrestha[4], H. David Jeong[5], and Doug D. Gransberg[6]

Abstract

A significant amount of time and effort is invested to collect and analyze DWR data. But, their current uses have been very limited mostly to contractor payment, progress monitoring, and dispute resolution. This study conducts literature review and two nationwide questionnaire surveys to identify current challenges of collecting and utilizing DWR data and develops a new framework to improve the scenario. The challenges identified in this study include the lack of automation for DWR data analysis, data quality issues, and duplication of efforts. A survey result shows that many benefits of DWR data such as production rate estimation, activity cost estimation, contractor evaluation, contract time determination, etc. are obtained by half or less of the respondents. The limited use of DWR data is statistically associated with the limited level of automation for various benefits that can be obtained by analyzing DWR data. An enhanced automation framework is developed to improve the scenario. It consists of three components a) data model, b) automation of data DWR data

[4] PhD Candidate; Dept. of Civil, Construction & Environmental Engineering, Iowa State University, Ames, IA 50011; email: shrestha@iastate.edu

[5] Associate Professor; Dept. of Civil, Construction & Environmental Engineering, Iowa State University, Ames, IA 50011; Phone: (515) 294-7271; email: djeong@iastate.edu

[6] Professor; Dept. of Civil, Construction & Environmental Engineering, Iowa State University, Ames, IA 50011; Phone: (515) 294-4148; dgran@iastate.edu

analysis and reporting, and c) technical aspects. The methods to automate select analysis are presented in mathematical form and in the form of Structured Query Language (SQL) queries. The framework is validated by DWR experts of the U.S. and a case study. The framework can be used to develop a new DWR system or to improve existing systems. It is expected to improve the utilization of DWR data for improved construction decision makings.

Key Words: daily-work-report, field-data, construction-data, data-driven-decision-making, big-data, data-analytics, visualization, automation.

Introduction and Background

Despite being one of the largest industries in the U.S., the use of the digital technologies in the construction industry has been limited. The speed of uptake of newer technologies in the construction industry has been much slower compared to the speed in other sectors such as healthcare, retail, automotive, and utilities (Holler et al. 2014). A significant growth in the volume, velocity, and variety of data being collected has been observed in these sectors (Lancy 2001). The collection of a large amount of data has become easier and cheaper because of the exponential growth in the processing and storage capacities. As per the Moore's law, the capacities of the electronic circuits have been doubling every year (Moore et al. 1999). Meanwhile, the amount of digital information has been increasing by 10 times every five years (The Economist 2010). Although slowly, the construction industry has also introduced various technologies to collect data which can be used for making data-drive decisions.

Construction projects are associated with the collection, processing, and exchange of large amount of data among project stakeholders (Cox et al. 2002). The Daily Work Report (DWR) data collected by inspectors and Resident Construction Engineers (RCEs) at the construction sites are the most important data collected by them (Alabama Department of Transportation (ALDOT) 2013). Inspectors and RCEs spend as much as 50% of their time in collecting the DWR data (McCullouch and Gunn 1993). This DWR data generally consist of construction activities, labor hours, equipment hours, material stockpiles, weather data, and significant communications with contractors. Traditionally, DWR data was collected and stored in paper-based systems (Cox et al. 2002). The paper based systems have their own challenges. For example, a four year project may have over 1,000 DWR forms which make it challenging to utilize the collected data for any decision making such as claims and dispute resolution (ASCE Task Committee on Application of Small Computers in Construction of the Construction Division 1985). The paper based DWR systems are also inefficient and time consuming (Dowd 2011).

Even if those data are initially collected in the paper based systems, a considerable amount of time can be saved while analyzing and utilizing the data, if those data are transferred and stored in a digital system (Cox et al. 2002). Many digital DWR systems are developed since 1990s which include the state-specific DWR systems developed by Vermont, Utah, Michigan, Kansas, and the AASHTO developed AASHTOWare SiteManager (AASHTO 1999; ExeVision 2012; KDOT 1999; MDOT 2005; Rogers 2013).

Those digital DWR and contract management systems have enabled state DOTs to save millions of dollars. For example, MDOT reported savings of $22 million by automating the previously paper-based, error prone, slow, and intensively manual process of DWR data collection, material tracking, and contract payment (MDOT 2005). Rogers (2013) also documented savings of 20 hours on each time sheet data entry, increased accuracy of data collected, improved communication, and reduction in paper works by using Maintaining Assets for Transportation System (MATS). McCullouch (1991) estimated possible savings of over two million dollars because of reduced paper uses.

Despite those time and cost savings in data collection observed by various studies and despite the growth in DWR data being collected, the use of DWR data is still very limited – possibly because of the minimal recognition of the usability of the data, lack of in-house resources to analyze the data, insufficient data for any meaningful analysis, non-standard data format, and poorly defined procedures and mechanism use to extract, process, and analyze the data and generate usable information and knowledge to assist highway project decision makers (Woldesenbet et al. 2014). Much of the reported benefits are the result of the ease in DWR data access rather than from the better analysis and utilization of the data. There is an emerging need to develop a framework for proper DWR data collection and active utilization of DWR data.

Research Methodology

This study a) reviews the current practices of collecting and utilizing DWR data, b) identifies the benefits that can be obtained from DWR data, c) investigates the challenges for better collection and utilization of DWR data, and d) presents an enhanced framework to overcome the challenges identified. The study consists of an extensive literature review, phone interviews, and two nationwide questionnaire surveys (Figure 8). The literature review is focused on the utilization of existing DWR data and the studies to improve current DWR systems.

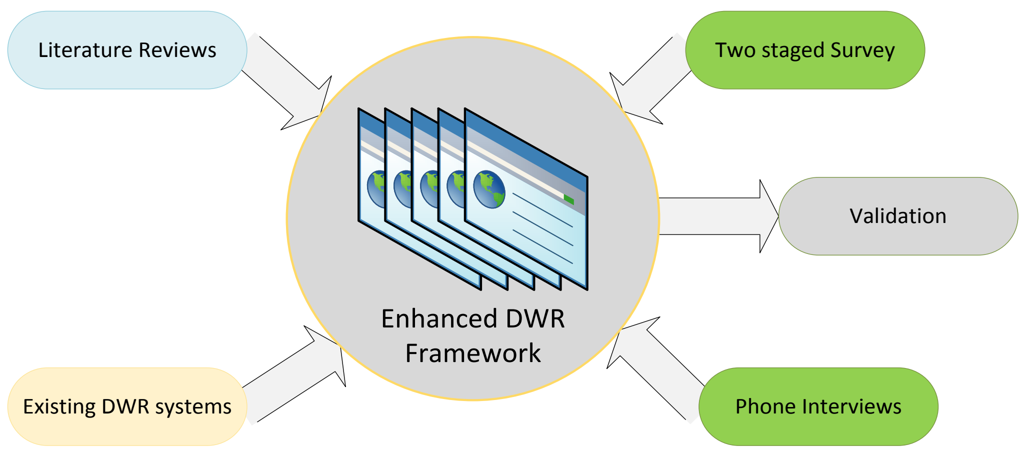

Figure 8 Research methodology to develop an enhanced DWR framework

Phone interviews are conducted to understand the current practices of DWR data collection and analysis in Iowa DOT. Two nationwide surveys are conducted to identify the national practices of DWR data collection, utilization, level of automation of various analysis, and challenges improving the current practices. Pilot surveys are conducted before each survey. The first survey is conducted in spring 2014 which received 151 responses out of 433 state DOT representatives contacted (34.87% response rate). The respondents represent 40 states out of 50 states contacted. It focuses on identifying the current practices of collecting DWR data its benefits. The second survey was conducted in fall 2014 and received 44 responses out of 115 state DOT representatives contacted (38% response rate). The respondents represent 27 states. It focuses on understanding the current level of automation of DWR data analysis. Based on the findings of the study, a framework is developed to enhance existing DWR systems or develop a new one. The proposed framework is validated by seven DWR experts from the U.S. and a case study.

Prior Studies

Prior studies about DWR can be broadly classified into a) the studies related to the use of DWR data and b) studies conducted to improving existing DWR systems.

DWR data utilization

The DWR data can be used for various purposes including progress monitoring, as-built schedule development, quantifying construction staffing needs, production rate estimation, and claim settlement. The start and finish dates of activities from DWR data can be used to monitor a construction progress (Chin et al. 2005; Elazouni and Salem 2011; Navon and Haskaya 2006). Navon and Haskaya (2006) developed a tool to build as-built schedules and track construction progress using DWR data. They argue that DWR data can be used for many other purposes such as to get early warnings if construction is not progressing as expected, developing a database for better future planning, improve litigation process, and monitor various construction resources.

Colvin (2008) used DWR data from South Carolina DOT’s DWR system (AASHTOWare SiteManager) to determine staffing needs for construction inspection. South Carolina DOT utilizes internal staff as well as outsource some of the inspection services to consulting firms. The DOT identifies future inspection staffing needs based on historical data and shares the results with consulting community so that consultants can plan and adjust their business plans accordingly.

The DWR data is also very important for construction arbitration (ALDOT 2013; Iowa DOT 2004; Kangari 1995). Iowa DOT project documentation guideline instructs their RCEs to collect sufficiently detailed data so that important events can be reconstructed later as they actually occurred. It further recommends collecting any data that may be useful to determine appropriate compensation for claims or disputes. Alabama DOT (2013) also instructs that sufficient details should be collected in its DWR system so that it can be used for legal issues to substantiate a just claim and disprove an unjust one. DWR data must be completed every day and should document important conversation with the contractors. If the contractor notifies about its intent to claim, more detailed data should be collected about the contractor's labor, equipment, and material usage. Kable (2006) analyzed the equipment usage data from Caltrans DWR system to estimate the emissions generated from its highway projects. The study notes the possible differences between the actual equipment operation hours and equipment hour data recorded in its DWR system.

Some states have conducted studies to utilize their DWR data to estimate production rates of controlling activities and determine contract times for future projects (Jeong and Woldesenbet 2010; Taylor et al. 2013). Realistic production rates form a basis for determining reasonable contract times. Aziz (2009) concluded that better contract times can be obtained by considering the effect of weather on the production rates. Several studies have been conducted to utilize the weather data collected in DWR systems to evaluate the effect of weather (e.g. rainfall) in production rates and to develop working day charts for every month of a year (El-Rayes and Moselhi 2001; Kenner et al. 1998; Sims et al. 2009). Another study noted that despite spending so much time and effort on DWR data collection, historical production rates with activity level granularity are not available to state DOTs (Williams et al. 2007).

Improving existing DWR systems

The findings of various studies related to improving construction data collection and visualization can be incorporated to improve existing DWR systems. Those studies reviewed here are focused on utilization of various digital technologies such as laser scanning, digital photography, mobile systems, teleconferencing.

In 1993, Russell developed a computerized DWR system as an improvement to the existing paper-based DWR system (Russell 1993). At that time, the system was developed to ease the retrieval of construction progress data. The system was also expected to improve the response time in dealing with problems and claims. Chin et al. (2005) argue that existing DWR systems are manual, time consuming, and there is a lack of information structure to represent activity information. To overcome the limitations, the study developed a check-list-based DWR system that can track macro level activities. The system aided in faster communication of activities needed to be done among the contractors and subcontractors. McCullouch (2008) developed a system for Indiana DOT (INDOT) to record material delivery ticket details using bar codes. The system transfers the ticket information electronically to the material supplier, trucking company, and INDOT personnel and contractor. The study further discussed the possibility of integrating the system with AASHTOWare SiteManager.

McCullouch (2000) previously recommended the use of teleconferencing, digital cameras, internet tools, and additional software programs for field data collection and construction project management for Indiana DOT. Those tools were also recommended in addition to the DWR module of AASHTOWare SiteManager that is already being used in the DOT. Hwang et al. (2003) synthesized existing technologies that can be used for DWR data collection. The technologies reviewed include 3D laser scanning, digital close range photogrammetry, sensors, mobile computing, wireless communication, video conferencing, remote collaboration, and project application service providers for data storage and management. The study concluded that the new technologies will improve the efficiency and enhance quality in collecting field data. Similarly, Leung et al. (2008) combined the data from a long-range wireless network, network cameras, and web-based collaborative platform to capture real-time images and videos for progress monitoring. Trimble Navigation Limited (2014) developed a contractor oriented DWR system called Trimble Proliance system. It has a rich visualization and analytical capabilities to show various construction statistics such as construction progress and earned value to date.

Other studies relevant to this study are focused on the use of non-structured data such as construction photographs and semi-structured laser scan data to develop as-built schedules for construction project monitoring, productivity analysis, and safety analysis (Golparvar-Fard et al. 2009; Turkan et al. 2012).

Current Practices of DWR Data Collection and Utilization

The current practices of DWR data collection and utilization are identified from literature review, phone interviews, and nationwide surveys. The current practices are presented in four sections: a) existing DWR systems, b) DWR data attributes, c) benefits of DWR data, and d) use of DWR data among various teams within sate DOTs. Finally, the challenges of DWR data collection and utilization is summarized.

Existing DWR systems

Although state DOTs are constantly pursuing to utilize more digital systems, many of them are still using paper-based systems for DWR data collection. Out of the 40 state DOTs that responded, over half of them (23) are using hybrid DWR systems, i.e. both paper-based systems and digital systems. Only 14 DOTs are using paper-less digital DWR systems. The remaining three DOTs are still completely relying on paper-based DWR systems. Analyzing such paper-based data will be very *labor-intensive*. At the same time, transferring DWR data from paper-based to digital system will result in the *duplication of efforts*. There is much room for error when manually transferring the data from paper-based to digital systems – resulting in the *data quality issues*.

A number of digital DWR systems have been developed over time by state DOTs. The AASHTOWare SiteManager is the most popular DWR system and is used by 22 state DOTs. The AASHTOWare FieldManager and Maintaining Assets for Transportation Systems (MATS) are two other systems developed, maintained, and used by more than one state DOTs' effort. Other state specific DWR systems developed and maintained by single state DOTs include PennDOT CDS NeXtGen, Utah Project Development Business System (PDBS) Field Book, Delaware FieldOps, Arizona DOT Pen, south Dakota Construction Measurement & Payment System (CM&P), and Kansas Construction Management System (KCMS).

Table 5 compares the capabilities of those DWR systems in terms of structured DWR data collection. The data attributes are classified into 9 categories. Most of the current systems are capable of collecting the fundamental information about the work quantities, contractor's presence, work suspension status, weather details, work location, and labor and equipment details. However, the level of granularity that can be collected about those information varies. For example, AASHTOWare SiteManager, AASHTOWare FieldManager, CDS NeXtGen, and PDBS Field Book can be used to collect low and high temperature of a day. The Delaware Field Data Collection (FDC) is developed to collect the temperature by time which enables more granular temperature data collection. However, temperature data cannot be collected in structured format in CM&P, Pen, and KCMS. Those systems also do not have functionality to collect AM and PM weathers separately.

The FieldOps and AASHTOWare SiteManager can be used to collect work suspension duration data but other systems do not have such functionality. Similarly, while AASHTOWare SiteManager, AASHTOWare FieldManager, CDS NeXtGen, PDBS Field Book, and FDC have structured fields to collect equipment data, FieldOps, Pen, CM&P, and KCMS lack such feature. Overall, the DWR systems developed and maintained by multiple state DOTs are relatively more flexible and powerful in terms of data collection and analysis capabilities.

Table 5 Data attributes that can be collected in DWR systems used by state DOTs

Attributes	**AASHTOWare SiteManager**	**MATS**	**AASHTOWare FieldManager**	**CDS NeXtGen**	**PDBS Field Book**	**FDC**	**FieldOps**	**Pen**	**CM&P**	**KCMS**
General										
Date	X	X	X	X	X	X	X	X	X	X
Day charging	X	-	X	-	X	-	X	-	-	X
Weather	-	-	X	-	X	X	-	X	-	-
Weather by time	-	-	-	-	-	-	-	-	-	-
AM weather	X	-	-	X	-	-	-	-	-	-
PM weather	X	-	-	X	-	-	-	-	-	-
Low temperature	X	-	X	X	X	-	-	-	-	-
High temperature	X	-	X	X	X	-	-	-	-	-
Temperature by time	-	-	-	-	-	X	-	-	-	-
Sunset	-	-	X	-	-	-	-	-	-	-
Sunrise	-	-	X	-	-	-	-	-	-	-
Work status	-	-		-	X	-	-	X	-	-
Work suspended from	X	-	-	-	-	-	X	-	-	-
Work suspended to	X	-	-	-	-	-	X	-	-	-
Accident indicator	-	X	-	-	-	-	-	-	-	-
Work activities										
Location	X	X	X	X	-	X	-	X	X	-
Installation station	-	-	-	-	-	X	X	-	-	-
Installation station from	X	X	X	X	X	-	-	-	-	-
Installation start town	-	X		-	-	-	-	-	-	-
Installation station to	X	X	X	X	X	-	-	-	-	-
Installation end town	-	X	-	-	-	-	-	-	-	-
Offset	-	X	-	-	-	-	-	-	-	-
Route direction		X	-	X	-	-	-	-	-	-
Item	X	X	X	X	X	X	X	X	X	X
Installed item quantity	X	X	X	X	X	X	X	X	X	X
Item measurement indicator	X	-	-	-	-	-	-	-	-	-
Controlling item indicator	X	-	X	-	-	-	X	-	-	-
Item needs attention flag	-	-	X	-	-	-	-	-	-	-
Item completion status	-	-	X	-	X	-	-	-	-	-
Material stockpile										
Stockpile quantity	-	-	X	-	-	-	X	X	-	X
Material source	-	-	X	-	-	X	-	X	-	-
Material manufacturer	-	-	-	-	-	-	-	X	-	-
Audit/approval status	X	-	-	-	-	-	X	X	-	-
Contractor details										
Contractor	X	-	X	X	X	X	-	-	-	-
Contractor presence	X	-	-	-	-	-	-	-	-	-
Daily staff presence	X	-	-	-	-	-	-	-	-	-
Contractor working status	-	-	X	-	-	-	-	-	-	-
Contractor hours worked	-	-	X	X	-	X	X	-	-	-
Labor details										
Personnel type	X	-	X	X	X	X	-	X	-	-
Personnel number	X	-	X	X	X	-	-	X	-	-
Personnel hours	X	-	X	-	X	-	-	X	-	-

Table 5 continued

Attributes	AASHTOWare SiteManager	MATS	AASHTOWare FieldManager	CDS NeXtGen	PDBS Field Book	FDC	FieldOps	Pen	CM&P	KCMS
Equipment details										
Equipment type	X	-	X	X	X	X	-	-	-	-
Equipment number	-	-	X	X	-	-	-	-	-	-
Equipment hours	X	-	X	-	X	-	-	-	-	-
Equipment standby hours	-	-	-	-	X	X	-	-	-	-
Utility details										
Utility personnel type	-	-	-	-	-	X	-	-	-	-
Utility equipment type	-	-	-	-	-	X	-	-	-	-
Utility equipment standby time	-	-	-	-	-	X	-	-	-	-
					-	-	-	-	-	-
DOT staff/inspector	X	X	X	-	-	X	-	-	X	-
DOT staff hours	-	X	-	-	-	-	-	-	-	-
DOT staff time from	-	-	-	-	-	X	-	-	-	-
DOT staff time to	-	-	-	-	-	X	-	-	-	-
DOT resources										
Vehicle mileage	X	-	-	X	-	-	-	-	-	-
DOT/Rental equipment	-	X	-	X	-	-	-	-	-	-
DOT/Rental equipment hours	-	X	-	-	-	-	-	-	-	-
DOT/Rental equipment mileage	-	-	-	X	-	-	-	-	-	-
Miscellaneous										
Force account details	X	-	-	X	-	-	-	X	-	-
Visitors	-	-	-	-	X	X	-	-	-	-

State DOTs have limited resources available to maintain the DWR systems and some DOTs reported that their DWR systems are old, not well maintained, and in in need of an update. The *lack of sufficient resources* is one of the factors that have forced state DOTs to use outdated systems. State DOTs have reported that partnering with other state DOTs have enabled them to combine their limited resources to develop a DWR system that meets their common needs (Fowler 2010).

DWR data attributes

In the nationwide survey, respondents are asked about the major DWR data attributes being collected, their perceived importance to obtain the benefits of DWR data, and the current method to collect those data (paper-based or digital). Irrespective of the importance of the data attributes, most of the data attributes are collected by about 150 respondents. There is no clear pattern between the average ratings and the data collection methods. It might be a better option to collect the more important data attributes in digital systems to ease its analysis. For example, the agency's quality assurance tests are considered to be the second most important data, but currently about half of the respondents (74) are collecting it in paper-based systems. The crew and equipment details for each day are not considered as important, but it is mostly collected in digital systems. The features available in their DWR systems and state DOT policies are probably the reason associated with collection of important data attributes in the paper-based systems. If appropriate policies and DWR systems are implemented, the important data attributes such as traffic control reports, contractor's quality assurance tests, and safety and incident reports can be recorded in digital format.

The link between the activity and equipment/crew is necessary to calculate the production rate. But, while many state respondents are collecting the equipment and crew details for each day (150 and 145 respondents), fewer respondents are collecting the equipment and crew details for each pay-item activity (98 and 98 respondents). This is possibly because of the lack of utilization of DWR data for production rate estimation in many state DOTs. Based on the survey, only about 25% of additional effort is required to collect resources data linked to the activities compared to the resources data collection without the links.

