TOTAL QUALITY MANAGEMENT AND INTEGRATION OF HETEROGENEOUS SEWER INSPECTION DATABASES FOR ASSET MANAGEMENT

By

MOHAMMAD HOSSEIN KHALEGHIAN

Bachelor of Science in Civil Engineering
Amirkabir University of Technology
Tehran, Iran
2006

Master of Science in Environmental Engineering
Amirkabir University of Technology
Tehran, Iran
2009

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2021

TOTAL QUALITY MANAGEMENT AND

INTEGRATION OF HETEROGENEOUS SEWER

INSPECTION DATABASES FOR ASSET

MANAGEMENT

Dissertation Approved:

Dr. Yongwei Shan

Dissertation Adviser

Dr. John Veenstra

Dr. Phil Lewis

Dr. Chenang Liu

ACKNOWLEDGEMENTS

This work would not have been possible without the support of my advisor, Dr. Yongwei Shan, whose sincerity and encouragement I will never forget. His invaluable assistance and insights led me to the writing of this dissertation. Also thanks to my committee members, Dr. John Veenstra, Dr. Phil Lewis, and Dr. Chenang Liu, who offered guidance and support.

I would like to present my sincere gratitude to my dear father, who passed away this past October, for his great role in my life and his numerous sacrifices for me. I'm also very thankful for my late mother, who played a great part in my life, as well as shaping me into the person I am today. Also, many thanks to my brother, Reza, and my sister, Negin, for their tremendous support and hope they have given to me.

Name: MOHAMMAD HOSSEIN KHALEGHIAN

Date of Degree: JULY, 2021

Title of Study: TOTAL QUALITY MANAGEMENT AND INTEGRATION OF HETEROGENEOUS SEWER INSPECTION DATABASES FOR ASSET MANAGEMENT

Major Field: CIVIL ENGINEERING

Abstract: Understanding the current condition of sewer pipe networks is a critical step in improving national wastewater systems. Several studies have attempted to develop deterioration models for sewer pipes, and a common concern raised by those studies was data availability and reliability. Due to quality issues in current data collection practices, approximately one-third of the data is not usable. These data records are often associated with severe defects in the pipes that cause the interruption of the inspection process. The most common data cleaning process was to eliminate missing and duplicate data which helps to avoid misinterpretation of the data. To address the reliability problems, data quality evaluation tools were developed. Data quality evaluation is a multi-dimensional concept that includes both subjective perceptions and objective measurements to be evaluated. Five data quality metrics were defined to assess different quality dimensions of the sewer inspection data including Accuracy, Consistency, Completeness, Uniqueness, and Validity. Moreover, sewer pipes condition assessment databases with more than 90,000 inspections provided by different municipalities across the nation were examined to develop a data quality assurance tool. The quality assurance process consists of three steps: 1) Formulating a quality assurance framework, 2) Detecting problematic data, and 3) Resolving problematic data. The results show that, by applying the developed quality assurance tool, the percentage of good quality inspection data increased from 50%-75% (pre-process) to 95% (post-process). Also, it has been noticed that the data has been collected in so many different formats. As a result, a data mapping tool was developed to address this problem by transforming data into the PACP data structure while keeping the integrity of the database. By implementing this tool one of the major issues in the industry will be addressed and the data can be viewed, modified, and analyzed independently from the generating software. Also, a normalized dimensional Sewer Inventory Schema (SIS) was developed for integrated national sewer inventory. This research contributes to the overall body of knowledge by providing a robust data quality evaluation and integration process for sewer pipes inspection data, which will result in quality data for sewer asset management endeavors.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

BACKGROUND

The general condition of America's infrastructure is alarmingly poor. According to the American Society of Civil Engineers (ASCE) 2021 Infrastructure Report Card, the cumulative grade for the overall infrastructure is C^-. It is estimated that $2.6 trillion is needed for the next ten years to restore the nation's infrastructure systems to good condition. Among these systems, wastewater received a grade of D^+ and needs more than $270 billion in improvements over the next 10 years. Sewer pipelines are the primary component of wastewater systems and they consume approximately 80% of the capital investment for wastewater. There are nearly 800,000 miles of public sewer pipelines and many of them are at the end of their service life (ASCE 2021). Therefore, understanding the current condition of the sewer pipe network is a critical step for infrastructure asset management strategies and improving national wastewater systems.

The United States Environmental Protection Agency (EPA) defined asset management as "managing infrastructure capital assets to minimize the total cost of owning and operating them while delivering the service levels customers desire" (USEPA 2002). Asset management provides agencies with a broad view of the infrastructure capital assets before making any strategic plan (FHWA 2010). Infrastructure assets are the capital assets that are operated as a system to deliver their intended service to the customers, such as roads, bridges, dams, and sewer collection systems. Although infrastructure asset management is considered a crucial process in managing public infrastructures, it is considered more broadly in the transportation sector (USEPA 2002).

Since asset inventory is the first step in the asset management process, it is rarely possible to get the advantages of this process without easily accessible and reliable data (FHWA 2010). Some agencies are more advance in asset data collection processes. Initiated in 1972, structural and condition assessment data of all bridges in the United States have been collected in a unified standard database established by the Federal Highway Administration (FHWA). This database is known as the National Bridge Inventory (NBI) and it contains over 600,000 bridges from across the nation (Weseman 1995). The NBI is considered the main resource for condition assessment of the national bridges and provides the data for bridge management. In this regard, wastewater systems are way behind. Currently, there is no national standard sewer inventory for collecting sewer data across the country.

A sustainable national sewer inventory is needed because it will publicly present empirical data that accurately demonstrates the condition and trends for underground sewer infrastructure across the United States. The rationale is that a unified national inventory of quality assured sewer data will provide credence to widespread but rarely documented claims of failing sewer infrastructure as reported in the ASCE 2021 Infrastructure Report Card. A national sewer inventory will permit the sewer infrastructure community to speak with a common language regarding sewer infrastructure management and future research for benchmarking the sewer pipes' service life.

In 2002, the National Association of Sewer Service Companies (NASSCO) developed the Pipeline Assessment and Certification Program (PACP). PACP is a standard data collection format for Closed-Circuit Television (CCTV) inspections (NASSCO 2010). Prior to PACP, there was no standardized protocol in the United States for the collection and management of data related to pipelines' internal inspections. The primary purpose of PACP is to assure that all data describing the conditions within a pipe are collected and coded in a consistent and reliable manner. PACP became the industry standard for sewer condition data and it was implemented by more than 200 cities and utility districts in the United States and Canada (Thornhill and Wildbore 2005). While different environmental factors, such as soil type or ground water level, may affect sewer pipe conditions, the objective of PACP is to evaluate the internal structural and operational condition of sewer pipelines. Also, PACP provides a grading system to quantify pipe conditions based on the most severe defects (Quick Rating) or the average severity of grades (Pipe Rating Index). The PACP grading system is widely used to evaluate deficient pipes within the sewer network and is considered as a decision support tool for sewer asset management (Opila and Attoh-Okine 2011; Islam et al. 2009; Garrett 2005). Since PACP is a widely accepted standard in the sewer industry, it was the foundation for this research.

PROBLEM STATEMENT

Quality data that documents the current condition of sewer pipelines is fundamental for the development of sewer asset management tools and strategies. Significant efforts have been made to evaluate the condition rating of sewer pipes and determine the factors affecting them. Several different deterioration models have been developed to assess pipe conditions including Markov Chains, artificial neural network (ANN), fuzzy logic, and logistic regression (Abraham and Wirahadikusumah 1999; Ariaratnam et al. 2001; Chughtai and Zayed 2008; Khan et al. 2009). These models determined the main factors that have significant effects on pipe condition such as age, depth, length, soil type, location, size, and material; however, a common concern raised by those studies was data availability and data quality (Ariaratnam et al. 2001; Khan et al. 2009; Opila and Attoh-Okine 2011; Scheidegger et al. 2011). Scheidegger et al. (2011) concluded that previously developed pipe deterioration models suffer from a lack of standardized sewer data; thus, a standard database of sewer condition data is needed for evaluation of current deterioration models as well as developing new and accurate prediction models. Recently, artificial intelligence (AI) and machine learning have become a common approach to analyze the data; however, in most cases, the required data infrastructure is not built to implement such tools.

A solid foundation is required for the data before implementing AI and machine learning strategies and getting effective results (Rogati 2017). Figure 1 shows the data science needs. Data collection is at the bottom of the pyramid. Then, reliable data flow and structured data storage are needed to make it accessible. Data quality management, an under-rated side of data science, and data preparation is the next step to make it reliable for optimization and analytics. These two steps are fundamental to data science; however, they are not usually addressed properly.

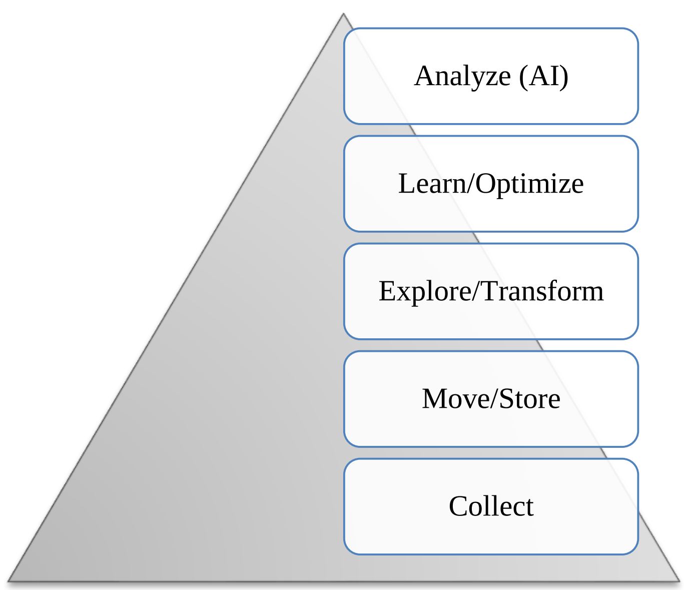


Figure 1. Hierarchy of Needs in Data Science (Rogati 2017)

The accessibility and reliability of the current sewer inspection data are questionable due to several different factors such as inspector experience, PACP versions, and data collection software. Due to the quality issues in the current data collection practices, more than one-third of the data is not usable for asset management purposes based on the current data quality improvement practices (Salman 2010). In order to construct a reliable database, it was required to develop a comprehensive data quality assurance and integration process that facilitates the growth of the unified national sewer inventory in the future.

RESEARCH SCOPE

Developing a reliable and accessible national sewer inventory was an initiative to develop and publically present empirical data that accurately demonstrates the condition and trends for underground sewer infrastructure across the United States. By documenting the condition of the total sewer infrastructure, this initiative gained support from all groups with a vested interest in the underground sewer infrastructure market. This initiative can permit the sewer infrastructure industry to speak with “one voice.”

For the purposes of this study, the PACP data was considered for developing the national sewer inventory. The PACP certified programs have been widely adopted by municipalities across the nation. This data standard made this research more feasible in terms of data exchange. In order to construct a reliable database, it was required to develop a comprehensive data quality assurance process that facilitates the growth of the national sewer inventory in the future. As part of the research effort, seven sewer pipes inspection databases that were provided by different municipalities were evaluated and common data quality issues were found. The major quality issues of the databases were: a) Violating PACP rules, b) Pipe referencing and numbering issues, c) Duplicate inspections for the same pipe in a short period of time, d) Miscalculation of the Pipe Rating Index (PRI).

After implementing the quality assurance process, the integration process was developed to transform the collected databases into a unified data format. This national sewer inventory includes quality assured sewer pipes data which can be used by municipal wastewater agencies, wastewater contractors and equipment manufacturers, operators, consulting firms, corporations, and others who are conducting sewer infrastructure research that requires real-world and quality-assured data.

RESEARCH OBJECTIVES

The overarching goal of this research was to support the development of the national sewer inventory. To attain this goal, the primary objective of this research effort was to determine the specific needs for developing a national sewer inventory. The success of this primary objective was based on the development of a prototype database developed from a limited amount of data gathered from selected municipalities and engineering firms. In order to accomplish the primary objective, the following secondary objectives were addressed:

1. Developing data quality metrics to evaluate the quality of the current databases.
2. Developing a data quality assurance flowchart to solve the current data quality issues.
3. Developing a data transformation and integration process to provide the national inventory prototype based on the collected sewer inspection databases across the country.

By accomplishing the research objective, the outcomes of this research are:

1. A framework for building a national sewer inventory based on real-world, quality-assured data. This framework included protocols for data collection, quality assurance, analysis, and dissemination.
2. A prototype of a national sewer inventory database that enhances research and asset management practices for the sewer industry.
3. Benchmarks for the assessment of the change of pipeline condition over time.
4. An established foundation for developing new industry standards for acceptable levels of service which may lead to a framework for regulatory monitoring of sewer conditions.

SIGNIFICANCE OF THE RESEARCH

This research has significant impacts on the industry. One of the benefits of having a unified sewer inventory at the national level is the provision of a benchmark for utility sewer infrastructure performance in the nation. It can help identify trends in the data for utility districts similar to their own in size, type, and location, which helps identify strategies for infrastructure condition improvement. It may also be used to justify additional or new funding for sewer infrastructure from local, state, and federal agencies. Specific improvements and opportunities that may accrue from this research include:

- Improved awareness and recognition of sewer infrastructure at a national level based on real documented condition data;
- Improved support for industry through data-driven documentation of sewer conditions;
- Improved understanding of how and why pipe deterioration occurs, based on local conditions;
- Improved data availability in a central repository including a valid database for all sewer pipeline condition information.

REFERENCES

Abraham, D., & Wirahadikusumah, R. (1999). "Development of prediction models for sewer deterioration." Durability of Building Materials and Components, 8.

Ariaratnam, S. T., El-Assaly, A., & Yang, Y. (2001). "Assessment of infrastructure inspection needs using logistic models." Journal of Infrastructure Systems, 7(4), 160-165.

ASCE (2021). 2021 Report Card for America’s Infrastructure. American Society of Civil Engineers, www.infrastructurereportcard.org/, Information viewed on June 10, 2021.

Chughtai, F., & Zayed, T. (2008). "Infrastructure condition prediction models for sustainable sewer pipelines." Journal of Performance of Constructed Facilities, 22(5), 333-341.

Federal Highway Administration (FHWA) (2010). Data Integration Primer, Office of Asset Management, Washington D.C.

Garrett, C. (2005). "Asset Management and PACP: A Cost-Effective Solution." Proceedings of the Water Environment Federation, 2005(13), 2994-3001.

Islam, M., Ali, A., and Purtell, J. (2009). "Enhanced condition assessment methodologies of buried infrastructure." Pipelines 2009: Infrastructure's Hidden Assets, 1417-1426.

Khan, Z., Zayed, T., and Moselhi, O. (2009). "Structural condition assessment of sewer pipelines." Journal of performance of constructed facilities, 24(2), 170-179.

National Association of Sewer Service Companies (NASSCO) (2010). Pipeline Assessment and Certification Program (PACP): Reference manual, Volume 6.0.1, Marriottsville, MD

Opila, M. C., and Attoh-Okine, N. (2011). "Novel approach in pipe condition scoring." Journal of pipeline systems engineering and practice, 2(3), 82-90.

Rogati, M. (2017). "The AI hierarchy of needs." Hacker Noon.

Salman, B. (2010). Infrastructure management and deterioration risk assessment of wastewater collection systems, Doctoral dissertation, University of Cincinnati, Cincinnati, OH

Scheidegger, A., Hug, T., Rieckermann, J., & Maurer, M. (2011). “Network condition simulator for benchmarking sewer deterioration models.” Water Research, 45(16), 4983-4994.

Thornhill, R., & Wildbore, P. (2005). "Sewer defect codes: Origin and destination." U-Tech Underground Construction Paper.

USEPA (2002). Fact Sheet—Asset Management for Sewer Collection Systems, Office of Wastewater Management, Washington D.C.

Weseman, W. A. (1995). “Recording and coding guide for the structure inventory and appraisal of the nation's bridges.” Federal Highway Administration, USA.

CHAPTER II

DEVELOPMENT OF QUALITY EVALUATION TOOLS FOR SEWER INSPECTION DATA

ABSTRACT

The increasing amount of data and the growing use of them in the information era have raised a question about the quality of data and its impact on the decision-making process. Currently, the importance of data with high quality is widely recognized by researchers and decision-makers. Sewer inspection data has been collected for over a decade, but the reliability of the data was questionable. It was estimated that between 25% to 50% of data is eliminated due to data quality problems. In order to address the reliability problems, data quality evaluation tools were developed. Data quality evaluation is a multi-dimensional concept that includes both subjective perceptions and objective measurements to be evaluated. Five data quality metrics were defined to assess different quality dimensions of the sewer inspection data including Accuracy, Consistency, Completeness, Uniqueness, and Validity. These data quality metrics were calculated for the collected sewer inspection data and it was found that consistency and uniqueness are the major problems based on the current practices. In addition, problematic issues with current commercial pipe rating software programs were identified. As a result, the rating systems in the current PACP certified data collection software overestimates the current condition of the sewer systems.

INTRODUCTION

Data quality evaluation is usually considered in response to the problems that occurred in the decision-making process. This reactive approach may address the issues observed in the current database, however; it will not eliminate the source of the problem and avoid future quality issues in the database. Data quality assessment is a continuous effort because the flaws in the data can occur at any time which affects the quality of the data (Loshin 2006).

Data quality should be considered through the intended use of the data and will get its definition based on the relevancy to the context the data is to be used (USEPA 2000). Data quality is a long-lasting issue in the field of civil infrastructure condition assessment (Westin and Sein 2013). Civil infrastructure data is collected to support the asset management decision-making process and knowing the quality issues of the data can minimize decision mistakes. Most of the previous research on infrastructure assets has been focused on infrastructure deterioration models and few efforts have been done on evaluating the quality of the data (Buchheit 2002).

As it is mentioned before, the focus of this research is on sewer inspection data and specifically PACP data. PACP inspection is collected through certified software which is supposed to implement PACP rules and requirements in the inspection process. However, it has been recognized that the PACP data collected with one software is not compatible with other software, due to the software's violation of the PACP rules. Also in some cases, the exported database is not in a proper PACP database format to make it possible to import the data into another software. In one data sample of 6,500 inspection records, 47,000 incompatibility errors have been found. In other words, more than 7 errors in each inspection. This incompatibility will result in several problems in developing a unified national sewer inventory. It is estimated that between 25% to 50% of data is eliminated to make the database ready for analysis (Caradot et al. 2018). This approach will result in underestimating the severity of the current condition of the system and false outputs.

In order to address this issue in the sewer inspection data, it has been recognized that the data quality of each database should be evaluated. The goal of data quality evaluation is to reach a high level of accuracy in the inspection data and make it consistent with other datasets. This process is a significant step in developing a sewer data inventory by integrating existing datasets. Any problems with current datasets must be identified and resolved before data integration (Pipino et al. 2002).

To construct a reliable database, it was required to develop a comprehensive data quality evaluation process that facilitates the growth of the national sewer inventory in the future. A set of data quality metrics has been developed to assess different quality dimensions of the sewer inspection database and these metrics have been presented in the quality dashboard. Also, the sewer pipes structural grading calculated by PACP certified software has been evaluated to determine their accuracy. Although PACP provides an internal structural and operational condition of the sewer pipes, the focus of this research was on the structural condition of the pipes. For this purpose, the collected data were reviewed to detect any data quality problems. The major quality issues of the PACP databases are as follows:

1. PACP rules are not fully applied during the data collection process.
2. Data redundancy (the same data exists in multiple places)
3. Miscalculation of the Pipe rating index (PRI)
4. Pipe referencing and numbering issues
5. Several duplicated inspections were recorded for the same pipe in a short period of time.

The first three problems which are related to each individual inspection record are addressed in this chapter. In other words, each inspection record has been evaluated separately for quality issues and the proposed solution has been provided. The last two quality problems have been evaluated within the database based on the relation between different inspection records and will be discussed in the next chapter.

In order to identify the quality issues, it was required to define the database rules and develop a set of data quality metrics. The data quality metrics were developed based on the literature and sewer inspection data requirements. Then, the data were evaluated based on the defined metrics to determine the quality problems within the database. The results were reported and the root cause of each quality flaw has been determined to provide the correction process and implement the resolution.

BACKGROUND

Municipalities are recording multiple forms of data on sewer pipe conditions including Closed Circuit TV, Sonar, Laser, Acoustic, etc. The data has been collected for several years and considered as a basis for asset management plans. However, the real value of the data can only be obtained if combined with quality. In the previous studies on the quality issues of sewer inspection data, it is concluded that the quality problems mainly occurred due to the inspector's level of experience (Fischer et al. 2006; Th et al. 2007; Dirksen et al. 2013). Fischer et al. (2006) explained that the quality of the inspection data depends on the level of skill and motivation of the inspector. The comparison of the sewer pipes' inspections by different inspectors showed that only 16% of the 307 inspections were similar in the number of defects recognition. Also, to improve the quality of the sewer pipes' inspections, the following suggestions were made (Dirksen et al. 2013):

1. The inspection coding system should be simplified to avoid misclassification of the defects.
2. The defect image should be evaluated with the defects information to avoid the misinterpretation of the defects.
3. The sewer inspectors should be provided with reliability feedback on their inspections evaluations.

Although these suggestions will improve the quality of the inspection records, they are not addressing the current issues within the sewer inspection databases.

While the use of data analysis of the collected infrastructure data is growing in decision making and asset management process, improper decisions and poorly performing decision models as a result of poor data quality will adversely affect the infrastructure (Veregin 1999). Since data quality is not defined quantitatively, it is difficult to evaluate the data quality. Moreover, data quality should be improved according to the context of the data (Buchheit 2002). Data quality issues may result in the following problems (Lin et al. 2006):

a. Extra costs for rework
b. The delay in the construction process
c. The infrastructure failure due to incorrect data
d. Poor deterioration models

The data quality evaluation consists of three steps: a) identify, b) measure, and c) resolve. This process provides guidance for data quality management and decision-making. Figure 2 shows the process of data quality evaluation (Wang 1998; Nousak and Phelps 2002).

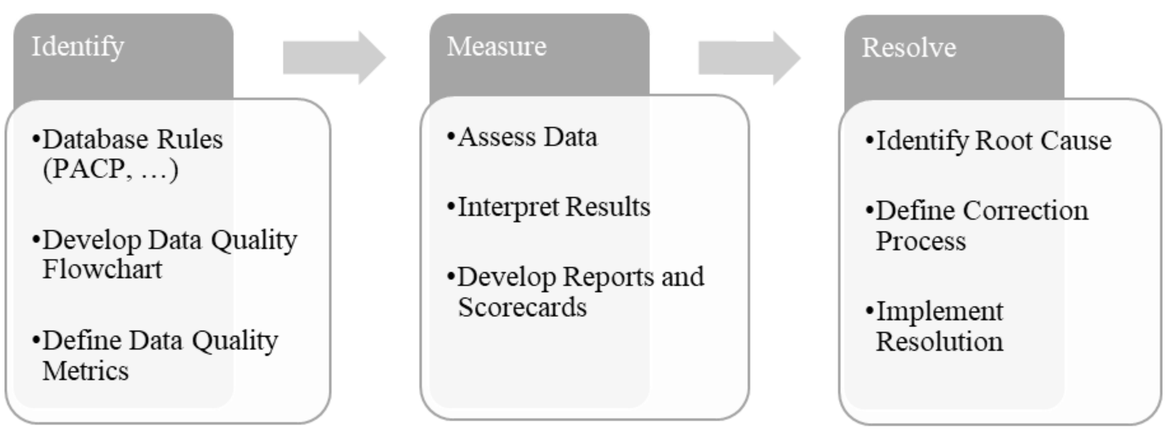


Figure 2. Data Quality Evaluation Process

Data quality evaluation has become the center of attention specifically in business and healthcare institutions where data analysis is the main decision support tool (Ardagna et al., 2018; Pezoulas et al., 2019; Weiskopf et al., 2017; Xia, 2012). Developing an evaluation process and defining a set of data quality metrics is a regular practice in many academic and professional fields.

Data quality is a rational approach to define a set of dimensions to measure and improve the quality of data. Defining the data quality dimensions for the database is the first difficult step (Batini, Cappiello, Francalanci, & Maurino, 2009). The dimensions should consider the specific applications and uses of the data.

Data quality is a multi-dimensional concept that includes both subjective perceptions and objective measurements to be evaluated. The experience of the individuals involved with the data forms the subjective assessment of the data quality. Objective assessment can be divided into two categories: 1) task-dependent or 2) task-independent. Task-independent metrics are developed without considering database rules or restrictions, while task-dependent metrics include them. Pipino et al. (2002) provided three functional forms for objective data quality metrics that considering the objective and subjective assessments:

1. Simple Ratio: the ratio of the positive outcome to the total outcome is a simple way to measure different dimensions. It considers that 1 or 100% is the total desired outcome and the ratio will show the positive outcomes.
2. Min or Max Operations: this form is used when the data quality dimension is a combination of several variables. The min or max will be compared to the preassigned values.
3. Weighted average: the weighted average can be calculated for dimensions with multiple variables. Each variable will be weighted according to its importance between 0 to 1 with the sum of 1. This form will provide an appropriate measurement if precisely developed.

In previous studies, several different sets of data quality metrics have been developed for data quality evaluations. Table 1 shows the most common data quality metrics.

Table 1. Data Quality Metrics Examples.

References	Data Quality Metrics
Pipino et al. 2002	Accessibility, Appropriate Amount of Data, Believability, Completeness, Concise Representation, Consistent Representation, Ease of Manipulation, Free-of-Error, Interpretability, Objectivity, Relevancy, Security, Timeliness, Understandability, Value-Added
Piprani and Ernst 2008	Accuracy, Completeness, Consistency, Precision, Reliability, Temporal Relatability, Timeliness, Uniqueness, Validity
Nousak and Phelps 2002	Validity, Completeness, Consistency, Uniqueness, Timeliness, Accuracy, Precision
Loshin 2006	Uniqueness, Accuracy, Consistency, Completeness, Timeliness, Currency, Conformance

These metrics are used to assess different data quality dimensions. In order to measure these metrics, definitions should be provided and the measurement techniques should be defined. These techniques can be quantitative or qualitative based on the provided definition (Xia, 2012). In this chapter, the data quality metrics have been developed to evaluate the sewer inspection data.

SEWER INSPECTION DATA

Municipalities have collected sewer inspection data for over a decade. The main inspection method is known as Closed-Circuit Television (CCTV) that uses the camera to record the video of the internal pipe and then the defects are coded based on different guidelines or standards. The Pipeline Assessment Certification Program (PACP) is the widely accepted standard in the US.

NASSCO's mission is to set industry standards for the rehabilitation and assessment of underground infrastructure and to assure the continued acceptance and growth of trenchless technologies. NASSCO is a leader in the industry in the delivery of high-quality products through education, technical resources, and industry advocacy. In 2002, NASSCO established PACP to provide a standard and consistent approach to evaluate underground infrastructure. PACP was developed to provide users with a standardized method to code defects, assess asset conditions, and plan for the rehabilitation or replacement of sewer systems (Thornhill and Wildbore 2005). The data generated through PACP standards used by hundreds of utilities has resulted in a need to research relevant approaches and develop a national benchmark and repository for this massive amount of sewer pipeline condition data.

The objective of PACP is to evaluate the internal structural and operational condition of sewer pipelines. General information about the inspection and the pipe segment is required in the "Header" section of the CCTV inspection. The header includes 44 different fields. Although it is not required to fill out all the fields in the header section, some of them are mandatory (Table 2). The inspection observations and defects are entered in the lower part of the form which is known as the "Details" section (NASSCO 2010).

Table 2. Fields Description of the Header Section in PACP

Field Number	Name	Description	Type of Data	Requirement
Field 01	Surveyed by	Name or initials of the operator or surveyor	Alphanumeric	Mandatory
Field 01a	Certificate Number	The number of the surveyor’s certificate	Alphanumeric	Mandatory
Field 02	Owner	Owner of the asset (municipality or utility district)	Alphanumeric	
Field 03	Customer	If different from the owner (consultant or contractor)	Alphanumeric	
Field 04	Drainage Area	Common name for drainage area	Alphanumeric	
Field 05	Sheet Number (Inspection ID)	Forms are numbered consecutively	Numeric	Mandatory
Field 06	P/O Number	Customer’s Purchase Order number	Alphanumeric	
Field 07	Pipe Segment Reference	Unique pipe segment reference number	Alphanumeric	
Field 08	Date	Survey date in order of year, month, and day	Numeric	Mandatory
Field 09	Time	Time at the beginning of the survey	Alphanumeric	
Field 10	Street	Street number and name of upstream manhole	Alphanumeric	Mandatory
Field 10a	City	The city or town where the sewer is located	Alphanumeric	Mandatory
Field 11	Location Details	Further details on the location if appropriate	Alphanumeric	
Field 12	Upstream MH	Upstream manhole number	Alphanumeric	Mandatory
Field 13	Upstream Rim to Invert	Distance between the rim of MH and the invert level	Numeric (feet and tenths of the foot)	
Field 14	Upstream Grade to Invert	Distance between the ground level and the invert	Numeric (feet and tenths of the foot)	
Field 15	Upstream Rim to Grade	Distance between the rim of MH and the ground	Numeric (feet and tenths of the foot)	
Field 16	Downstream MH	Downstream manhole number	Alphanumeric	Mandatory
Field 17	Downstream Rim to Invert	Distance between the rim of MH and the invert level	Numeric (feet and tenths of the foot)	
Field 18	Downstream Grade to Invert	Distance between the ground level and the invert	Numeric (feet and tenths of the foot)	

Field 19	Downstream Rim to Grade	Distance between the rim of MH and the ground	Numeric (feet and tenths of the foot)	
Field 20	Sewer Usage	Codes to define the use of the sewer	Alphabetic	
Field 21	Direction	The direction of the survey in relation to flow	Alphabetic	Mandatory
Field 22	Flow Control	Description of controlling flow during the inspection	Alphabetic	
Field 23	Height (dimeter)	Pipe diameter if circular or height if noncircular	Numeric (nearest inch)	Mandatory
Field 24	Width	Maximum sewer width if not circular	Numeric (nearest inch)	Mandatory
Field 25	Shape	Describe the sewer shape (defined categories)	Alphabetic	Mandatory
Field 26	Material	Predominate pipe material (original construction)	Alphabetic	Mandatory
Field 27	Lining Method	If the sewer was relined describe the method	Alphabetic	
Field 28	Pipe Joint Length	The average length between the joints	Numeric (feet to the nearest tenth)	
Field 29	Total Length	Distance between the access point and the entrance of the finish	Numeric (feet to the nearest tenth)	
Field 30	Length surveyed	The distance surveyed (complete or abandoned survey)	Numeric (feet to the nearest tenth)	
Field 31	Year laid	The construction year provided by the client (YYYY)	Numeric	
Field 32	Year Renewed	The rehabilitation year provided by the client (YYYY)	Numeric	
Field 33	Media Label	The media label assigned by the inspection firm	Alphanumeric	
Field 34	Purpose	Purpose of the survey based on defined categories	Alphabetic	
Field 35	Sewer Category	Critically rating of the particular sewer line	Alphabetic	
Field 36	Pre-Cleaning	Cleaning method before the survey	Alphabetic	Mandatory
Field 36a	Date Cleaned	Date of pre-cleaning (YYYYMMDD)	Numeric	
Field 37	Weather	Weather at the time of survey based on defined categories	Numeric	

Field 38	Location Code	Appropriate code for the ground cover above the pipe surveyed	Alphabetic	
Field 39	Additional info	Any other appropriate details about the survey	Alphanumeric	
Field 40	Work Order	Work order number to track associated costs and activities	Alphanumeric	
Field 41	Project	The project name the inspection belongs	Alphanumeric	
Field 42	Pressure Value	Minimum testing pressure (three decimal places)	Numeric	

The "Structural Family" of defects describes the types of defects where the pipe has been damaged or otherwise is defective. The structural family groups are as follows (NASSCO 2010): (defects 1 through 6 are ranked in order of severity with crack being the least severe and collapsed the most serious.)

1. Crack:

   The crack code is used when a crack line is visible on the surface but is not visibly open. There is no gap between the edges of the crack.

2. Fracture:

   The fracture is a crack that has become visibly open and a gap can be seen through the sections of pipe. The sewer wall is still in place and not able to move.

3. Broken:

   Broken refers to a pipe where pieces are noticeably displaced and have moved from their original position at least ½ the thickness of the pipe.

4. Hole:

   A hole is a condition where the pipe material is missing and the surrounding soil is exposed.

5. Deformed:

   The rigid pipe is damaged to the point that the original cross-section of the sewer line is noticeably altered. For rigid pipe materials like clay or concrete pipes, deformation is the last stage of severity before a collapse.

6. Collapse:

   A collapse is a case where deformation is so great and there has been a complete loss of integrity of the sewer with more than 40% of the cross-sectional area lost.

7. Joint:

   This group is used to describe defective displacements at joints. It can provide a pathway for groundwater, soil, and roots, and the result is pipe failure.

8. Surface Damage:

   A group of codes is used to describe a wide range of pipe material surface damage failures.

9. Buckling:

   Buckling is the deformation of flexible pipes such as plastic, pitch fiber pipes, metal, etc. without loss of visible structural integrity.

10. Lining Features:

    This group is used to describe features in renewed sewers.

11. Weld Failure:

    A failure in a weld of the pipe fabric. Weld failure is used also to describe welds that do not have uniform patterns.

12. Point Repair:

    A group of codes is used to record conditions where part of the pipe has been repaired or replaced.

13. Brickwork:

    A group of codes is only to be used for brick sewers.

Table 3 shows the requirements of some structural defects in PACP. The complete summary of defects description can be found in Appendix A.1. For each defect, there could be two variables, which are:

- Continuous defect: The continuous defect feature is used when the defect extends more than three feet or 1 meter. In some cases, circumferential cracks may occur frequently at joints and the repeating continuous defect feature is used if 75% of the joints are affected.
- Joint: The joint shall be applied to the cracks that are associated with a joint and do not extend to more than 8 inches from the joint.

If the collected data is in PACP format, it should obey the PACP rules. However, it is been recognized that it is not the case in all databases. This problem might occur because of the PACP certified software issues or the inspector's experience. Also, the data format could be different from the PACP data structure.

Data Collection

In order to raise awareness among stakeholders, a formal mission statement, objectives, and benefits of the research were drafted. These elements were presented in a one-page flyer to promote the research project to the sewer infrastructure community. A confidentiality agreement was prepared to ensure anonymity to data providers. Moreover, data transfer protocols were developed to ensure the security of the data storage. In addition, we reached out to municipalities, PACP software vendors, and sewer pipe inspection consultants to request PACP data. As a result, seven datasets have been collected (Table 4). In addition, two different PACP data collection software programs were evaluated to identify data compatibility issues in terms of data exchange. In Table 4, the number of inspections refers to the number of each inspection record in the database, and the number of conditions refers to the total number of codes associated with the inspections.

Table 3. PACP Requirements for Structural Family of Defects

R: Required, NR: Not Required, O: Optional

Group	Descriptor	Modifier	Code	Continuous	Value S/M/L	Value 1st	Value 2nd	Value %	Joint	Clock At/From	Clock To
Crack (C)	Circumferential (C)		CC	O (Joint or Length)	NR	NR	NR	NR	R	R	R
Crack (C)	Longitudinal (L)		CL	O (Joint or Length)	NR	O (Length)	NR	NR	R	R	NR
Crack (C)	Multiple (M)		CM	O (Joint or Length)	NR	NR	NR	NR	R	R	R
Crack (C)	Hinge (H2)		CH2	O (Joint or Length)	NR	NR	NR	NR	R	R	R
Crack (C)	Hinge (H3)		CH3	O (Joint or Length)	NR	NR	NR	NR	R	R	R
Crack (C)	Hinge (H4)		CH4	O (Joint or Length)	NR	NR	NR	NR	R	R	R
Crack (C)	Spiral (S)		CS	O (Joint or Length)	NR	NR	NR	NR	R	R	R
Fracture (F)	Circumferential (C)		FC	O (Joint or Length)	NR	NR	NR	NR	R	R	R
Fracture (F)	Longitudinal (L)		FL	O (Joint or Length)	NR	O (Length)	NR	NR	R	R	NR
Fracture (F)	Multiple (M)		FM	O (Joint or Length)	NR	NR	NR	NR	R	R	R
Fracture (F)	Hinge (H2)		FH2	O (Joint or Length)	NR	NR	NR	NR	R	R	R
Fracture (F)	Hinge (H3)		FH3	O (Joint or Length)	NR	NR	NR	NR	R	R	R
Fracture (F)	Hinge (H4)		FH4	O (Joint or Length)	NR	NR	NR	NR	R	R	R
Fracture (F)	Spiral (S)		FS	O (Joint or Length)	NR	NR	NR	NR	R	R	R
Pipe Failures (Silent)	Broken (B)		B	O (Joint or Length)	NR	O	NR	NR	R	R	O
Pipe Failures (Silent)	Broken (B)	Soil Visible (SV)	BSV	O (Joint or Length)	NR	O	NR	NR	R	R	O
Pipe Failures (Silent)	Broken (B)	Voide Visible (VV)	BVV	O (Joint or Length)	NR	O	NR	NR	R	R	O
Pipe Failures (Silent)	Hole (H)		H	O (Joint or Length)	NR	O	NR	NR	R	R	O
Pipe Failures (Silent)	Hole (H)	Soil Visible (SV)	HSV	O (Joint or Length)	NR	O	NR	NR	R	R	O
Pipe Failures (Silent)	Hole (H)	Voide Visible (VV)	HVV	O (Joint or Length)	NR	O	NR	NR	R	R	O

Table 4. PACP Databases

	No. of Inspections	No. of Conditions	PACP Version	Data Structure
City A	5,232	84,785	PACP 6	PACP 6
City B	3,503	30,609	PACP 6	PACP 6
City C	46,091	365,659	PACP 4	PACP 4
City D	40,966	522,400	PACP 6	PACP 6
City E	212	1,916	PACP 6	Software Preference
City F	7,587	99,596	PACP 6	Software Preference
City G	72	724	N/A	City Preferance

Based on these seven datasets, differences in data management practices among the data providers were identified. For instance, while some providers stored their data in a single data repository, others kept their data in separate datasets based on the year of inspection. This practice resulted in several small datasets which made the proposed quality assurance process more complicated. Although most of the databases were in PACP format, some interoperability issues occurred because the data were exported from different software, which was proved to be a common problem.

DATA QUALITY METRICS

Lebied (2018) provided data quality management techniques in 5 pillars:

1. The people: the quality of data relies on the individuals who implement it.
2. Data profiling: Reviewing data and comparing data to its own metadata
3. Defining data quality: Developing data quality rules and metrics based on the context and use of data (business rules)
4. Data Reporting: identifying data errors and reporting for the resolution process
5. Data repair: addressing data error in the most efficient way

In order to develop an effective data quality management approach for sewer inspection data, more than 100,000 inspection records were reviewed with NASSCO's consultants and more than 50 industry experts provided their feedback on the quality of the collected sewer data (Pillar 1). The main data quality problems were identified based on the industry needs and PACP requirements (Pillar 2). Then, data quality metrics were developed based on previous findings (Pillar 3) and reported (Pillar 4). Finally, a practical resolution for different data quality issues has been provided and implemented (Pillar 5).

As it is mentioned above, a set of data quality metrics is required to assess the data quality of the collected sewer inspection data based on the PACP rules and database requirements. The following rules have been considered during the development process of data quality metrics (Nousak and Phelps 2002):

1. Metrics should be insensitive to changes in the number of records in the database;
2. Metrics should accurately reflect the degree of the data quality requirements;
3. Metrics should be independent of each other;
4. The number of metrics chosen should be kept to a reasonable number;
5. Metrics should address database rules

Defining a proper set of data quality metrics simplifies the measurement of the quality of the data and provides a quantitative structure for data quality evaluation. Data quality rules are integrated into the quality metrics and provide a tool for data quality management (Loshin 2006). Table 5 shows the proper set of data quality metrics developed based on the five rules described above.

Table 5. Data Quality Metrics

Name	Description
Accuracy	Data element values are properly assigned
Consistency	Data element is free from variation and contradiction based on the condition of another data element
Completeness	Data element is required based on the condition of another data element and database rules (required and optional data)
Uniqueness	Data element is unique
Validity	Data element passes all requirements for acceptability (PACP Rules)

The description of each metric is as follows:

- Accuracy:

  Accuracy indicates if the data has significant errors. It can be measured by the source documentation or comparison of the attributes in the database. The metric is defined as the number of errors divided by the total number of attributes subtracted from 1:

$$Accuracy = 1 - \frac{(total\ number\ of\ errors)}{(total\ number\ of\ attributes)}$$

- Consistency:

  Consistency indicates if the data is presented in the same format. The metric is defined as the number of violations divided by the total number of consistency checks subtracted from 1:

$$Consistency = 1 - \frac{(total\ number\ of\ violations)}{(total\ number\ of\ consistency\ checks)}$$

- Completeness:

  Completeness indicates if there are any missing values in the database. Completeness is defined based on the database rules as not all the attributes are required for database fields. Table 3 shows different types of requirements in each field. The metric is calculated by the ratio of the incomplete units to the total number of units and subtracting from 1:

$$Completeness = 1 - \frac{(incomplete\ units)}{(total\ number\ of\ units)}$$

- Uniqueness:

  Uniqueness indicates if the data is represented uniquely in the database and no entity exists more than once. In other words, it captures the redundancy in the database. Redundancy measures the occurrence of data and uniqueness is calculated by subtracting redundancy from 1:

$$Uniquness = 1 - \frac{(Number\ of\ occurrance)}{(total\ number\ of\ entities)}$$

- Validity:

  Validity indicates whether the database complies with PACP. This metric is only applicable to those databases that have been collected in the PACP standard. To measure this metric, all the PACP requirements should be evaluated to calculate the validity of the database. For this purpose, a referential database was created that included all PACP requirements as described in Appendix A.1. This database includes 5 tables as shown in Figure 3.

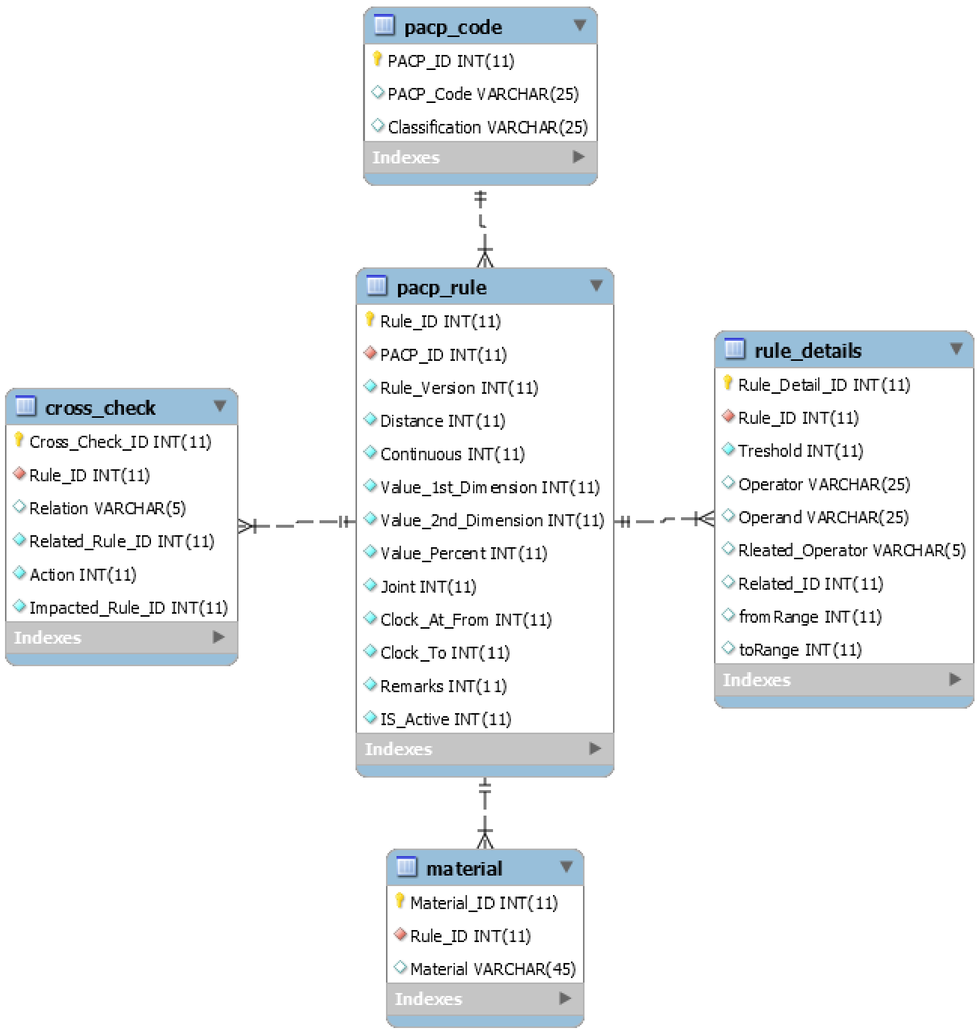


Figure 3. PACP Referential Database

In order to evaluate each database for validity metrics, the `condition_rule` procedure was developed in MySQL. This procedure that consists of over 400 lines of codes checks all PACP requirements in the database. The part of the SQL code that evaluates the “continuous” field is as follows:

```
-- evaluate the correctness of the continuous filed based on the PACP requirement
-- continuous defect should be more than 3 feet
-- distance F% - S% > 3
IF (continuous_Var=2 AND continuous_Var_c IS NULL)  OR (continuous_Var=0 AND cont
inuous_Var_c IS NOT NULL) THEN
        SET continuous_Var_r=0;
ELSE
    BEGIN
        IF continuous_Var_c LIKE 'S%' THEN
                SELECT  distance INTO distance_f_Var
                FROM test_data.conditions ccc
                WHERE ccc.continuous = replace(continuous_Var_c,'S','F')
                AND ccc.pacp_code=pacp_code_var_c
                AND ccc.inspectionid=inspection_ID_Var;
                IF distance_f_Var - distance_var_c < 3 THEN
                        SET continuous_Var_r=0;

                ELSE
                        SET continuous_Var_r=1;

                END IF;

       ELSE IF continuous_Var_c LIKE 'F%' THEN
                SELECT  distance INTO distance_f_Var
                FROM test_data.conditions ccc
                WHERE ccc.continuous = replace(continuous_Var_c,'F','S')
                AND ccc.pacp_code=pacp_code_var_c
                AND ccc.inspectionid=inspection_ID_Var;
                IF distance_var_c - distance_f_Var < 3 THEN
                        SET continuous_Var_r=0;

                ELSE
                        SET continuous_Var_r=1;

                END IF;

         ELSE
                SET continuous_Var_r=1;
          --
        END IF;

        END;
END IF;
```

Also, a sample database has been developed for the validation process of the developed code for all types of material and all the structural defects in the PACP manual. The developed code detected 100% of the invalid attributes in the sample database and can be applied to collected sewer inspection data. The metric is calculated as follows:

$$Validity = 1 - \frac{(Number\ of\ invalid\ attributes)}{(total\ number\ of\ attributes)}$$

RESULTS OF QUALITY METRICS EVALUATION

In order to evaluate the data quality metrics that have been developed for sewer inspection data, the data from the City of G is used. Data is not in PACP format and was hardly accessible by the city. The city can only extract 2.2 miles of inspections from the downtown area. The accessibility problem will be discussed later in chapter four. The city has a total of 775 miles of sanitary sewer lines (ranging from 6" to 54"), 9,100 manholes, and 70 lift stations. Figure 4 shows the data structure preferred by the city.

`ASSET` table includes data related to the sewer pipe constant features. `MANHOLE` table provides manhole IDs. 1MAIN_INSPECTION` stores the inspection data each time it is done and the `OBSERVATION` table contains the pipe internal condition for each inspection. These tables are connected based on the primary key of each table as a point of reference. The data format is completely different from the PACP format and will be discussed later. In order to evaluate the data quality of this database, each table is assessed separately. The percentages show the metric values and not applicate (N/A) represent the fields that could not be calculated due o lack of references and more data should be collected to calculate these metrics.

ASSET								
KEY	SEGMENTID	UPSTREAM_MANHOLE	DOWNSTREAM_MANHOLE	PIPE_TYPE	PIPE_SHAPE	HEIGHT	WIDTH	ADDRESS
566	Sta. Maria & Water M/H.	1043	1044	PVC	Circular	8	8	Sta. Maria & Water

MAIN_INSPECTION							
KEY	ASSET	REVERSED	SCHEDULED_DATE	WEATHER	COMMENT	PROJECT	OPERATOR
695	566	-1	6/25/2014 7:15:33 AM	Dry	Started of on Sta. Maria towards Juarez upstream towards Water.	86	Raul Alfaro

MANHOLE	
KEY	MANHOLE_ID
1044	Water & Sta. Maria

OBSERVATION							
DISTANCE	CODE	LENGTH	CLOCK_FROM	CLOCK_TO	SEVERITY	INSPECTION	COMMENT
0	START AGAINST FLOW					695	

Figure 4. Data Structure for City of G

`ASSET` Table (72 rows)

This table provides the pipe features including manhole information, material, shape, dimension, and length. Each field has been evaluated separately and the result is described in Table 6. The ‘KEY’ represents the primary key of the table and only the completeness and uniqueness metrics were applied to it. Any data error in this field can result in data integrity issues. The evaluation of other fields are as follows:

- SEGMENTID:

  The current information is redundant data from other sources. Also, the data is not consistent with each other and the source field (MANHOLE_ID). This field should be redefined to assign a unique ID to each individual pipe. Table 6 provides the analysis of this table.