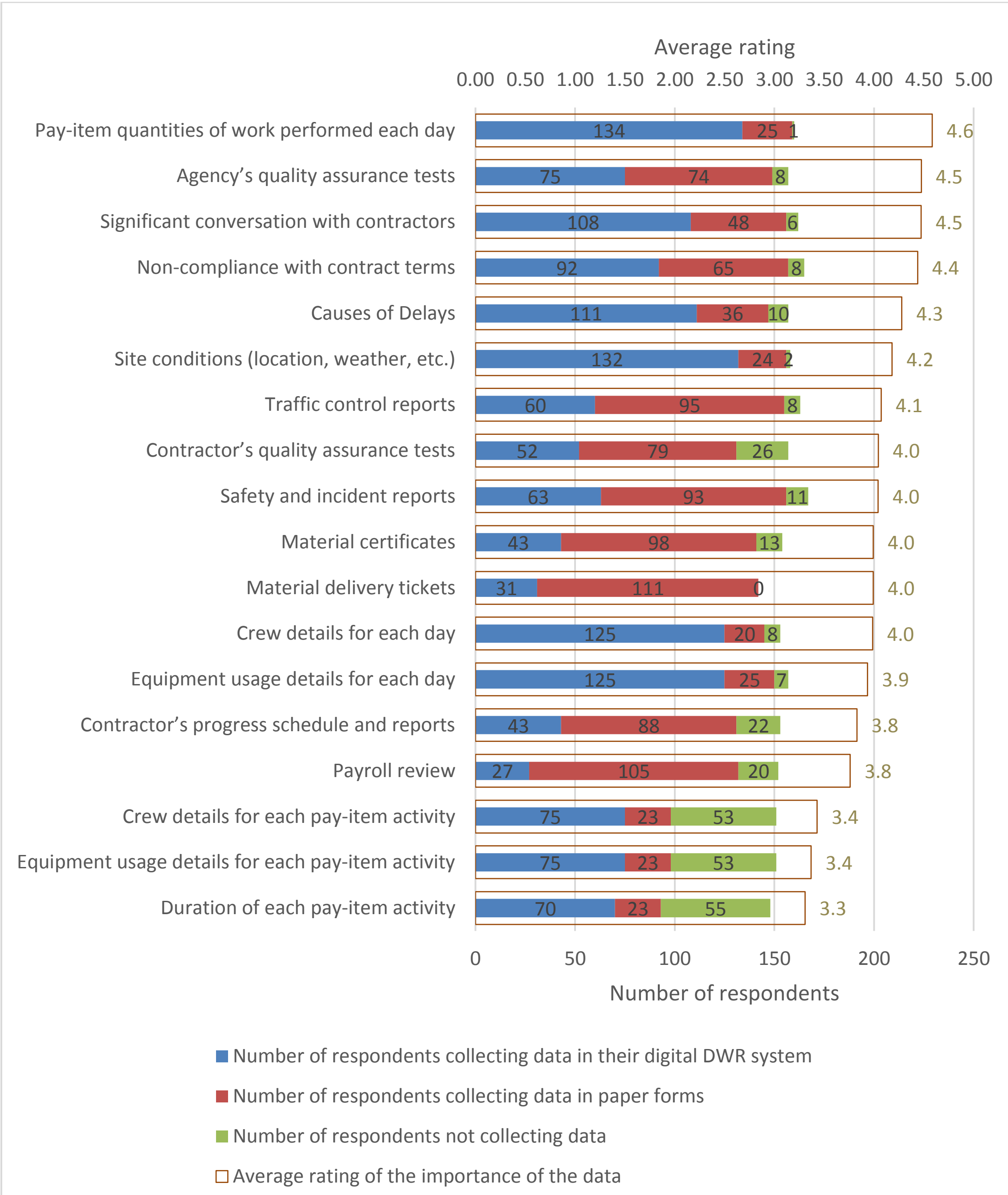

Figure 9 DWR data attribute collection practices

Benefits of DWR data

In the survey, the respondents are asked to rate the importance of various benefits of DWR data and the current level of automation to obtained those benefits. The DWR data are mostly being used for progress monitoring (92% respondents), dispute resolution (88% respondents), and contractor payment (91% respondents) (Figure 10). Those benefits are perceived to be very important by the respondents and are rated 4.1 out of 5 for progress monitoring, 4.1 for dispute resolution, and 4.6 for contractor payment. The other applications such as activity cost estimation, production rate estimation, contractor evaluation, and contract time determination are used by only about half or less than half of the respondents. The importance of DWR data is realized but the current level of benefits obtained is very limited. The dashed orange line shows the average path of the automation ratings and the dashed blue line shows the average path of the importance ratings. There are opportunities and areas for improvement to obtain more benefits from the DWR data as indicated by the gap between those two lines.

The survey results show that when there is a lower *level of automation*, the benefits are obtained by fewer respondents. On one side, the progress monitoring is rated with the highest rating of all (3.3 out of 5 on average) and the benefits of DWR data for progress monitoring is obtained by the highest percentage of the respondents (92%). On the other side, the automation rating for safety analysis is only 1.7 and such analysis are performed by only 29% of the respondents.

A Pearson's correlation coefficient (r) is calculated using the level of automation of various benefits as an independent variable and the percentage of respondents who obtained the corresponding benefits as a dependent variable to understand the relationship between those variables. "The Pearson's correlation coefficient is the product moment correlation coefficient, r, a dimensionless index that ranges from -1.0 to 1.0 inclusive and reflects the extent of a relationship between two data sets" (Microsoft Corporation n.d.). Mathematically, the Pearson's correlation coefficient can be represented as:

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{(x-\bar{x})^2}\sqrt{(y-\bar{y})^2}} \tag{11}$$

where x and y are variables under study and $\bar{x}$ and $\bar{y}$ are corresponding mean values. A positive coefficient indicates that an increase in one variable is associated with an increase in another variable. The higher the value of the coefficient, the stronger is the relationship. However, a correlation does not necessarily mean a causal relation.

There is a good correlation (0.59) between the level of automation and the percentage of the respondents obtaining the benefits. Thus, if the level of automation of as-built information (schedule, cost, etc.) generation is improved, more state DOTs may start to generate as-builts from the DWR data. Thus, there is a need to *automate the analysis* for obtaining various benefits using DWR data that is already collected. In other words, a proper methodology and algorithms should be developed for those analysis. The authors are working on another paper to automate the as-built schedule development using DWR data.

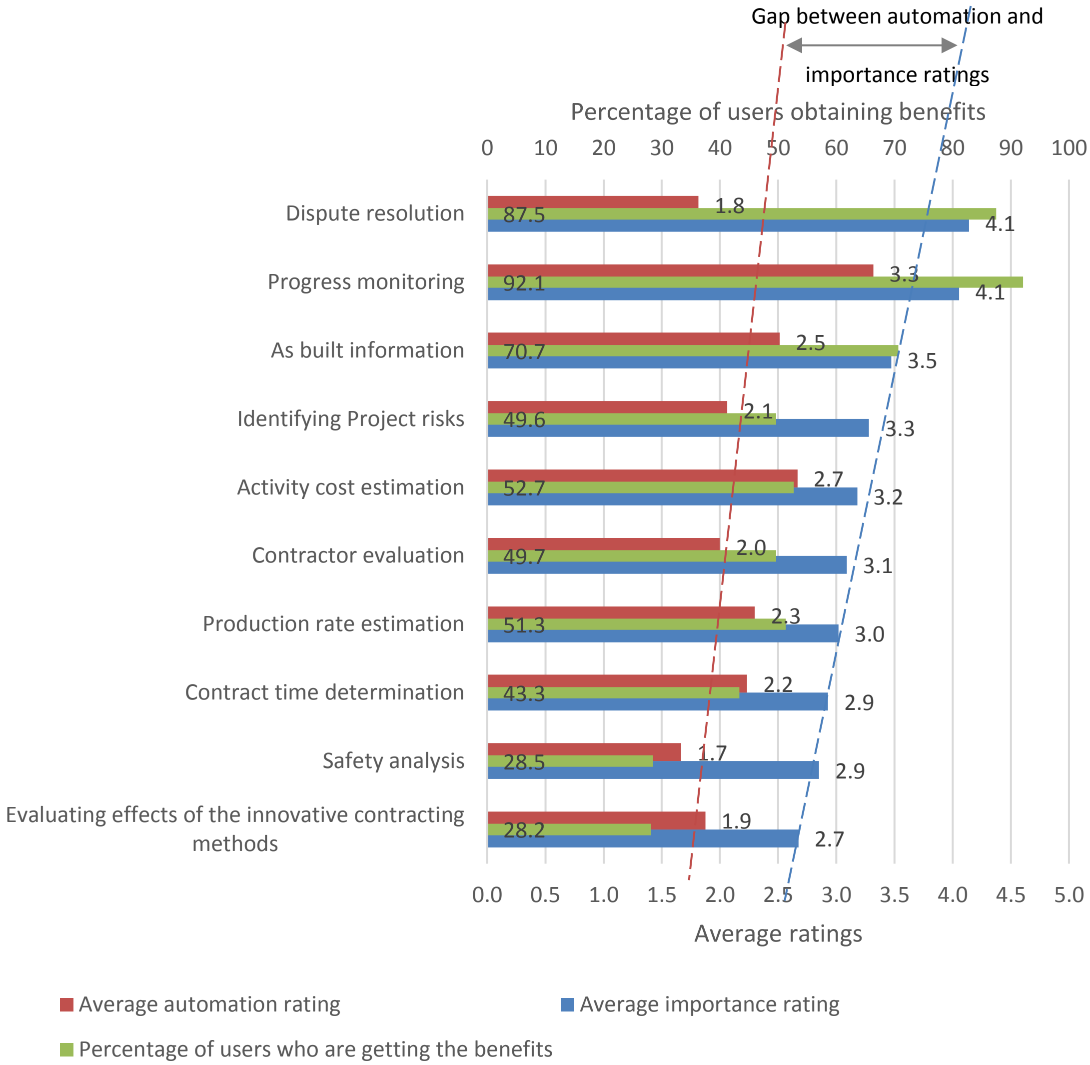

Figure 10 Application benefits of DWR data

Use of DWR data among various teams within state DOTs

The survey further investigated the reasons behind the limited use of DWR data, departments that can benefit, and departments that are actually benefiting from DWR data. A consistent gap was found between the two as shown in Figure 11 by the dotted lines. For example, 64% (23/36) of cost estimation teams who can possibly benefit from DWR data are not actually benefiting from it. Overall, more than half the teams (56%) are not benefiting from DWR data despite its potential benefits. The results indicate that there is a *lack of awareness* about DWR data and/or a *lack of automation* for various analyses required to obtain specific benefits by those teams.

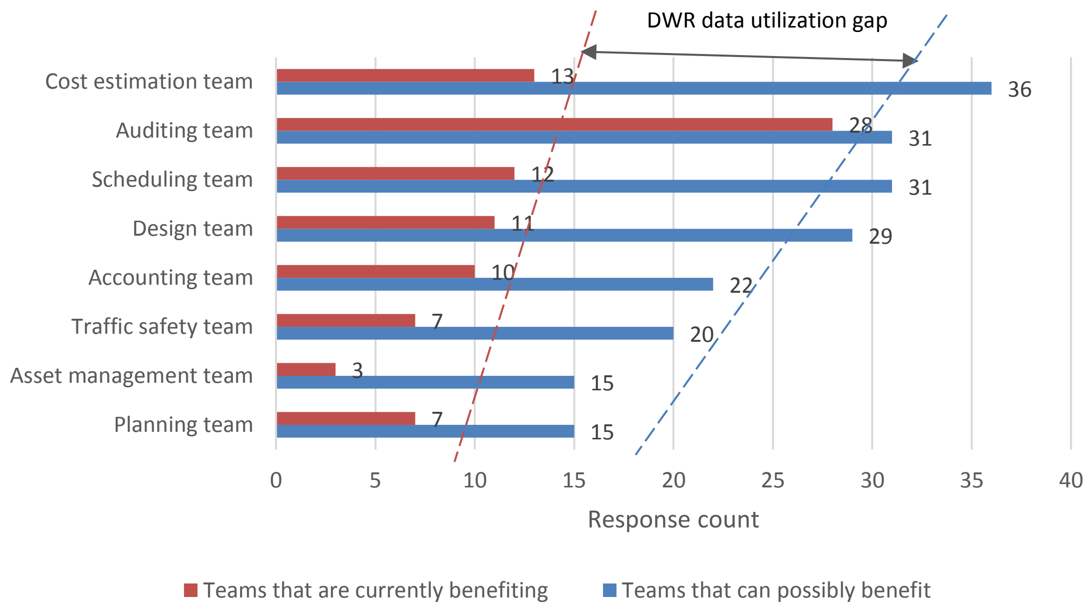

Figure 11 Teams that are possibly- and actually-benefiting from DWR data

Challenges of DWR data collection and utilization

Multiple challenges are identified for better collection and utilization of DWR data. Current DWR data collection practices are *labor intensive* and result in *duplication of efforts* and *data quality issues*. State DOTs have *limited resources* and *high turnover of experienced personnel*. They are using *slow and outdated hardware and software* that need to be updated.

Various teams within the state DOTs are not aware of the detailed data being collected in their systems. The teams that are aware about the data *lack the methodologies* to obtain various benefits or the methodologies are not automated. A considerable amount of data is collected in linguistic format which is challenging to analyze. A framework is developed to overcome those challenges and improve DWR data collection and utilization practices and is presented in the next section.

Enhanced Framework for Better Collection and Utilization of DWR Data

An enhanced framework for better collection and utilization of DWR data is developed based on the findings from the literature review, survey questionnaires, phone interviews, and review of existing systems. It consists of three major components: a) data model, b) automation of DWR analysis and reporting, and c) technical aspects.

Data model

A data model consists of data attributes, data entities, and the relationships between the data entities. The current DWR systems have a varying level of capabilities in terms of the data attributes that can be collected using the systems. Much focus is given in collecting linguistic data in many of the systems. Those unstructured linguistic data such as remarks often contains valuable information but is often not as easy to automate their analysis as other structured data.

Three methods can be used to develop structured fields to substitute or complement the unstructured linguistic data fields. The first one is based on text mining the linguistic data that is already collected. Text mining can reveal the frequently collected and major values for each remark type. For example, if two of the major reasons of change orders are a design error and effect of weather based on the analysis of historical remarks, then the linguistic remark field for the cause of change orders can be converted to a structured field with design error, weather, and others as choice values. The second option is to identify those major values based on the experience of inspectors and RCEs. For example, temperature, humidity, rain, snow, and wind might be the major weather factors affecting the productivity. Thus, instead of providing the linguistic weather remarks field, numerical fields for temperature and nominal fields for weather severities can be developed. Finally, state DOTs can also learn from other DWR systems currently being used by other state DOTs. Table 5 presents the various DWR data attributes currently recorded in the existing DWR systems.

Lack of the proper relationship between the data attributes will result in the limited usability of the data for detailed analysis. For example, when the labor hours are collected without any link to a particular activity, then a realistic estimation of production rates and study of effect of various factors on the production rates becomes probabilistic.

The Entity-Relation (ER) data model can be used to develop a data model for a DWR system. It is one of the popular methods to develop database systems. An ER model can visually represent data attributes, entities, and the relationships between the entities. An entity represents an item about which data is to be stored (Jan L. Harrington 2009). In case of a DWR system, equipment can be taken as an example of an entity. The equipment type, equipment name, number of equipment, etc. would be the examples of data attributes corresponding to the entity. For example, a data attribute – number of equipment – can have any integer value and hence its domain is in integer values.

Automation of DWR data analysis and reporting

Figure 12 presents 14 applications or benefits of DWR data that should be automated to ensure reduced time and efforts for decision makings. The quantities of work activities and the corresponding date cam be used to automatically generate as-built quantity, cost, and schedule information. Those as-builts can be compared against as-planned quantity, cost, and schedule to monitor construction progress and identify any deviation from a planned cost and schedule. The progress information becomes the basis for making contractor payments.

The deviations identified can be used to identify the possible risks of quantity, cost, and schedule overruns. The as-built quantities can also be used to calculate actual production rates of various activities. The production rates from historical projects, progress of the current project, and deviations from planned schedules can be used to identify the impact of the deviation in the overall schedule. This will enable state DOTs to take corrective actions before the schedule of the whole project is impacted. The claims for extra works performed can be resolved based on the activity costs based on historical projects. The claims related to the weather and site conditions related delays can be resolved by analyzing the weather effect on the productivity in past projects. Similarly, a delay analysis can also be used to identify the types of projects or project activities that are more likely to be delayed than others. In other words, the project risks in terms of schedule delay can be analyzed for future projects based on historical projects.

Similarly, a DWR system can be tied to an asset management system for asset management decisions such as construction project prioritization. For example, roughness index data collected after the completion of a project to check the quality of the pavement is an important data to plot pavement condition degradation curve. It can also be tied to DOT resource management systems to estimate the inspection resources required based on the time spent by inspectors for various work activity inspections in previous projects. Some activities like clearing and grubbing may not require much inspection, but other activities like base course installation may require more detailed inspection.

Further, DWR data can be used to check historical compliance records about regarding the civil rights such as minimum wage rate requirements and use of certified materials. The timely completion, number and severity of various issues encountered with a contractor, cost and schedule overrun/underrun, quality of final pavement, etc. can be used to automate the contractor ratings in parts. The cost and schedule overrun/underrun can be further used to evaluate the innovative contracting methods. The production rates and resources employed can be used as another method to estimate activity costs for future project cost estimation. Current practices are largely based on the use of historical bid data and per lane mile based parametric method. Finally, ongoing work activities and traffic control setup data can be analyzed with work zone crash data to evaluate its effect on work zone safety.

Select applications are presented in detail in the following subsections. Methods to automate those analyses are presented as mathematical equations and/or using a Structured Query Language (SQL). "SQL is a set-oriented programming language that is designed to allow people to query and update tables of information" (IBM 2013). To develop SQL queries a DWR table with four DWR data attributes are used. The data attributes, its descriptions, and its data type are presented in the Table 6.

Table 6 Data attributes in ‘work_activity’ table

Data attribute	Description	Data type
PRJ_NBR	Project Number	Short Text
DWR_DT	Daily Work Report Date	Date/Time
ITM_CD	Item code	Short Text
RPT_QTY	Reported Quantity	Number

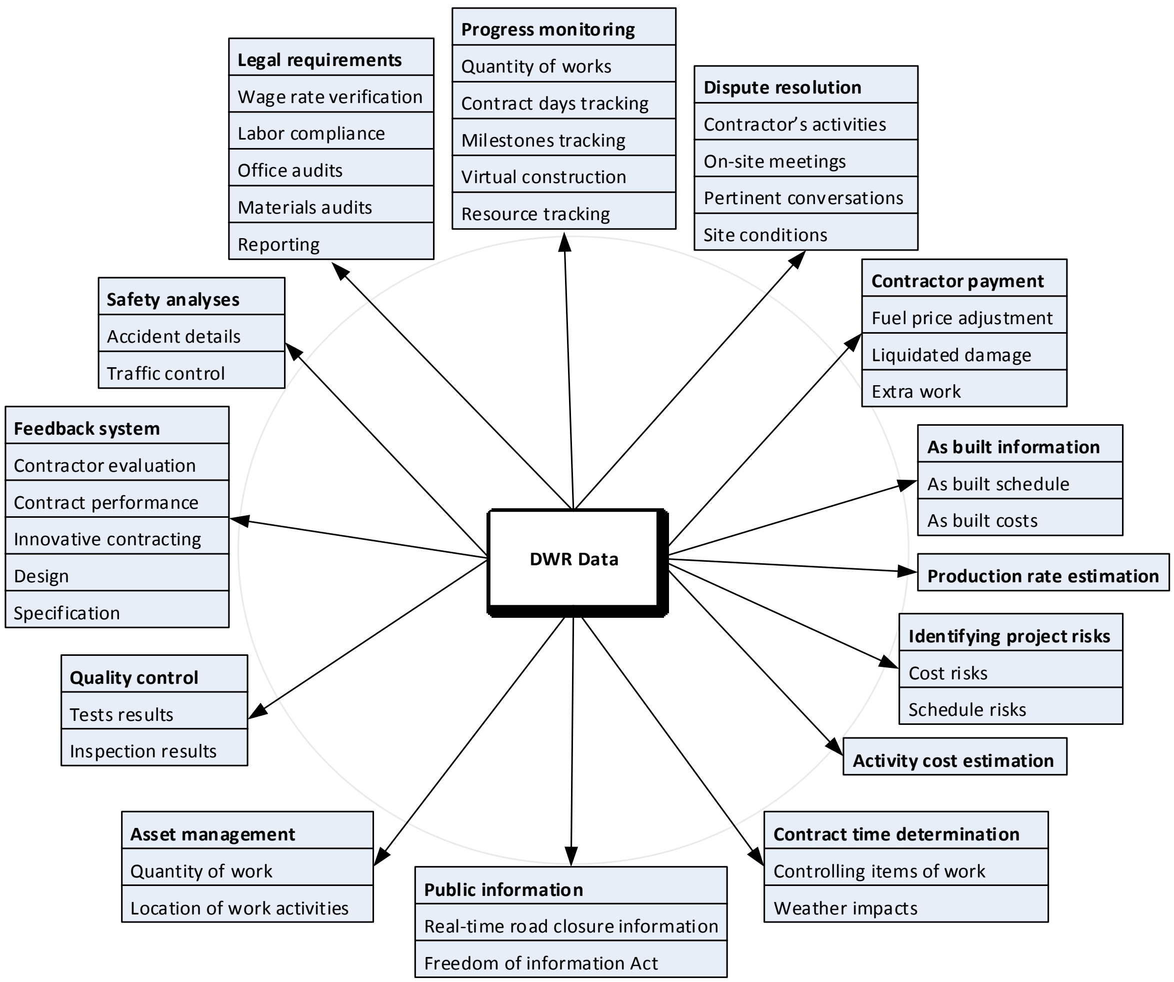

Figure 12 Applications of DWR data

As-built information

As-built information such as as-built quantities are one of the primitive information required to obtain many other benefits. Inspectors and RCEs collect work quantities every day and each record in the Work-activity entity indicates the quantity of works performed in that particular day. To obtain the total quantities of works performed to date, the quantities from all previous records corresponding to the project are be added. Mathematically,

$$\text{As} - \text{built quantity (ABQ)} = \sum_{i=1}^{t} \text{Quantities of work completed in day i} \qquad (12)$$

Where ‘t’ is the current date. In terms of SQL, an as-built quantity can be obtained using the query presented in Figure 13.

```
SELECT SUM(RPT_QTY) AS ASBUILTQTY
FROM work_activities
WHERE PRJ_NBR = 'pid'   AND
    ITM_CD = 'iid'   AND
    DWR_DT <= #yyyy/mm/dd#;
```

Figure 13 SQL query to extract as-built quantity

Where ‘pid’ is a unique project identification number for the project under consideration, ‘iid’ is a unique id of the item under consideration, and ‘yyyy/mm/dd’ is the date up to which the work quantity is to be calculated. Such a query can be executed for all the work items used in a project to generate as-built quantities for the whole project.