Table 6. `ASSET` Table

Column	Description	Completeness	Consistency	Uniqueness	Accuracy
KEY	Primary Key	100%	N/A	100%	N/A
SEGMENTID	Developed based on MAHONHOLE_ID	100%	40%	40%	26%
UPSTREAM MANHOLE	Manhole number Ref to MANHOLE Table	100%	N/A	N/A	N/A
DOWNSTREAM MANHOLE	Manhole number Ref to MANHOLE Table	100%	N/A	N/A	N/A
WIDTH	The width is not required for circular pipes (all pipes in the database)	100%	N/A	0%	N/A
ADDRESS	Based on SEGMENTID	100%	30%	0%	94%
ASSET_LENGTH	The total length of the pipe	83%	N/A	N/A	0%

- UPSTREAM_MANHOLE/DOWNSTREAM_MANHOLE:

  The current information is continuous numbers assigned to the manholes and each manhole should have its own identification number to be distinguished from each other. This field cannot be evaluated because the reference to each manhole is not available.

- WIDTH:

  For the circular pipes, the width is considered redundant information and can be eliminated.

- ASSET_LENGTH:

  For all the inspections SURVEYED_FOOTAGE has been assigned as ASSET_LENGTH. Since these two variables are different from each other, it can be concluded that this field is inaccurate.

- ADDRESS:

  The current information is repeated data from another field (SEGMENTID).

  Also, the data is not consistent with each other and the source data.

`MANHOLE` Table (143 rows)

This table includes the Manhle ID that should be unique for each manhole. Table 7 provides the analysis of this table.

Table 7. `MANHOLE` Table

Column	**Description**	**Completeness**	**Consistency**	**Uniqueness**	**Accuracy**
KEY	Primary Key	100%	N/A	100%	N/A
MANHOLE_ID	It represents the manhole address	100%	75%	40%	N/A

- Key:

  A unique number has been assigned to each manhole to identify them.

- MANHOLE_ID:

  The ID is not well-defined. There is redundancy, inaccuracy, and inconsistency.

`MAIN_INSPECTION` Table (72 rows)

This table contains the general information on the inspection such as operator, weather, date, direction, and comments. It can be considered as the header in PACP. Table 8 provides the analysis of this table.

Table 8. `MAIN_INSPECTION` Table

Column	Description	Completeness	Consistency	Uniqueness	Accuracy
KEY	Primary Key	100%	N/A	100%	N/A
ASSET	Foreign Key to ASSET Table	100%	100%	100%	100%
REVERSED	It shows the reverse setup	100%	N/A	N/A	N/A
COMMENT	Direction and Location	100%	70%	18%	87%
OPERATOR	The name of the inspector	100%	N/A	N/A	N/A
REASON	The purpose of the inspection	98%	N/A	N/A	N/A
SURVEYED FOOTAGE	The length of the surveyed segment	98%	100%	N/A	100%
DATE_START DATE_END	The date and time of inspection	100%	100%	N/A	100%

- REVERSED

  The reversed setup was not found. So, it is not possible to evaluate duplicate inspections because of the current manhole numbering system.

- DATE_START/DATE_END

  The date and time of the inspection can be extracted from these columns. DATE_START will be considered as a reference.

- COMMENT

  The location and direction are considered to be redundant information. 100% accurate direction can be extracted from the `OBSERVATOIN` Table. Some other information on the pipe condition and material is useful. This information can be extracted.

`OBSERVATION` Table (724 rows)

This table provides some information on the condition including defect type, distance, clock positions, and severity. Table 9 shows the data quality evaluation of this table. As it can be noticed, the data quality of the observations could not be assessed through the same approach applied before. This table is similar to the `Condition` table in the PACP database. In order to have a comprehensive data quality evaluation of this table, each defect type should be evaluated separately. It has been noticed that for all the codes, no information has been provided on the length, width, or percentage. Figure 5 shows the distribution of different code groups in the city of G. 77% of the observations are related to construction and miscellaneous features such as manhole location, water level, direction, etc. Each code has been evaluated separately and the results are presented in Table 10. Accuracy is the only metric that could not be evaluated for the codes based on the current database as it requires another reference document. Figure 6 shows the total data quality evaluation of the `OBSERVATION` table. It can be noticed that 66% redundancy has been calculated in the table. This amount of redundancy is also affecting the accuracy of the data.

Table 9. `OBSERVATION` Table

Column	**Description**	**Completeness**	**Consistency**	**Uniqueness**	**Accuracy**
KEY	Primary key	100%	N/A	100%	N/A
DISTANCE	The distance of the occurrence	100%	N/A	N/A	N/A
CODE	The observation	100%	N/A	N/A	N/A
LENGTH	The length of the observation	0%	N/A	N/A	N/A
CLOCK_FROM CLOCK_TO	The clock positioning of the observation	N/A	N/A	N/A	N/A
SEVERITY	The description of the CODE	N/A	N/A	N/A	N/A
INSPECTION	Foreign Key to MAIN_INSPECTIONS	100%	N/A	N/A	N/A
COMMENT	The comment by the operator	N/A	N/A	N/A	N/A

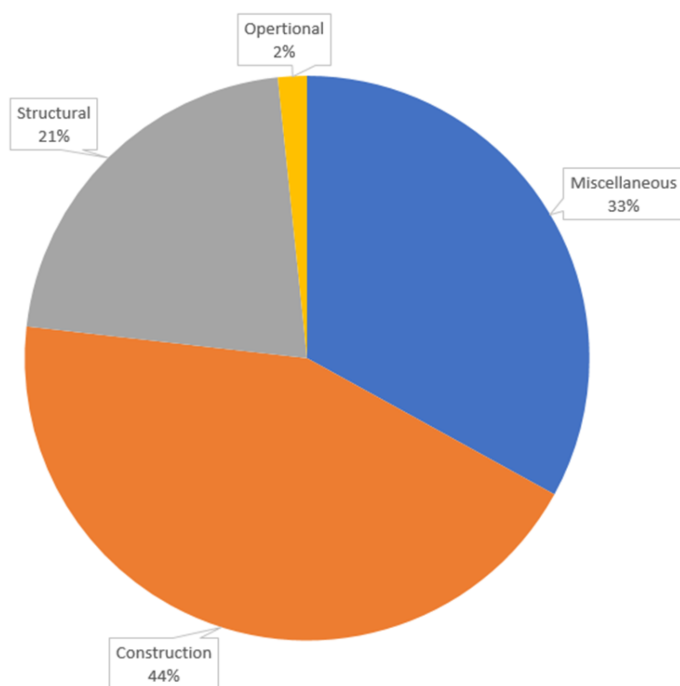


Figure 5. Code Groups of City of G

Table 10. Data Quality Evaluation of `OBSERVATION` Table

Code	Completeness	Consistency	Uniqueness	Accuracy
Abandoned Survey	75%	51%	75%	61%
STOP	100%	98%	0%	100%
Water Level	57%	50%	28%	N/A
Water Mark	75%	25%	25%	N/A
Camera Under Water	0%	100%	50%	N/A
General Observation	88%	35%	12%	N/A
Pipe Type	81%	52%	52%	N/A
MANHOLE	91%	78%	13%	91%
Lateral	97%	88%	9%	N/A
Lateral Connection Problem	100%	66%	0%	N/A
Intruding Sewer Tap	100%	75%	0%	N/A
Broken	65%	60%	90%	N/A
Crack	37%	37%	100%	N/A
Joint Offset	46%	46%	100%	N/A
Joint Separated	27%	27%	100%	N/A
Sag	50%	50%	100%	N/A
Root in Lateral/ Root in Joint	66%	66%	100%	N/A
Obstecle	100%	100%	100%	N/A

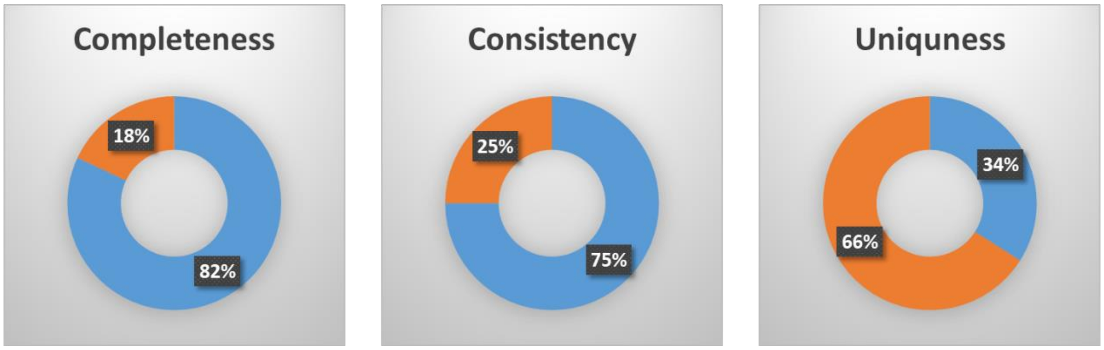


Figure 6. Total Quality Metrics of `Observation` Table

Validity

PACP inspection is collected through certified software which is supposed to implement PACP rules and requirements (Appendix A.1) in the inspection process. However, it has been recognized that the PACP data collected with one software is not compatible with another software, due to the software's violation of the PACP rules. In one of the studied cities, 22,084 input errors out of 12,115 Inspections were found, which raises serious doubts about the reliability of the inspection process. Table 11 shows some errors within the database. The numbers on the left represent the total amount of errors for the related PACP requirement for each defect. These errors will cause serious issues in the asset management process.

In order to evaluate the PACP compatibility and validity of the collected data, the developed tool was applied to the data from City A. City A provided the only database that was developed based on PACP 6 coding system and implemented all PACP data structures. Table 12 shows that the validity metric for City A is 82.2% based on the number of valid attributes.

Table 11. A Summary of PACP Rules Violation in One PACP Database

Defects	Count of Errors
CC	**299**
Should not have a value in Value_Percent	299
CH2	**7**
Required field Clock_To missing	4
Should not have a value in Value_Percent	3
CL	**474**
Should not have a value in Value_Percent	474
CM	**543**
Should not have a value in Value_Percent	543
CS	**5**
Should not have a value in Value_Percent	5
FC	**462**
Should not have a value in Value_Percent	462
FH2	**16**
Required field Clock_To missing	12
Should not have a value in Value_Percent	4
FH3	**10**
Required field Clock_To missing	10
FH4	**5**
Required field Clock_To missing	5
FL	**342**
Should not have a value in Value_Percent	342
FM	**475**
Should not have a value in Value_Percent	475
FS	**16**
Should not have a value in Value_Percent	16

Table 12. Validity Metric for City A

Total Number of Attributes	26,260
Total number of Valid Attributes	21,589
Total number of Invalid Attributes	4,671
Validity	82.2%

PACP GRADING SYSTEM

The PACP rating system focuses on the structural and operational condition of sewer pipes. Defects are classified into four different families (structural, operational, construction, and other) and graded from 1 to 5 based on the severity of each defect. A grade of 1 represents the least severe condition and 5 is the worst condition. The Pipe Rating Index (PRI) is the average of the grades within a pipe and is calculated by adding all the grades in the same categories (structural or operational) and dividing the sum by the total number of defects, as shown in the following equation (NASSCO 2010).

$$RI = \frac{\sum Grades\ in\ the\ Pipe}{\sum Defects\ in\ the\ Pipes}$$

PACP also uses the Quick Rating Index to show the worst defects in the pipes. It is a 4-digit index in which the first and third digits represent the worst grades in the pipe and the second and fourth digits are the frequency of those defects. For example, a sewer pipe with two defects with grade five and three defects with grade four will have a quick rating of 5243. This rating helps identify the most severe defects which may lead to pipe failures.

Although the PACP inspection process is a standardized process assisted with the use of PACP certified software, the PACP rating calculations in the software are based on an algorithm specified in the PACP manual. Many previous deterioration models were based on the PRI; however, the outputs of those models may be somewhat questionable. During this study, it has been discovered that the ratings calculated by one of the PACP certified software programs included some PACP codes that have no effect on pipe condition (i.e. there are not any grades assigned to these codes); thus, the software underestimates PRI by inflating the value of the denominator by counting those codes as defects. These codes are used just to provide information on the pipe features like the Mean Water Level (MWL). Also, there were some inconsistencies in calculating ratings for the defects which have various criteria such as continuous, clock position, and percentage to define their grades; hence, a new rating calculation tool was developed based on the PACP manual by NASSCO. This tool strictly complied with the defect coding system specified in the manual. This new tool calculates pipe ratings in a more reliable way which can be used for pipe condition analyses.

In order to demonstrate the PACP grading system, two PACP inspection datasets (City A and City B) were used. These two datasets include 8,228 individual inspections. Although PACP ratings provide scores for the structural and operational conditions of the pipe, only the structural score was assessed at this point. The structural pipe rating (total number of structural grades in a pipe) and defects (total number of defects in a pipe) and PRI were calculated for both databases by a PACP certified software and a pipe rating tool developed in this study. As anticipated, the pipe ratings for both approaches were very similar. Among the 5,232 inspections, 36 inspections had discrepancies in structural pipe ratings calculated by the commercial software and the tool developed in this research. Those discrepancies were a result of different interpretations of the defects criteria. For structural defects, these criteria were grouped into two categories as follows:

1. Continuous defects: Although the PACP manual clearly explained how to calculate the grading of continuous defects by dividing the length of the defect by 5 and rounding the fraction to the nearest whole number, the commercial software arbitrarily rounded the fraction up or down. This miscalculation will affect the pipe rating.
2. Pipe Failure PACP group: For the defects which were graded based on their clock position (Broken or Hole), the software calculated them for their clock input instead of clock position (e.g. if the defect started from 2 to 3, the software assumed the two clock position instead of one). This problem occurred for clock positions 1 and 2, since values more than 3 were graded 5.

Although there were some discrepancies in the structural pipe ratings, they rarely affected the PRI. A maximum discrepancy of 0.5 was observed in 13 inspections among the 5,232 data records in City A; however, the difference in the number of defects in the pipe had noticeably altered the PRI which were caused by:

1. Continuous defects: For the same reason mentioned previously, discrepancies in the total number of defects calculated by the commercial software and the research team's tool were observed in 28 inspections that had continuous defects out of the 5,232 inspections in City A.
2. Defects with 0 grade: A 0 grade should not be considered as a defect, but the commercial software counted it as a defect, which overestimated the total number of defects in a pipe. Although 0 grades do not affect the pipe structural rating, the PRI was underestimated since the denominator (the total number of defects) was overestimated. City A included 209 inspections.

Table 13 and Table 14 present a comparison of the PRI calculated by a PACP certified software and the PRI calculated by the tool developed in this research. The ratings calculated by the commercial software and the tool developed by this research were denoted as PRI and PRI*, respectively. The last row in each table shows the total number of pipes for each grade category. By comparing the total number of pipes for PRI and PRI*, the PACP software apparently underestimated the PRI. Since the PRI is often used for evaluation of remaining service life, the difference between these two calculations is large enough to change the condition category of a pipe and make results in the previous studies debatable (Opila and Attoh-Okine 2011; Islam et al. 2009; Garrett 2005).

Table 13. Structural Pipe Rating Index in City A

Material	Pipes	Pipes with grades 4-5		Pipes with grades 2-3		Pipes with grades 0-1	
		PRI*	PRI	PRI*	PRI	PRI*	PRI
Asbestos Cement	317	19	19	69	67	229	231
Brick	31	1	1	8	8	22	22
Cast Iron	5			1	1	4	4
Clay Tile	17	2	2	5	4	10	11
Concrete Pipe (non-reinforced)	68	11	11	21	21	36	36
Corrugated Metal Pipe	46	10	9	6	7	30	30
Ductile Iron Pipe	10					10	10
Fiberglass Reinforced Pipe	9					9	9
Not Known	2			2	2		
Other	21			7	6	14	9
Plastic/Steel Composite	1	1	1				
Polyethylene	21	3	3			18	18
Polypropylene	14			2	2	12	12
Polyvinyl Chloride	575	13	12	25	26	537	537
Reinforced Concrete Pipe	1,384	86	84	376	375	922	925
Reinforced Plastic Pipe (Truss Pipe)	100	4	4	5	5	91	91
Segmented Block	3	1	1			2	2
Vitrified Clay Pipe	2608	164	158	1131	1108	1313	1342
Grand Total	**5,232**	**315**	**305**	**1,658**	**1,632**	**3,259**	**3,295**

PRI*: The ratings calculated by this research PRI: The ratings calculated by the commercial software

Table 14. Structural Pipe Rating Index in City B

Material	Pipes	Pipes with grades 4-5		Pipes with grades 2-3		Pipes with grades 0-1	
		PRI*	**PRI**	**PRI***	**PRI**	**PRI***	**PRI**
Asbestos Cement	824	26	22	68	66	730	736
Brick	1					1	1
Cast Iron	4			2	2	2	2
Clay Tile	1			1	1		
Corrugated Metal Pipe	4					4	4
Ductile Iron Pipe	2					2	2
Not Known	2					2	2
Other	2					2	2
Polyvinyl Chloride	568	7	7	30	30	531	531
Reinforced Concrete Pipe	15	1	1	6	6	8	8
Reinforced Plastic Pipe (Truss Pipe)	119	8	6	28	28	83	85
Transite Pipe	5			1	1	4	4
Vitrified Clay Pipe	1,449	41	31	359	277	1,049	1,141
Grand Total	**2,996**	**83**	**67**	**495**	**411**	**2,418**	**2,518**

PRI*: The ratings calculated by this research   PRI: The ratings calculated by the commercial software

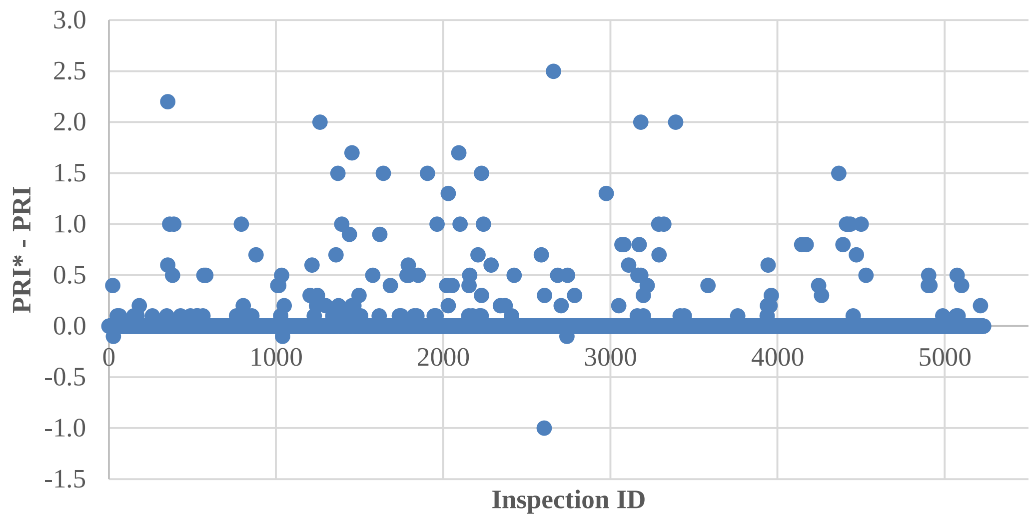


Figure 7. Distribution of Difference of PRI Calculations for City A

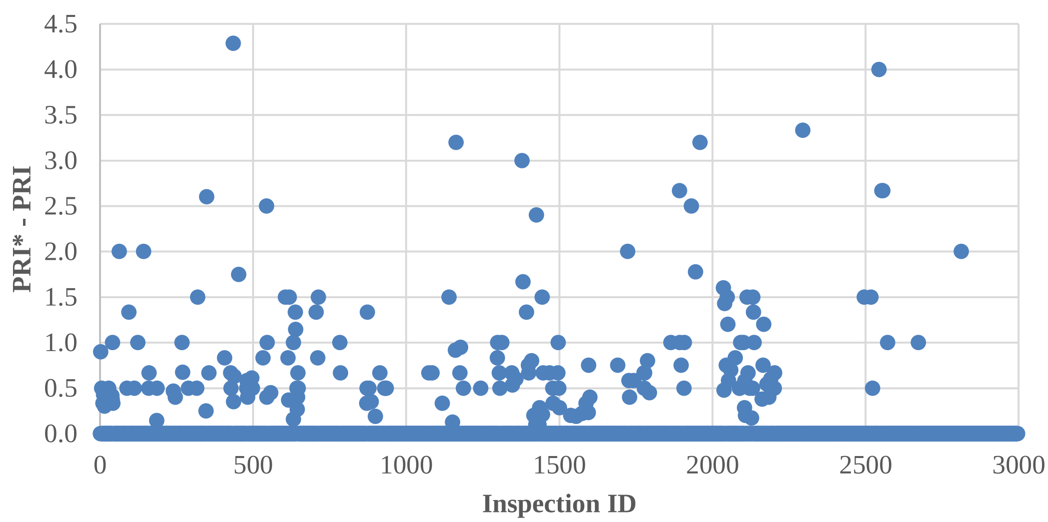


Figure 8. Distribution of Difference of PRI Calculations for City B

Figure 7 and Figure 8 depict the difference between PRI and PRI*. The horizontal axis represents the inspection ID (individual inspections) and the vertical axis represents the difference between PRI* and PRI. While there were only four PRI greater than PRI* in City A, all PRI* were higher than PRI for City B. Also, the maximum difference was 2.5 for City A and 4.3 for City B, which may result in the misclassification of a deficient pipe in the good condition category. The tool developed in this study strictly complies with PACP rating systems, thus providing more reliable pipe ratings compared to those calculated by the PACP certified software.

CONCLUSIONS AND RECOMMENDATIONS

Municipalities are collecting CCTV inspection data on 10% of their sewer systems every year. Although this data can be used to support the asset management decision-making process, reliability problems can increase the number of decision mistakes. Most of the previous research on infrastructure assets has been focused on infrastructure deterioration models and few efforts have been done on evaluating the quality of the data (Buchheit 2002). Data quality evaluation has become the center of attention specifically in business and healthcare institutions where data analysis is the main decision support tool (Ardagna et al., 2018; Pezoulas et al., 2019; Weiskopf et al., 2017; Xia, 2012). However, very few studies have been done on civil infrastructure data. Currently, it is estimated that between 25% to 50% of data is eliminated to make the database ready for analysis (Caradot et al. 2018). This approach will result in underestimating the severity of the current condition of the system and false outputs.

The focus of this study was to develop a data quality evaluation process to identify, measure, and resolve different quality dimensions of sewer inspection data. A set of five data quality metrics were developed based on the literature to evaluate the most common data quality errors. Through implementing these metrics, it was concluded that consistency and uniqueness are the significant issues in the data that can be addressed mainly by resolving redundancy problems in the database. Moreover, the PACP grading system, which is widely used to determine pipe conditions in a sewer network, was evaluated to avoid inconsistencies in the calculation of ratings. In many cases, PRI was overestimated due to the miscalculation of the total defects in the pipe.

This study provides a quantitative analysis of the quality problems in the sewer inspection data and for the first time provides tools for industry stakeholders to address these problems. Further research is needed to improve the data quality resolution process that requires more data to be analyzed.

REFERENCES

Ardagna, D., Cappiello, C., Samá, W., & Vitali, M. (2018). Context-aware data quality assessment for big data. Future Generation Computer Systems, 89, 548-562.

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. ACM computing surveys (CSUR), 41(3), 1-52.

Bergdahl, M., Ehling, M., Elvers, E., Földesi, E., Körner, T., Kron, A., Lohauß, P., Mag, K., Morais, V., and Nimmergut, A. (2007). "Handbook on data quality assessment methods and tools." Handbook on Data Quality Assessment Methods and Tools, 9-10.

Buchheit, R. B. (2002). "Vacuum: Automated procedures for assessing and cleansing civil infrastructure data." Carnegie Mellon University Pittsburgh, PA.

Caradot, N., Rouault, P., Clemens, F., & Cherqui, F. (2018). Evaluation of uncertainties in sewer condition assessment. Structure and Infrastructure Engineering, 14(2), 264-273.

Dirksen, J., Clemens, F., Korving, H., Cherqui, F., Le Gauffre, P., Ertl, T., Plihal, H., Müller, K., and Snaterse, C. (2013). "The consistency of visual sewer inspection data." Structure and Infrastructure Engineering, 9(3), 214-228.

Fischer, B., Hunger, W., Lehmann, T., Muller, K., and Schafer, T. (2006). "Objective condition establishment of sewer systems." Proc., Proceedings of 2nd Conference on Sewer Operation and Maintenance, Vienna, Austria (Ertl T, Pressel A, Kretschmer F and Haberl R (eds)). SIG Eigenverla, Vienna, Austria, 207-215.

Garrett, C. (2005). "Asset Management and PACP: A Cost-Effective Solution." Proceedings of the Water Environment Federation, 2005(13), 2994-3001.

Islam, M., Ali, A., and Purtell, J. (2009). "Enhanced condition assessment methodologies of buried infrastructure." Pipelines 2009: Infrastructure's Hidden Assets, 1417-1426.

Lebied, M. (2018). The Ultimate Guide to Modern Data Quality Management (DQM) For An Effective Data Quality Control Driven by The Right Metrics. URL https://www. datapine. com/blog/data-quality-management-and-metrics.

Lin, S., Gao, J., and Koronios, A. (2006). "The need for a data quality framework in asset management." Proc., 1st Australasian Workshop on Information Quality, Citeseer.

Loshin, D. (2006). "Monitoring data quality performance using data quality metrics." Informatica Corporation.

National Association of Sewer Service Companies (NASSCO) (2010). Pipeline Assessment and Certification Program (PACP): Reference manual, Volume 6.0.1, Marriottsville, MD.

Nousak, P., and Phelps, R. (2002). "A Scorecard approach to improving Data Quality." Proceedings of Data Warehousing and Enterprise Solutions, Sugi-27, Orlando, Florida.

Opila, M. C., and Attoh-Okine, N. (2011). "Novel approach in pipe condition scoring." Journal of pipeline systems engineering and practice, 2(3), 82-90.

Pezoulas, V. C., Kourou, K. D., Kalatzis, F., Exarchos, T. P., Venetsanopoulou, A., Zampeli, E., . . . Tzioufas, A. G. (2019). Medical data quality assessment: On the development of an automated framework for medical data curation. Computers in biology and medicine, 107, 270-283.

Pipino, L. L., Lee, Y. W., and Wang, R. Y. (2002). "Data quality assessment." Communications of the ACM, 45(4), 211-218.

Piprani, B., and Ernst, D. (2008). "A model for data quality assessment." Proc., On the Move to Meaningful Internet Systems: OTM 2008 Workshops, Springer, 750-759.

Th, E., Gangl, G., Bölke, K., and Kretschmer, F. (2007). "Implementing quality management and EN 13508-2 for CCTV sewer inspection in Austria." NOVATECH 2007.

Thornhill, R., and Wildbore, P. (2005). "Sewer defect codes: Origin and destination." U-Tech Underground Construction Paper.

USEPA (2000). Guidance for Data Quality Assessment Practical Methods for Data Analysis, Office of Environmental Information, Washington D.C.

Veregin, H. (1999). "Data quality parameters." Geographical information systems, P. A. G. Longley, Michael F.; Maguire, David J.; Rhind, David W., ed., Wiley, New York, 177-189.

Wang, R. Y. (1998). "A product perspective on total data quality management." Communications of the ACM, 41(2), 58-65.

Weiskopf, N. G., Bakken, S., Hripcsak, G., & Weng, C. (2017). A data quality assessment guideline for electronic health record data reuse. Egems, 5(1).

Westin, S., and Sein, M. K. (2013). "Improving data quality in construction engineering projects: An action design research approach." Journal of Management in Engineering, 30(3), 05014003.

Xia, J. (2012). Metrics to measure open geospatial data quality. Issues in Science and Technology Librarianship, 68.

CHAPTER III

DEVELOPING A CONTEXTUAL DATA QUALITY ASSURANCE PROCESS FOR SEWER INSPECTION DATA

ABSTRACT

Due to quality issues in current sewer inspection data, approximately one-third of the data are not usable for asset management purposes. Those data primarily are related to pipes that are in deficient conditions and when the inspection process is interrupted by severe defects in the pipes. The objective of this chapter is to present a quality assurance process for the sewer inspection data which may be used to develop a unified and useful database for further analysis. PACP databases with more than 90,000 inspections provided by different municipalities across the nation were examined to identify common data quality problems. The quality assurance process consists of three steps: 1) Formulating a quality assurance framework; 2) Detecting problematic data; and 3) Resolving problematic data. The results show that, by applying the proposed quality assurance process, the percentage of good quality inspection data increased from 50%-75% (pre-process) to 95% (post-process). This research contributes to the overall body of knowledge by providing a robust data quality assurance process for underground sewer pipe inspection data, which will result in quality data for sewer asset management endeavors.

INTRODUCTION

Data quality is one of the crucial issues in the asset management process because the decision makers rely more on data to implement their plans. In the civil engineering industry, data error rates of 75% have been reported, while errors up to 30% are common (Baskarada et al. 2005). Thus, there is a need to address data quality issues in infrastructure asset management. After evaluation of the data quality of each individual inspection of sewer pipes, it is required to apply the quality assurance process to the whole database to cross-check the accuracy and consistency of the inspection records. Although the sewer inspection data might be in compliance with the PACP data rules through the implementation of the data quality process explained in the previous chapter, the inspection data may still have complex data quality problems, such as multiple inspections of the same pipe within a short period of time, uncompleted inspections, inaccurate Pipe Segment References (PSR) and manhole numbers.

In most of the previous studies on the condition of the pipes, researchers tried to eliminate the duplicate inspections which happened in a short period of time with different defects information or pipe rating index (Salman 2010). This practice is a quick approach to compile a dataset for the analysis of the pipe conditions; however, this practice may eliminate the pipes with severe defects which caused an inspector to do the reverse set-up. The cases of several inspections of the same pipe that occurred in a short period of time can make the analysis of the pipe condition complicated because most of the deterioration models were developed based on the age of pipes as the main variable (Abraham and Wirahadikusumah 1999; Ariaratnam et al. 2001; Chughtai and Zayed 2008; Khan et al. 2009). Thus, those duplicate and inconsistent inspections records for the pipes with the same inspection date were eliminated from the analysis database, which possibly resulted in less accurate models due to the omission of pipes with potentially significant problems.

In order to address this issue, a quality assurance framework was developed. A quality assurance framework is a step-by-step approach that provides the flow of actions required to find a solution to the existing problems. Data quality assessment framework is suggested through several previous studies on data quality (English 1999; Lee et al. 2002; Pipino et al. 2002). The goal of these frameworks is to enhance the quality of data by identifying the errors and providing the correction process (Westin and Sein 2013).

BACKGROUND

The main purpose of asset management is to deliver a proper level of service by managing the infrastructure with efficient rehabilitation and replacement strategies. These strategies are developed based on the structural and the hydraulic performance of the sewer network, while the structural performance is considered as the main factor for budget allocation (Caradot et al. 2018; Van Riel et al. 2014). Currently, Closed Circuit Television Inspection (CCTV) is the main source of information, more than 60%, for defining maintenance and rehabilitation projects (van Riel et al. 2016). As a result, the quality of the collected CCTV data plays an important role in the accuracy of the final decisions. Caradot et al. (2018) evaluated the data quality of the collected CCTV data for the city of Braunschweig in Germany. It has been reported that the probability to overestimate a pipe in bad condition is 20% whereas the probability to underestimate a pipe in good condition is 15%. It has been noticed that to prepare the data for this evaluation, only 45,049 inspections out of 69,384 total inspections have been analyzed. In another word, 35% of data has been neglected due to inconsistency, incompleteness, and lack of reference keys. This data elimination practice will result in underestimating the severity of the system by neglecting the assets that could have the more severe condition in the system.

Incorrect and inconsistent data always leads to false results. While the amount of collected data is increasing rapidly, evaluating the data quality is becoming a big issue (Kleindienst 2017). Any data analytics tools, charts, and algorithms will be worthless if they have been developed based on low-quality data. The data quality assurance process needs to be carefully developed based on the context and be economically feasible to be applicable in the real world.

Data errors can be very costly as they will result in poor decisions. Although high-quality data will not guarantee good decisions, low-quality data is associated with poor decisions. This statement has been noticed in several studies and it has been concluded that addressing the data quality issues and developing an effective data quality improvement plan is necessary for the decision making process (Even and Shankaranarayanan 2007; Kleindienst 2017; Shah et al. 2012).

It has been mentioned that the data quality evaluation consists of three steps: a) identify, b) measure, and c) resolve (Nousak and Phelps 2002; Wang 1998). So, data quality should be improved repeatedly through the data quality assurance process to address the identified quality problems including inconsistency, redundancy, and data errors (Heinrich et al. 2007; Kleindienst 2017). Also, data quality assurance is a stepwise approach and having a structured framework provides consistency in data quality improvement (Hamilton et al. 2020).

The literature on data quality improvement frameworks is very limited. The low data quality has been studied mainly in business and healthcare research fields as poor decisions in these fields have significant consequences. However, very few studies have provided practical data improvement plans and one of the main reasons is known to be the complexity of each database (Kraft et al. 2015; Taylor et al. 2014; Walshe and Freeman 2002).

The effectiveness of these data quality improvement plans depends on the context in which they are used and how they are implemented. Walshe and Freeman (2002) had reported that there are three important aspects to each data quality assurance effectiveness:

1. Approach: choose the plan carefully based and maintain it
2. Performance: realize how and why the plan works
3. Evaluation: the plan should be evaluated and monitored

Data quality assurance is a context-oriented process. This will result in logical approaches to improve the quality of data and data cleaning based on logical dependencies. Quantitative data cleaning is a common method that is based on expected statistical distributions and the final database will be statistically desirable; however, this method could not address subjective and contextual issues within the databases. It is almost impossible to develop a unified method to address data quality problems and each method should be developed according to the characteristics of the data and its usage (Alipour-Langouri et al. 2018; Berti-Equille et al. 2011; Dasu and Loh 2012; Prokoshyna et al. 2015). Bertossi and Milani (2018) have described the contextual approach as the logical mapping method to transform the raw database (D) into quality versions of D.

The contextual data quality assurance approach is usually implemented through timely manual data cleaning based on user experience that will result in human errors. Therefore, automated methods are needed to address data quality problems in large databases (Bertossi and Milani 2018). In the sewer industry, there is no defined data quality assurance process and the data is mainly evaluated manually by the inspection team. Since this process is very time-consuming, less than 10% of the data has been evaluated in the best-case scenario. Municipalities have collected CCTV data for nearly two decades and while this amount of data is very valuable, the historical data is not usable at this time. So, contextual data quality assurance can provide an automated solution to address this significant industry problem.

DATA

Duplicate Inspections

Currently, the sewer inspection data is collected through PACP certified software which applies the PACP rules and requirements to the inspection process. Some of the general requirements of the PACP manual are as follows (NASSCO 2010):

1. A separate inspection form is required for each pipe segment between access points.

2. If an intermediate access point is noticed during the inspection, a new form should be used for the remaining pipe segment.
3. A separate inspection form is also needed for a reverse inspection.

The first two rules require that inspection records start and finish at access points without any intermediate access points, such as manholes, between them. The third rule is applied to incomplete inspections which are abandoned due to difficulties during the inspection process, such as the presence of rocks, roots, debris, or any other type of obstacle in the pipe. In order to complete the inspection, the inspector may request cleaning of the pipe or continue the inspection in a reverse direction, as shown in Figure 9. Either of these solutions creates duplicate inspections in the datasets. The third requirement is the major cause of duplicate inspections in the database. If two separate inspections were done on the same pipe within a short period of time, they both represent the condition of the pipe at the time of inspection; thus, to understand the condition of that particular pipe segment, the duplicate inspections must be properly combined. This process is usually neglected due to its complexity over time. As a result, duplicate inspection data has often been deleted in the data cleaning process for later analysis (Salman 2010).

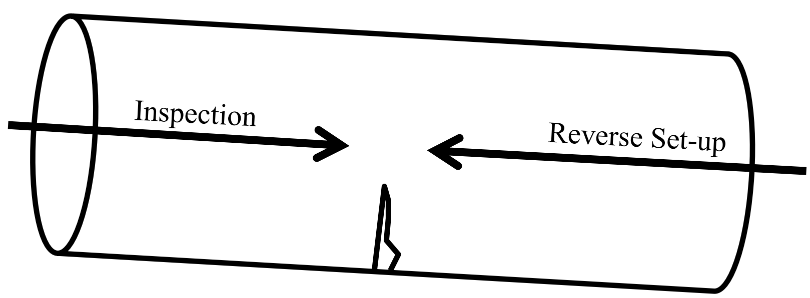


Figure 9. Inspection of the Pipe in Two Different Directions Due to Obstruction

Pipe ID

In order to identify different pipe segments in the network, it is important to develop unique identifications (ID) for each of them. Standard practice for defining unique IDs for all the sewer pipe assets is needed to avoid duplicate inputs and reduce the confusing information in the database and facilitate the process of data integration. It is beneficiary to maintain an asset inventory of sewer systems to manage the information collected through inspection, maintenance, repairs, and rehabilitation (USEPA 2016). There are two approaches to assign the unique IDs to sewer pipes:

Manhole Numbers

Manholes are considered as the foundation of sanitary sewer systems by connecting the sewer pipes. Since manholes are required by the PACP manual and considered unique for linking sewer pipes, they could be proper identifiers for the pipe segments. Therefore, assigning unique numbers to the manholes should be the first step in developing an asset inventory for the sewer systems. It is very important to implement a systematic method to assign IDs to the manholes to avoid any duplicates in the asset inventory. There are some factors to consider for developing manhole numbers (USEPA 2016):

- system flow patterns
- geographical or topographical features
- the mainline, trunk line, or force main segments
- existing parcel maps that divide the city or town into districts
- gridding the mapped area by dividing it into squares with alphabetic characters on one axis and numeric characters on the other (more commonly used by large systems)

Pipe Segment Reference (PSR)

PACP manual has suggested defining a unique Pipe Segment Reference (PSR) number for each pipe segment. PSR is mainly developed based on the upstream manhole number and it is suggested to use a suffix to distinguish between sewer pipes that exit from the same manhole (usually A, B, or C). However, PSR is not mandatory in PACP which makes it an unreliable ID for pipe segment (NASSCO 2010).

Since upstream and downstream manhole numbers are required by PACP, which are the unique IDs developed for each manhole by the utilities, manhole numbers were considered as the main referencing IDs for the pipe segments. However, after evaluation of the collected PACP sewer data, it has been noticed that there could be inaccuracies in manhole numbering in the sewer pipe systems. Thus, PSR remains the secondary checkpoint to differentiate the pipe segments in the system. The advantage of using both numbering systems is described in the next section. As a result, a new Pipe ID was developed for each pipe segment based on the upstream and downstream manhole numbers:

Pipe ID = "USMH – DSMH"

where,

USMH = Upstream Manhole Number

DSMH = Downstream Manhole Number

This coding system would help in the identification of the pipe segments and evaluation of the inspection data.

DATA QUALITY ASSURANCE FRAMEWORK

The goal of data quality assurance (QA) is to reach a high level of accuracy in the sewer inspection data and make it consistent with other datasets. This process is a significant step in developing a sewer data inventory by integrating existing datasets. Any quality problems with current datasets must be identified and resolved prior to data integration (Pipino et al. 2002). In order to implement the QA process for the collected sewer inspection data, a framework was developed in the form of a flowchart to identify and resolve the problems in the collected sewer inspection data based on a contextual approach. The flowchart can identify, measure, and resolve the data quality problems within the database and the output will be divided into three different categories as described in Figure 10.

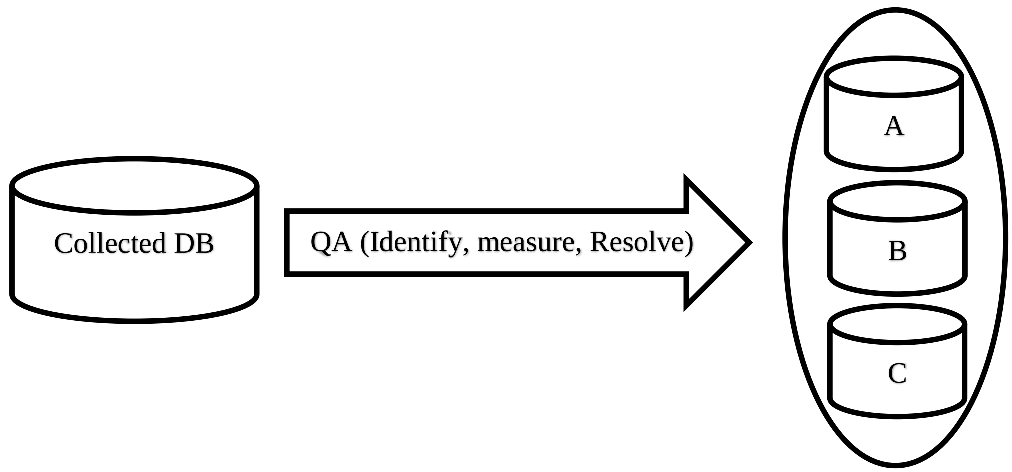


Figure 10. Data quality assurance process

A = Original data will be kept
B = Data requires further evaluation
C = Data quality resolution has been applied

PACP Data Evaluation

In order to develop proper QA procedures for PACP datasets, the collected data were reviewed to detect any data quality problems. By reviewing 100,000 inspection records with the help of NASSCO consultants, it was discovered that duplicate inspections for the same pipe are a major issue that must be addressed. It is possible that the pipe was inspected several times on different occasions; however, some inspections occurred within a very short period of time that has been defined as within 60 days. Also, some other issues were recognized in PACP datasets such as:

1. Upstream and downstream manhole numbers were used in reverse order
2. Different PSRs were assigned to the pipes with the same Upstream and downstream manhole numbers
3. Same PSRs were assigned to the pipes with different Upstream and downstream manhole numbers
4. Mislabeled upstream and downstream manhole numbers

Flowchart Description

After the evaluation of the collected PACP databases, the quality assurance flowchart has been introduced to categorize inspections and identify the proper solutions for each quality issue within the inspection data (Figure 11). The first step was to check if there was any pipe segment with reversed manhole numbers detected in the database (Step 1). Although the Pipe ID was defined as “USMH – DSMH," some pipe segments had the same manhole numbers, but in reverse order, such as USMH of Pipe1 = DSMH of Pipe2, and DSMH of Pipe1 = USMH of Pipe2.

In most cases, the pipes with similar manhole numbers in reverse order were not the same, and further evaluation of those inspections was required by the data owner to determine the cause of this issue. This problem usually happens on old inspections, which accounted for less than 2% of the inspections in the datasets obtained for this research. This issue can be easily prevented by developing a proper manhole numbering for the sewer system. Table 15 shows an example of inspection records with reverse manhole numbering.

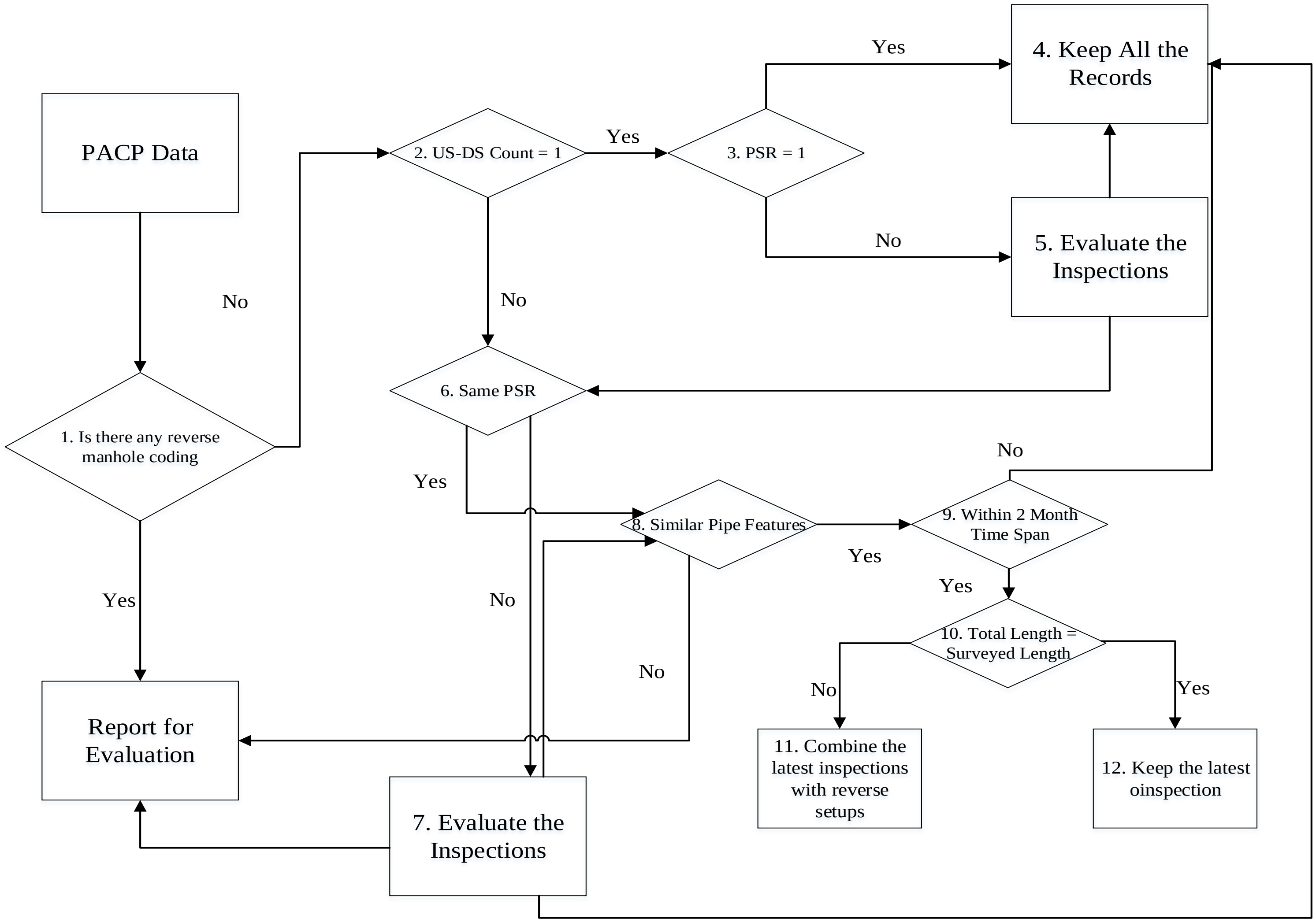


Figure 11. Data Quality Assurance Flowchart

Table 15. Inspection Records with Reverse Manhole Numbering

Inspection ID	**668**	**670**	**821**	**822**
Pipe Segment	V30 022 X	V30 022 X	V30 022 A	V30 022 A
Date	20080411	20080411	20080624	20080624
Time	9:10	9:55	10:55	11:10
Street	Elm Drive	Elm Drive	Elm Drive	Elm Drive
USMH	**V30 021**	**V30 021**	**V30 022**	**V30 022**
DSMH	**V30 022**	**V30 022**	**V30 021**	**V30 021**
Sewer Use	Sanitary	Sanitary	Sanitary	Sanitary
Direction	Downstream	Upstream	Downstream	Upstream
Height	8	8	12	12
Shape	Circular	Circular	Circular	Circular
Material	RCP	RCP	RCP	RCP
Length Surveyed	141.7	150.4	140.3	212.5
Location Code	Yard	Yard	Highway	Highway

Note: The first column of the table represents the table header.

In Step 2 of the QA process, the data is divided based on single inspections and duplicate inspections by comparing the new Pipe ID, "USMH – DSMH". The single inspections are those that have only one inspection record (Step 3). Since there are not any available criteria to evaluate the accuracy of these records, the data record is kept and only the length of the pipe can be verified by evaluation of the total surveyed length (Step 4). However, there are some inspections in this category that have similar PSRs with other inspections although the manhole numbers are different (Table 16). These inspections were labeled by the similar PSR and in many cases, there were different pipes with mislabeled PSRs due to similar upstream manholes (Step 5). These pipes were considered individual inspections and the PSR was updated for them, as seen in Inspection 90 in Table 16.

Pipes with several inspections were also evaluated based on the similarity of PSRs (Step 6). Although manhole numbers were unique numbers assigned to the pipes, there were several cases in which the pipes with the same manhole numbers had a different PSR (Table 17).

These inspections were evaluated if either the PSR was mislabeled or if they were different pipe segments (Step 7). The evaluation process was based on the comparison of pipe features such as material, dimensions, and length, and location details. After the evaluation process, the inspection records were categorized as single or duplicate inspections, or eliminated from the database and reported for further investigation by municipalities.

To evaluate duplicate inspections, the accuracy and validity of the duplicate inspections had to be verified (Step 8). All the data in the inspection table were compared to validate the fixed attributes of the data, such as length and location. Table 18 illustrates invalid data of the inspection based on the length of the pipe with InspectionID of 2358.