Progress monitoring

The previous query gives the total as-built quantity of an item at one instance of time. To monitor the progress of a project over time, the total cumulative quantities of works performed by day can be generated directly using the following SQL query (Figure 14).

```
SELECT w1.DWR_DT,
(SELECT SUM(w2.RPT_QTY) FROM work_activities w2
    WHERE w2.DWR_DT <= w1.DWR_DT AND
    w1.ITM_CD = w2.ITM_CD AND
    w1.PRJ_NBR = w2.PRJ_NBR) AS ASBUILTQTY
FROM work_activities AS w1
WHERE PRJ_NBR = 'pid' AND w1.ITM_CD = 'iid';
```

Figure 14 SQL query to generate progress monitoring information

Here, the result of the query will be the dates and cumulative quantities corresponding to each date. The output can be plotted to generate an S-curve for visual progress monitoring. If the query is executed for all the items in a project, the results can be used to generate an as-built schedule for the whole project.

Contractor payment

The contractor payment can be calculated using the unit prices and quantities of works done to date. If the contract has a retention clause, a percentage of the total cost to date should be deducted for retention (equation (13)).

$$\text{Contractor Payment (CP)} = (1 - r) * \sum_{i=1}^{n} q * u \tag{13}$$

Where r is the retention percentage, q is the quantity of work inspected and approved, u is the unit rate of the corresponding item, and I is an index representing items from 1 to n. The above is a basic equation to calculate the contractor payment. Additional consideration should be given to several contractual terms. For example, consideration should be given to the mobilization that is paid in advance and the retention that will be paid at the end of the project. Similarly, payments can be made in advance for the purchase of stockpiled materials and items. The liquidated damages, if any, should be deducted from the total payment.

Daily production rate

The DWR data collected in the field can also be used to generate daily production rates. A daily production rate (PR) can be defined mathematically as:

$$\text{Production Rate (PR)} = \frac{\text{Quantity of work (q)}}{\text{Total number of days (d)}} \tag{14}$$

A good estimation of production rate is required for better as-planned schedule development and contract time determination. While extracting the number of days (d), care must be taken to ensure that only days when that particular work is being conducted are counted. If more accurate production rates are to be determined, the number of workers working and equipment employed to complete the task may also be considered as crews of different sizes and equipment of different capacities can have different level of productivity. The weather data recorded in a DWR system can also be used to analyze the impact of weather on the production rates. The relationship can be then used to justify the contract time extension and/or settle claims related to production rates.

Pattern mining

The construction data can be analyzed to identify various patterns in the construction work performances. For example, the types of projects that are generally associated with higher change orders, frequent types of disputes, work items that are often associated with the quantity underrun/overrun, contractors with frequent claims and disputes, scheduler performance of various contractors, traffic control setups associated with the various levels of crash severities, etc. Similarly, the trend of the number of disputes over the years, scheduler performance improvements over time, etc. can also be studied using DWR data. Such analysis can be performed using various data mining techniques such as association mining and time series analysis.

Figure 15 shows an example of pattern mining using the concept of knowledge discovery in databases (KDD). The left side shows the various steps of KDD and the right side shows how it can be used to analyze the relationship between the production rates and various site conditions. The process of calculating production rate is already presented in a previous section. Here, in addition to the dates and pay-item quantities, the resources and site conditions are also used as input data for the selection stage. The data is then processed to generate cumulative quantities which can be transformed to generate production rates for various site conditions. Finally, the pattern analysis techniques can be used to analyze the relationship between the site conditions and production rates.

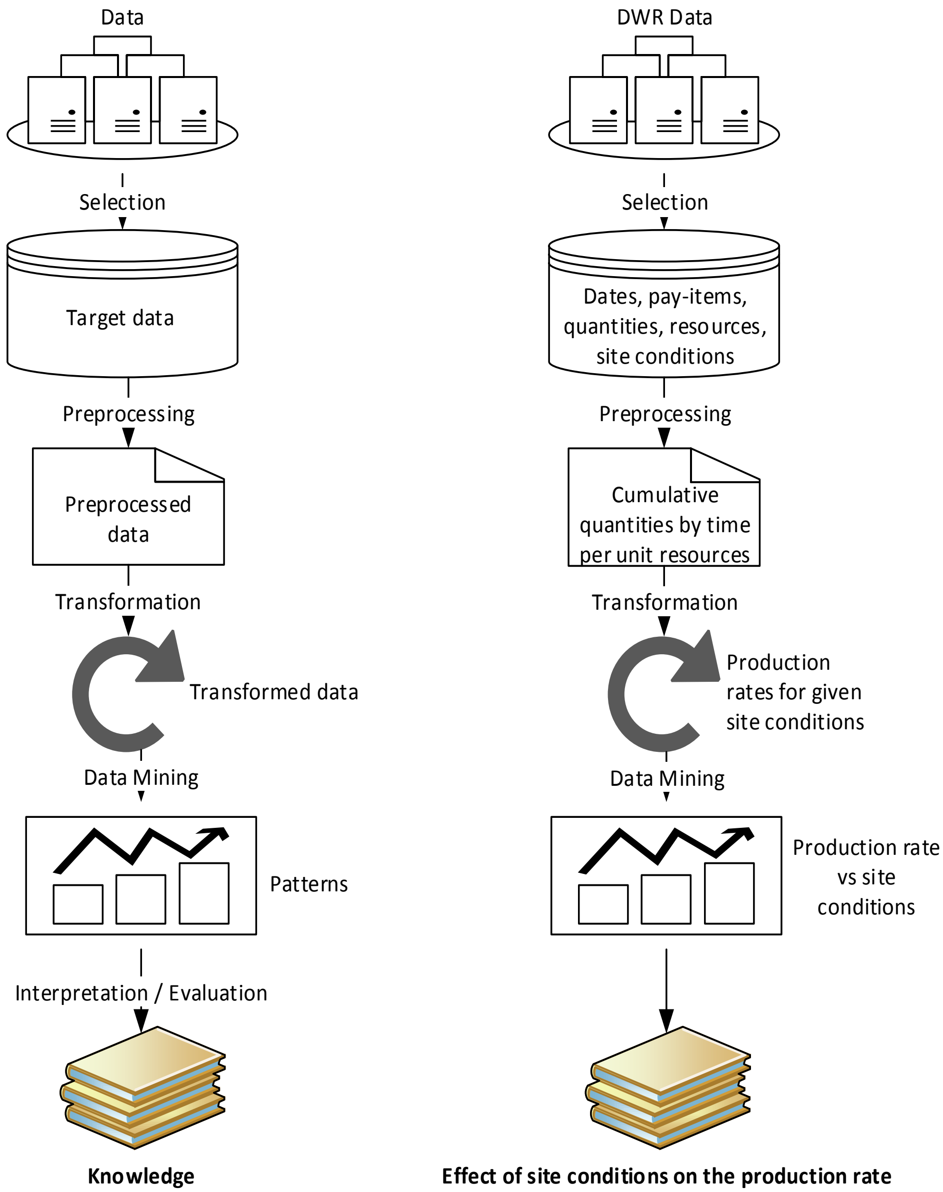

Figure 15 Application of knowledge discovery in databases (KDD) for pattern mining

Technical aspects

A DWR system should be technically robust in multiple aspects. It should be interoperable, should have an intuitive user interface, and good visualization techniques. Automated data collection can be implemented to ease DWR data collection.

A DWR system should be interoperable with other systems used by the DOT including cost estimation, asset management, contract time determination, bidding, contractor payment, and traveler's information. An interoperable system allows a seamless exchange of project data with other systems throughout the project development and execution. It further enables integrated and in-depth analysis of data and timely exchange of information. For example, if a DWR system is interoperable with traveler's information website, a real-time road closure information can be provided to the road users. Similarly, weather information of a construction site can be retrieved automatically from weather service providers' websites.

The system developed should also have intuitive user interface to ease DWR data recording and retrieval. A DWR system should have a proper login and access privilege system for data security. Existing departmental accounts can be used for the login system. The system should suggest default values for various data attributes like DWR data recording date. Also, the data entered in the system should be validated for proper format (such as numerical value) before recording in the system. A search functionality can be provided to ease the retrieval of previously recorded DWR data. The system should support a digital signature mechanism to reduce the duplication of effort resulting from the "wet ink" requirement. The system should be scalable to allow the collection of frequent and larger amount of data.

A DWR system developed should also have a good visualization system to present the tabular data in intuitive graphs and charts. Some other visualization systems that can be connected to a DWR system include Tableau, Pentaho Instaview and Treemap (AEC Big Data Inc. 2013; Pentaho Corporation 2013; Tableau Software 2013). A Geographical Information System (GIS) is another visualization system that can be connected to a DWR system to present various data such as construction progress.

Once a DWR system is developed, proper hardware such as portable laptops and tablets should be provided. Additional automated data collection systems such as Radio Frequency Identification (RFID), bar codes, LiDAR, Geographical Information System (GIS), equipment sensors, and camera can be used to facilitate DWR data collection. As one of the state DOT representative imagined, all DWR data will be collected automatically in the future without needing to manually enter the data in computer systems.

Validation

The enhanced framework developed in this paper is validated using two approaches. First, the framework is also validated by seven DWR experts from the U.S. Second, a case study is conducted to show the progress monitoring aspect of the framework.

Validation by DWR experts

In this case, the validation is performed via a questionnaire survey and it focused on soliciting expert opinions on an overall advancement provided by the framework over existing DWR systems. In the questionnaire, the experts are asked to rate various aspects of the framework on the scale of 1 to 10 – 1 being "poor" or "strongly disagree" and 10 being "excellent" or "strongly agree."

The DWR experts provided overall positive responses along with some constructive feedback. The experts commented that the framework is promising to improve existing DWR systems. This is also indicated by the average rating of 8.0 out of 10 in question 3 (Table 7). They supported the concept of an integrated and web based DWR system provided in the framework.

Table 7 Average ratings of the proposed DWR framework

S.N.	Question	Average Rating
1	This framework proposes some useful advancement over current DWR systems.	7.3
2	This framework can aid in tackling current challenges (listed below) of getting benefits from current DWR systems:	-
2.1	Lack of proper data attributes	6.4
2.2	Resources limitation	6.3
2.3	Duplication of efforts	7.2
2.4	Technical limitations	6.8
2.5	Current business practices	6.3
3	Current DWR systems can adapt some parts of the framework to improve their existing system.	8.0
4	The framework can be adapted to develop a DWR system if a state DOT does not have one.	8.3
5	The framework is comprehensive in terms of its scope.	8.5
6	The framework is easy to comprehend.	7.7

The ratings also show that the framework proposes some useful advancement over current DWR systems (7.3 out of 10). The framework is of high level and is comprehensive in terms of its scope for that level (8.5 out of 10). It is also easy to comprehend by DWR experts (7.7 out of 10). Finally, it is fairly good to overcome existing challenges that were identified in this study as indicated by the ratings for items 2.1 through 2.5 (rating of over 6 out of 10).

Case study

A sample Entity-Relation (ER) data model for DWR system is presented in Figure 16. It consists of nine entities or tables to represent major data attributes. The lines show between the entities show the connection between the entities. For example, the work activity entity is connected with the majority of entities including the equipment, labor, and weather. This link is missing in the existing DWR systems.

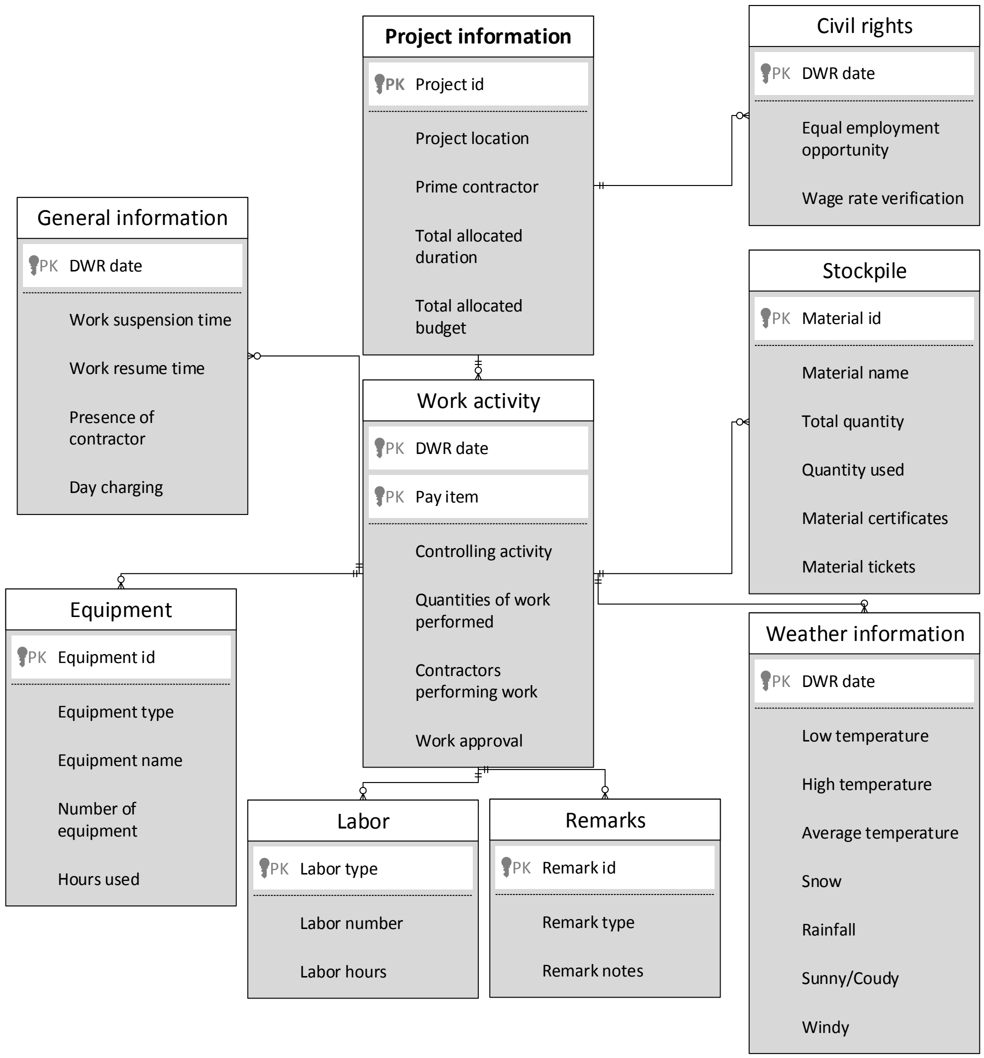

Figure 16 ER data model for proposed DWR system

A sample work activity data for a single project was extracted from a DWR database to demonstrate an application automation. The project contains 76 items and was let and awarded in 2014. A "class S concrete roadway" item was selected to present the analysis results for as-built and progress monitoring. Table 8 shows the results of the SQL queries. Each row in the table indicates as-built quantity completed to the date. The date, as-built pairs represents the progress of the concrete roadway work item over time which can also be plotted to generate an S-curve.

Table 8 Progress monitoring using cumulative as-built quantity

Date	Cumulative as-built quantity
7/17/2014	14 CY
8/1/2014	27 CY
8/6/2014	52 CY
8/13/2014	66 CY

Based on the successful validation results, it can be concluded that the proposed framework can be used for development of a new DWR system and/or updating current DWR systems that is focused on the automation of various decision making analysis.

Conclusions

Although, the importance of DWR data has been widely realized, there are multiple challenges faced by state DOTs that limits the current utilization of DWR data. Some of the challenges include data quality issues; duplication of efforts; and lack of resources, awareness of data being collected, methodologies to analyze data, automation of those analysis, and proper hardware. Many existing systems can collect fundamental data such as work quantities, contractor’s presence, location, and labor data with limited details. But, those systems lack the capability to collect detailed structured data such as temperature, work suspension time, equipment usage, incident report, traffic control. Also, the important links such as an activity-resource link are not present in those DWR data collection systems.

The study also identified a notable gap between the possible and current benefits obtained by state DOTS. Many benefits such as production rate estimation, activity cost estimation, contractor evaluation, contract time determination, etc. are obtained by half or less of the respondents. The benefits are also obtained by fewer teams within state DOTs than all teams who can possibly benefit from it. The limited use of DWR data is statistically associated with the limited level of automation of the benefits. Thus, DWR analysis should be automated to improve the use of DWR data.

A framework is developed to overcome the challenges identified for DWR data collection and utilization. The framework consists of a) data model, b) automation of DWR analysis and reporting, and c) technical aspects. Three methods to identify and develop a proper data model along with an example data model is presented. Fourteen application benefits and examples to automate analysis required for those benefits are presented using mathematical form and as Structured Query Language (SQL) queries. Under the technical aspects, interoperability, intuitive user interface, visualization, and automated data collection techniques are presented. The framework is validated by DWR experts from the U.S. The framework can be used to develop a new DWR system or to improve existing systems and is expected to aid in data-driven decision makings to manage construction projects.

Acknowledgements

The authors would like to acknowledge Mid-American Transportation Center for funding this study. We would also like to thank state DOT representatives for providing their valuable opinions for the study.

CHAPTER 4

COMPUTATIONAL ALGORITHM TO AUTOMATE AS-BUILT SCHEDULE DEVELOPMENT USING DIGITAL DAILY WORK REPORTS

K. Joseph Shrestha[7] and H. David Jeong[8]

Abstract

As-built schedules prepared during and after construction are valuable tools for State Highway Agencies (SHAs) to monitor construction progress, evaluate contractor's schedule performance, ensure successful execution of a project, and defend against potential legal disputes. However, previous studies in this area indicate that current as-built schedule development methods are manual and rely on the information scattered in various field diaries, meeting minutes, and progress reports. SHAs have started to use digital Daily Work Report (DWR) systems to store field activity data in structured format that include sufficient data to develop as-built schedules. These valuable data have great potential to automatically generate as-built schedules if a proper methodology and its computational algorithm are designed and developed. This study directly addresses this issue and develops a complete methodology and its computational algorithm that can generate project level and activity level as-built schedules during and after construction. A standalone prototype system 'As-built Schedule System (ABSS)' is also developed to automate the entire process and develop

[7] PhD Candidate, Iowa State University, Dept. of Civil, Construction & Environmental Engineering, Iowa State University, Ames, IA 50011, shrestha@iastate.edu
[8] Associate Professor, Dept. of Civil, Construction & Environmental Engineering, Iowa State University, Ames, IA 50011, djeong@iastate.edu (515) 294-7271

visualized as-built schedules. A real highway project's DWR data was obtained from a state highway agency to successfully demonstrate the ability of ABSS in generating As-built schedules and provide insights for empowering future project schedulers and project managers. The outcomes of this study is expected to significantly aid SHAs in making better use of already collected DWR data, facilitate as-built schedule development and visualization, monitor construction progress with higher granularity, and utilize as-builts for productivity analysis and resolving potential issues causing delays.

Key Words: as-built-schedule, as-built-to-date schedule, daily-work-report, project schedule-monitoring, big-data, data-analytics, visualization, automation.

Introduction

As-built schedules represent the actual sequences and durations of construction activities of a project and take account of the change orders and schedule changes from the originally planned schedule (Hegazy et al. 2005; Henschel and Hildreth 2007; Knoke and Jentzen 1996; Vandersluis 2013). For highway construction projects, contractors are generally required to submit the originally planned schedule before the construction of a project starts and update the project progress during the construction. As the owner of a highway project, State Highway Agencies (SHAs) also collect and document various work progress related information from the construction site on a daily basis to make a monthly payment to the contractor and to be prepared for resolving any possible claims.

There are different types of as-built schedules in terms of the timing of as-built schedule development and the level of details as shown in Figure 17. As-built schedules can be categorized into *as-built to date schedule* and the *final as-built schedule*. As the terms suggest, an *as-built to date schedule* is developed during construction as a real time check on schedule performance. Once the construction is complete, the final as-built schedule can be developed, which includes actual sequences and durations of all activities of the project.

Based on the level of detail, as-built schedules can be categorized into a *project level as-built schedule* and an *activity level as-built schedule*. A project level as-built schedule can be presented as a bar chart and it shows the progression of work activities throughout construction. An activity level as-built schedule can be presented as cumulative quantities of work over time and hence it can show the progression of a specific work item. Since a typical highway project involves a small number of repetitive activities on a long stretch of a roadway, an activity level as-built schedule can play a significant role in assessing the project's schedule performance.

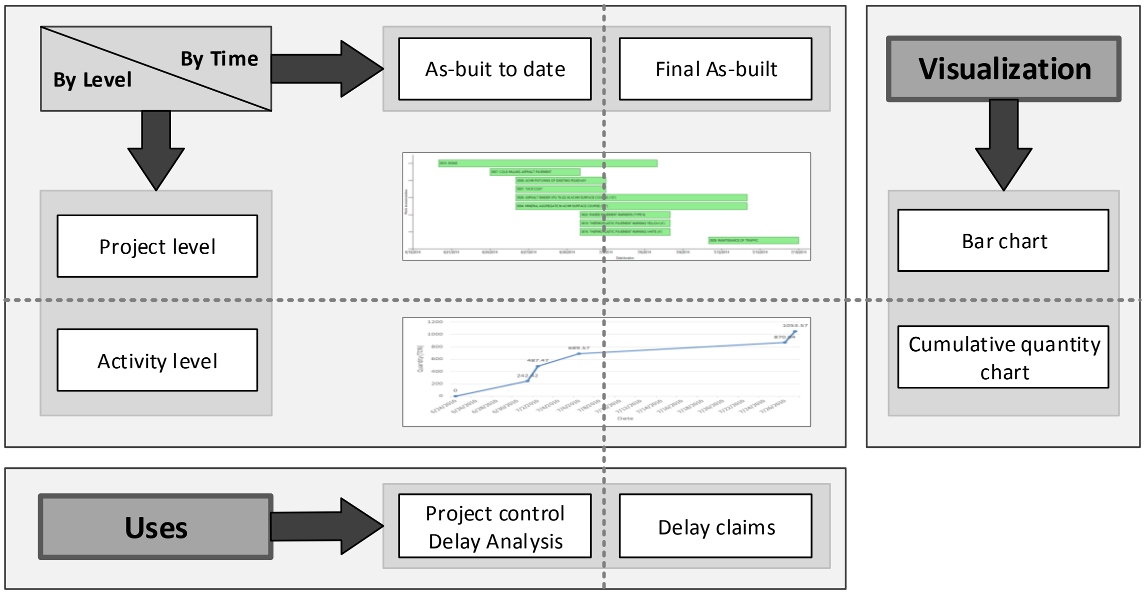

Figure 17 Different types of as-built schedules

A project level as-built schedule to date is an important tool to ensure that a project will be completed within the contact time (Knoke and Jentzen 1996). It can be used to verify contractors’ progress report on ongoing activities (Obr 2015). Delays can be identified by comparing the as-built schedule to date with the planned schedule (Avalon 2014; Knoke and Jentzen 1996). If any delay is identified early in the project, corrective actions can be taken to bring the project back on track to complete it on time. When, a delay occurs during construction, an as-built to date schedule can be used to validate the contractors’ claim for delay compensation or the request for time extension. The final as-built schedule is a documentation of durations and sequences of all activities. As such, it can also aid new schedulers in developing schedules for new projects (Knoke and Jentzen 1996).