Table 16. Inspections with Same PSR and Different Manhole Numbers

InspectionID	36	86	90
Pipe_Segment_Reference	**M15 119 X**	**M15 119 X**	**M15 119 X**
Date	20061116	20061115	20061117
Time	11:26	11:22	9:10
Street	4820 S LAFAYETTE	4820 S LAFAYETTE	MAPLE GROVE
Upstream_MH	**M15 119**	**M15 119**	**M15 119**
Downstream_MH	**M15 124**	**M15 124**	**M15 120**
Sewer_Use	Combined	Combined	Combined
Direction	Downstream	Upstream	Downstream
Height	8	8	8
Shape	Circular	Circular	Circular
Material	Clay Tile	Clay Tile	VCP
Total_Length	151.9		247.2
Length_Surveyed	151.9	36.4	247.2
	Same Pipes		**Different Pipe**

Table 17. Pipes with same Manhole Numbers and Different PSR

InspectionID	228	3457	3458	3897	3898
PSR	**L11 090 X**	**L11 090 X**	**L11 090 X**	**L11 090 F**	**L11 090 F**
Date	20070807	20120328	20120328	20120905	20120905
Time	11:26	9:34	9:34	12:21	12:21
Street	4103 South Wayne	4103 South Wayne	4103 South Wayne	4039 South Wayne	4039 South Wayne
Upstream_MH	**L11 090**	**L11 090**	**L11 090**	**L11 090**	**L11 090**
Downstream_MH	**L11 096**	**L11 096**	**L11 096**	**L11 096**	**L11 096**
Sewer_Use	Combined	Combined	Combined	Combined	Combined
Direction	Upstream	Downstream	Upstream	Upstream	Downstream
Height	12	12	12	12	12
Shape	Circular	Circular	Circular	Circular	Circular
Material	VCP	VCP	VCP	VCP	VCP
Length_Surveyed	414.8	295.5	119.6	121.6	292.7

Table 18. Length Comparison of the Similar Pipes

InspectionID	2350	2351	2357	2358
Pipe_Segment_Reference	**M15 177 X**	**M15 177 X**	**M15 177 X**	**M15 177 X**
Date	20100810	20100810	20100816	20100816
Time	10:48	13:50	8:56	10:24
Upstream_MH	**M15 178**	**M15 178**	**M15 178**	**M15 178**
Downstream_MH	**M15 177**	**M15 177**	**M15 177**	**M15 177**
Sewer_Use	Combined	Combined	Combined	Combined
Direction	Upstream	Downstream	Upstream	Downstream
Height	8	8	8	8
Shape	Circular	Circular	Circular	Circular
Material	VCP	VCP	VCP	VCP
Total_Length				**56.6**
Length_Surveyed	**342**	**24**	**389.1**	**56.6**

After verification of duplicate inspections and elimination of the odd inspections, the remaining inspections were evaluated to either keep the original data or combine it with the reverse inspection (Step 9). The reverse inspection was recognized through the direction of the inspection (Figure 9). In many cases, the inspector mentioned “reverse set-up” in the remark section in the PACP database. The duplicate data were assessed based on a comparison of the total length of the pipe and the total length of the survey (Step 10). If these two were equal, the latest complete inspection was kept in the dataset (Step 11); otherwise, the date of the inspection identified the relationship among the duplicate data (Step 12). The time span for comparing the inspection records was within two months (60 days). Although several reverse inspections may have been completed within several days, the severity of the defects that caused the abandonment of the inspection required time to resolve. Table 19 summarizes the time spans for the completion of the inspections. Also, the cause of the incomplete inspections and abandoning the process was evaluated for a proper process of combining inspections.

FRAMEWORK IMPLEMENTATION AND RESULTS

As it was mentioned before, the contextual data quality assurance framework is valuable if it can be automated to detect the errors in the database and resolve them. Manual data cleaning is time-consuming and will result in human errors. Therefore, automated algorithms have been developed to apply the flowchart to the collected inspection databases. The SQL queries have been developed for the MySQL environment and automated in Python that resulted in nearly 4,000 lines of codes. This process will be explained in the following sections.

Data Migration

The sewer inspection data is collected through different software platforms and usually stored in MS-ACCESS format (.mdb). There are some other formats but we found that the .mdb format is used more than 95% of the time. Based on the limitations of the MS-ACCESS it was decided to move the databases into MySQL which is an open-source relational database management system that can facilitate data manipulation and data warehousing. Also, MySQL is known to be one the best platforms for a web database and application that was the final goal of this research.

Table 19. Time Span for a Complete of Sewer Pipe Inspection

InspectionID	565	569	573	686	687
PSR	Q02 169 X	Q02 169 X	Q02 169 X	Q02 169 X	Q02 169 X
Date	**20080304**	**20080303**	**20080307**	**20080417**	**20080417**
Time	9:24	12:42	11:47	8:59	8:59
Upstream_MH	Q02 169	Q02 169	Q02 169	Q02 169	Q02 169
Downstream_MH	Q02 165	Q02 165	Q02 165	Q02 165	Q02 165
Sewer_Use	Combined	Combined	Combined	Combined	Combined
Direction	Downstream	Downstream	Downstream	Upstream	Downstream
Height	24	24	24	24	24
Shape	Circular	Circular	Circular	Circular	Circular
Material	VCP	VCP	VCP	VCP	VCP
Total_Length	333.1	333.1	333.1	333.1	333.1
Length_Surveyed	54.8	30	86	186.5	333.1

Since the MySQL Connector/ODBC did not transfer data due to incompatibility issues of data types, data rules, indexes, and references, the data has been transferred through third-party software, “BullZip MS ACCESS to MySQL”. Though the migration process was simply done, the software dropped all the relationships, keys, and data type features to make the transfer process possible. So the first step was to re-establish the relationships and make the data the same as the original format. Then the tables related to LACP (Lateral Assessment and Certification Program) were dropped because there were not related to the study and all the fields were Null. The sample of the codes is as follows:

```
-- Delete LACP Tables
DROP TABLE
   `valid_lacp_codes`,
   `valid_lacp_lining_methods`,
   `valid_lacp_locations`,
   `valid_lacp_materials`,
   `lacp_conditions`,
   `lacp_custom_fields`,
   `lacp_custom_labels`,
   `lacp_inspections`,
   `lacp_media_conditions`,
   `lacp_media_inspections`,
   `valid_start_manhole`;

-- Inspection Table Connections
ALTER TABLE `inspections`
  ADD FOREIGN KEY (`Purpose`) REFERENCES `valid_purposes` (`Value`);
ALTER TABLE `inspections`
  ADD FOREIGN KEY (`Weather`) REFERENCES `valid_weather` (`Value`);
ALTER TABLE `inspections`
  ADD FOREIGN KEY (`GPS_Accuracy`) REFERENCES `valid_accuracyofgps` (`Value`);
ALTER TABLE `inspections`
  ADD FOREIGN KEY (`Pre_Cleaning`) REFERENCES `valid_pre_cleaning` (`Value`);
ALTER TABLE `inspections`
  ADD FOREIGN KEY (`Direction`) REFERENCES `valid_survey_directions` (`Value`);
ALTER TABLE `inspections`
  ADD FOREIGN KEY (`Flow_Control`) REFERENCES `valid_flow_controls` (`Value`);
ALTER TABLE `inspections`
  ADD FOREIGN KEY (`Shape`) REFERENCES `valid_shapes` (`Value`);
ALTER TABLE `inspections`
  ADD FOREIGN KEY (`Lining_Method`) REFERENCES `valid_lining_methods` (`Value`);
ALTER TABLE `inspections`
  ADD FOREIGN KEY (`Material`) REFERENCES `valid_materials` (`Value`);
ALTER TABLE `inspections`
  ADD FOREIGN KEY (`Location_Code`) REFERENCES `valid_locations` (`Value`);
ALTER TABLE `inspections`
  ADD FOREIGN KEY (`Sewer_Use`) REFERENCES `valid_sewer_uses` (`Value`);
```

Python Automation

Each task has been written as a "Procedure" in MySQL. It means that queries will be stored in each database and then will be called to make the required changes. For this purpose, a template has been defined in Python to facilitate this process and automate the procedure calls. This is a complicated process that connects to MySQL through "mysql.connector" and keeps the connection and executes statements through cursor(). Each procedure with its call query was stored separately. Finally, they executed in a sequence to create the procedures first and then call them to make changes. Using this approach, the data quality assurance framework can be applied to the sewer data automatically. The sample of codes are as follows:

```
import mysql.connector
from mysql.connector.cursor import MySQLCursor
from pprint import pprint

conn = mysql.connector.connect(user= u, password= p,
                    host='127.0.0.1',
                    database= db,use_pure=True)

cursor = conn.cursor()

# query with format parameter style

sql1 = """CREATE PROCEDURE `Sample` ()
BEGIN

.
.
.

END"""

sql2 = """CALL `PACP Relationship`()"""

queries = [sql1, sql2]
```

```
# returns an iterator
results = cursor.execute(";".join(queries),  multi=True)

count = 1

for result in results:

    # result is a cursor object i.e result == cursor
    # so we have access to all cursor attributes and methods

    print("Query {0} - {1} :".format(count, result.statement))

    # does query has result?
    if result.with_rows:
        for row in result:
            print(row)
        count = count + 1
    else:
        print("No result found")
        count = count + 1

    print()

cursor.close()
conn.close()
```

Defining the flowchart

It was required to define the flowchart in Figure 11 in the SQL environment. SQL is the standard language for relational database management systems (RDBMS). RDBMS is a common type of database that stores structured data in tables that can be used in relation to other stored databases (ANSI 1986). These tables are divided into rows known as records and columns known as fields. In order to apply the developed flowchart on the collected databases, each step was implemented in a procedure and the results were been stored in a new table. So, it was required to define the required tables before executing the procedures (Table 20). All the tables were created in one procedure and modified by the related queries.

Table 20. Flowchart Tables

Table	Table Description
`inspections_Total_Length`	This table will have data for corrected Total Length from the Conditions table based on the PACP_code=AMH filter
`inspections_Upstrm_dwnstrm`	This contains data with new columns by concatenating upstream_downstream and downstream_upstream columns
`inspections_Upstrm_dwnstrm_COMB`	This table contains data of `UpstrmMH_DstrmMH` = `DstrmMH_UpstrmMH` OR `DstrmMH_UpstrmMH` = `UpstrmMH_DstrmMH` (Report for Evaluation to check Is there any reverse manhole coding)
`inspections_Upstrm_dwnstrm_nt_equal`	This table contains data of `UpstrmMH_DstrmMH` <> `DstrmMH_UpstrmMH` OR `DstrmMH_UpstrmMH` <> `UpstrmMH_DstrmMH`
`Cnt_UpstrmMH_DstrmMH`	This is the intermediate table (which has the count for upstream_dwnstreams) and created to build the next table which has Upstream_downstream count =1
`ins_up_dwn_cnt_1`	This is the intermediate table(which has the count for inspections,upstream_dwnstreams)and created to build the next table which has Upstream_downstream count =1
`inspections_Upstrm_dwnstrm_cnt_1`	This table contains the data which has Upstream_downstream count=1
`Cnt_PSR`	This is the intermediate table (which has the count for Pipe_Segment_Reference) and created to build the next table which has Pipe_Segment_Reference =1
`ins_psr_cnt_1`	This is the intermediate table (which has the count for inspections, Pipe_Segment_Reference) and created to build the next table which has PSR count =1
`inspections_psr_eql1_3_4`	This table contains the data which has PSR count = 1
`inspections_psr_nt_eql1_5`	This table contains the data which has Pipe_Segment_Reference count <> 1
`inspections_upstrm_dwnstrm_nteql1`	This table contains the data which has Upstream_downstream count <> 1
`sud_access`.`Cnt_PSR_gtr_1`	These 3 tables are intermediate tables which are created to build the below 2 tables which check for upstream_downstream not equal with PSR and upstream_downstream equal with PSR
`sud_access`.`Cnt_udstrm_gtr_1`	
`sud_access`.ins_updstrm_psr_cnt_chk	
`sud_access`.`ins_updtrm_psr_nt_match_7`	This table contains data that has upstream_downstream count <>1 and does not have the same PSR(Evaluate the inspections)
`sud_access`.`ins_updstrm_psr_cnt_match_6`	This table contains data that has upstream_downstream count <>1 and has the same PSR
`sud_access`.`ins_simlr_pipe_ftrs_8`	This table contains data that has upstream_downstream count <>1 and has the same PSR and has similar pipe features.
`sud_access`.`ins_date_time_9`	This table contains the data which has inspections within a 2-month timespan
`sud_access`.`ins_Totlength_10`	This table contains the data where TotLength >= 1% of Surveyed Length OR TotLength <= 1% of Surveyed Length OR TotalLength equal to Surveyed Length
`sud_access`.`ins_Totlength_10_nt_eql`	This table contains the data where TotLength <> 1% of Surveyed Length AND TotLength <> 1% of Surveyed Length AND TotalLength <> Surveyed Length

`ins_Totlength_labl_Dirctns`	This table contains the data for duplicate records that have been identified by the same label
`ins_New_Totlength_labl_Dirctns`	This table defines the new total length for duplicate entries based on the length_surveyed
`ins_Totlength_revrse_Dirctns`	These are the intermediate table to calculate the new distance of the reverse inspection and assign the new total length to the combining process of duplicate records
`ins_Totlength_revrse_Dirctns_eql`	
`ins_Totlength_revrse_Dirctns_grtr`	
`ins_Totlength_revrse_Dirctns_null`	
`ins_Totlength_revrse_Dirctns_cals`	
`ins_New_Totlength_revrse_Dirctns`	
`inspections_flg`	This table contains the data that flags the duplicate entries to combine the condition records based on the new total length
`inspections_mrn_thn_2`	This table contains data where there are more than 2 inspections and the inspections are not in `ins_New_Totlength_revrse_Dirctns`
`conditions_new`	This table contains the new condition data based on the integration of duplicate inspections

Also, two variables are required to be calculated. It has been noticed that the total length of the pipe could have errors or missing values in the database and it can be updated if there is a complete survey from manhole to manhole. Moreover, the number of surveys for each pipe in 60 days time span needs to be calculated for duplicate evaluation. As a result, two separate codes have been written to calculate these two variables. Since “mysql.connector” didn’t work properly for these calculations, “sqlalchemy” has been used along with “pandas”, “numpy”, and “pymysql”. The results have been stored in `'inspections_totlength` and `datetime_9`.

Results

Table 21 shows the distribution of the sewer inspections for the four case studies based on the proposed quality assurance flowchart. In the previous studies, only the inspections with unique upstream and downstream manhole numbers were evaluated for deterioration models, which was 50% to 75% of the database. However, by applying the proposed data quality assurance process, approximately 90% of the data was available for condition assessment of the sewer pipes.

The total lengths of pipes are being calculated by the GPS location attributes. These attributes can always be miscalculated due to human errors. The length surveyed is the actual length that a CCTV camera travels to complete the inspection. So, by assigning the length surveyed of the complete inspection (manhole to manhole), an accurate length of the pipe can be determined. For example, for the database from City E with 212 inspections, 136 total lengths were updated which was nearly 65% of the data. Eight of them were data errors and the rest were missing values (Table 22).

Table 21. Distribution of the Inspections in Four Cities Based on the QA Flowchart

	City A	City B	City C	City D
No. of Pipes with Reverse Manhole Coding	38 (1%)	13 (0%)	230 (0%)	298 (1%)
USMH-DSMH Count = 1 & PSR = 1	3835 (73%)	2644 (75%)	24449 (53%)	27263 (67%)
USMH-DSMH Count = 1 & PSR > 1	12 (0%)	0 (0%)	9 (0%)	108 (0%)
Duplicate Inspections with & Different PSR	461 (9%)	148 (4%)	1056 (2%)	506 (1%)
Duplicate Inspections with & Same PSR	886 (17%)	698 (20%)	20347 (44%)	12791 (31%)
Total	5232	3503	46091	40966

Table 22. Total length errors

InspectionID	Updated_Length	Total_Length	Difference
1	364.6	355.5	9.1
2	92.8	92.4	0.4
3	185.3	188	2.7
4	220	219.8	0.2
5	222	223.3	1.3
8	382.5	359.4	23.1
9	258	252.6	5.4
12	590	600.9	10.9

In order to evaluate the quality assurance flowchart, the collected sewer data in PACP format from one city with 212 inspections were used. A unique ID was developed for each inspection based on the manhole codes. According to upstream and downstream manhole comparisons, 160 inspections were categorized under step 3 (unique manhole numbers and PSR). None of these inspections were done on the same pipe and the process kept all the inputs for those corresponding inspections.

To evaluate the other 52 inspections, the dates and lengths of inspections were compared. There were 2 pipes with 3 separate inspections and 23 pipes with 2 inspections (Table 23). Table 24 shows the results of the data evaluation process based on flowchart steps. Steps 10 and 12 are the result of a single inspection, while Step 11 is the combination of 46 duplicated inspections. All these inspections were done in different directions due to obstruction occurrences during the inspection (Figure 9). Thus, there were not any duplicate condition inputs in the final dataset. As a result, only 2 first inspections of Step 12 will be deleted from the final database since the latest inspection has been performed on the entire pipe and the other 2 inspections are not useful for evaluating pipe condition (Table 24).

As a result of implementing the proposed QA process, the final dataset consisted of 187 (160+27) inspections, which provides high-quality data for further sewer asset management processes. In previous studies, the most common data cleaning process was to eliminate any data with missing values and duplicates which help to avoid misinterpretation of the data. However, those studies neglected some pipes with serious structural and operational problems. By applying the quality assurance process developed in this study, there were just two deleted inspections compared to 52 deleted inspections based on previous QA practices.

Table 23. Duplicate Records

ID	PSR	Date	USMH	DSMH	Direction	Height	Shape
6	A006_A005	3/23/2011	A006	A005	Downstream	10	Circular
7	A006_A005	3/23/2011	A006	A005	Upstream	10	Circular
10	A009_A008	2/24/2011	A009	A008	Downstream	10	Circular
11	A009_A008	2/24/2011	A009	A008	Upstream	10	Circular
20	A017_A016	2/22/2011	A017	A016	Downstream	8	Circular
21	A017_A016	2/22/2011	A017	A016	Upstream	8	Circular
50	A122_A121	4/4/2011	A122	A121	Upstream	8	Circular
51	A122_A121	4/5/2011	A122	A121	Downstream	8	Circular
54	A126_A125	4/4/2011	A126	A125	Downstream	8	Circular
55	A126_A125	4/4/2011	A126	A125	Upstream	8	Circular
66	A137_A136	8/26/2004	A137	A136	Downstream	8	Circular
211	A137_A136	8/26/2004	A137	A136	Upstream	8	Circular
212	A137_A136	8/26/2004	A137	A136	Downstream	8	Circular
69	A140_A139	4/5/2011	A140	A139	Downstream	8	Circular
70	A140_A139	4/5/2011	A140	A139	Upstream	8	Circular
74	A144_A143	3/29/2011	A144	A143	Upstream	8	Circular
75	A144_A143	3/29/2011	A144	A143	Downstream	8	Circular
101	A222_A208	3/28/2011	A222	A208	Upstream	8	Circular
102	A222_A208	3/28/2011	A222	A208	Downstream	8	Circular
103	A223_A222	3/28/2011	A223	A222	Downstream	8	Circular
104	A223_A222	3/28/2011	A223	A222	Upstream	8	Circular
105	A224_A223	3/28/2011	A224	A223	Downstream	8	Circular
106	A224_A223	3/28/2011	A224	A223	Upstream	8	Circular
108	A226_A225	4/5/2011	A226	A225	Upstream	8	Circular
109	A226_A225	4/5/2011	A226	A225	Upstream	8	Circular
110	A227_A209	4/5/2011	A227	A209	Downstream	8	Circular
111	A227_A209	4/5/2011	A227	A209	Upstream	8	Circular

117	A403_A402	3/22/2011	A403	A402	Downstream	8	Circular
118	A403_A402	3/22/2011	A403	A402	Upstream	8	Circular
121	A501_A500	3/22/2011	A501	A500	Downstream	8	Circular
122	A501_A500	3/22/2011	A501	A500	Upstream	8	Circular
132	M009_M008	2/22/2011	M009	M008	Downstream	8	Circular
133	M009_M008	2/22/2011	M009	M008	Upstream	8	Circular
138	M013_M012	12/20/2010	M013	M012	Downstream	8	Circular
139	M013_M012	12/20/2010	M013	M012	Upstream	8	Circular
144	M018_M017	12/17/2010	M018	M017	Upstream	6	Circular
145	M018_M017	2/16/2011	M018	M017	Downstream	6	Circular
147	M0H3_MOH2	2/21/2011	M0H3	MOH2	Downstream	8	Circular
148	M0H3_MOH2	2/21/2011	M0H3	MOH2	Upstream	8	Circular
155	M201_M200	12/6/2010	M201	M200	Upstream	8	Circular
156	M201_M200	12/6/2010	M201	M200	Downstream	8	Circular
210	M201_M200	9/22/2011	M201	M200	Upstream	6	Circular
160	M303_M302	2/16/2011	M303	M302	Downstream	8	Circular
161	M303_M302	2/16/2011	M303	M302	Upstream	8	Circular
169	M402_M401	12/17/2010	M402	M401	Upstream	6	Circular
170	M402_M401	12/17/2010	M402	M401	Downstream	6	Circular
171	M403_M401	12/17/2010	M403	M401	Upstream	8	Circular
172	M403_M401	12/17/2010	M403	M401	Downstream	8	Circular
178	MPS2_M001	12/6/2010	MPS2	M001	Upstream	8	Circular
179	MPS2_M001	2/16/2011	MPS2	M001	Downstream	8	Circular
203	R103_R102	12/20/2010	R103	R102	Upstream	8	Circular
204	R103_R102	12/21/2010	R103	R102	Downstream	8	Circular

Table 24. Summary of Duplicate Inspections

QA Groups	**Number**
Step 10	3
Step 11	23
Step 12	3

CONCLUSIONS & RECOMMENDATIONS

The lack of high-quality data is mentioned by several studies on deterioration models of the sewer systems (Scheidegger et al. 2011). Thus, developing a standard data protocol to collect sewer pipes inspection data and a quality assurance process to improve the accuracy of the databases is vital for sewer systems asset management. In previous studies, the most common data cleaning process was to eliminate missing and duplicate data which helps to avoid misinterpretation of the data. However, this "heavy-handed" approach neglected some pipes with serious structural and operational problems. The Quality Assurance procedure presented here is a unique approach to develop a final database with a high level of accuracy. This framework is developed based on contextual evaluation of the current database and providing a logical approach to address the main problems in sewer inspection databases. The goal of this process is to preserve most of the sewer inspection data and improve the quality of the data for further sewer asset management processes. Four PACP databases with more than 90,000 inspections were evaluated to determine the major data quality issues. After implementation of the data quality assurance process, the percentage of good quality inspection data increased from 50%-75% (pre-process) to 95% (post-process). Also, the quality assurance flowchart is implemented on a small database to evaluate its effectiveness.

In order to evaluate the efficiency of the data quality assurance framework, more databases should be collected to validate the proposed data quality assurance flowchart. This flowchart will facilitate the quality improvement process for sewer inspection data specifically PACP data and make the PACP data integration feasible. By implementation of this flowchart, the information about the pipes with severe defects that should actually be included in the dataset for a more thorough analysis can be included in the asset management.

REFERENCES

Abraham, D., and Wirahadikusumah, R. (1999). "Development of prediction models for sewer deterioration." Durability of Building Materials and Components, 8.

Alipour-Langouri, M., Zheng, Z., Chiang, F., Golab, L., and Szlichta, J. "Contextual data cleaning." Proc., 2018 IEEE 34th International Conference on Data Engineering Workshops (ICDEW), IEEE, 21-24.

ANSI, X. (1986). "American National Standard for Information Systems: Database Language SQL." American National Standards Institute, NY, 135.

Ariaratnam, S. T., El-Assaly, A., and Yang, Y. (2001). "Assessment of infrastructure inspection needs using logistic models." Journal of Infrastructure Systems, 7(4), 160-165.

Baskarada, S., Gao, J., Lin, S., Yeoh, G. S., and Koronios, A. (2005). "Data Quality Enhancing Software for Asset Management-State of the Art Evaluation."

Berti-Equille, L., Dasu, T., and Srivastava, D. "Discovery of complex glitch patterns: A novel approach to quantitative data cleaning." Proc., 2011 IEEE 27th International Conference on Data Engineering, IEEE, 733-744.

Bertossi, L., and Milani, M. (2018). "Ontological multidimensional data models and contextual data quality." Journal of Data and Information Quality (JDIQ), 9(3), 1-36.