An activity level as-built schedule enables monitoring of activity level progress and quantity underrun/overrun if any. Contractors' payment can be adjusted based the quantity underrun/overrun. It can be used to calculate the observed production rates that can be compared with the expected and/or standard production rates to evaluate the contractors' performances. Once the project is completed, the production rate obtained from the project can be used as a historical production rate to determine contract time for similar future projects (Woldesenbet et al. 2012).

Despite the importance of as-built schedules, as-built to date schedules are not typically developed and maintained throughout the project (Knoke and Jentzen 1996). The final as-built schedule is developed at the end of the project based on memory and information scattered in various forms and field diaries that may be outdated (Hegazy and Ayed 1998). Such methods involves manual efforts and are often inaccurate as some useful information may be lost before the end of the project (Elazouni and Salem 2011; Memon et al. 2006). Developing and maintaining as-built schedules throughout the project could be a cost effective approach as they will enable the project team to resolve any potential delay issues as they occur–which can avoid costly claims at the end of the project (Knoke and Jentzen 1996). However, there is a lack of a systematic methodology to generate both as-built schedule to date and the final as-built schedule (Hegazy et al. 2005).

Moreover, existing commercial scheduling systems do not allow the collection and recording of actual activity progresses over time, but only allow recording of the latest status of the project (Kahler 2012; Knoke and Jentzen 1996). As such, current systems are not very useful for as-built schedule development.

SHAs have recently started to use digital daily work report systems that are customized to collect structured site activity data. These systems offer great potential to automatically develop as-built schedules when proper data extraction and computational methods are employed. Existing studies in this area are focused on either utilizing unstructured data to manually develop as-built schedules or developing a new data collection system that can be used for developing as-built schedules (Hegazy and Ayed 1998; Knoke and Jentzen 1996; Navon 2007). Further, those studies are often focused only on a project level as-builts, disregarding the importance of activity level as-builts.

The overall goal of this study is to develop an automation methodology to develop as-built schedules for progress monitoring with higher granularity, support project schedule control decisions, and aid in resolving claims. The specific objectives are to a) develop a systematic methodology to generate project level and activity level as-built schedules during and after construction using structured site data already collected by SHAs, and b) develop a prototype to automate the computational process and visualize the as-built schedules

Prior Studies

Prior studies conducted on developing as-built schedules from site records are focused on two areas: a) developing as-builts manually based on available unstructured information and b) developing site data collection systems that can be used to collect specific data required for an as-built schedule development. The first two studies reviewed here including Kahler (2012) and Knoke and Jentzen (1996) relied on unstructured information while the next two studies conducted by Hegazy et al. (2005) and Navon and Haskaya (2006) focused on developing Spreadsheet based tools to collect the required data for as-built schedule development.

Kahler (2012) discussed the possibility of developing as-built schedules from site data and presented a conceptual methodology of developing various databases and tools to achieve the goal. The study claimed that developing as-built schedules from site data would not add any significant workload to the site personnel. However, the study failed to realize that such data are generally collected in various paper-based documents and are not structured. As such, additional effort will be required to extract the data and develop as-built schedules.

Knoke and Jentzen (1996) also discussed the possibility of utilizing scattered site records such as daily reports, correspondence, meeting minutes, progress reports, payment applications, testing records, submittal logs, material delivery tickets, and change orders to manually extract useful information to develop as-built schedules. The relevant information that can be manually extracted for as-built schedule includes start and finish dates of activities and project milestone dates. The study suggested that if the start date of any activity is close to the end date of another activity, those activities might have a finish to start relationship that can potentially be used to develop a critical path diagram. However, it is up to the scheduler to determine if such relationships are practically valid. The start and finish dates of activities are used to develop a bar chart using a commercial scheduling system. The study reported that while this method could be useful in defending against delay claims, it could be costly and time consuming because of the manual processing of the site records.

Hegazy et al. (2005) pointed out the importance of collecting structured site data instead of relying on existing site information collected through various site records. They developed a spreadsheet based file to collect information about the percentage of various work activities completed every day in a format similar to a bar chart (Figure 18).

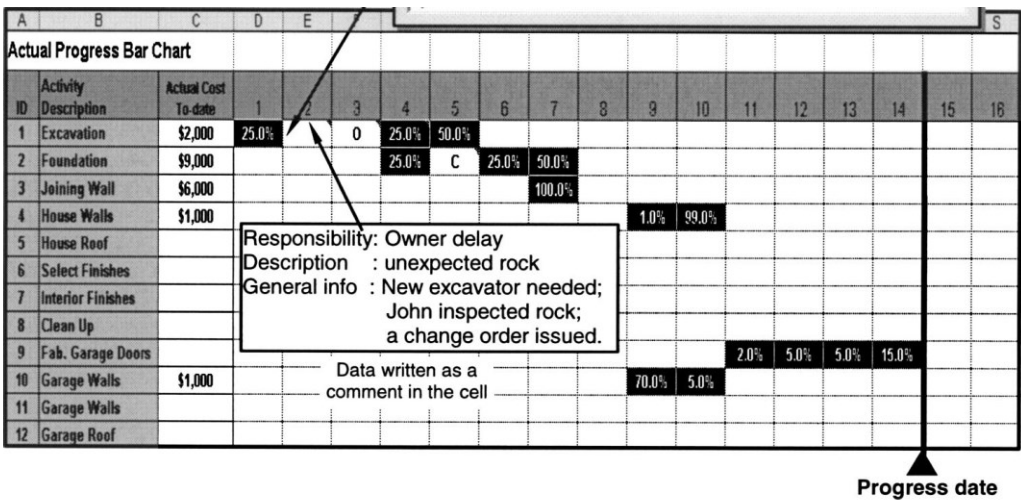

Figure 18 Spreadsheet based site record collection and as-built schedule

(Source: Hegazy et al. (2005))

If a contractor does not perform work in a particular day, a site engineer needs to enter a reason for the delay as a remark in the spreadsheet cell. This file can serve as semi-structured site records as well as an as-built bar chart schedule. The limitation of the study, as recognized by them, is that the spreadsheet-based program is not very practical to handle large projects and it cannot effectively store and analyze data from multiple projects. Navon and Haskaya (2006) conducted a similar study. First, information about the percentage of work completed is recorded in a spreadsheet file. Then, this data is exported to Microsoft

Project program to develop an as-built bar chart schedule. It used a simple hypothetical project consisting of only three activities to demonstrate the process.

In those previous studies, either information from various existing sources needs to be processed manually or relevant data and information need to be collected directly into a system to reduce manual efforts. Those studies did not clearly recognize the possibility of using already collected information, which might be attributed to field data collection practices that their research was based on. This thought is echoed by Elazouni and Salem (2011) and Memon et al. (2006) who reported that even if as-built schedules are possible to be developed, current methods are manual, slow, inaccurate, and expensive. Kahler (2012) also reported that as-built schedules are prepared mostly based on an outdated information and only after construction is completed. Further, current methods are just focused on the project level as-built schedules and they are not designed to develop activity level as-built schedules.

With the continuous evolution in digital technologies, SHAs in the U.S. have started collecting various site records in a structured system known as a Daily Work Report (DWR) system. SHAs started using such systems as early as 1990s. As such, a vast amount of data is already collected in those systems and are readily available for developing as-built schedules if a proper methodology is developed to utilize them.

Daily Work Report Data

SHAs collect a significant amount of data such as ongoing construction activities, labor hours, types of equipment used, equipment hours, weather data, and significant communications with contractors in a Daily Work Report (DWR) system (Shrestha et al. 2015). Site inspectors spend as much as 40% time in collecting those data (McCullouch 1997). SHAs have developed various electronic DWR systems over time including AASHTOWare SiteManager, AASHTOWare FieldManager, MATS, Next Generation, and Field Operations (Shrestha et al. 2015). Currently, 37 SHAs are using various electronic DWR systems. Figure 19 shows a screenshot of AASHTOWare SiteManager.

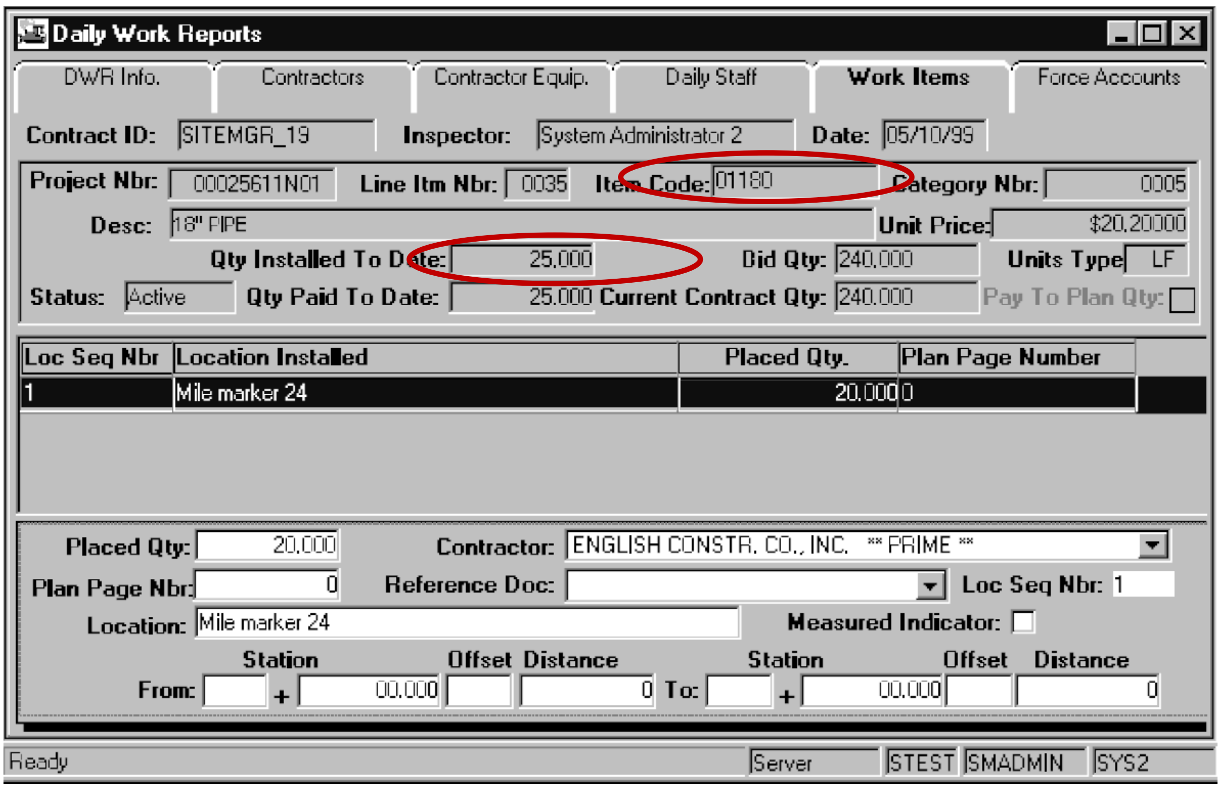

Figure 19 AASHTOWare SiteManager (ASM)

DWR systems have been developed and used with the main objective of making correct payment to contractors and documenting field activity records as preparation for potential claims and disputes. Moreover, the data attributes recorded in the DWR system have potential to be utilized for other purposes such as as-built schedule development, production rate and activity cost estimation, contract time determination, and contractors' performance evaluation (Shrestha et al. 2015). However, most SHAs have not benefited from those potential applications possibly because of the lack of knowledge on those potential benefits, enabling methodologies, and automation processes. Shrestha et al. (2015) found that more users of DWR systems obtain benefits when the level of automation is higher.

DWR data attributes are typically linked to a work item. In the U.S. highway industry, SHAs have developed an extensive list of work items that are primarily developed to facilitate the bidding process under unit price contracting mechanism. Those work items are also used to develop a project schedule as work activities. SHAs have developed and maintained specification manuals that provide detailed specification of each work item. For example, an item code '01180' in red circle in Figure 3 indicates a work item "supply and installation of a mile marker." A typical set of data attributes collected in DWR systems can be classified into six categories: general information, work activities, weather information, equipment, labor, and remarks (Figure 20).

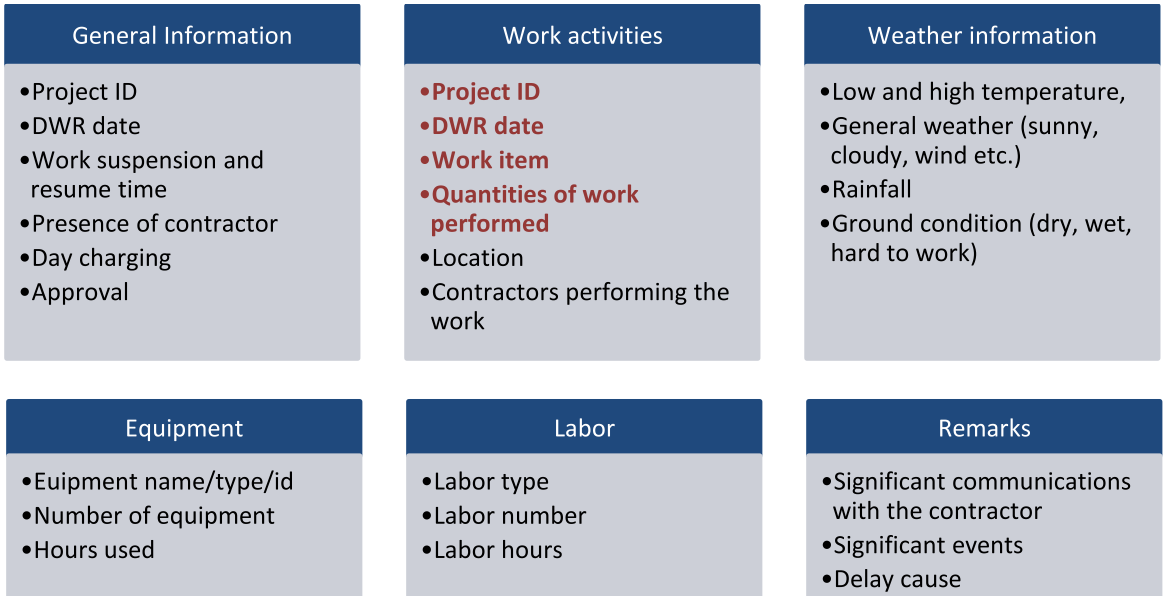

Figure 20 Typical data attributes collected in DWR systems

Among these six categories, it is important to note that the category of work activities contains directly relevant and sufficient data needed for developing as-built schedules. Mattila and Bowman (2004) used such work activity data to verify the accuracy of contractors' schedule. First, they developed a list of controlling activities by date. Then, they compared each date's actual progress of controlling activities against activities noted in the original planned schedule. If the controlling activities between as-built schedule and as-planned schedule match, it is noted as 'accurate', if not it is noted as 'inaccurate'. The ratio of the total number of 'accurate' days to the 'inaccurate' days is considered the accuracy of the contractors' schedule. The study found that schedule accuracies ranged from very low (10%) to high (80%) for 22 projects analyzed. They concluded that contractors consistently tended to optimistically estimate durations of controlling activities in those projects. A framework is presented in the next section that can aid SHAs in their decision-makings including reducing such erroneous reporting of project progresses.

Framework for Automatic As-built Schedule Development

The overall framework to generate and visualize as-built schedules is presented in Figure 21. The framework can be divided into five components: a) database development, b) project selection, c) data processing, d) project performance evaluation, e) visualization of as-built schedules.

First, a required dataset is obtained from an existing DWR system. Then, a project is selected for which as-built schedules are to be developed. The data about quantities and dates are extracted for all work items associated with the project. Then, the production rates for those work items are calculated based on the quantities and dates. The production rates are used to evaluate the contractor's performance on each activity and calculate time remaining to complete a work item. Finally, activity level and project level as-built schedules are developed and visualized.

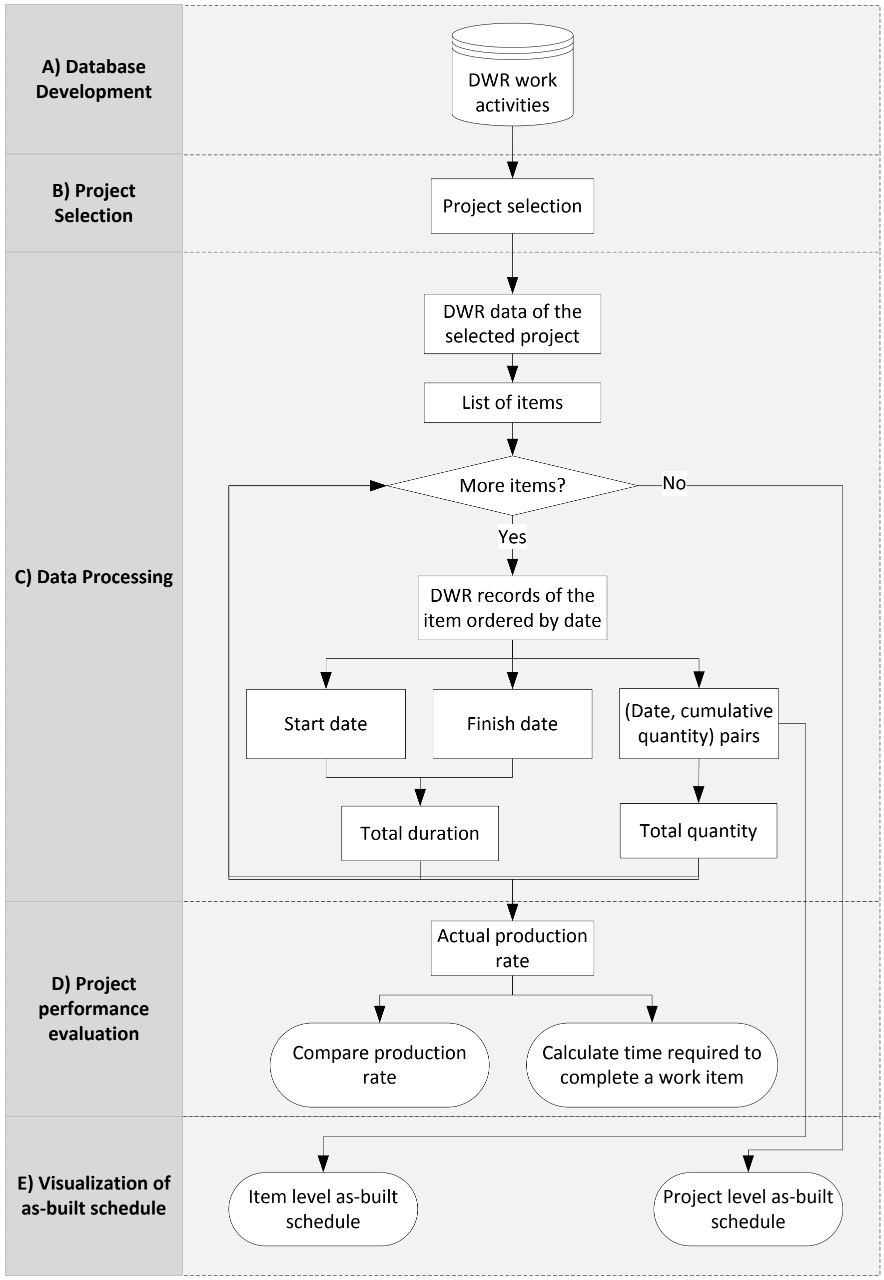

Figure 21 Methodology to develop as-built schedule

Database development

The framework requires full work activity data that can be obtained from a DWR system used by a SHA. At minimum, the dataset should contain project ID, DWR date, work item, and quantity of work done on the recorded date. Assume a set of project ID (P) containing ‘n’ project IDs defined as:

$$P = \{p_1, p_2, p_3, \dots\ p_n\} \tag{15}$$

An uppercase letter denotes a set while the corresponding lowercase letter denotes its elements and the subscripts indicate the element numbers. Assume another set, DWR (D) whose elements (d_i) are vectors of Project ID (P), DWR date (T), work item code (W), and quantity (Q) as elements (equations (16) and (17)). This set contains one record for each work item for each day.

$$D = \{d_1, d_2, d_3, \dots, d_n\} \tag{16}$$

$$d_i = (p_i, t_i, w_i, q_i) \tag{17}$$

Project selection

In this component, a project of interest is selected (p_i). The project can be an ongoing project whose progress can be checked and monitored using the as-built schedule to date, or it can be an already completed project whose final as-built schedule can be generated.

Data processing

This is the core component of the framework that generates the data required for project level and activity level as-built schedule development and production rate calculation. First, a subset of the DWR set ($S \subseteq D$) pertaining to the selected project is generated. This can be mathematically expressed as:

$$d_i \in S \mid d_i(p) = p_i \tag{18}$$

Where, d_i (p) represents the project ID of the particular DWR element and the symbol '|' indicates 'such that' or 'conditional' statement.