Caradot, N., Rouault, P., Clemens, F., and Cherqui, F. (2018). "Evaluation of uncertainties in sewer condition assessment." Structure and Infrastructure Engineering, 14(2), 264-273.

Chughtai, F., and Zayed, T. (2008). "Infrastructure condition prediction models for sustainable sewer pipelines." Journal of Performance of Constructed Facilities, 22(5), 333-341.

Dasu, T., and Loh, J. M. (2012). "Statistical distortion: Consequences of data cleaning." arXiv preprint arXiv:1208.1932.

English, L. P. (1999). Improving data warehouse and business information quality: Methods for reducing costs and increasing profits, Wiley, New York.

Even, A., and Shankaranarayanan, G. (2007). "Utility-driven assessment of data quality." ACM SIGMIS Database: the Database for Advances in Information Systems, 38(2), 75-93.

Hamilton, S., Jennings, A., and Forster, A. J. (2020). "Development and evaluation of a quality improvement framework for healthcare." International Journal for Quality in Health Care, 32(7), 456-463.

Heinrich, B., Kaiser, M., and Klier, M. (2007). "How to measure data quality? A metric-based approach."

Khan, Z., Zayed, T., and Moselhi, O. (2009). "Structural condition assessment of sewer pipelines." Journal of performance of constructed facilities, 24(2), 170-179.

Kleindienst, D. (2017). "The data quality improvement plan: deciding on choice and sequence of data quality improvements." Electronic Markets, 27(4), 387-398.

Kraft, S., Carayon, P., Weiss, J., and Pandhi, N. (2015). "A simple framework for complex system improvement." American Journal of Medical Quality, 30(3), 223-231.

Lee, Y. W., Strong, D. M., Kahn, B. K., and Wang, R. Y. (2002). "AIMQ: A methodology for information quality assessment." Inf. Manage., 40(2), 133–146.

National Association of Sewer Service Companies (NASSCO) (2010). Pipeline Assessment and Certification Program (PACP): Reference manual, Volume 6.0.1, Marriottsville, MD

Nousak, P., and Phelps, R. (2002). "A Scorecard approach to improving Data Quality." Proceedings of Data Warehousing and Enterprise Solutions, Sugi-27, Orlando, Florida.

Pipino, L. L., Lee, Y. W., and Wang, R. Y. (2002). "Data quality assessment." Communications of the ACM, 45(4), 211-218.

Prokoshyna, N., Szlichta, J., Chiang, F., Miller, R. J., and Srivastava, D. (2015). "Combining quantitative and logical data cleaning." Proceedings of the VLDB Endowment, 9(4), 300-311.

Salman, B. (2010). Infrastructure management and deterioration risk assessment of wastewater collection systems, Doctoral dissertation, University of Cincinnati, Cincinnati, OH

Scheidegger, A., Hug, T., Rieckermann, J., & Maurer, M. (2011). "Network condition simulator for benchmarking sewer deterioration models." Water Research, 45(16), 4983-4994.

Shah, S., Horne, A., and Capellá, J. (2012). "Good data won't guarantee good decisions." Harvard business review, 90(4), 23-25.

Taylor, M. J., McNicholas, C., Nicolay, C., Darzi, A., Bell, D., and Reed, J. E. (2014). "Systematic review of the application of the plan–do–study–act method to improve quality in healthcare." BMJ quality & safety, 23(4), 290-298.

USEPA (2016). "Developing a Manhole or Catch Basin Numbering System." Eliminating Sanitary Sewer Overflows in New England, <https://www3.epa.gov/region1/sso/manhole-id.html>. (Information viewed on May 21, 2021.

Van Riel, W., Langeveld, J., Herder, P., and Clemens, F. (2014). "Intuition and information in decision-making for sewer asset management." Urban Water Journal, 11(6), 506-518.

Van Riel, W., van Bueren, E., Langeveld, J., Herder, P., and Clemens, F. (2016). "Decision-making for sewer asset management: Theory and practice." Urban Water Journal, 13(1), 57-68.

Walshe, K., and Freeman, T. (2002). "Effectiveness of quality improvement: learning from evaluations." BMJ Quality & Safety, 11(1), 85-87.

Wang, R. Y. (1998). "A product perspective on total data quality management." Communications of the ACM, 41(2), 58-65.

Westin, S., and Sein, M. K. (2013). "Improving data quality in construction engineering projects: An action design research approach." Journal of Management in Engineering, 30(3), 05014003.

CHAPTER IV

DEVELOPING INTEGRATED DATA STRUCTURE FOR SEWER PIPES ASSET MANAGEMENT

ABSTRACT

Availability of accessible and reliable asset inventory is crucial for infrastructure asset management. Although there are several improvements in data management of the transportation sector at the national level, there is no national standard database for sewer pipes inspection data. Data integration is a key component of effective asset management. New standard formats should be defined to translate the current data into a common format that can be interpreted by all the stakeholders. There is no unified data format for collected sewer inspection data. While some municipalities are using the PACP coding system, their final database is not always PACP compatible. This can happen for many reasons, such as coding modification, software issues, etc. As a result, the final inspection database could be in so many different formats. A data mapping tool was developed to address this problem by transforming data into the PACP data structure while keeping the integrity of the database. By implementing this tool one of the major issues in the industry will be addressed and the data can be viewed, modified, and analyzed independently from the generating software. Also, a normalized dimensional Sewer Inventory Schema (SIS) was developed for integrated national sewer inventory.

INTRODUCTION

The United States Environmental Protection Agency (EPA) defined infrastructure asset management as “managing infrastructure capital assets to minimize the total cost of owning and operating them while delivering the service levels customers desire” (USEPA 2002). Municipalities are confronting several issues regarding the infrastructure assets including a) deterioration of the assets, b) increasing demand level, c) new regulations, and d) inadequate maintenance and renewal budget. (Danylo and Lemer 1998; Grigg 1999; Halfawy 2004). Thus, municipalities are trying to implement effective asset management strategies through improving their infrastructure inventory (Halfawy 2008) as the first step in asset management plan development is collecting infrastructure data. This approach will help the asset management process to move from reactive to proactive manners.

Several tools have been developed to manage the infrastructure data; however, these tools often store data in an isolated format and information fragments are created, which affects the consistency and integrity of data. In order to have optimized asset management strategies, it is required to develop an integrated process to make it feasible to access the infrastructure data (Lemer 1998; Grigg 1999; Halfawy et al. 2002). Some agencies are more advanced in the integrated asset data collection process. Initiated in 1972, structural and condition assessment data of all bridges in the United States have been collected in a unified standard database established by the Federal Highway Administration (FHWA). This database is known as the National Bridge Inventory (NBI) and it contains over 600,000 bridges from across the nation. The inspection data have been collected from the states and stored by FHWA in a relational database including significant information about the bridges such as load, geometry, age, location, functional classification, average daily traffic, material and design types, structural deficiency, and functional obsolescence (Weseman, 1995). NBI is considered the main resource for condition assessment of the national bridges and enables all bridge stakeholders to access the data for bridge management and research purposes. Moreover, the data integration process has been implemented in several different transportation sectors. On the contrary, wastewater systems are way behind the highway infrastructure systems in data inventory at the national level. Currently, there is no national standard database for collecting sewer data across the country.

There is a need for a national inventory to present empirical data that accurately represents the condition of sewer systems. The benefits of such an inventory include:

1. empirical data to justify increased spending on failing infrastructure;
2. benchmarks of the national sewer infrastructure by region;
3. identification of deterioration mechanisms in sewers; and
4. a means for national dissemination of the data.

The objective of this chapter is to propose a solution to facilitate the development of the sustainable national sewer inventory. The integrated database will address the problem of interoperability of sewer inspection data developed by different PACP certified software. Moreover, the schema for the centralized database will be discussed to integrate the information required for sewer infrastructure asset management.

BACKGROUND

Previous studies on data integration of sewer systems are mainly focused on integrated data of one municipality to develop a final database that can be used in the asset management process of that municipality (Berkeley et al. 2015; Halfawy et al. 2006; Halfawy 2008; Park and Kim 2013). Figure 12 shows an example of a database diagram for the integrated sewer network data model (Halfawy 2008). Although these studies address the data gap issues within each municipality, they are not providing the national perspective of the sewer infrastructure data.

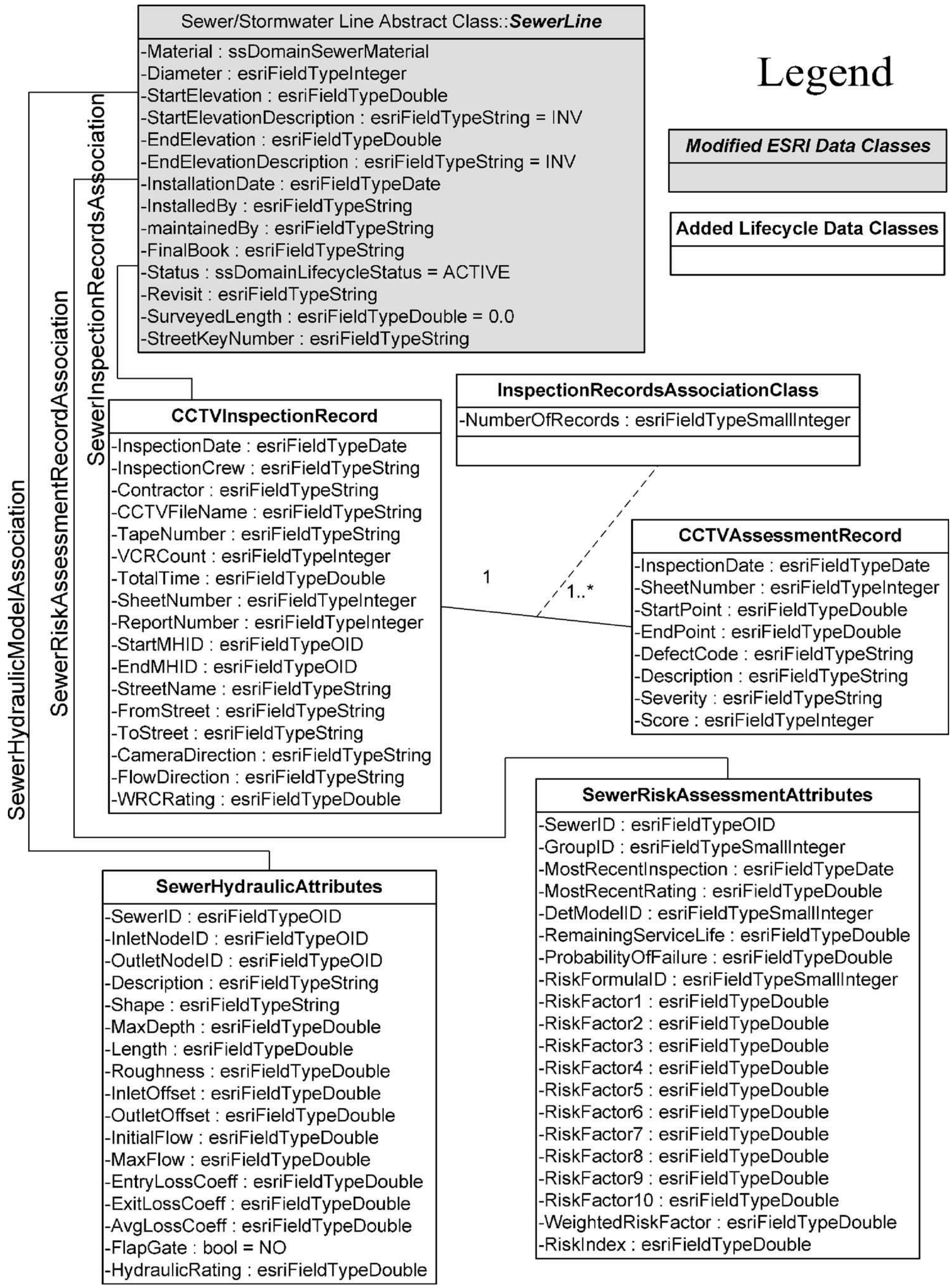


Figure 12. Database diagram for the Integrated Sewer Network Data Model (Halfawy 2008)

Infrastructure asset management is the integrated approach to sustain public infrastructure assets. The American Society of Civil Engineers classified the infrastructure assets into four categories (Water & Environment, Transportation, Public Facilities, and Energy) and 16 sub-categories. The asset management concepts have been successfully implemented in urban centers and large regional sewer collection systems in order to enhance operational, environmental, and financial performance. Many of these plans are based on sophisticated information systems and considerable personnel resources. However, much more effort is needed to put it extensively into action in small utilities. Asset management is a continuous improvement planning which can be practiced based on the current resources and can be extended as the program progresses (USEPA 2002). Asset management helps agencies develop a set of concepts to regulate the policy-making process and resource allocation. The core principles of asset management are defined as follows (NCHRP 2006):

- Policy-Driven: A well-defined set of goals and policies should be defined to indicate the proper infrastructure condition and level of service based on the available resources.
- Performance-Based: Objectives are converted into performance measures and evaluated through the monitoring process.
- Analysis of Options and Tradeoffs: Since there is always some limited funding, the outcome of different alternatives should be evaluated to select the best options.
- Decisions Based on Quality Information: Decisions are made using reliable data. Various tools are required to collect, evaluate, and analyze the data.
- Monitoring to Provide Clear Accountability and Feedback: Performance measures are monitored and reported to assess the effects of the plans. The goals and policies were reevaluated based on the actual performance measures.

The infrastructure asset management process is illustrated in Figure 13. The flowchart shows the components of asset management and their relationships. After defining the goals, objectives, and policies of the agency based on the desired infrastructure condition, the asset inventory data is the first step to make short-term and long-term plans. After collecting required data, performance modeling and condition assessment of the system provide an accurate evaluation of the current assets. The Decision-making process will be done after the prioritization of the assets and then the budget will be evaluated to determine the feasible plans.

Since asset inventory is the first step in the asset management process, it is rarely possible to get the advantages of this process without easily accessible and reliable data. Although the importance of fast access to vital data is obvious to the agencies in this "information age", the infrastructure data is still dispersed and disjointed.

In 2002, NASSCO developed the Pipeline Assessment and Certification Program (PACP) to become a standard for the evaluation of the sewer pipes condition through Closed-Circuit Television (CCTV) inspection on an industry-wide basis (NASSCO 2010). The PACP became the only industry standard for the sewer condition data and was implemented by more than 200 cities and utilities in the US and Canada (Thornhill and Wildbore 2005). However, sewer data integration is not practiced by most agencies. Also, several agencies have an adequate amount of data but the required information is not easily accessible (USEPA 2015). The benefits of asset management in sewer collection systems are as follows (USEPA 2002):

- Protecting the system components from unexpected failure through pre-planned maintenance.
- Developing proactive maintenance programs to reduce the overall costs.
- Demand forecasting and a better understanding of future needs.
- Optimizing the budget and resource allocations through alternatives evaluation using life-cycle costing and value engineering.
- Evaluating the implemented plans through performance monitoring.

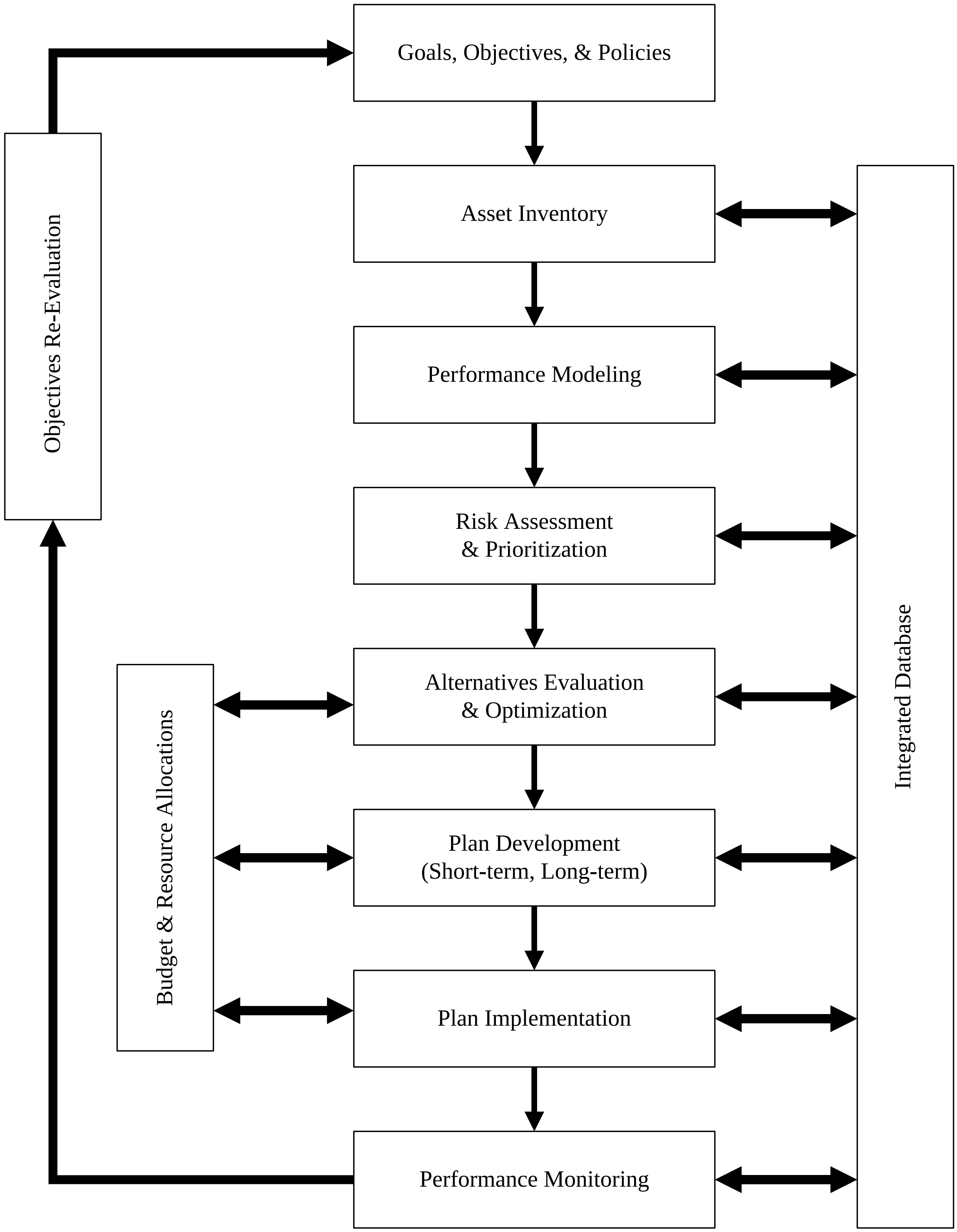


Figure 13. Infrastructure Asset Management Process (FHWA 2010)

In developing an effective sewer asset management program, understanding the current condition of the sewer systems is inevitable. Current data may not be easily accessible or in a proper format that makes it unusable for sewer pipes condition evaluation(Caradot, Rouault, Clemens, & Cherqui, 2018). A successful asset management program requires an organized and well-maintained database system. This database is also useful for understanding the trends of the sewer system's condition (USEPA 2015).

There are over 50,000 municipalities in the US that manage the collected sewer in different formats (ASCE 2000). Each data format is not compatible with other databases, and even cannot be used in other departments in the same municipalities (Sinha, 2019). During this research, it has been noticed that in most of the cases the collected inspection data is not useable for engineering departments because of accessibility and compatibility issues. Sinha et al. (2009) developed a GIS platform to collect water and wastewater data; however, the developed model mainly collects general information and not inspection data (Figure 14). The variety of data formats for sewer inspection data is the main challenge to develop a unified inventory.

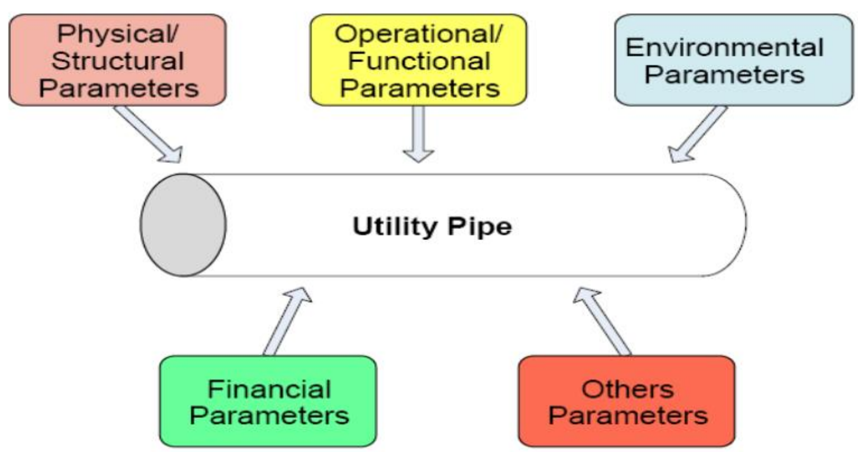


Figure 14. Pipe Data Parameters (Sinha, Dymond, Vemulapally, Dickerson, & Perry, 2009)

DATA INTEGRATION

Reliable and useful data is a foundation for effective infrastructure asset management. A large amount of data should be collected, evaluated, stored, and transferred to make a decision support tool. Thus, it is required to develop a data integration process to manage the information in a way that can be helpful for asset management. Data integration is defined as "the method by which multiple data sets from a variety of sources can be combined or linked to provide a more unified picture of what the data mean and how they can be applied to solve problems and make informed decisions" (FHWA 2010). Agencies handling large quantities of data to maintain the desired level of service through conventional operations. The accuracy and accessibility of the data are crucial for making appropriate decisions and avoiding operational problems. For sewer collection systems, lacking comprehensive sewer data leads to reactive mode – response to the system failures only. However, through evaluation of the reliable data, it might be possible to predict the pipe's failure, aging, and future conditions so that proactive plans can be developed and emergency responses can be avoided, hence lowering the maintenance cost of the system.

Data integration is a key component of effective asset management. New standard formats should be defined to translate the current data into a common format that can be interpreted by all the stakeholders. Data integration can be implemented at two different levels: 1) Agency level (City, County, and State) and 2) national level. At the agency level, several sources can be combined to provide the required information for the maintenance and renewal plans. At the national level which is the objective of this research, data from all over the country will be merged to form a comprehensive database of the national asset and show the current status of those assets. This database could potentially be used as a basis for the federal budget and resource allocations to optimize the total costs of maintenance nationwide.

Asset Management depends upon the availability of fully integrated data. The integration process is complex and challenging because different agencies are using their own database system which may not work with each other. The data integration process is showed in Figure 15. The description of each step is as follows:

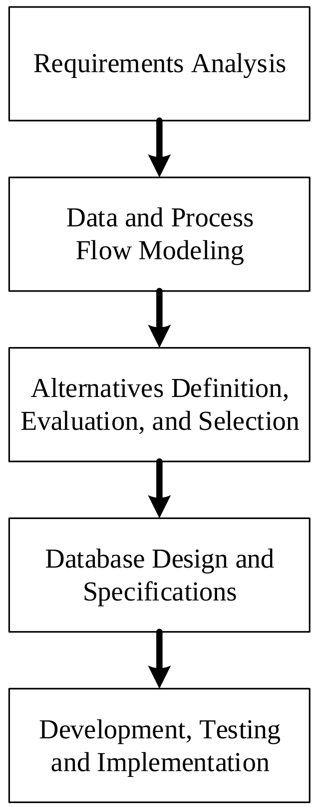


Figure 15. The Data Integration Process (FHWA 2010)

1. Requirement Analysis

   Depending upon the size and extent of integration, this can be a complex and time-consuming step. The purpose of the data integration is defined at this step and the required information will be characterized. All stakeholders should be involved at this point to evaluate the optimum process of data integration. Also, the proper integration strategy will be identified by analysis of the current databases. Also, the quality of the data should be evaluated to satisfy the ultimate goal of data integration.

2. Data and Process Flow Modeling

   At this step, the relationship between different sources is defined and a data flow diagram will be developed. Based on the requirement analysis, it should be examined how the required information will be collected and who owns that information.

3. Alternatives Definition, Evaluation, and Selection

   After requirements are evaluated and the data flow is determined, integration alternatives can be identified. Two general approaches are available: fused databases and interoperable databases. Data fusion (also known as data warehousing) integrates multiple sources of data with one-time access. In other words, fused data can be disconnected from the data source after updating its independent server. Ultimately, all fused data are stored in a single database server with substantial processing and data storage capacity. Interoperable databases (also known as federated or distributed systems) are made of several data sources that are connected through a defined platform. The data sources are located on different computers and required information is accessible through a multi-database query. Interoperable databases might be useful in data integration at the agency level because of feasible access to multi-databases and lower cost of storing data while data warehousing is inevitable at the national level.