The elements of this new set (S) are divided into several subsets (WS_j)–one for each work item (equation (19) and (20)). Thus, if there are 'm' numbers of work items, there will be 'm' numbers of subsets.

$$S = WS_1 \cup WS_2 \cup WS_3 \cup \ldots \cup WS_m \tag{19}$$

$$s_i \in WS_j \mid s_i(w) = w_j \tag{20}$$

The elements of each subset (WS) is sorted by DWR date (t) in ascending order. The sorted data is used for two purposes: identification of start and finish date of the current work item and generation of cumulative quantity over time (cq). The first and the final DWR records in the sorted lists are considered to be the start date (sd_j) and finish date (fd_j) of that particular activity (w_j). The work item, start date, finish date vector (w_j, sd_j, fd_j) is used later in the visualization component. From this vector, the days required to complete the work item (dt) can be computed using equation (21):

$$dt = fd_j - sd_j \tag{21}$$

The cumulative quantity (cq_t) for each date (t) can be calculated as

$$cq_t = \sum_{l=0}^{t} q_l \tag{22}$$

Where q_l indicates the quantity of that item reported in time 'l'. From this, (t, cq_t) pairs can be generated for all values of 't'. Such pairs are used in the next two components to evaluate the schedule performance of the project and visualize item level as-built schedules.

Project performance evaluation

In this component, an activity level performance of the project can be evaluated. First, a production rate (PRj) is calculated as the ratio of the final cumulative quantity of the work item to the total duration of the work item (23).

$$PRj = \frac{cq_{t,j}}{t} \tag{23}$$

Where $cq_{t,j}$ is the cumulative quantity for work item w_j in time 't'. If the work item is completed, then t = dt.

This activity level production rate information can be used for several purposes. First, in case of delay claims, the fluctuation of the production rate over time can be studied to identify the exact times when production rates were lower than expected. Second, this production rate can be compared with historical production rates from previous projects to assess the project team's performance. Third, once the project is completed, this production rate may be used in calculating the historical production rate for the work item. Finally, for an ongoing work item, this production rate information can be used to realistically predict the time required (TR_j) to complete the remaining quantity of work (equation (24)).

$$\text{Time required to complete the activity} = \frac{Q_j - cq_{t,j}}{PR_j} \qquad (24)$$

Where Q_j is the total bid quantity of the item.

Visualization of as-built schedules

Project level and activity level as-built schedules are visualized in this component using data obtained from the data processing component. First, the work item, start date, finish date vector (w_j, sd_j, fd_j) for all work items are sorted in ascending order by start date (sd_j). If multiple work items have the same start date, the items are further sorted by the finish date (fd_j) in ascending order so that an activity completed first comes first in the list. A project level as-built schedule is developed by plotting this sorted data for all work items in a bar chart. For the activity level as-built schedules, the pair of time and cumulative quantity (t, cq_t) for a specific activity is plotted as a cumulative quantity chart.

Prototype Development

A prototype, namely, As-Built Schedule System (ABSS) is developed with MS Access database (Figure 22) and a Visual C#.NET frontend (Figure 23) to implement the framework and automate the entire computation process. Four data tables are created using

Entity-Relations Model (ERM) to optimize the database (Stephens 2010). The table 'a_contract' contains a list of contracts, 'a_contract_items' contains the list of work items for a given contract, 'a_itm_master' contains the standard specification of the work item, and 'a_work_items' contains DWR data for each work item.

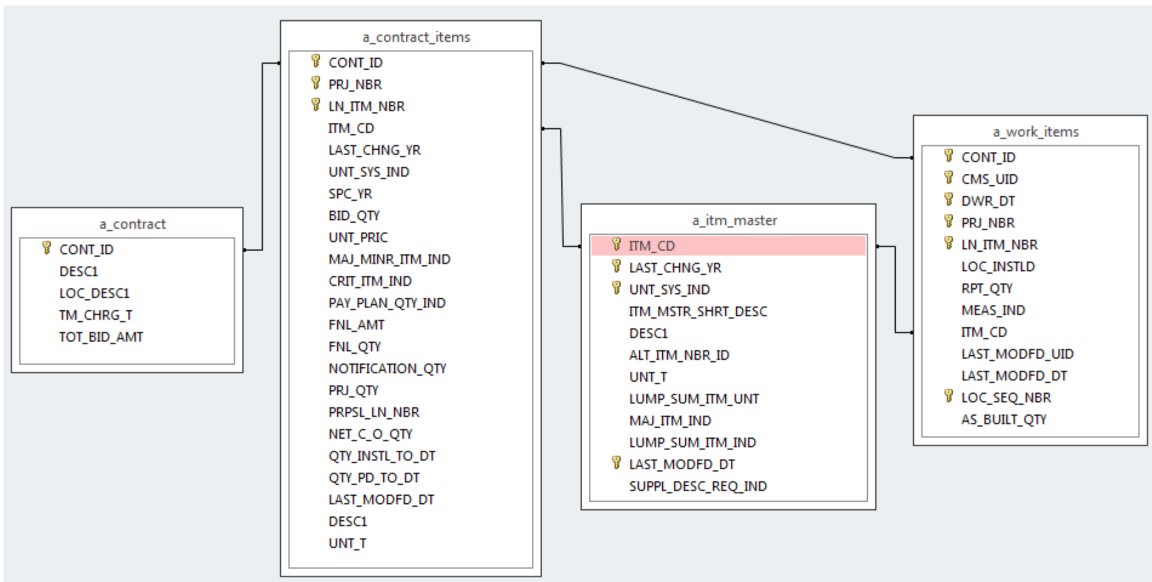

Figure 22 MS Access database of ABSS

Structured Query Language (SQL) commands are used to process the data to generate relevant information using a Graphical User Interface (GUI) or the frontend. SQL is a set-oriented programming language that is designed to allow people to query and update tables of information (IBM 2013). Table 9 presents important data attributes and their descriptions.

The SQL code presented in Figure 23 logically binds data attributes from three data tables to generate a dataset required for a project level as-built schedule. This list is generated for a particular project with project ID that equals to 'pid'. The output includes line item number, item code, item description, start date, and finish date, bid quantity, and actual reported quantity to date.

Table 9 Descriptions of important data attributes for as-built schedule development

Code	Name	Description
prj_nbr	Project number	A unique identifier for the project.
ln_itm_nbr	Line item number	A unique project specific code that indicates a particular work item.
itm_cd	Item code	An agency defined universal code used to identify a particular work item.
itm_mstr_shrt_desc	Item master short description	Short textual explanation of the work item.
bid_qty	Bid quantity	Total quantity of work for the work item
rpt_qty	Reported quantity	Quantity of work completed in a particular date
dwr_dt	DWR date	The date the Daily work report was created.
last_chng_yr	Last change year	A system-generated key which indicates the last time specification was changed.
spc_yr	Specification year	The year of the specification book in which the work item appears.

```
SELECT DISTINCT *
FROM (SELECT * FROM
    (SELECT c.LN_ITM_NBR, c.ITM_CD, c.BID_QTY, c.PRJ_QTY, c.QTY_INSTL_TO_DT,
        c.QTY_PD_TO_DT, c.CONT_ID, c.PRJ_NBR, c.SPC_YR
        FROM a_contract_items c
        WHERE c.PRJ_NBR='"+pid+ @"') T1
    RIGHT JOIN
    (SELECT w.LN_ITM_NBR, Min(w.DWR_DT) AS START_DATE, Max(w.DWR_DT) AS END_DATE,
        Count(w.DWR_DT) AS DWR_COUNT, Sum(w.RPT_QTY) AS REPORTED_QUANTITY
        FROM a_work_items w
        WHERE w.PRJ_NBR='" + pid + @"'
        GROUP BY w.LN_ITM_NBR) T2
    ON T1.LN_ITM_NBR =  T2.LN_ITM_NBR)  AS T3 LEFT JOIN (SELECT i.ITM_CD,
        i.LAST_CHNG_YR, i.DESC1, i.UNT_T
    FROM a_itm_master i)  AS T4 ON (T3.SPC_YR = T4.LAST_CHNG_YR) AND (T3.ITM_CD = T4.ITM_CD)
ORDER BY START_DATE DESC, END_DATE DESC
```

Figure 23 SQL code to extract project level as-built schedule information

Figure 24 presents another SQL code to generate quantity of work to date which is processed further to generate the pair of time, and cumulative quantity (t, cq_t).

```
SELECT DWR_DT, SUM(RPT_QTY) AS DAILY_QUANTITY
    FROM a_work_items
    WHERE PRJ_NBR='" + pid + "' AND (LN_ITM_NBR)='" + cLI + @"'
    GROUP BY DWR_DT
    ORDER BY DWR_DT ASC
```

Figure 24 SQL code to extract activity level as-built schedule information

The frontend of ABSS provides a tabular view on the left and a project level as-built bar chart schedule on the right (Figure 25). An additional window is displayed to show the activity level as-built schedule. The system flow can be explained with four major steps as indicated by numbers 1 to 4 in Figure 9. Users can 1) select a year and 2) click the contracts and projects from the selected year to load a list of projects on the tabular view. On double clicking any of the row in the tabular view, its project ID will be copied to the project ID field in the bottom (3). Alternately, a user can directly enter a project ID in the text entry field (3). After that, the user can click on the 4) generate as-built schedule button to generate and load the as-built information on the left. It also generates a bar chart on the right side in a similar fashion as MS Project and Oracle Primavera. Details such as bid item quantity, start date, end date, and the production rates of an activity can be observed by placing the cursor on top of the desired activity. Finally, double clicking in the desired bar in the chart generates an activity level as-built schedule for that work item.

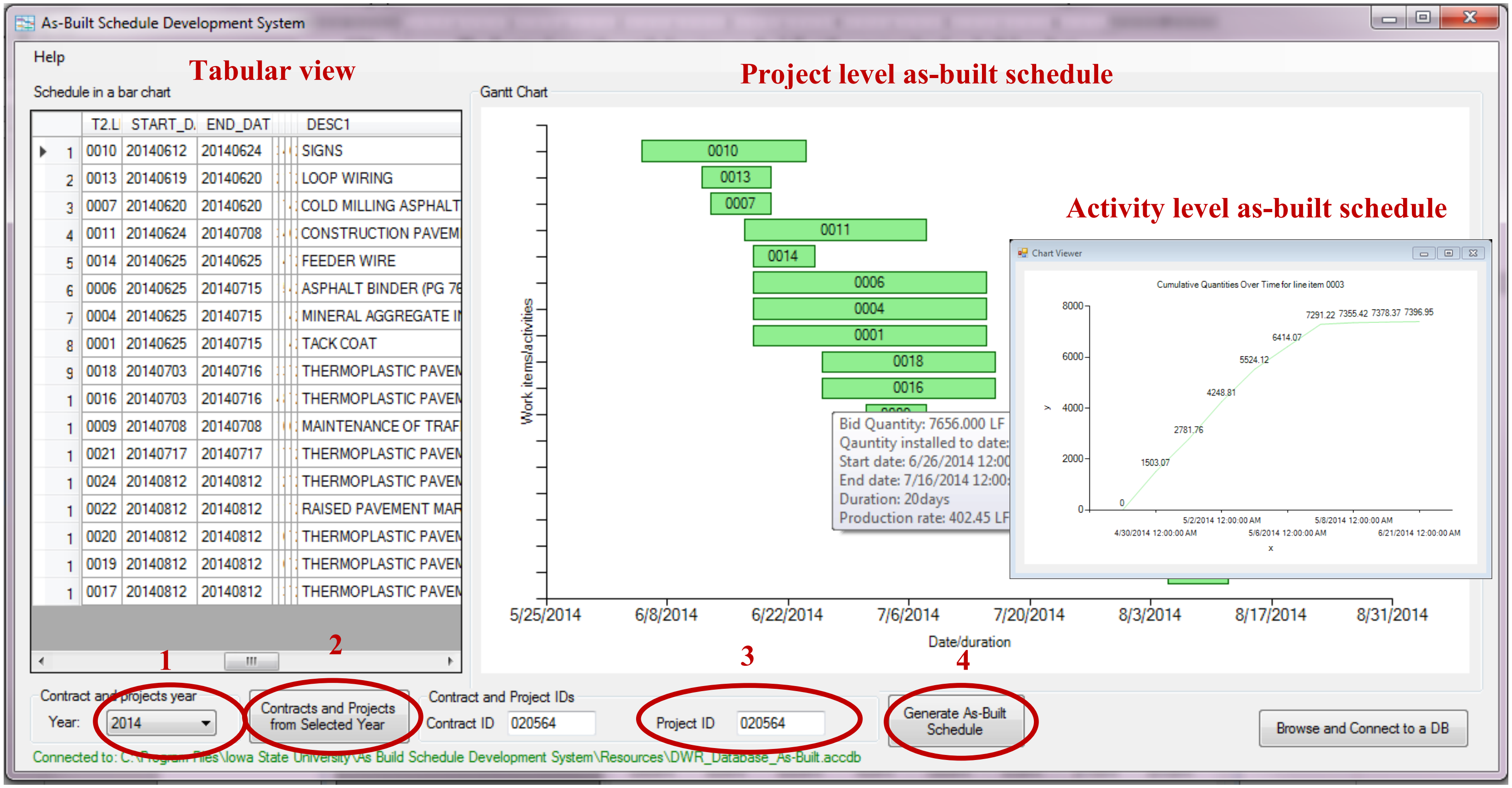

Figure 25 Visual C#.NET frontend of ABSS

Demonstration of ABSS and Discussions

A DWR data extracted from AASHTOWare SiteManager is collected from a SHA (anonymous) that includes data of 3,017 projects let from 2001 to 2014. It contains 646,488 DWR records that show the quantities of various activities for completed projects during the period. A cable median barrier installation project with a contract value of $3,574,783.47 is selected from the obtained database is selected and used to demonstrate ABSS and the as-built schedules developed from ABSS. Cable median barriers are safety barriers installed on the median to reduce crossover crashes.

Project level as-built

The bar chart in Figure 10 shows the as-built schedule successfully developed through ABSS using the DWR data for the project. Based on the as-built schedule, aggregate base course installation (004) was the most time consuming activity. A portable changeable message (011) was installed in the early phase of construction to alert road users about the ongoing construction activity. Other work items conducted in the early phase included installation of signs, traffic drums, and advance warning arrow panel. In some areas, additional traffic maintenance was performed (007). Various work items associated with temporary seeding, water, much cover, etc. were performed as closing activities.

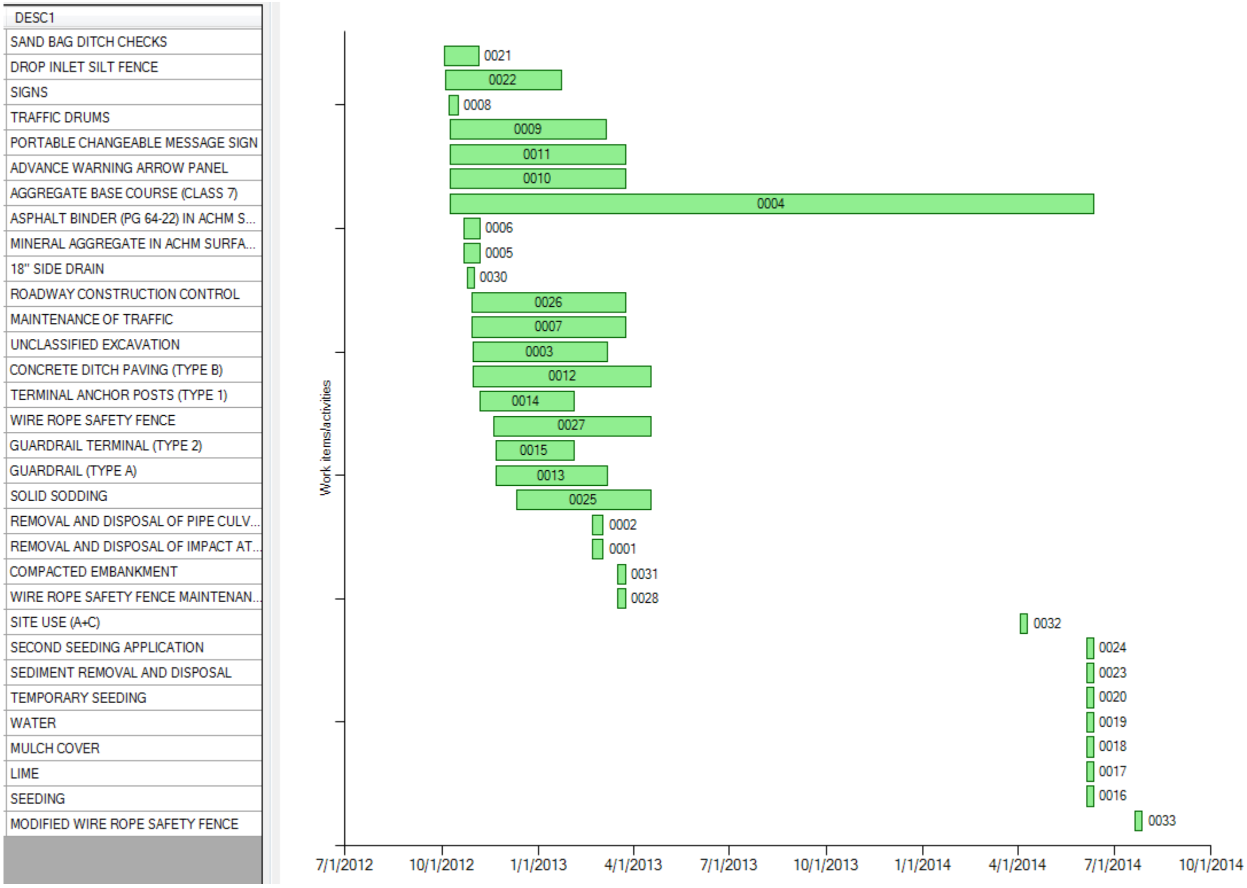

Figure 26 Project level as-built schedule example

Item level as-built schedule

Item number 0012: Concrete ditch paving (Type B) is used to generate an activity level as-built schedule and explored in detail (Figure 27). The actual cumulative quantity curve (solid line) does not align closely with the constant productivity line (dashed line) which shows that the productivity of the item was not consistent throughout the project. Based on this as-built schedule, it can be seen that the work was paused for some time from December of 2012 to January of 2013 which resulted in the uneven production rates. This is most likely because of the winter holidays. The work was also paused after February 2013 until mid-April 2013 probably due to inspection and final approval. If those two pauses are not considered, the production rate appears to be relatively steady throughout the project.

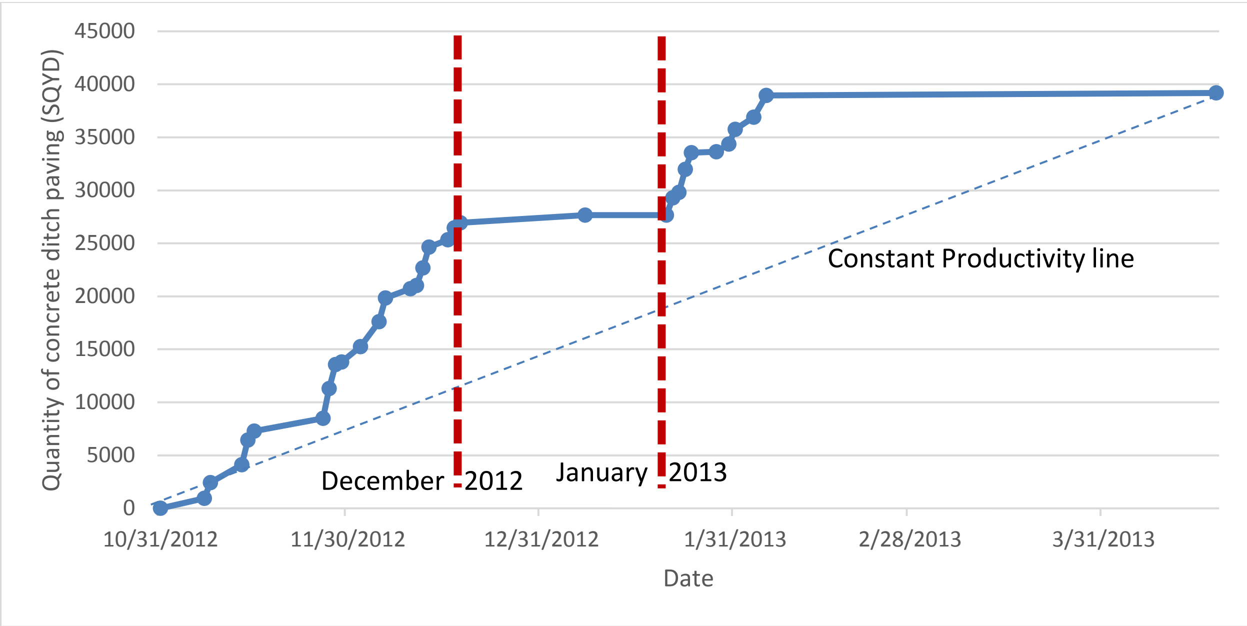

Figure 27 Activity level as-built schedule example

This demonstration shows that the ABSS can successfully develop both project level as-built schedule and activity level as-built schedule and it provides ideas that these visualized as-built schedules can empower project schedulers and project managers in the future. These as-built schedules can serve as an important tool to ensure that a project is progressing in desired pace to complete it within the given contracted time and take corrective actions in a timely manner. In addition, automatically generated final as-built schedules can serve as evidence to defend for possible legal disputes. Inexperienced schedulers can also use final as-built schedules of previous projects to understand the sequencing of the activities. An item level as-built schedule can be used to evaluate the performance of contractors over time as well. This study utilizes the first and last record date for each activity as its start and finish date. The activity may be conducted intermittently between those dates as shown in Figure 11. These gaps are currently neglected in the project level as-built schedule. However, the gaps can be observed in the activity level as-built schedule.

Conclusions

The importance of as-built schedules is widely recognized, but there is a lack of a computational methodology to generate as-built schedule information from existing data. This study developed a computational algorithm to automatically extract Daily Work Report data and visualize project level and activity level as-built schedules. A prototype- ABSS is developed to automate the entire computational process of the algorithm and is used to demonstrate its ability to generate both project level and activity level as-built schedules. The DWR data of a construction project in progress can also be used to generate as-built schedules in real-time (as-built schedules to date). The as-built schedules can be used to monitor the progress with higher granularity, identify schedule deviations such as delays and accelerations, evaluate project performance, anticipate the construction completion date, take corrective actions to complete the project on time, and aid in settling claims and disputes. Further, inexperienced schedulers can learn by studying the historical as-built schedules to learn activity sequences to prepare a schedule for a future project with similar characteristics.

As future studies, the historical as-built schedules can be used as a database of construction activity sequencing and productivity analysis. Using a proper sequential pattern mining technique, construction activity sequences can be extracted to be used for as-planned schedule generation for future projects.

Acknowledgements

This study was conducted as an extension of the study funded by Mid-American Transportation Center (MATC). The authors would like to acknowledge MATC for funding the study that prompted the need of this study. We would also like to thank a state highway agency's representative for providing the DWR dataset for the analysis.

CHAPTER 5

DISCOVERING PRECEDENCE RELATIONSHIPS OF ACTIVITIES USING SEQUENTIAL PATTERN MINING TO SUPPORT SCHEDULE DEVELOPMENT

K. Joseph Shrestha[1] and H. David Jeong[2]

Abstract

A realistic schedule must be developed before the start of a project to ensure its successful execution. The sequencing of construction activities is one of the most important, challenging, and complex parts of any project schedule development. It requires a significant amount of field experience and knowledge of a scheduler on construction processes, construction means and methods, production rates of activities, site logistics, and resource allocation. The knowledge of precedence relationships of activities obtained from historical projects would empower schedulers in gaining confidence in developing a new project's schedule and it would help an inexperience scheduler develop a schedule with strong evidence. Recently, many project owners, such as state highway agencies have started to use digital daily work reports that contain data about each activity's detailed progress throughout the project duration. This rich activity level data provides a great potential to discover the precedence relationships of activities when a proper data extraction and analysis method is applied. This study builds a computational algorithm to extract schedule related data of activities from a digital daily work report system,

---

[1] PhD Candidate, Iowa State University, Dept. of Civil, Construction & Environmental Engineering, Iowa State University, Ames, IA 50011, shrestha@iastate.edu
[2] Associate Professor, Dept. of Civil, Construction & Environmental Engineering, Iowa State University, Ames, IA 50011, djeong@iastate.edu (515) 294-7271

transform the extracted data into a format suitable for applying a Sequential Pattern Mining (SPM) algorithm to generate an extensive list of sequential rules or precedence relationships of activities. Those precedence relationships can form a rich reference knowledge base to help schedule a new project. The algorithm that was developed in this study was applied to a real daily work data obtained from a state highway agency. The daily work system contains over 2,000 highway projects from 2001 to 2014. The algorithm has successfully discovered 12,643 precedence relationships

Key Words: precedence-relation, sequential-pattern-mining, schedule-overrun, daily-work-report, field-data, activity-sequencing, schedule, big-data, automation, data-analytics, visualization.

Introduction

Schedule development is a vital part of construction planning and delivery (Cherneff et al. 1991; Douglas 2009; Echeverry et al. 1991; Fischer and Aalami 1996). Schedules aid in communicating and coordinating activities among the construction stakeholders (Cashman and Tayam 2010). An optimized schedule enables the contractors and owners to complete a project on time with the minimum resources (Cashman and Tayam 2010; Jaśkowski and Sobotka 2006). Once construction starts, the project schedule can be used to systematically track construction progresses and identify any delays (Cashman and Tayam 2010; Contreras and Van De Werken 2005). A schedule is also a valuable tool to analyze and quantify the potential impact of delays on the overall project schedule (Henschel and Hildreth 2007).

Developing a realistic schedule is challenging for inexperienced as well as experienced schedulers (Fischer and Aalami 1996). It is a complex process and requires knowledge of construction methods, materials, and labor productivity (Bruce et al. 2012). An as-planned schedule development typically consists of three steps: a) identification of activities through work breakdown structure development for the given project, b) determination of the duration of each activity, and c) determination of logical and realistic sequence of activities (Clough and Sears 1991; Fischer and Aalami 1996). Many academic studies have focused on the first two steps; a) identifying construction activities and b) determining activity durations using historical and reliable production rates to enhance the accuracy of schedule development (Kim et al. 2013; Woldesenbet et al. 2012). However, on determining the sequence of activities, many studies pointed out heavy reliance on the experience and knowledge of experienced schedulers as the single most important source (Bruce et al. 2012; Jeong et al. 2009).

In the highway industry sector, specifically, many owners such as state highway agencies (SHAs) have developed a set of scheduling templates that store a typical sequence of activities for a specific type of project to facilitate their scheduling activities. This approach helps capture and take advantage of experience schedulers' knowledge even after their retirement. However, such approach has as least three major limitations. First, it relies heavily on the experience of a scheduler to develop such templates. Second, various types of projects may have different construction sequences. As such, multiple templates will need to be developed manually—one for each project type. Third, construction means and methods may evolve over time and such static sequencing templates may become outdated.

Thus, in addition to these schedule templates, if there is a systematic approach to discovering the knowledge of precedence relationships of activities from historical projects and having the knowledge available for schedulers, it would greatly empower schedulers in gaining confidence in developing a new project's schedule and also help an inexperience scheduler to develop a schedule with strong evidence.

More than 37 SHAs have started to use digital DWRs which contain rich project progress and performance data at each activity level. This digital dataset can be directly used to discover various sequences of construction activities when an appropriate data analytics is employed. This study applies a powerful Sequential Pattern Mining (SPM)) algorithm to easily available historical Daily Work Report (DWR) data that contains historical project schedule information to discover precedence relationships of activities.

Prior Studies

Prior studies suggest that schedules are mostly developed manually (Kim et al. 2013). Manual inputs are needed especially in activity sequence development. Existing studies on generating activity sequences are focused on utilizing the expertise of experienced schedulers or logical assumptions about construction activity sequences. For example, Jeong et al. (2009) developed 14 different highway scheduling templates based on Oklahoma Department of Transportation (DOT) schedulers' experience. Bruce et al. (2012) utilized a list of controlling activities and developed templates based on experts' inputs for 12 types of road and bridge construction projects. They also studied the resident engineers' project diaries and other project documentations to develop the templates.

In the vertical construction, some studies are conducted to improve scheduling practices that can be conceptually adopted to the horizontal construction. Cherneff et al. (1991) developed a systematic approach to generating activity sequences by assigning various component-constraints. An example of such component constraint is that a door can only exist in a wall. As such, wall must be constructed before a door is installed. Echeverry et al. (1991) used the similar approach with four types of logical assumptions: a) physical relationships between building components, b) interaction of construction trades, c) interference-free path for the movement of construction equipment and materials, and d) safety considerations. Fischer and Aalami (1996) followed the Echerverry et al.'s (1991) method by generating component-constrained and activity-constrained relations. Component constraints are physical constraints based on the construction components whereas activity constraints are based on activity types. In a recent study by Kim et al. (2013), identification of construction activities is improved by utilizing Building Information Models (BIMs), but, activity sequencing is still based on a set of static sequencing templates similar to the previous studies.

Thus, existing studies have a limitation of being static and being dependent on the knowledge of experienced schedulers. The next two sections provide discussions on the DWR data and SPM algorithm.

Daily Work Report Data

SHAs collect a significant amount of data such as ongoing construction activities, labor hours, types of equipment used, equipment hours, weather data, and significant communications with contractors in a Daily Work Report (DWR) system (Shrestha et al. 2015). Site inspectors and resident engineers spend as much as 40% time in collecting those data (McCullouch 1997). Currently, 37 SHAs in the U.S. are using digital DWR systems. Figure 19 shows a screenshot of the AASHTOWare SiteManager which is the most popular DWR system among SHAs.

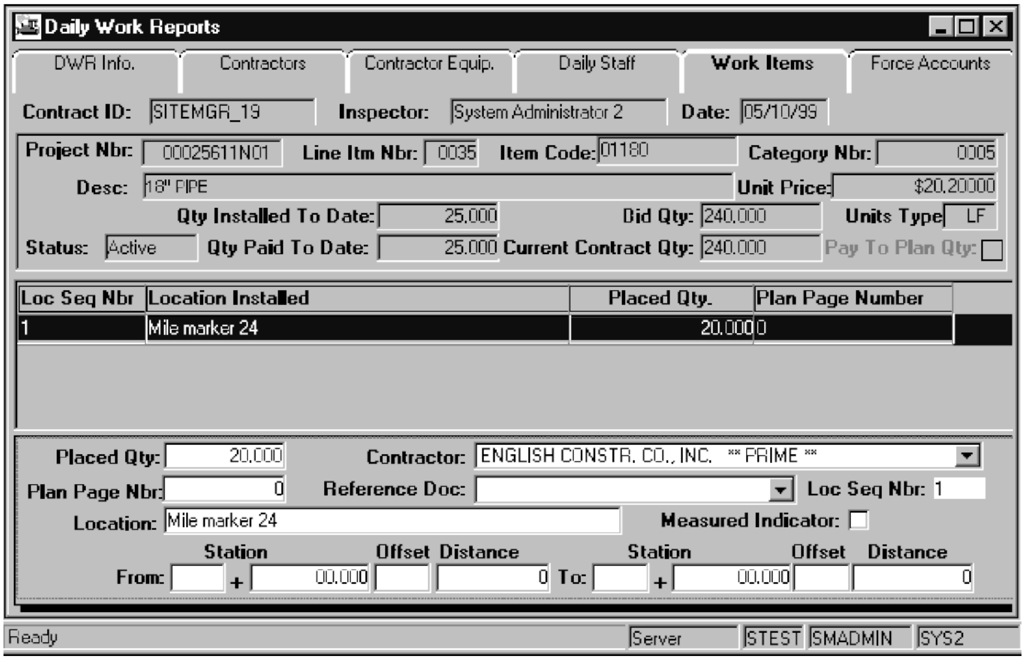

Figure 28 AASHTOWare SiteManager screenshot

DWR systems have been developed and used with the main objective of making correct payment to contractors and documenting field activity records as preparation for potential claims and disputes. Moreover, the data attributes recorded in the DWR system have potential to be utilized for other purposes such as analysis of activity sequencing, as-built schedule development, production rate and activity cost estimation, contract time determination, and contractors' performance evaluation (Shrestha et al. 2015). However, most SHAs have not benefited from those potential applications possibly because of the lack of knowledge on those potential benefits, enabling methodologies, and automation processes.

DWR data attributes are typically linked to each work item. In the U.S. highway industry, SHAs have developed an extensive list of work items that are primarily developed to facilitate the bidding process under unit cost contracting mechanism. Those work items are also used to develop a project schedule since they are typically independent work activities. SHAs have developed and maintained specifications that provide a detailed description of each work item. For example, an item code '01180' in Figure 3 indicates a work item "supply and installation of a mile marker." A typical set of data attributes collected in DWR systems can be classified into six categories: general information, work activities, weather information, equipment, labor, and remarks (Figure 29).

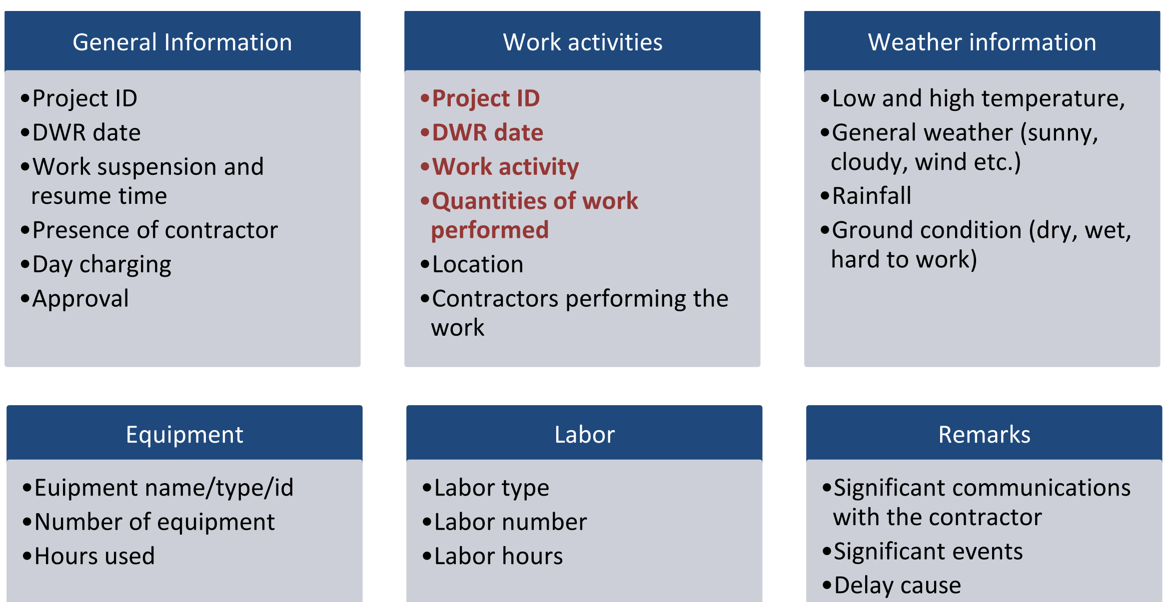

Figure 29 Typical data attributes collected in DWR systems

Among these six categories, it is important to note that the category of work activities contains directly relevant and sufficient data needed for this study. The DWR data and work activity can be used to extract start dates of each activity. Then, the start dates of various activities can be compared to identify the activity sequences for all historical projects. Those activity sequences can be identified using Sequential Pattern Mining (SPM).

Sequential Pattern Mining

A SPM identifies the hidden patterns from a set of sequential data that can be used to predict sequences for additional datasets (Masseglia et al. 2009). SPM algorithms have been used for DNA sequencing, medical treatment, consumer behavior, web access pattern, and stock market (Li et al. 2005; Masseglia et al. 2009; Wang 2005). For example, if a consumer buys a cell phone on an ecommerce website, a case for the phone can be recommended to the consumer, as the consumer is likely to purchase a cellphone case based on the SPM analysis of histories of other consumers who bought cell phones. If the available historical dataset gets larger, the accuracy of finding useful hidden sequences gets better.

In identifying the sequences of construction activities, as-built schedules of historical projects can be analyzed using a SPM algorithm to help determine the sequences of activities for a new project. The Sequential Rules Common to Several Sequences (CMRules) is an open-source algorithm that will be used for this study (Fournier-Viger et al. 2012).

Sequential rules common to several sequences

The CMRules can be used to identify the sequential pattern hidden in a Sequential Database (SD). The CMRules algorithm finds the sequential patterns that appear frequently in a sequence database and meets a minimum threshold value of confidence and support.

Consider a SD = $\{s_1, s_2, s_3 \ldots s_n\}$ and a set of items (or work activities) I = $\{i_1, i_2, i_3 \ldots. i_n\}$ where each sequence s_x is an ordered list of transactions (or work activities in a given project in case of scheduling) $s_x = \{X_1, X_2, X_3 \ldots., X_n\}$ such that

$$X_1, X_2, X_3 \ldots., X_n \subseteq I. \tag{25}$$

A sequential rule X⇒Y (X is followed by Y) is a relationship between two item sets X, Y such that:

$$X, Y \subseteq I \text{ and } X \cap Y = \emptyset. \tag{26}$$

The sequential support of a rule X⇒Y can be defined mathematically as:

$$SeqSup(X \Rightarrow Y) = Sup(X \blacksquare Y) / |SD|. \tag{27}$$

The sequential confidence of a rule X⇒Y can be defined mathematically as:

$$seqConf(X \Rightarrow Y) = sup(X \blacksquare Y) / sup(X). \tag{28}$$

Here, sup(X■Y) denotes the number of sequences from a sequence database where all items of X appear before all items of Y. |SD| denotes the number of sequences in the SD and sup(X) denotes the number of sequences that has X. The minimum support and confidence are set in the algorithm to ensure that the sequences identified by the algorithm occur frequently in the SD as well as the subset of SD containing the predictor item sets (X_i).

In the next section, a framework to develop a dynamic list of precedence relationships of activities by applying CMRules on DWR data is discussed.

Framework

The framework consists of six components: a) database preparation, b) project type selection c) data extraction, d) data transformation, e) application of CMRules, and f) visualization of activity precedence relations. First, DWR data is obtained from a current DWR system. As different types of projects may have different project sequencing, a desired project type and relevant DWR data are selected. From selected projects, start date information of each activity is extracted for each project. This data is then transformed to a format suitable for applying CMRules algorithm. CMRules identify and generate precedence relationship found in the DWR data. The precedence relationships of activities are then visualized to enable extraction of activity sequences for a new project. This precedence relations and the diagram becomes a knowledge base for extracting activity sequences for new projects.

Database preparation

In this component, historical DWR data are obtained from existing DWR systems such as AASHTOWare SiteManager, AASHTOWare FieldManager, MATS, Next Generations, and Field Operations. At minimum, the database should contain project type, DWR date, and work activity conducted in each DWR recording date.

Consider a DWR dataset (D) consisting of 'n' number of elements (equation (29)). An uppercase letter denotes a set while the corresponding lowercase letter denotes its elements and subscripts indicate the element numbers. Each element of this DWR dataset (d_i) is a vector of project type (y_i), project ID (p_i), DWR date (t_i), and work activity code (w_i) (equation (30)).

$$D = \{d_1, d_2, d_3, \ldots, d_n\} \tag{29}$$

$$d_i = (y_i, p_i, t_i, w_i) \tag{30}$$

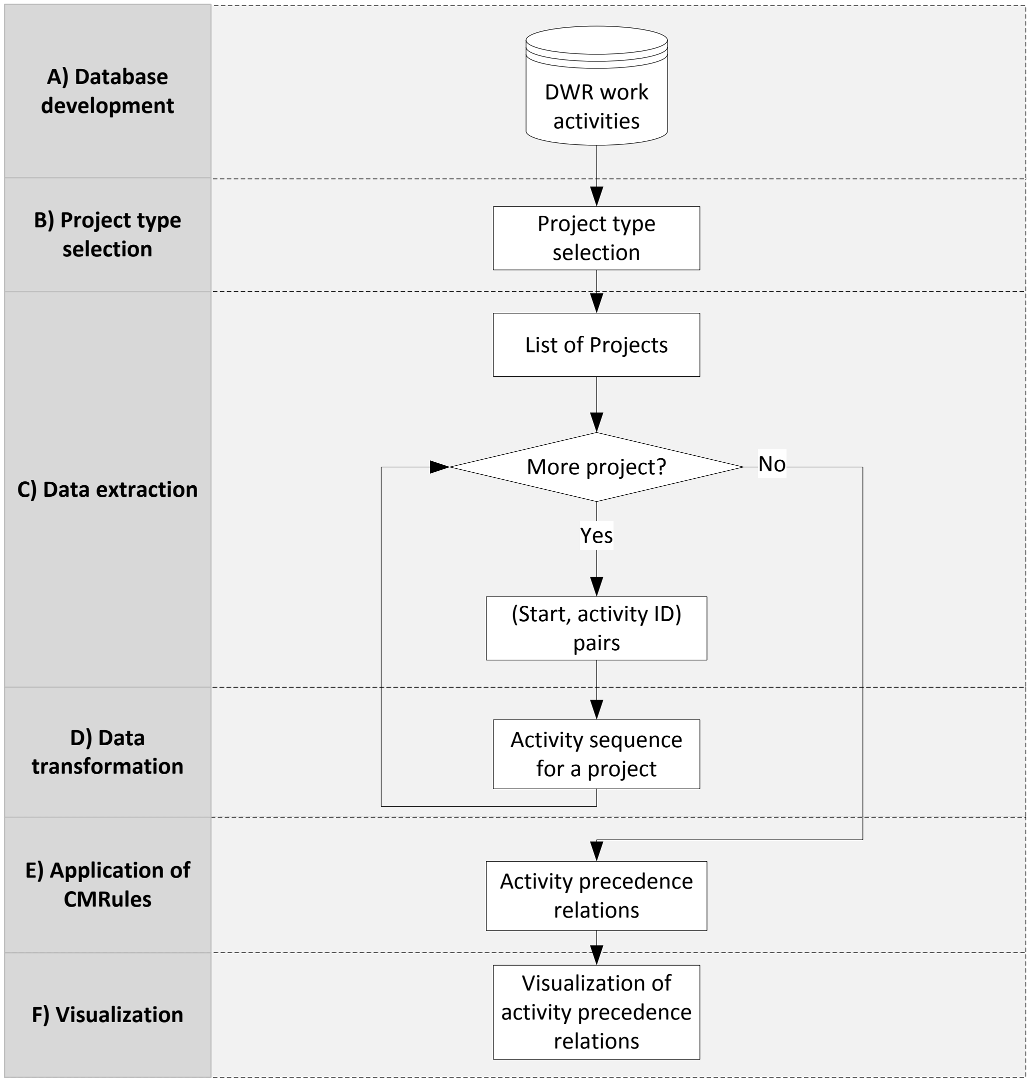

Figure 30 Framework to generate activity precedence knowledge base

Project type selection

A project type such as a ‘roadway widening’ is likely to have a different sequence of activities from that of other project types such as a ‘new roadway construction’. Thus, DWR data of only one project type should be obtained for further process. Mathematically, a subset of

DWR data ($SS \subseteq D$) corresponding to a specific project type (y_i) is extracted for further analysis and can be represented by equation (31).

$$d_i \in SS \mid d_i(y) = y_i \tag{31}$$

Where, d_i (y) represents the project type of a particular DWR record and the vertical bar symbol '|' indicates 'such that' or 'conditional' statement.