4. Database Design and Specification

   The detailed development plan is identified at this step. The main components of database design are as follows:

   - Data models, standards, and reference systems
   - Metadata and a data dictionary
   - Network communication requirements
   - Data management requirements

5. Development, Testing, Training, and Implementation

   The final stage in data integration includes prototype software development and implementation of the computer system.

An integrated database can be implemented in two different approaches: 1) centralized database or 2) distributed database. A centralized database stores the data in a single repository, while a distributed database manages the data in multiple repositories (Elmasri and Navathe 1999). A centralized database is more effective because it provides a single access point to manage the infrastructure data and improve data flow (Halfawy 2008). For this research, a centralized relational database management system was implemented using dimensional modeling. Dimensional modeling is a widely accepted technique for data warehouse design and the main benefit of this database design technique is its simplicity. Simplicity is crucial to make the data understandable and deliver results efficiently (Kimball and Ross 2013). Dimensional models are known as star schemas in the relational database management system (Figure 16).

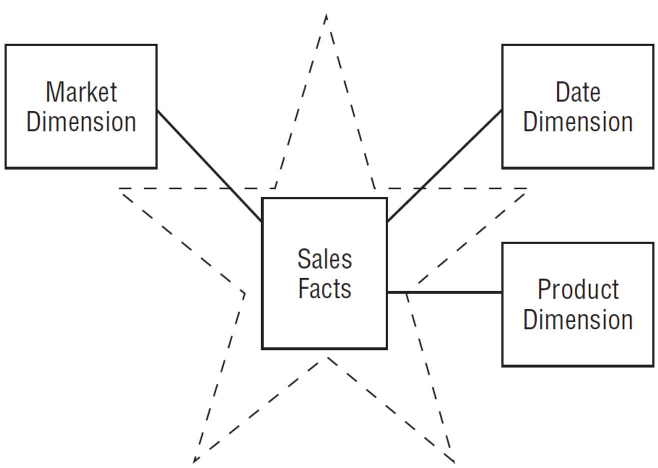


Figure 16. Star Schema (Kimball and Ross 2013)

The dimensional modeling consists of two main concepts:

1. Fact table:

   The fact table stores the performance data of the ongoing process (sewer pipes inspections). The fact tables have two or more foreign keys (FK) to connect to the dimension table's primary keys (PK).

2. Dimension table:

   Dimension tables contain the information related to the fact table. For example, the location table provides the address and other location detail of the associated inspection record. Each dimension table is defined by a primary key (PK) as a reference to join to the fact table.

One of the challenges of developing an integrated database is to determine what data should be included in the data warehouse. Two approaches are considered for determining the data source: a) availability-based, or b) need-based approach (Rujirayanyong and Shi 2006). The first approach is only considered the currently available infrastructure data, while the need-based approach determines the data required for future analysis and asset management planning programs. In this study, the data structure is developed based on the need-based approach to address the requirements of sewer asset management. Angkasuwansiri and Sinha (2013) reported 60 parameters affecting sewer pipes performance. Collecting and integrating all these parameters in a unified inventory can provide a strong basis for performance analysis of the system.

DATA MAPPING

As it was mentioned before, PACP is the only data standard for collected sewer inspection data. PACP is also providing a standard data structure to store the data and makes it compatible with other software programs. Figure 17 shows the PACP data structure and the relationships between the tables and the metadata is provided in Appendix A.2.

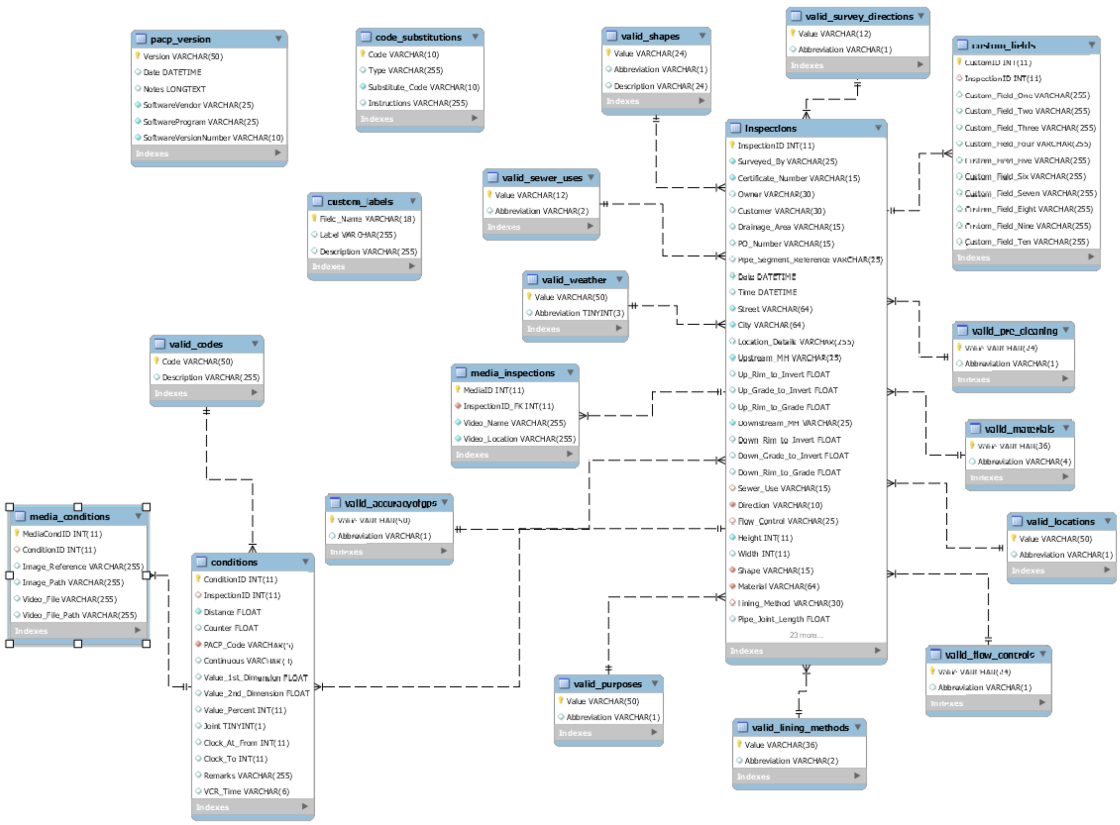
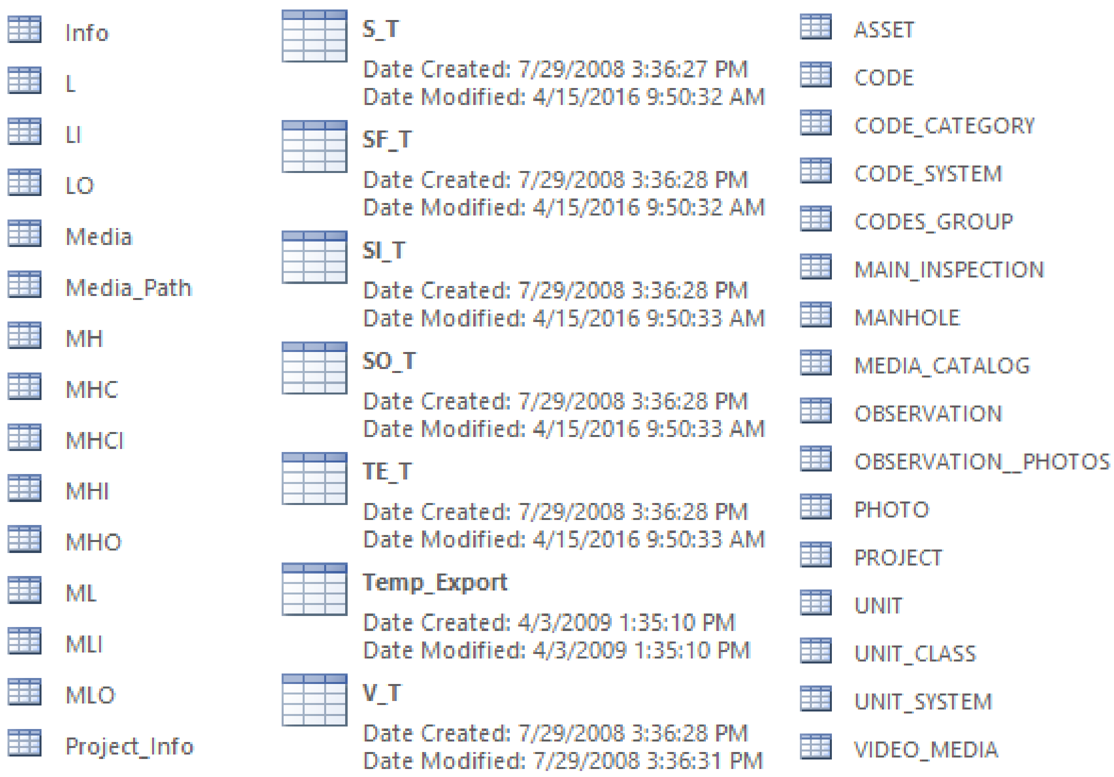


Figure 17. PACP Data Structure

While municipalities are using the PACP coding system, their final database is not always PACP compatible. This can happen for many reasons, such as coding modification, software issues, etc. As a result, the final inspection database could be in different formats. Figure 18 shows three different data structures in the collected data. Also, different data management practices have been noticed among the municipalities, such as integrating different PACP versions databases and separating databases based on year of collection, and pipe size categories.



City E City F City G

Figure 18. Different Data Structures of the Sewer Inspection Data

In order to address these issues, a data mapping tool was developed in the MySQL platform to transform the collected databases into a unified format that was defined based on the PACP data structure. Through implementing this tool, it was possible to extract information from all sewer inspection databases no matter what type of data structure, software, or coding system has been used. Figure 19 shows the data and its format for the City of F. The data was collected in three main tables and a complicated referencing system was used by the software program. As a result, the city could not use the data out of the software and while the data was collected through PACP coding more than 47,000 incompatibility errors were reported by the software in the process of exporting data into PACP data structure that can be used in other software programs.

S_ID	S_Project_ID	S_Mandantor	S_Print_ID	S_Sat_ID	S_AutoNumb
FDA33F963A6C}	8BA162405386}	1210667AB892}	54457B3C5B53}		4
F115DDEFD173}	8BA162405386}	1210667AB892}	6E3F8E8C503E}		3
CE9A4350487C}	8BA162405386}	1210667AB892}	573B75594C3D}		1
CA0592F8135B}	8BA162405386}	1210667AB892}	506A99444863}		2
14FE28ADABD8}	8BA162405386}	1210667AB892}	4753645B4D88}		5

SI_ID	SI_Section_ID	SI_Mandantc	SI_Sat_ID	SI_Drawing_I	SI_Text_ID
158DECB983A3}	CE9A4350487C}	1210667AB892}			
9F541AB75A22}	F115DDEFD173}	1210667AB892}			
12103F76254A}	FDA33F963A6C}	1210667AB892}			
C4A94EE660DF}	14FE28ADABD8}	1210667AB892}			
307293C5CE7D}	CA0592F8135B}	1210667AB892}			

SO_ID	SO_Inspecs_I	SO_Sat_ID	SO_AutoNum	SO_Counter	SO_Position
3DA5C45EBCAE}	C4A94EE660DF}		66	3	2.8042
340FBFF7BBA1}	158DECB983A3}		20	18	56.2966
C2F6ED6A32F0}	158DECB983A3}		7	5	19.5072
398EE71EB061}	158DECB983A3}		10	8	32.6136
9CDF62626CDC}	158DECB983A3}		11	9	43.3121

Figure 19. Data from City of F

InspectionID	Surveyed_By	Certificate_Number	Owner	Customer	Drainage_Area	PO_Number	Pipe_Segment_Reference	Date	Time
566	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	566	2014-06-25	07:25:43
567	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	567	2014-10-13	15:09:16
568	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	568	2014-10-14	09:41:17
578	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	578	2014-10-31	04:49:52
593	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	593	2014-11-07	05:49:09
597	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	597	2014-11-20	04:31:40
598	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	598	2014-11-20	05:24:46
599	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	599	2014-11-21	05:34:00
600	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	600	2014-11-21	07:06:03
601	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	601	2014-11-21	07:23:14
602	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	602	2014-11-21	09:17:07
609	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	609	2014-12-04	04:40:17
610	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	610	2014-12-04	05:52:55

ConditionID	InspectionID	Distance	Counter	PACP_Code	Continuous	Value_1st_Dimension	Value_2nd_Dimension	Value_Percent	Joint
5038	695	0	NULL	START AGAINST FLOW	NULL	NULL	NULL	NULL	0
5039	695	401.4	NULL	STOP	NULL	NULL	NULL	NULL	0
5040	695	66.5	NULL	Lateral	NULL	NULL	NULL	NULL	0
5041	695	72.6	NULL	Lateral	NULL	NULL	NULL	NULL	0
5042	695	94.1	NULL	Lateral	NULL	NULL	NULL	NULL	0
5043	695	147	NULL	Lateral	NULL	NULL	NULL	NULL	0
5044	695	165.9	NULL	Lateral	NULL	NULL	NULL	NULL	0
5045	695	262.1	NULL	Pipe Type	NULL	NULL	NULL	NULL	0
5046	695	264.7	NULL	Lateral	NULL	NULL	NULL	NULL	0
5047	695	284.2	NULL	Lateral	NULL	NULL	NULL	NULL	0
5048	695	323	NULL	Pipe Type	NULL	NULL	NULL	NULL	0
5049	695	343.7	NULL	Manhole	NULL	NULL	NULL	NULL	0
5050	695	343.8	NULL	General Observation	NULL	NULL	NULL	NULL	0
5051	695	366.8	NULL	Water Level	NULL	NULL	NULL	NULL	0

Figure 20. Transformed Data from City of F

The result of the data mapping tool is shown in Figure 20. Data was converted into two main tables of PACP data structure while the integrity of the attributes has been kept. The primary keys were updated to make the data understandable by the industry stakeholders. Also, the validity metric can be applied to the database to detect the PACP incompatibilities and resolve the issues. As a result, the engineering department could evaluate the collected sewer inspection data for the first time.

SEWER INVENTORY SCHEMA

In order to develop the integrated data warehouse for sewer pipes, the PACP sewer inspection data structure is considered as a core for the data integration process. As it is shown in Figure 17, PACP inspection data consists of two main tables: inspection table and condition table. Each row in the inspection table represents a single inspection record and its associated PACP codes are stored in the condition table. Since we have transformed all different types of data structures into PACP data format, loading data into the unified schema is facilitated.

The sewer inventory should be efficient to address the industry needs. Thus, the final schema is required to be normalized. Database normalization is a process of organizing a database according to normal forms to improve data integrity and reduce redundancy. Here are the requirements for the normalization process (Codd, 1972):

1. No redundancy of facts: To free the database from insert, update, and delete anomalies.
2. No cluttering of facts: To reduce the need for restructuring the database, as new types of data are introduced.
3. Must preserve information: To provide the same information on JOIN functions.
4. Must preserve functional dependencies: To make the database neutral to the query statistics.

A functional dependency is a relationship between two attributes in a table, usually between the primary key and other non-key attributes. In relation, the values of the keys are unique, so keys are used to enforce functional dependencies. For any relation R, attribute Y is functionally dependent on attribute X, if X uniquely determines Y (Watt & Eng, 2014):

$$X\ (determitant) \rightarrow Y\ (dependent)$$

There are a set of inference rules known as Armstrong's axioms to examine the functional dependencies in a relational database:

1. Reflexivity: if Y is a subset of X, then X determines Y
2. Augmentation: if X determines Y, then XZ determines YZ for any Z
3. Transitivity: if X determines Y, and Y determines Z, then X must also determine Z

As a result, the sewer inventory should preserve all the inspection attributes and their functional dependencies while reducing redundancies and restructuring the database for the new types of data that can be added in the future. This also addresses the other industry problem to accept new data generated by the new inspection technologies. The star schema was developed based on the important parameters affecting sewer pipes and was normalized to be considered as a basis for the sewer data management system (internal schema). The PACP data structure can be retrieved at any time as an external schema that can be used by industry stakeholders and imported to any other software program. Figure 21 shows the Sewer Inventory Schema (SIS).

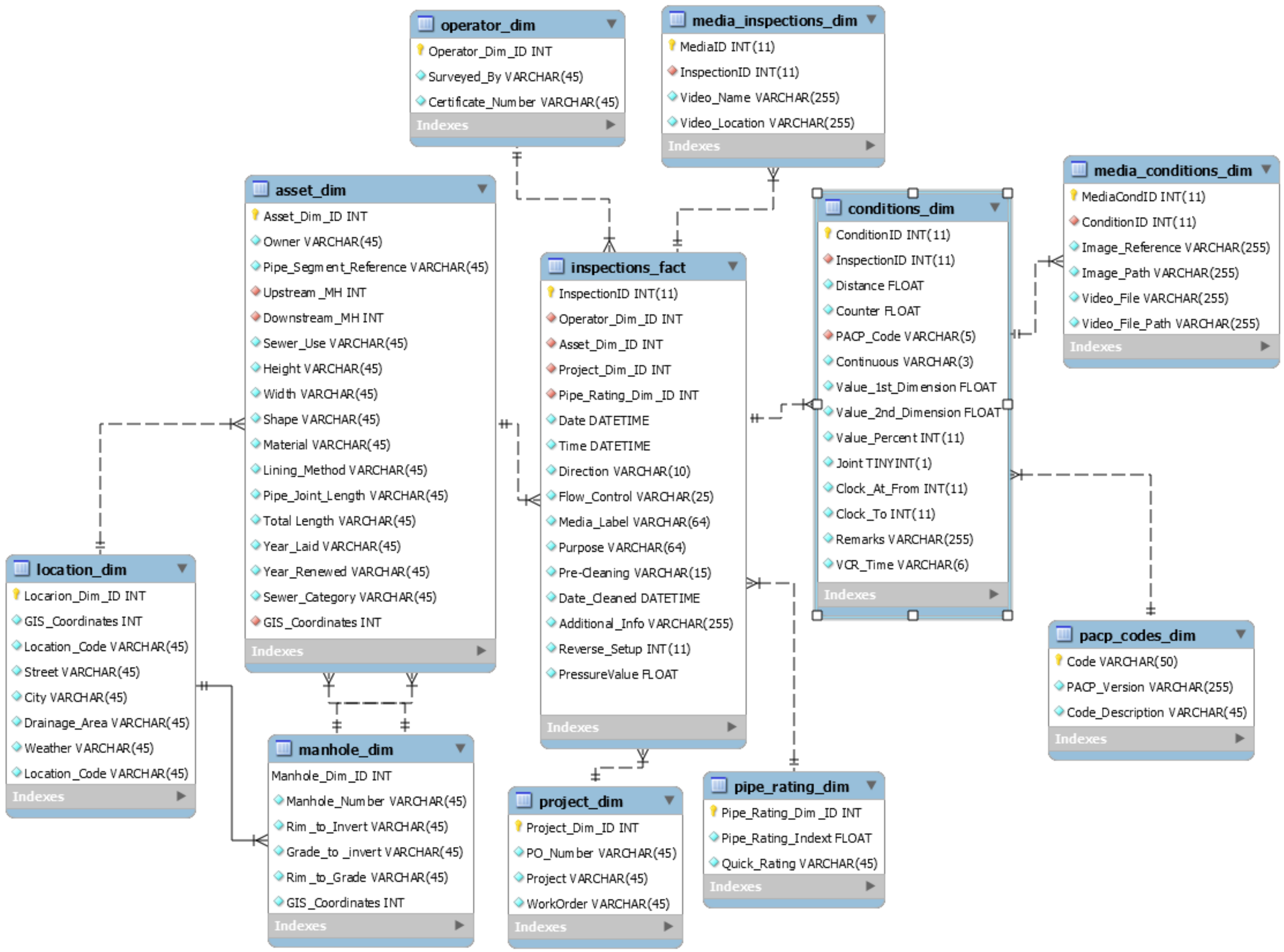


Figure 21. Sewer Inventory Schema (SIS)

The SIS can meet one of the main industry needs which is flexibility. The advantage of using dimensional modeling will be that the data warehouse is flexible for changes of dimension. As we mentioned, the data sources will be included in the need-based approach. According to the previous studies on deterioration models of sewer pipes, factors such as age, depth, length, soil type, frost depth, location, size, and material could affect the conditions of the pipes (Ariaratnam et al. 2001; Khan et al. 2009; Opila and Attoh-Okine 2011; Scheidegger et al. 2011). Some of these factors like soil type are not included in the CCTV databases. However, it could be possible to acquire this information from other resources and integrate them into the final inventory. The implementation of the SIS paves the way to developing national sewer inventory.

CONCLUSIONS AND RECOMMENDATIONS

Asset inventory is the first step in the asset management process and it is rarely possible to get the advantages of this process without easily accessible and reliable data. The integration process is complex and challenging because different data structures have been used which may not work with each other. Even for those municipalities that are using the PACP coding system, their final database is not always PACP compatible. As a result, the final inspection database could be in different formats. In order to address these issues, a data mapping tool was developed to transform the collected databases into a unified format that was defined based on the PACP data structure. Through the implementation of this tool, it was possible to extract information from all sewer inspection databases no matter what type of data structure, software, or coding system has been used. This tool can address compatibility issues which is one of the major concerns in the sewer industry. Also, a normalized dimensional Sewer Inventory Schema (SIS) was developed for integrated national sewer inventory. This research contributes to the overall body of knowledge by providing a robust data transformation and integration platform to solve the data accessibility and compatibility issues and improve the effectiveness of asset management plans.

REFERENCES

American Society of Civil Engineers (ASCE), (2000). “Wastewater Facilities Construction Funding,” Policy Statement 326, Reston, VA.

Angkasuwansiri, T., & Sinha, S. K. (2013). Comprehensive list of parameters affecting wastewater pipe performance. Technology Interface International Journal, 13(2), 68-79.

Ariaratnam, S. T., El-Assaly, A., & Yang, Y. (2001). "Assessment of infrastructure inspection needs using logistic models." Journal of Infrastructure Systems, 7(4), 160-165.

Berkeley, C., Jewell, C., Peck, J., Rowe, R., and von Zweck, P. (2015). "Boston Water and Sewer Commission: Data Integration to Support Asset Management." Pipelines 2015, 1036-1046.

Caradot, N., Rouault, P., Clemens, F., & Cherqui, F. (2018). Evaluation of uncertainties in sewer condition assessment. Structure and Infrastructure Engineering, 14(2), 264-273.

Codd, E. F. (1972). Further normalization of the database relational model. Database Systems, 6, 33-64

Danylo, N., and Lemer, A. (1998). “Asset management for the public works manager: Challenges and strategies, findings of the APWA Task Force on asset management.” American Public Works Association, Washington, D.C.

Elmasri, R., and Navathe, S. (1999). Fundamentals of database systems, 3rd Ed., Addison-Wesley, Boston.

Federal Highway Administration (FHWA) (2010). Data Integration Primer, Office of Asset Management, Washington D.C.

Grigg, N. S. (1999). “Viewpoint: Infrastructure: Integrated issue or Tower of Babel?” J. Infrastruct. Syst., 5(4), 115–117.

Halfawy, M. R. (2008). "Integration of municipal infrastructure asset management processes: challenges and solutions." Journal of Computing in Civil Engineering, 22(3), 216-229.

Halfawy, M. R., Vanier, D. J., and Froese, T. M. (2006). "Standard data models for interoperability of municipal infrastructure asset management systems." Canadian Journal of Civil Engineering, 33(12), 1459-1469.

Halfawy, M., Pyzoha, D., and El-Hosseiny, T. (2002). “An integrated framework for GIS-based civil infrastructure management systems.” Proc., Conf. of the Canadian Society for Civil Engineers, Canadian Society of Civil Engineers, Montréal.

Khan, Z., Zayed, T., and Moselhi, O. (2009). "Structural condition assessment of sewer pipelines." Journal of performance of constructed facilities, 24(2), 170-179.

Kimball, R., and Ross, M. (2013). The data warehouse toolkit: The definitive guide to dimensional modeling, John Wiley & Sons.

National Association of Sewer Service Companies (NASSCO) (2010). Pipeline Assessment and Certification Program (PACP): Reference manual, Volume 6.0.1, Marriottsville, MD.

National Cooperative Highway Research Program (NCHRP) (2006). Performance measures and targets for transportation asset management, Transportation Research Board, Vol. 551.

Opila, M. C., and Attoh-Okine, N. (2011). "Novel approach in pipe condition scoring." Journal of pipeline systems engineering and practice, 2(3), 82-90.

Park, T., and Kim, H. (2013). "A data warehouse-based decision support system for sewer infrastructure management." Automation in Construction, 30, 37-49.

Rujirayanyong, T., & Shi, J. J. (2006). “A project-oriented data warehouse for construction.” Automation in Construction, 15(6), 800-807.