Data extraction

The elements of this new set (SS) are divided into several subsets (WS_j)–one for each work activity (equation (32) and (33)). Thus, if there are 'm' numbers of work items in a project, there will be 'm' numbers of subsets.

$$SS = WS_1 \cup WS_2 \cup WS_3 \cup \ldots \cup WS_m \tag{32}$$

$$ss_i \in WS_j \mid ss_i(w) = w_j \tag{33}$$

Finally, only one element of each WS_j that has the earliest DWR date is selected to obtain our final extracted dataset (ED). This earliest date is considered the start date of the current work activity.

$$ed_i = (y_i, p_i, et_i, w_i) \tag{34}$$

Where, et_i is the earlies date of activity w_i in project p_i.

Data transformation

To apply CMRules, the final extracted dataset from the previous component must be transformed to a sequential database (SD) in which each element represents a sequences of all activities in a project. In order to accomplish this data transformation, first, work activities are selected and sorted in an ascending order by their start dates for each project. Then, a SD is defined as a set containing such sequences from all projects as an element of the SD (equations (35) and (36)).

$$SD = \{sd_1, sd_2, sd_3, \ldots., sd_p) \tag{35}$$

$$sd_i = (w_1, w_2, w_3, \ldots, w_q) \tag{36}$$

Where ‘p’ is the number of projects of the selected project type. Each elements of the sequential database, sd_i, represents a sequence of ‘q’ number of activities (w) for a particular project. The number of activities (q) varies by the project.

Application of CMRules

The CMRules is used to analyze the SD generated from the previous component. The CMRules will generate a list of sequential rules with corresponding support and confidence as:

$$(preceding\ activity\ set\ (X) \Rightarrow succeeding\ activity\ set(Y)), \quad SeqSup\ (count), \quad seqConf\ (\%)$$

Here, sequential support is expressed only in terms of the count (Sup(X■Y)) as all sequences have the same denominator (|SD|) for a given database. The sequential confidence has a different denominator for each rule (sup|X|) depending on the preceding activity set. As such, it is expressed as a decimal percentage.

Sequential rules can consist of one to one rules (e.g. $w_1 \Rightarrow w_2$), one to many (e.g. $w_1 \Rightarrow w_2, w_3$), many to one (e.g. $w_1, w_2 \Rightarrow w_3$), and many to many (e.g. $w_1, w_2 \Rightarrow w_3, w_4$). Different threshold values of support and confidence can be set to generate a smaller or a larger list of sequential rules as desired. The higher the values of confidence or support are, the smaller the list of sequential rules is. Moreover, if various projects have diverse activity sequences (e.g. sequences of two activities are reversed in different projects), fewer sequential rules will be obtained. Thus, to ensure that sufficient sequential rules are obtained that contains all sequential rules required to develop activity precedence diagram for a new project, varying level of support and confidence need to be experimented with.

Visualization of activity precedence relations

One to one sequential rules generated from the previous component can be visualized in a chart. Such visualization becomes an easy tool to understand and illustrate activity sequences.

Prototype

The framework is semi-automated using a prototype that uses MS Access database, Structured Query Language (SQL) queries, MS Excel, SPMF, and Gephi. First, a MS Access database consisting of required data attributes is prepared. An Entity Relations (ER) model of the database is presented in Figure 31.

Figure 31 DWR database for discovering precedence relationships of activities

Table 10 presents brief descriptions of important data attributes.

Table 10 Important data attributes to discover activity precedence relationships

Code	Name	Description
cont_id,	Contract ID	Primary identifier for a contract
prj_nbr	Project number	Unique identifier for a project
itm_cd	Item code	An agency defined code used to identify a particular work item.
itm_mstr_shrt_desc	Item master short description	Short textual explanation of the work item
dwr_dt	DWR date	Daily work report data collection date
last_chng_yr	Last change year	A system-generated key that represents the year the specification for a particular work item was last changed
spc_yr	Specification year	The year of the specification book in which the work item is based on

The project selection and data extraction is performed via a SQL query Figure 32. It logically binds four data tables ‘a_itm_master,’ ‘a_work_items,’ ‘a_contract_items,’ and ‘a_contract’ using data attributes ‘cont_id,’ ‘itm_cd,’ ‘spc_yr,’ and ‘last_chng_yr.’ It produces start date of each activity of all projects of project type ’04.’

```sql
SELECT a_work_items.cont_id AS Contract_ID
    ,a_work_items.prj_nbr AS Project_Number
    ,a_itm_master.itm_cd AS Item_Code
    ,a_itm_master.itm_mstr_shrt_desc AS Item_Description
    ,Min(a_work_items.dwr_dt) AS Start_Date
FROM (
    (
        a_itm_master INNER JOIN a_work_items ON a_itm_master.ITM_CD = a_work_items.ITM_CD
        ) INNER JOIN a_contract_items ON (a_itm_master.LAST_CHNG_YR = a_contract_items.SPC_YR)
        AND (a_work_items.ITM_CD = a_contract_items.ITM_CD)
        AND (a_work_items.CONT_ID = a_contract_items.CONT_ID)
    )
INNER JOIN a_contract ON a_contract_items.CONT_ID = a_contract.CONT_ID
WHERE (((a_contract.WRK_T) = "04"))
GROUP BY a_work_items.prj_nbr
    ,a_itm_master.itm_cd
    ,a_itm_master.itm_mstr_shrt_desc
    ,a_work_items.cont_id
    ,a_work_items.cont_id;
```

Figure 32 SQL Query to extract required dataset for CMRules

This extracted dataset is transformed to a format suitable to apply CMRules using MS Excel. Then, the CMRules is applied using SPMF (Fournier-Viger 2014). Finally, the sequential rules generated form the CMRules is visualized using Gephi (Bastin et al. 2009).

Validation

A DWR database is obtained from a SHA in the MS Access format. The database consists of project information of over 2,000 projects let from 2001 to 2014. Table 11 presents the top five project work type in terms of the total dollar amount. The project type–widening existing roadway, is the largest work type of all. This work type includes the addition of passing lanes to improve traffic flow and road safety conditions. It also has a large number of contracts (third largest out of 35 different work types in the database) and is selected for validating the framework.

Table 11 Distribution of contract costs and count by work type

Work type code	Work type	Total dollar amount	Number of contracts
04	Widening existing roadway	1,799,474,488	204
19	Structures and approaches	864,634,085	231
07	Overlay	806,625,107	1264
06	Rehabilitation	605,243,188	49
16	Grading and structures	482,557,003	42

The ‘widening existing roadway’ projects represent about $1.8 billion worth of projects. The CMRules algorithm is applied to the SD generated from the DWR data of the selected project type. Various values of minimum support and confidence are used to generated a varying number of sequential rules (Table 12).

Table 12 Number of sequences obtained from various settings for CMRules

Support	**Confidence**		
	0.9	**0.8**	**0.7**
0.6	2	9	9
0.5	2	220	643
0.4	290	6,412	12,643

As stated before, the higher values of minimum confidence and support result in fewer rules. When very high values are set for the minimum confidence (90%) and minimum support

(60%), only two rules are obtained. As the values are decreased to 70% for the confidence and 40% for the support, 12,643 rules are obtained from the same dataset. This knowledge base of 12,643 rules can be searched to identify the sequences between various activities and are used for the following discussions. In this study, this process of searching relevant activity sequences is semi-automated with MS Excel.

Two of the results obtained from the analysis are presented below–a simple one with two items (i.e. one-to-one relation) and a complex one with multiple items (many-to-many relation).

1. (603001 ⇒ 412001), 117, 0.78

This sequential rule indicates that out of 204 contracts, the activity 603001 (maintenance of traffic) starts before the activity 412001 (cold milling asphalt pavement) in 117 projects (support). In the remaining projects, either those activities are not included in the same project, or activity 412001 starts at the same time or before the activity 603001. The confidence of 0.78 shows that in 78% of the projects that include the preceding activity 603001, the succeeding activity follows the preceding activity. In the remaining 22% of the projects that include the preceding activity, either the succeeding activity is not included or it occurs before the preceding activity. Based on this sequential rule, if a new project includes those the two activities, the activity 412001 would be recommended to start before the activity 603001.

2. (210201,303107,604023 ⇒ 624001,719001,719101), 112, 0.78

This many-to-many relationship includes three preceding activities followed by three succeeding activities. The three preceding activities can start any time relative to each other; the three succeeding activities can also start anytime relative to each other. The three preceding activities are 210201 (unclassified excavation), 303107 (aggregate base course (class 7)), and

604023 (traffic drums); the three succeeding activities are 624001 (solid sodding), 719001 (thermoplastic pavement marking, white-100), and 719101 (thermoplastic pavement marking, yellow-100). The three preceding activities occurred before the succeeding activities in 112 projects. Further, in 78% of the projects with the preceding activities, the succeeding activities listed above started after the preceding activities.

Such many-to-many sequential rules provide additional insights as they group items that can occur simultaneously or in varying orders. For example, the rule does not explicitly indicate that activity 624001 to start before, after, or at the same time as activity 604023. Unless a one-to-one sequential rule is identified between those two activities, their sequences may vary depending on the project characteristics. For example, in some cases, 'traffic drums' may need to be installed before 'soil sodding.' In other cases, the 'traffic drums' may be necessary only near the final stretch of the construction and hence are installed after 'soil sodding.'

One-to-one sequences are visualized in a chart that can be used to visually extract the activity sequences for a new project (Figure 33). The chart shows the activities as nodes and sequences by arrows similar to an activity precedence diagram. For example, the first sequential rule discussed above about activity 603001 (maintenance of traffic) and activity 412001 (cold milling asphalt pavement) are highlighted in red in the Figure 33. The sequencing of two activities is indicated by the arrow connecting them (603001 ⇒ 412001). The widths of the arrow lines indicate the confidence of the sequence: the thicker the arrow line is, the higher the confidence is. Based on the activities of a new project, relevant portion of this chart can be extracted visually to develop activity precedence diagram for the project. Alternately, the knowledge base of precedence relationships of activities can be directly searched to develop activity precedence diagram.

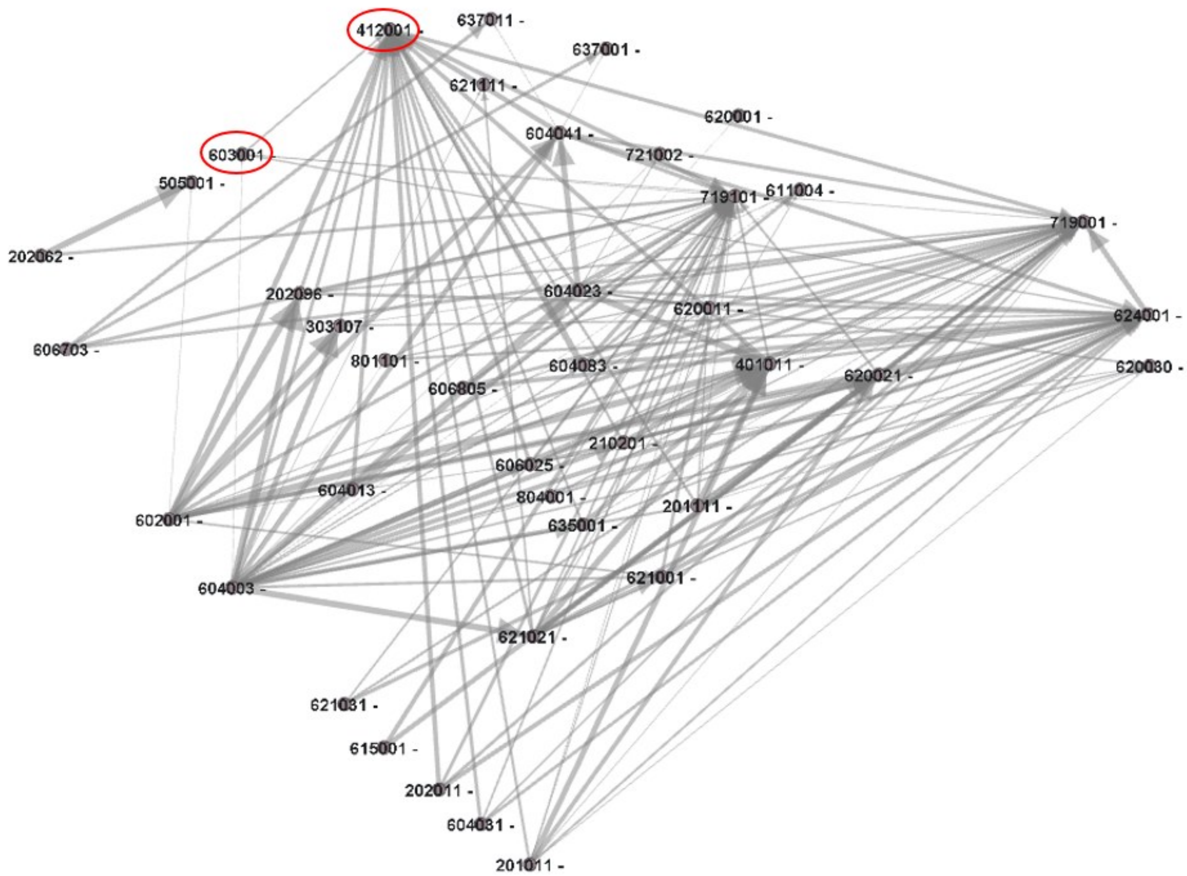

Figure 33 Visualization of precedence relationships of activities

Figure 34 shows construction sequences extracted for a hypothetical sample project consisting of seven pay item activities. The percentages in the arrow indicates the confidence that the activity on arrow head occurs after the activity on the arrow tail. The diagram presents a logical flow of the activities based on the predictive analysis. For example, traffic maintenance and unclassified excavation starts only after installation of work zone signs begins. Similarly, track coat is should be applied only after grubbing and unclassified excavations are starts. This diagram contains some redundant relationships as well: dashed arrow lines indicate such relations. For example, aggregate base course installation starts after unclassified excavation and track coat starts installation starts after aggregate base course starts. Thus, it automatically indicates that track coat installation starts after unclassified excavation starts. Such redundant relationships are removed in Figure 35.

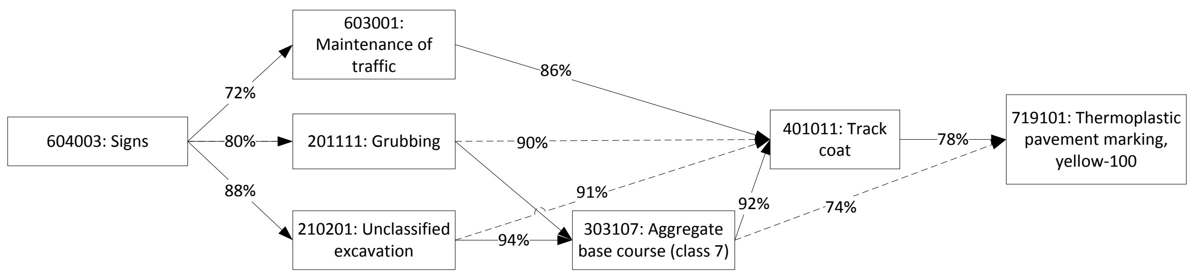

Figure 34 Preliminary activity precedence diagram including redundant relations

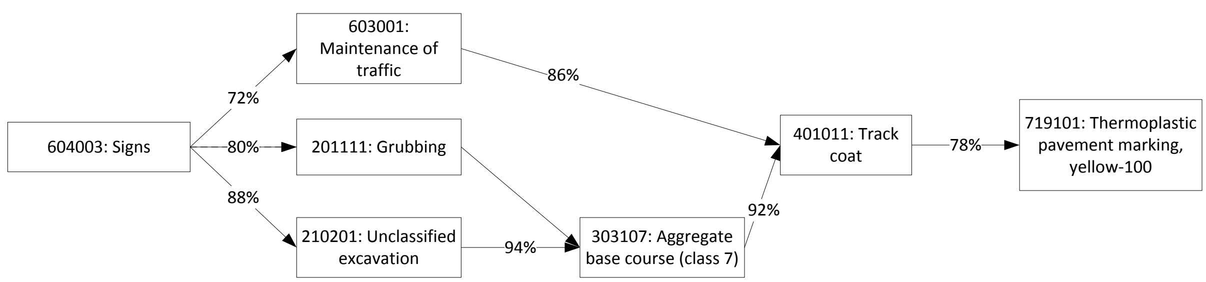

Figure 35 Final activity precedence diagram without redundant relations

Discussions

This study developed a framework to utilize DWR data to develop a knowledge base of precedence relationships of activities that can aid schedulers to develop schedules for future projects. The CMRules was applied to develop the knowledge base by using different values of minimum confidence and support. Over 12,000 rules were generated by analyzing the data with 70% confidence and 40% support. Such rules can answer questions such as, “which of the two activities in this new project should start first based on the historical records?” Further, the knowledge base will be updated dynamically to reflect the changes in construction methods and techniques as project data from new projects are added for the analysis.

Currently, sequences between every pairs of the activities need to be searched and then redundant relationships need to be removed manually. However, such process can be automated with C#.NET system. Such system will enable automation of not only the discovery of precedence relationships of activities, but also the use of the knowledge base to automatically suggest activity precedence diagram for future projects.

This framework will enable the inexperienced and experienced schedulers to develop activity precedence diagram with their limited inputs. It will save time and resources for the schedulers and have potential to develop more realistic construction schedules than traditional experience based methods as it is based on the historical data.

Limitations and Future Research Work

The study developed a methodology to produce activity precedence diagram based on DWR data collected by SHAs. The knowledge base of precedence relationships of activities can be used to identify which of the two activities of interests should start first. While a scheduler can make an assumption based on the activity grouping in the many-to-many sequential rules that two activities should start together instead of one after another, it does not confirm such relation. Future studies should focus on developing frameworks to enable such relations.

Finally, existing measures of certainty of the sequencing rules are not sufficient to describe the reliability of such rules for this study. The measure of support and confidence computes the certainty of the sequencing rules with respect to all the records and with respect to the records containing the precedent activities, respectively. This may result in an erroneous interpretation of results, as the values of the support and confidence may be low when measured against such parameters (all records). The confidence levels indicates with respect to the number of records that has both precedent activity set and succeeding activity set would be a more reliable indicator of confidence of the sequential rules. Such indicator of confidence will indicate whether an activity 'X' starts before an activity 'Y' or vice versa. The remaining confidence value (100% - confidence of (X■Y) – confidence of (Y■X)) will indicate the confidence that those activities starting together.

Conclusions

While some industries have heavily utilized their data to obtain data-drive insights, the construction industry is lagging behind. This study developed a methodology to discover precedence relationships of activities based on daily work report data collected by State Highway Agencies (SHAs). Currently, state SHAs and contractors rely completely on engineer's experience to develop activity precedence diagram, which is very time consuming and complex. In this study, a novel framework based on the Sequential Pattern Mining (SPM) is developed to automate such process. This framework will aid schedulers in quickly identifying precedence relationships of activities, which is the most complex part of schedule development. Thus, it provides a guidance to the schedulers based on the historical data, minimizes their inputs, and improves their confidence.

A dataset from a SHA was analyzed to validate the framework. The SPM generated 12,643 sequential rules that becomes a knowledge base to generate activity precedence diagram for new projects. A hypothetical project consisting of seven activities was used to test the framework. The framework successfully identified the sequential relationships between various activities of the project which are used to develop an activity precedence diagram for the project.

Overall, the framework will improve the utilization of data collected in the construction industry to improve current scheduling practices.

Acknowledgements

This study was conducted as an extension of the study funded by Mid-American Transportation Center (MATC). The authors would like to acknowledge MATC for funding the study that prompted the need of this study. We would also like to thank a state highway agency's representative for providing the DWR dataset for the analysis.

CHAPTER 6

CONCLUSIONS

While many industries such as health care, energy, and agriculture have utilized their digital data collected from various sources to make data-driven decisions to improve their business processes, the construction industry has been significantly lagging behind. This study developed and automated methodologies to analyze bid dataset, Daily Work Report (DWR) dataset, and project characteristics dataset to aid State Highway Agencies (SHAs) in improving existing Highway Construction Cost Index (HCCI) systems and as-built schedule development practices.

The first paper developed a concept of Dynamic Item Basket (DIB) to improve the coverage of an Item Basket (IB) used to calculate HCCIs. Then, a concept of multidimensional HCCIs was developed to enable more granular overview of the market conditions. A framework was developed by integrating those two concepts to generate a multi-level of HCCIs. A prototype, the Dyna-Mu-HCCI system was developed to implement and automate the framework. The framework and the prototype will serve as a guide and a tool to SHAs that desire to update their current methodologies. An analysis of bid data obtained from Montana Department of Transportation (MDT) using the Dyna-Mu-HCCI system showed a dramatic improvement in terms of item coverage as a result of DIB implementation: more than 8 times higher in terms of the number of bid items used and at least 20% higher in terms of the total project costs covered. Further, multidimensional HCCIs revealed that specific market segments such as bridge construction have a different trend over time compared to the overall market conditions. These granular and more accurate HCCIs are expected to aid SHAs in assessing their market conditions accurately and develop more customized business plans for different project types and sizes in different locations.

The second paper identified that DWR data is one of the largest but highly underutilized datasets collected by SHAs. Inspectors and resident engineers spend a significant amount of time and efforts to collect DWR data on site, but the use of the DWR data is very limited to contractor payment, progress monitoring, and dispute resolution. Other benefits such as as-built schedule development, production rate estimation, activity cost estimation, contractor evaluation, and contract time determination can be obtained if proper methods are applied, but most SHAs have not obtained those benefits. A statistical analysis showed that the limited use of DWR data to obtain a specific benefit is associated with the limited level of automation. To resolve this issue, an enhanced framework was developed for effective collection and utilization of the DWR data. It consists of three components: a) data model, b) automation of DWR data analysis and reporting, and c) technical aspects. Potential methods to automate some analysis processes were presented in mathematical forms and in the form of Structured Query Language (SQL) queries. DWR experts from DOT engineers across the U.S. validated the framework. The framework can be used to develop a new DWR system or to improve existing systems.

The third and fourth papers utilized the DWR data from existing systems to develop and analyze as-built schedules. In the third paper, a systematic methodology was developed to extract and visualize project level and activity level as-built schedules that can be used to monitor construction progress, evaluate contractors’ performances, defend against claims, and ensure successful execution of a project. A standalone prototype system, namely, ABSS was developed to automate the process and visualize the as-built schedules. Its performance was tested with actual DWR data obtained from a SHA. The methodology and the tool is expected to aid SHAs in making better use of already collected DWR data, facilitate as-built schedule development and visualization, monitor construction progress with higher granularity, and utilize as-built schedule for productivity analysis.

The fourth paper identified that current schedule development process was highly dependent on the experience of a scheduler. There is high potential to reduce such dependency by utilizing information obtained from as-built schedules for previous projects and identifying the patterns of activity sequences. Sequential Pattern Mining (SPM) algorithm called CMRules was used to detect such sequencing patterns from as-built schedules of previous projects. Frequent patterns were then visualized in a chart similar to a precedence diagrams. The sequencing patterns and the diagrams can serve as a knowledge base to aid inexperienced schedulers in developing schedules for future projects as well as providing experienced schedulers with confidence in their schedules. A schedule for a project consisting of seven activities was developed successfully based on the sequencing patterns to validate the framework.

Overall, this study developed various methodologies to improve SHAs' practices of collecting and utilizing various digital datasets. This study will aid SHAs in transforming into data-driven business decisions.

REFERENCES

AEC Big Data Inc. (2013). "Visualizing big data with treemaps." *AEC Big Data*, <http://www.aecbigdata.com/2013/02/26/visualizing-big-data-with-treemaps/> (Aug. 6, 2013).

Alabama Department of Transportation (ALDOT). (2013). *Construction Manual*.

Alavi, S., and Tavares, M. (2009). *Highway Project Cost Estimating and Management*. Montana Department of Transportation.

American Association of State Highway and Transportation Officials (AASHTO). (1999). "AASHTO Transport Software Management Solution."

American Association of State Highway and Transportation Officials (AASHTO). (2015). "AASHTOWare Project ™ licensing status." *AASHTOWare Project News*.

American Association of State Highway and Transportation Officials (AASHTO). (2016). "AASHTOWare Catalog."

Anastasopoulos, P. C., Tarko, A. P., and Mannering, F. L. (2008). "Tobit analysis of vehicle accident rates on interstate highways." *Accident Analysis & Prevention*, 40(2), 768–775.

Anderson, S., Molenaar, K. R., and Schexnayder, C. (2007). *NCHRP 574: Guidance for Cost Estimation and Management for Highway Projects During Planning, Programming, and Preconstruction*. Transportation Research Board of the National Academies, Washington, DC.

Arkansas Highway and Transportation Department (AHTD). (2013). "Arkansas Transportation Planning Conference."

<https://www.arkansashighways.com/PowerPoints/2013/051613_SEB_TransPlanningConf.pdf> (May 18, 2016).

ASCE Task Committee on Application of Small Computers in Construction of the Construction Division (Ed.). (1985). "Application of Small Computers in Construction." *Journal of Construction Engineering and Management*, 111(3).

Avalon, A. (2014). "Calculating the As-built Critical Path." AACE International, Inc.

Aziz, A. (2009). "Time Prediction for Highway Pavement Projects Using Regression Analysis." *Construction Research Congress 2009*, American Society of Civil Engineers, 896–905.

Bastin, M., Heymann, S., and Jacomy, M. (2009). "Gephi: an open source software for exploring and manipulating networks."

Bordat, C., McCullouch, B., and Sinha, K. (2004). "An analysis of cost overruns and time delays of INDOT projects." *Joint Transportation Research Program*, 11.

Bruce, R., Slattery, D., Slattery, K. T., and McCandless, D. (2012). "An Expert Systems Approach to Highway Construction Scheduling." *Technology Interface International Journal*, 13(1), 21–28.

Bureau of Labor Statistics (BLS). (2015). "Consumer Price Index Frequently Asked Questions." <http://www.bls.gov/cpi/cpifaq.htm#Question_2> (Jun. 21, 2016).

Bureau of Labor Statistics (BLS). (n.d.). "Consumer Expenditure Survey." <http://www.bls.gov/cex/> (Jun. 9, 2016).

Cashman, J., and Tayam, R. (2010). "Introduction to Critical Path Method Scheduling using Primavera P6.1 Client for Construction." New York State Department of Transportation.

Cherneff, J., Logcher, R., and Sriram, D. (1991). "Integrating CAD with Construction-Schedule Generation." *Journal of Computing in Civil Engineering*, 5(1), 64–84.

Chin, S., Kim, K., and Kim, Y. (2005). "Generate-Select-Check Based Daily Reporting System." *Journal of Computing in Civil Engineering*, 19(4), 412–425.

Clough, R., and Sears, G. (1991). *Construction project management*. John Wiley & Sons, Inc., New York, N.Y.

Collins, L., and Pritchard, T. (2013). "Applying Chain Index Methods in Volatile Industries: A Closer Look at the Ohio Chained Fisher Construction Cost Index (Working Paper)."

Colorado Department of Transportation (CDOT). (2015). "Colorado Construction Cost Index Report."

Colvin, L. D. (2008). "Construction staffing baseline."

Contreras, M. P., and Van De Werken, A. (2005). "Using CPM Scheduling to Successfully Manage & Complete Projects."

Cox, S., Perdomo, J., and Thabet, W. (2002). "Construction field data inspection using pocket PC technology." *International Council for Research and Innovation in Building and Construction, Aarhus, Denmark*.

Crossett, J., and Hines, L. (2007). *Comparing State DOT's Construction Project Cost and Schedule Performance: 28 Best Practices from 9 States: AASHTO Standing Committee on Quality*. AASHTO.

Dodier, V. (2014). "Adjusting Fuel Tax for Inflation and Fuel Economy."

Douglas, E. E. (2009). "Scheduling Claims Protection Methods."

Dowd, B. (2011). "AASHTO's SiteManager Tames Contract Documentation - Vol. 61· No. 6 - Public Roads." <http://www.fhwa.dot.gov/publications/publicroads/98may/aashto.cfm> (Jun. 13, 2014).

Echeverry, D., Ibbs, C. W., and Kim, S. (1991). "Sequencing knowledge for construction scheduling." *Journal of Construction Engineering and Management*, 117(1), 118–130.

Elazouni, A., and Salem, O. A. (2011). "Progress monitoring of construction projects using pattern recognition techniques." *Construction Management & Economics*, 29(4), 355–370.

El-Rayes, K., and Moselhi, O. (2001). "Impact of Rainfall on the Productivity of Highway Construction." *Journal of Construction Engineering and Management*, 127(2), 125–131.

Erickson, R., and White, K. (2011). "Description of Federal Highway Administration's National Highway Construction Cost Index."

ExeVision. (2012). "Utah Department of Transportation Case Study."

Federal Highway Administration (FHWA). (2014). "National Highway Construction Cost Index (NHCCI)." <https://www.fhwa.dot.gov/policyinformation/nhcci/desc.cfm> (Sep. 13, 2015).

Fischer, M. A., and Aalami, F. (1996). "Scheduling with Computer-Interpretable Construction Method Models." *Journal of Construction Engineering and Management*, 122(4), 337–347.

Fournier-Viger, P. (2014). "A Java Open-Source Data Mining Library." *SPMF*, <http://www.philippe-fournier-viger.com/spmf/index.php?link=documentation.php> (Sep. 24, 2014).

Fournier-Viger, P., Faghihi, U., Nkambou, R., and Nguifo, E. M. (2012). “CMRules: Mining sequential rules common to several sequences.” *Knowledge-Based Systems*, 25(1), 63–76.

Fowler, D. (2010). “Tri-State Partnership: Managing Assets for Transportation Systems (MATS).”

Ghosh, A., and Lynn, R. (2014). “DOD Area Cost Factors (ACF).” U.S. Army Corps of Engineers.

Golparvar-Fard, M., Peña-Mora, F., and Savarese, S. (2009). “Application of D4AR–A 4-Dimensional augmented reality model for automating construction progress monitoring data collection, processing and communication.” *ITCon*, 14(Special Issue: Next Generation Construction IT: Technology Foresight, Future Studies, Roadmapping, and Scenario Planning), 129–153.

Gransberg, D. D., and Diekmann, J. (2004). “Quantifying Pavement Life Cycle Cost Inflation Uncertainty.” *AACE International Transactions*, RI81-RI811.

Guerrero, P. (2003). “Comparison of States’ Highway Construction Costs.”

Hanna, A. S., Whited, G. C., Pashouwer, J. J., and Alsamadani, R. M. (2011). “New Methodology for Developing Cost Indexes for Highway Construction.”

Hegazy, T., and Ayed, A. (1998). “Neural Network Model for Parametric Cost Estimation of Highway Projects.” *American Society of Civil Engineers*, 210–218.

Hegazy, T., Elbeltagi, E., and Zhang, K. (2005). “Keeping Better Site Records Using Intelligent Bar Charts.” *Journal of Construction Engineering and Management*, 131(5), 513–521.

Hendrickson, C., and Au, T. (2008). “Organization and Use of Project Information.” *Project Management for Construction*,

<http://pmbook.ce.cmu.edu/14_Organization_and_Use_of_Project_Information.html> (Dec. 9, 2013).

Henschel, B. A., and Hildreth, J. C. (2007). *Schedule Impact Analysis Using CPM Schedules*. Virginia Department of Transportation.

Holler, J., Tsiatsis, V., Mulligan, C., Avesand, S., Karnouskos, S., and Boyle, D. (2014). *From Machine-to-Machine to the Internet of Things: Introduction to a New Age of Intelligence*. Academic Press.

Hwang, S., Trupp, T., and Liu, L. (2003). "Needs and trends of IT-based construction field data collection." *4th Joint International Symposium on Information Technology in Civil Engineering*, 15–16.

IBM. (2013). "IBM Knowledge Center." CT701, <http://www-01.ibm.com/support/knowledgecenter/SSPK3V_6.3.0/com.ibm.swg.im.soliddb.sql.doc/doc/tables.rows.and.columns.html> (Aug. 4, 2015).

Institute on Taxation and Economic Policy. (2013). *A Federal Gas Tax for the Future*.

International Labour Organization (ILO), International Monetary Fund (IMF), Organisation for Economic Co-operation and Development (OECD), United Nations Economic Commission for Europe (UNECE), and World Bank. (2004). *Producer price index manual: theory and practice*.

International Monetary Fund (IMF). (2010). *Export and Import Price Index Manual: Theory and Practice*. OECD Publishing.

Iowa Department of Transportation (IADOT). (2004). "Project Documentation Guidelines."

Iowa Department of Transportation (IADOT). (2012a). "Letting Guidelines."

Iowa Department of Transportation (IADOT). (2012b). “Cost Estimating Database Details.”

Iowa Department of Transportation (IADOT). (2013). *Price Trend Index for Iowa Highway Construction*.

Jain, D., Shrestha, K. J., and Jeong, H. D. (2015). “Spatiotemporal Analysis of Unit Prices of a Major Cost Item for Transportation Projects across Iowa Using Geographic Information Systems (GIS).” Busan, Korea.

Jan L. Harrington. (2009). *Relational Database Design and Implementation clearly explained*. Morgan Kaufmann.

Jaśkowski, P., and Sobotka, A. (2006). “Scheduling construction projects using evolutionary algorithm.” *Journal of Construction Engineering and Management*, 132(8), 861–870.

Jech, T. (1978). *Set Theory*. Academic Press.

Jeong, D., and Woldesenbet, A. (2010). *Development of an Improved System for Contract Time Determination - Phase III*. Oklahoma Transportation Center.

Jeong, H. S., Atreya, S., Oberlender, G. D., and Chung, B. (2009). “Automated contract time determination system for highway projects.” *Automation in Construction*, 18(7), 957–965.

Joint Federal Government/Industry Cost Predictability Taskforce. (2012). *Guide to Cost Predictability in Construction: An Analysis of Issues Affecting the Accuracy of Construction Cost Estimates*.

Kable, J. M. (2006). “Collecting construction equipment activity data from Caltrans project records.” University of California Davis.

Kahler, D. (2012). “Automated Development of As-Built Construction Schedules.” Fort Worth, Texas.

Kangari, R. (1995). “Construction Documentation in Arbitration.” *Journal of Construction Engineering and Management*, 121(2), 201–208.

Kansas Department of Transportation (KDOT). (1999). “CMS Procedures.”

Kenner, S., Johnson, R. L., Miller, J. R., Salmen, J. A., and Matt, S. A. (1998). *Development of Working Day Weather Charts for Transportation Construction in South Dakota*.

Kim, H., Anderson, K., Lee, S., and Hildreth, J. (2013). “Generating construction schedules through automatic data extraction using open BIM (building information modeling) technology.” *Automation in Construction*, 35, 285–295.

Knoke, J. R., and Jentzen, G. H. (1996). “Developing an as-built schedule from project records.” *Transactions of AACE International*.

Lacho, M. (2015). “Estimates Engineer – 4th Quarter 2014 Notes.”

Lancy, D. (2001). “3D Data Management: Controlling Data Volume, Velocity, and Variety.” META Group Inc.

LeCorps, R. (2001). *Microsoft Access Fundamentals: A Practical Workbook for Small Businesses*. RGL Learning & Publishing.

Lee, J., and McCullouch, B. (2008). “Automating Material Delivery Records.” *JTRP Technical Reports*.

Leung, S., Mak, S., and Lee, B. L. P. (2008). “Using a real-time integrated communication system to monitor the progress and quality of construction works.” *Automation in Construction*, 17(6), 749–757.

Li, T.-R., Xu, Y., Ruan, D., and Pan, W. (2005). “Sequential Pattern Mining*.” *Intelligent Data Mining*, Studies in Computational Intelligence, P. D. D. Ruan, P. D. G. Chen, P. D. E. E. Kerre, and P. D. G. Wets, eds., Springer Berlin Heidelberg, 103–122.

Luo, Z. (2009). “Analysis of the California Highway Construction Cost Index.” California Department of Transportation.

Mack, J. W. (2012). “Accounting for Material-Specific Inflation Rates in Life-Cycle Cost Analysis for Pavement Type Selection.” *Transportation Research Record: Journal of the Transportation Research Board*, 2304(1), 86–96.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A. H. (2011). “Big data: The next frontier for innovation, competition, and productivity.”

Masseglia, F., Teisseire, M., and Poncelet, P. (2009). *Sequential Pattern Mining.*

Mattila, K. G., and Bowman, M. R. (2004). “Accuracy of highway contractor’s schedules.” *Journal of construction engineering and management*, 130(5), 647–655.

McCullouch, B. (1997). “Automating field data collection in construction organizations.” *Construction Congress V: Managing Engineered Construction in Expanding Global Markets*, 957–963.

McCullouch, B. (2000). “Construction Data Management 2000.” *JTRP Technical Reports*.

McCullouch, B. G. (1991). “Automated construction data management system.”

McCullouch, B. G., and Gunn, P. (1993). “Construction field data acquisition with pen-based computers.” *Journal of Construction Engineering and Management*, 119(2), 374–384.

McKinsey Center for Business Technology. (2012). “Perspectives on digital business.”

Memon, Z. A., Majid, M. Z. A., Yusoff, N. I. M., and Mustaffar, M. (2006). “Systematic Approach for Developing As-Built Schedule For Construction Projects.” *Jurnal Kejuruteraan*, 18, 135–146.

Michigan Department of Transportation (MDOT). (2005). “2005 Innovations Awards Program Application: FieldManager.”

Microsoft Corporation. (n.d.). “PEARSON function.” <https://support.office.com/en-us/article/PEARSON-function-0c3e30fc-e5af-49c4-808a-3ef66e034c18> (Aug. 7, 2015).

Minnesota Department of Transportation (MnDOT). (2009). “Minnesota Cost Index Notes.”

Montana Department of Transportation (MDT). (2006). *Standards and Specifications for Road and Bridge Construction*.

Moore, C. J., Miles, J. C., and Williams, T. P. (1999). “Predicted Cost Escalations in Competitively Bid Highway Projects.” *Proceedings of the ICE - Transport*, 135(4), 195–199.

Nassereddine, H., Whited, G. C., and Hanna, A. S. (2016). “Developing a Chained Fisher Construction Cost Index for a State Highway Agency.” Washington, D.C.

Navon, R. (2007). “Research in automated measurement of project performance indicators.” *Automation in Construction*, 16(2), 176–188.

Navon, R., and Haskaya, I. (2006). “Is detailed progress monitoring possible without designated manual data collection?” *Construction Management & Economics*, 24(12), 1225–1229.

Nebraska Department of Roads (NDOR). (2015). *Price Index of Highway Construction Costs*.

Nygaard, R. (2010). “Chain Drift in a monthly chained superlative price index.” *article presented at the joint UNECE/ILOs workshop on scanner data, Geneva, Switzerland*.

Obr, J. F. (2015). "Chapter 10: Prosecution and Progress, Section 2: Progress Schedules." *Construction Contract Administration Manual (CCAM)*, Texas Department of Transportation (TxDOT).

Pentaho Corporation. (2013). "Pentaho Instaview – Instant and Interactive Big Data Analytics." *Pentaho Big Data Analytics Center*, <http://www.pentahobigdata.com/ecosystem/capabilities/instaview> (Aug. 6, 2013).

Rogers, S. (2013). "Managing Assets for Transportation Systems (MATS)." Vermont Department of Transportation.

Rueda, J. (2013). "Develop a price escalation method for Minnesota Department of Transportation indefinite delivery/indefinite quantity contracts: AxE bidding." *Graduate Theses and Dissertations*.

Rueda, J., and Gransberg, D. D. (2014). "Indefinite Delivery-Indefinite Quantity Contracting." *Transportation Research Record: Journal of the Transportation Research Board*, 2408(1), 17–25.

Rueda, J., and Gransberg, D. D. (2015). "Suitability Analysis of Existing Construction Cost Indexes for The Minnesota Department of Transportation Construction Projects." Washington, DC.

Russell, A. (1993). "Computerized Daily Site Reporting." *Journal of Construction Engineering and Management*, 119(2), 385–402.

Shahandashti, S. M. (2014). "Analysis of construction cost variations using macroeconomic, energy and construction market variables."

Shane, J. S., Molenaar, K. R., Anderson, S., and Schexnayder, C. (2009). “Construction Project Cost Escalation Factors.” *Journal of Management in Engineering*, 25(4), 221–229.

Shrestha, K. J., Jeong, H. D., and Gransberg, D. D. (2015). “Current Practices of Collecting and Utilizing Daily Work Report Data and Areas for Improvements.” Busan, Korea.

Shrestha, K. J., Jeong, H. D., and Gransberg, D. D. (2016). “Current Practices of Highway Construction Cost Index Calculation and Utilization.” San Juan, Puerto Rico, 351–360.

Sims, B., Ford, G., and Patterson, J. (2009). “How to Determine Construction Project Rain Delay Times Using Local Rainfall Databases in Asheville, NC.” *Construction Research Congress 2009*, American Society of Civil Engineers, 380–385.

Slone, S. (2009). “Transportation & Infrastructure Finance in the States.”

Stephens, R. (2010). *Beginning Database Design Solutions*. John Wiley & Sons.

Strickland, T., and Beasley, J. G. (2007). *2008 - 2009 Business Plan*. Ohio Department of Transportation.

Tableau Software. (2013). “Data Visualization.” *Tableau Software*, <http://www.tableausoftware.com/solutions/data-visualization> (Oct. 30, 2013).

Taylor, T., Goodrum, P., Brockman, M., Bishop, B., Shan, Y., Sturgill, R., and Hout, K. (2013). “Updating the Kentucky Contract Time Determination System.” *Kentucky Transportation Center Research Reports*.

The Economist. (2010). “Data, data everywhere.”

Tobler, W. R. (1970). “A Computer Movie Simulating Urban Growth in the Detroit Region.” *Economic Geography*, 46, 234–240.

Trimble Navigation Limited. (2014). *Proliance Analytics*.

Turkan, Y., Bosche, F., Haas, C. T., and Haas, R. (2012). "Automated progress tracking using 4D schedule and 3D sensing technologies." *Automation in Construction*, 22, 414–421.

Turner, V., Reinsel, D., Gantz, J. F., and Minton, S. (2014). "The Digital Universe of Opportunities: Rich Data & the Increasing Value of the Internet of Things." EMC Digital Universe.

Vandersluis, C. (2013). "The as-built schedule."

Wactlar, H. (2012). "Big Data R&D Initiatives." National Science Foundation.

Walters, J., and Yeh, D. (2012). "Transportation Literature Search & Synthesis Report: Research and State DOT Practice on Construction Cost Indices." Wisconsin Department of Transportation (WisDOT).

Wang, W. (2005). *Mining Sequential Patterns from Large Data Sets*. Kluwer international series on advances in database systems 28, Springer-Verlag.

West Virginia Division of Highways (WVDOH). (2015). "West Virginia Division of Highway's Highway Construction Price Index Calendar Year 2014 Update."

White, K., and Erickson, R. (2011). "New Cost Estimating Tool." *Public Roads*, 75(1).

Williams, R., Hildreth, J., and Vorster, M. (2007). "BIDDS: A Bid Item Level Performance Time Database Management System." *Computing in Civil Engineering (2007)*, American Society of Civil Engineers, 143–150.

Wilmot, C., and Cheng, G. (2003). "Estimating Future Highway Construction Costs." *Journal of Construction Engineering and Management*, 129(3), 272–279.

Wilmot, C. G. (1999). *Trends in highway construction costs in Louisiana*. Louisiana Transportation Research Center.

Woldesenbet, A., Jeong, D., and Oberlender, G. D. (2012). “Daily Work Reports–Based Production Rate Estimation for Highway Projects.” *Journal of Construction Engineering and Management*, 138(4), 481–490.

Woldesenbet, A., Jeong, H. D., and Park, H. (2015). “Framework for Integrating and Assessing Highway Infrastructure Data.” *Journal of Management in Engineering*, 0(0), 4015028.

Woldesenbet, A. K., Jeong, D. H. S., and Lewis, P. (2014). “Three-Tiered Data and Information Integration Framework: Case Study of Transportation Project Daily Work Report.” Washington, D.C.

Zaniewski, J. P., Butler, B. C., Cunningham, G., Elkins, G. E., and Paggi, M. S. (1982). “Vehicle Operating Costs, Fuel Consumption, and Pavement Type and Condition Factors.”