Scheidegger, A., Hug, T., Rieckermann, J., & Maurer, M. (2011). “Network condition simulator for benchmarking sewer deterioration models.” Water Research, 45(16), 4983-4994.

Sinha, S. K. (2019). Collection and Compilation of Water Pipeline Performance Data. In Pipelines 2019: Multidisciplinary Topics, Utility Engineering, and Surveying (pp. 207-216): American Society of Civil Engineers Reston, VA.

Sinha, S., Dymond, R., Vemulapally, R., Dickerson, T., & Perry, S. (2009). Development of a National GIS Database for Municipal Water and Wastewater Pipe Infrastructure System. Paper presented at the World Environmental and Water Resources Congress 2009: Great Rivers.

Thornhill, R., & Wildbore, P. (2005). "Sewer defect codes: Origin and destination." U-Tech Underground Construction Paper.

USEPA (2002). Fact Sheet—Asset Management for Sewer Collection Systems, Office of Wastewater Management, Washington D.C.

USEPA (2015). Condition Assessment of Underground Pipes, Water Infrastructure Outreach, Washington D.C.

Watt, A., & Eng, N. (2014). Database design.

Weseman, W. A. (1995). "Recording and coding guide for the structure inventory and appraisal of the nation's bridges." Federal Highway Administration, USA.

CHAPTER V

CONCLUSIONS

The development and rapid growth of the population in cities have accelerated the aging and deterioration of the already old sewer systems. It is estimated that more than $270 billion is needed for the next ten years to restore the wastewater systems to good condition. There are more than 800,000 miles of public sewer pipes and 500,000 miles of private pipes in the US and many of them are at the end of their service life (ASCE 2021). Therefore, understanding the current condition of the sewer system is a critical step for infrastructure asset management strategies and improving national wastewater systems.

Infrastructure asset management is the integrated approach to sustain public infrastructure assets. Asset management is a continuous improvement planning which can be practiced based on the current resources and can be extended as the program progresses (USEPA 2002). Asset inventory and condition assessment are the first steps in the asset management process and it is rarely possible to get the advantages of this process without easily accessible and reliable data (NCHRP 2006). Proactive asset management relies on smooth data exchange as well as the accessibility of quality condition data collected on sewer systems. Although collecting sewer inspection data is an industry norm and municipalities are collecting data on 10% of the sewer system every year, data accessibility and reliability are questionable (DeBoda and Bayer 2015).

Due to the quality issues in the current data collection practices, more than one-third of the data is not usable for asset management purposes based on the current data quality improvement practices (Salman 2010). To address data reliability and accessibility issues, it was required to develop a comprehensive data quality assurance and integration process that facilitates the growth of the unified national sewer inventory.

The final output of this research provides a solid platform for integrated data management of sewer systems that can help municipalities to understand their sewer systems, analyze the collected data and become proactive in their maintenance strategies. Through the availability of integrated data and national sewer inventory, it would be possible to understand how and why the pipe deterioration occurs based on all affecting parameters.

MAIN CONCLUSIONS AND CONTRIBUTIONS TO THEORY AND PRACTICE

The first study evaluated the quality of the collected sewer inspection databases base on the developed quality evaluation tools. Data quality evaluation is a multi-dimensional concept that includes both subjective perceptions and objective measurements to be evaluated. Five quantitative data quality metrics were defined to assess different quality dimensions of the sewer inspection data including Accuracy, Consistency, Completeness, Uniqueness, and Validity. These data quality metrics were calculated for the collected sewer inspection data and it was found that consistency and uniqueness are the major problems based on the current practices. This was the first quantitative study on the data quality of the sewer inspection databases. In addition, problematic issues with current commercial pipe rating software programs were identified. As a result, the rating systems in the current PACP certified data collection software overestimates the current condition of the sewer systems.

The second study provided a contextual data quality assurance tool to identify, measure, and resolve the common quality problems in the collected databases. The quality assurance process consists of three steps: 1) Formulating a quality assurance framework, 2) Detecting problematic data, and 3) Resolving problematic data. The results show that, by applying the proposed quality assurance process, the percentage of good quality inspection data increased from 50%-75% (pre-process) to 95% (post-process). This research contributes to the overall body of knowledge by providing a robust data quality assurance process for underground sewer pipe inspection data, which will result in quality data for sewer asset management endeavors.

The third study addressed the data accessibility and compatibility problems related to sewer inspection databases. There is no unified data format for collected sewer inspection data. Eve those municipalities that are using the PACP coding system, their final database is not always PACP compatible. This can happen for many reasons, such as coding modification, software issues, etc. As a result, the final inspection database could be in so many different formats. A data mapping tool was developed to address this problem by transforming data into the PACP data structure while keeping the integrity of the database. By implementing this tool one of the major issues in the industry will be addressed and the data can be viewed, modified, and analyzed independently from the generating software. Also, a normalized dimensional Sewer Inventory Schema (SIS) was developed for integrated national sewer inventory. An integrated database is considered a basis for proactive asset management.

It can be noticed that all these studies are connected and the real value of the developed tools is obtained by integrating them into one platform. So, a web-based platform, “One-Voice”, has been developed based on this research effort to address data compatibility, data quality, and data accessibility issues. This platform is developed in a Django environment that works flawlessly with Python and MySQL. The interface is developed through HTML, CSS, and JavaScript. Also, a data server has been purchased to host the platform. Figure 22 shows the user interface.

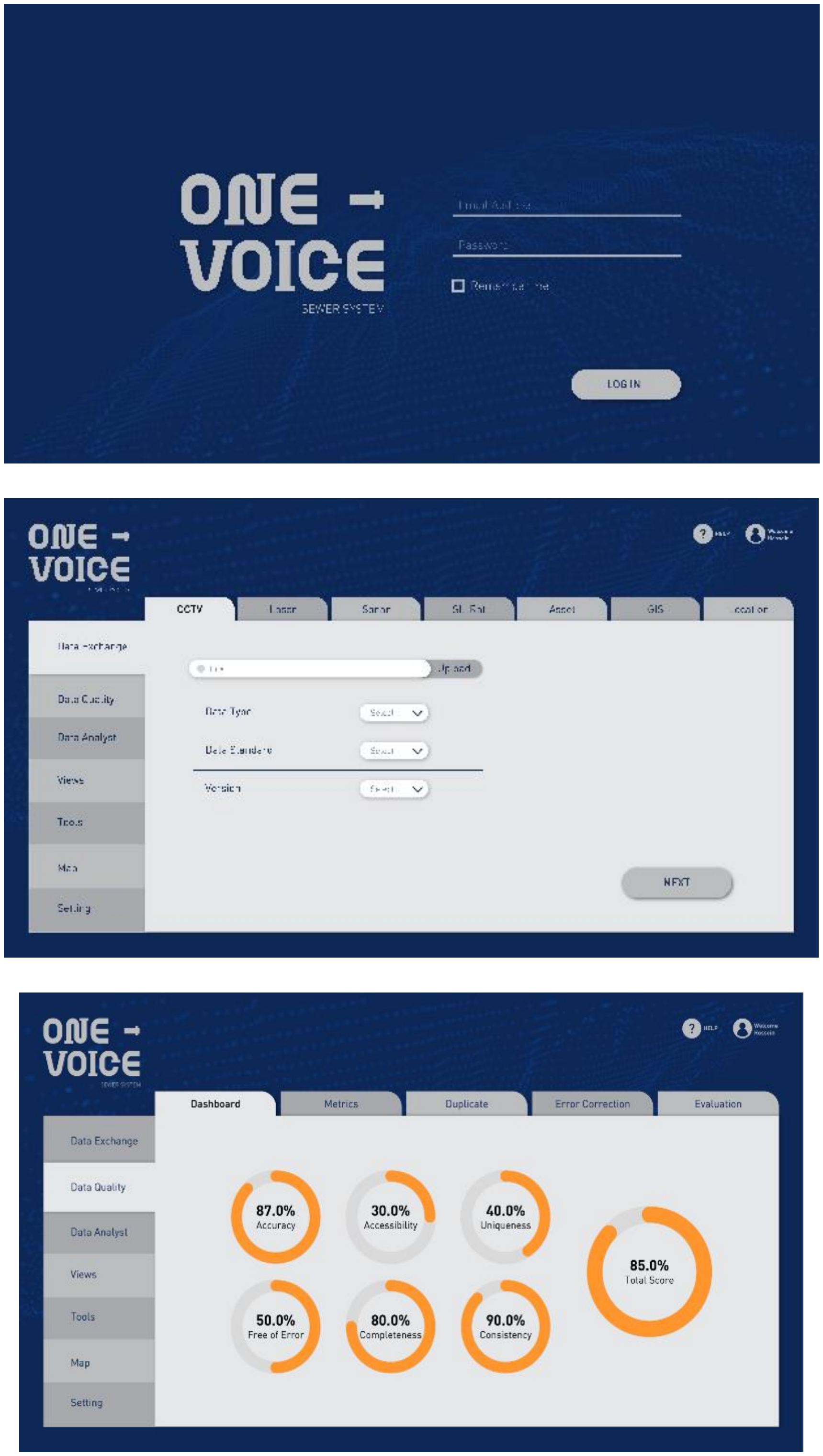


Figure 22. One-Voice Platform

SUGGESTIONS FOR FUTURE RESEARCH

Several topics for further study were identified through this research. To improve the efficiency and accuracy of data quality evaluation tools, studying larger databases can provide valuable findings and show the weaknesses points. Also, studying the effectiveness of the data quality assurance process in the municipalities' sewer management system can be worthwhile. Finally, including AI and machine learning approaches in the integration and quality assurance process will provide a robust platform to target unknown data errors.

REFERENCES

ASCE (2021). 2021 Report Card for America's Infrastructure. American Society of Civil Engineers, www.infrastructurereportcard.org/, Information viewed on June 10, 2021.

DeBoda, T., and Bayer, J. (2015). "Benefits of PACP® Version 7.0 Update NASSCO." Pipelines 2015, 878-886.

National Cooperative Highway Research Program (NCHRP) (2006). Performance measures and targets for transportation asset management, Transportation Research Board, Vol. 551.

Salman, B. (2010). Infrastructure management and deterioration risk assessment of wastewater collection systems, Doctoral dissertation, University of Cincinnati, Cincinnati, OH

USEPA (2002). Fact Sheet—Asset Management for Sewer Collection Systems, Office of Wastewater Management, Washington D.C.

APPENDICES

Appendix A.1: PACP 6 Structural Codes Description

Crack (C)

General: The crack code is used where a crack line is visible on the surface but is not visibly open. There is no gap between the edges of the crack.

Continuous Defect: The continuous defect feature is used when the cracks extend more than 3 feet or 1 meter. In some cases, circumferential cracks may occur frequently at joints and the repeating continuous defect feature is used if 75% of the joints are affected.

Value: For visual inspection enter value in feet to two decimal places and meters to two decimal places. Digital measurements allowed three decimal places.

Joint: The joint code shall be applied to cracks that are associated with the joint and are within 8 inches or 200 mm of the joint.

Cracks													
ConditionID	InspectionID	Distance	Counter	PACP_Code	Continuous	Value_1st_Dimension	Value_2nd_Dimension	Value_Percent	Joint	Clock_At_From	Clock_To	Remarks	VCR_Time
☑	☑	☑	☐	CL	☐	☐	☒	☒	☐	☑	☒	☐	☐
☑	☑	☑	☐	CC	☐	☐	☒	☒	☐	☑	☑	☐	☐
☑	☑	☑	☐	CM	☐	☐	☒	☒	☐	☑	☑	☐	☐
☑	☑	☑	☐	CS	☐	☐	☒	☒	☐	☑	☑	☐	☐
☑	☑	☑	☐	CH2	☐	☐	☒	☒	☐	☑	☑	☐	☐
☑	☑	☑	☐	CH3	☐	☐	☒	☒	☐	☑	☑	☐	☐
☑	☑	☑	☐	CH4	☐	☐	☒	☒	☐	☑	☑	☐	☐

Fracture (F)

General: A fracture is a crack that has become visibly open and a gap can be seen although the sections of the pipe wall are still in place and not able to move.

Continuous Defect: The continuous defect feature is used when the fractures extend more than 3 feet or 1 meter. In some cases, circumferential cracks may occur frequently at joints and the repeating continuous defect feature is used if 75% of the joints are affected.

Value: For visual inspection enter value in feet to two decimal places and meters to two decimal places. Digital measurements allowed three decimal places.

Joint: The joint code shall be applied to fractures that are associated with the joint and are within 8 inches or 200 mm of the joint.

Conditions													
ConditionID	InspectionID	Distance	Counter	PACP_Code	Continuous	Value_1st_Dimension	Value_2nd_Dimension	Value_Percent	Joint	Clock_At_From	Clock_To	Remarks	VCR_Time
☑	☑	☑	☐	FL	☐	☐	☒	☒	☐	☑	☒	☐	☐
☑	☑	☑	☐	FC	☐	☐	☒	☒	☐	☑	☑	☐	☐
☑	☑	☑	☐	FM	☐	☐	☒	☒	☐	☑	☑	☐	☐
☑	☑	☑	☐	FS	☐	☐	☒	☒	☐	☑	☑	☐	☐
☑	☑	☑	☐	FH2	☐	☐	☒	☒	☐	☑	☑	☐	☐
☑	☑	☑	☐	FH3	☐	☐	☒	☒	☐	☑	☑	☐	☐
☑	☑	☑	☐	FH4	☐	☐	☒	☒	☐	☑	☑	☐	☐

Broken (B)

General: Broken refers to a pipe where pieces are noticeably displaced and have moved from their original position at least ½ the thickness of the pipe.

Continuous Defect: The continuous defect feature is used when the breaks extend more than 3 feet or 1 meter.

Value: For visual inspection enter value in feet to two decimal places and meters to two decimal places. Digital measurements allowed three decimal places.

Joint: The joint code shall be applied to breaks that are associated with the joint and are within 8 inches or 200 mm of the joint.

Conditions													
ConditionID	InspectionID	Distance	Counter	PACP_Code	Continuous	Value_1st_Dimension	Value_2nd_Dimension	Value_Percent	Joint	Clock_At_From	Clock_To	Remarks	VCR_Time
☑	☑	☑	☐	B	☐	☐	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	BSV	☐	☐	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	BVV	☐	☐	☒	☒	☐	☑	☐	☐	☐

Hole (H)

General: A Hole is when the pipe material is missing and the surrounding soil is exposed.

Continuous Defect: The continuous defect feature is used when the holes extend more than 3 feet or 1 meter.

Value: For visual inspection enter value in feet to two decimal places and meters to two decimal places. Digital measurements allowed three decimal places.

Joint: The joint code shall be applied to holes that are associated with the joint and are within 8 inches or 200 mm of the joint.

Conditions													
ConditionID	InspectionID	Distance	Counter	PACP_Code	Continuous	Value_1st_Dimension	Value_2nd_Dimension	Value_Percent	Joint	Clock_At_From	Clock_To	Remarks	VCR_Time
☑	☑	☑	☐	H	☐	☐	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	HSV	☐	☐	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	HVV	☐	☐	☒	☒	☐	☑	☐	☐	☐

Deformed (D)

General: The Rigid pipe is damaged to the point that the original cross-section of the sewer is noticeably altered. It is possible to have a small amount of deformation in flexible pipes.

Continuous Defect: The continuous defect feature is used when the deformation extends more than 3 feet or 1 meter.

Value: Estimate vertical/horizontal changes as a percentage of the original diameter/height and give it in increments of 5% in the Value % field.

Conditions													
ConditionID	**InspectionID**	**Distance**	**Counter**	**PACP_Code**	**Continuous**	**Value_1st_Dimension**	**Value_2nd_Dimension**	**Value_Percent**	**Joint**	**Clock_At_From**	**Clock_To**	**Remarks**	**VCR_Time**
☑	☑	☑		D	☐	☒	☒	☑	☒	☒	☒	☐	☐
☑	☑	☑		DH	☐	☒	☒	☑	☒	☒	☒	☐	☐
☑	☑	☑		DV	☐	☒	☒	☑	☒	☒	☒	☐	☐

--- If deformation occurs, no other structural defect codes are required unless the deformation is at a point repair.

Collapse (X)

General: A collapse is where the deformation is so great that there has been a complete loss of structural integrity of the sewer with more than 40% of cross-sectional are lost.

Value: The percentage of cross-sectional area (>40%) lost should be entered in the Value % field.

Conditions													
ConditionID	InspectionID	Distance	Counter	PACP_Code	Continuous	Value_1st_Dimension	Value_2nd_Dimension	Value_Percent	Joint	Clock_At_From	Clock_To	Remarks	VCR_Time
☑	☑	☑		XP	☒	☒	☒	☑	☒	☒	☒	☐	☐
☑	☑	☑		XB	☒	☒	☒	☑	☒	☒	☒	☐	☐

--- The MSA code will follow the collapse!

Joint (J)

General: This group is used to describe defective displacements at joints.

Continuous Defect: The repeated continuous defect facility can be applied to those codes if they meet the requirement that 75% of joints are defective.

Value: If precision measurement tools are available the amount of offset or separation can be entered in the Value 1st field in inches or millimeters up to three decimal places. The angular displacement may also be entered in the value 1st column in degrees up to three decimal places optionally for Joint Angular defects.

Conditions													
ConditionID	InspectionID	Distance	Counter	PACP_Code	Continuous	Value_1st_Dimension	Value_2nd_Dimension	Value_Percent	Joint	Clock_At_From	Clock_To	Remarks	VCR_Time
☑	☑	☑		JAL	☐	☐	☒	☒	☒	☒	☒	☐	☐
☑	☑	☑		JAM	☐	☐	☒	☒	☒	☒	☒	☐	☐
☑	☑	☑		JOL	☐	☐	☒	☒	☒	☒	☒	☐	☐
☑	☑	☑		JOM	☐	☐	☒	☒	☒	☒	☒	☐	☐
☑	☑	☑		JSL	☐	☐	☒	☒	☒	☒	☒	☐	☐
☑	☑	☑		JSM	☐	☐	☒	☒	☒	☒	☒	☐	☐

JOM: 1.0<Value 1^{st} =<1.5

JOL : Value 1^{st} >1.5

JSM: Value 1^{st} =<1

JSL : Value 1^{st} >1

JAM: 5<Value 1^{st} =<10 or percentage: 90 degree is 100%

JAL : Value 1^{st} >10 or percentage: 90 degree is 100%

Surface Damage (S)

General: The group of codes is used to describe a wide range of pipe material surface damage failures.

Continuous Defect: In many instances, it will be necessary to code surface damage as a truly continuous defect.

Joint: The joint code can be applied to surface damage that is associated with and is within 8 inches of a joint.

Conditions													
ConditionID	**InspectionID**	**Distance**	**Counter**	**PACP_Code**	**Continuous**	**Value_1st_Dimension**	**Value_2nd_Dimension**	**Value_Percent**	**Joint**	**Clock_At_From**	**Clock_To**	**Remarks**	**VCR_Time**
☑	☑	☑	☐	SAM	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SAMC	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SAMM	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SAMZ	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SAP	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SAPC	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SAPM	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SAPZ	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SAV	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SAVC	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SAVM	☐	☒	☒	☒	☐	☑	☐	☐	☐

Conditions													
ConditionID	InspectionID	Distance	Counter	PACP_Code	Continuous	Value_1st_Dimension	Value_2nd_Dimension	Value_Percent	Joint	Clock_At_From	Clock_To	Remarks	VCR_Time
☑	☑	☑	☐	SAVZ	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SCP	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SMW	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SMWC	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SMWM	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SMWZ	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SRC	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SRCC	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SRCM	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SRCZ	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SRI	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SRIC	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SRIM	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SRIZ	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SRP	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SRPC	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SRPM	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SRPZ	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SRV	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SRVC	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SRVM	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SRVZ	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SSS	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SSSC	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SSSM	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SSSZ	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	SZ	☐	☒	☒	☒	☐	☑	☐	☑	☐
☑	☑	☑	☐	SZC	☐	☒	☒	☒	☐	☑	☐	☑	☐
☑	☑	☑	☐	SZM	☐	☒	☒	☒	☐	☑	☐	☑	☐
☑	☑	☑	☐	SZZ	☐	☒	☒	☒	☐	☑	☐	☑	☐

--- SCP in only for metal pipes (cast iron, ductile iron)

Buckling (K)

General: Deformations and other defects without loss of visible structural integrity in flexible pipes.

Continuous Defect: In many instances, it will be necessary to code for buckling damage as a truly continuous defect.

Value: the percentage column is used to approximate the reduction of cross-section area for inverse Curvature (KI).

Joint: The joint code shall be applied to all defects that are associated with the joint and are within 8 inches or 200 mm of the joint.

Conditions													
ConditionID	InspectionID	Distance	Counter	PACP_Code	Continuous	Value_1st_Dimension	Value_2nd_Dimension	Value_Percent	Joint	Clock_At_From	Clock_To	Remarks	VCR_Time
☑	☑	☑	☐	KW	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	KD	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	KI	☐	☒	☒	☑	☐	☑	☐	☐	☐

--- Flexible pipes: plastic, pitch fiber pipes, metal, …

Lining (LF)

General: This group is used to describe features in renewed sewers.

Continuous Defect: "truly" continuous defect facility will apply to several of the above codes.

Joint: The joint code shall be applied to all defects that are associated with the joint and are within 8 inches or 200 mm of the joint.

Conditions													
ConditionID	InspectionID	Distance	Counter	PACP_Code	Continuous	Value_1st_Dimension	Value_2nd_Dimension	Value_Percent	Joint	Clock_At_From	Clock_To	Remarks	VCR_Time
☑	☑	☑	☐	LFAC	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	LFAS	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	LFB	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	LFBK	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	LFBU	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	LFCS	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	LFD	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	LFDC	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	LFDE	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	LFDL	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	LFOC	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	LFPH	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	LFRS	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	LFUC	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	LFW	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	LFZ	☐	☒	☒	☒	☐	☑	☐	☐	☑

Weld Failure (WF)

General: The failure in a weld of the pipe fabric.

Continuous Defect: The continuous defect feature may apply to the longitudinal, multiple, and spiral codes.

Joint: The joint code shall be applied to all defects that are associated with the joint and are within 8 inches or 200 mm of the joint.

Conditions													
ConditionID	**InspectionID**	**Distance**	**Counter**	**PACP_Code**	**Continuous**	**Value_1st_Dimension**	**Value_2nd_Dimension**	**Value_Percent**	**Joint**	**Clock_At_From**	**Clock_To**	**Remarks**	**VCR_Time**
☑	☑	☑	☐	WFC	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	WFL	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	WFM	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	WFS	☐	☒	☒	☒	☐	☑	☐	☐	☐
☑	☑	☑	☐	WFZ	☐	☒	☒	☒	☐	☑	☐	☐	☑

--- Can occur in

1. Large diameter plastic spirally wound welded pipes or butt-fused pipes
2. Metallic pipes

Point Repair (RP)

General: This group of codes is used to record where parts of the pipes have been repaired or replaced.

Continuous Defect: The continuous defect feature may be applicable to point repairs that extend over three feet or one meter in length.

Joint: The joint code shall be applied to all defects that are associated with the joint and are within 8 inches or 200 mm of the joint.

Conditions													
ConditionID	InspectionID	Distance	Counter	PACP_Code	Continuous	Value_1st_Dimension	Value_2nd_Dimension	Value_Percent	Joint	Clock_At_From	Clock_To	Remarks	VCR_Time
☑	☑	☑	☐	RPL	☐	☒	☒	☒	☐	☒	☒	☑	☑
☑	☑	☑	☐	RPLD	☐	☒	☒	☒	☐	☒	☒	☑	☑
☑	☑	☑	☐	RPP	☐	☒	☒	☒	☐	☑	☐	☑	☑
☑	☑	☑	☐	RPPD	☐	☒	☒	☒	☐	☑	☐	☑	☑
☑	☑	☑	☐	RPR	☐	☒	☒	☒	☐	☒	☒	☑	☑
☑	☑	☑	☐	RPRD	☐	☒	☒	☒	☐	☒	☒	☑	☑
☑	☑	☑	☐	RPZ	☐	☒	☒	☒	☐	☒	☒	☑	☑
☑	☑	☑	☐	RPZD	☐	☒	☒	☒	☐	☒	☒	☑	☑

Brickwork

General: This group of codes is only to be used on brick sewers.

Continuous Defect: The continuous defect feature may be applicable where the defect is over three feet or one meter in length.

Conditions													
ConditionID	InspectionID	Distance	Counter	PACP_Code	Continuous	Value_1st_Dimension	Value_2nd_Dimension	Value_Percent	Joint	Clock_At_From	Clock_To	Remarks	VCR_Time
☑	☑	☑	☐	DB	☐	☒	☒	☒	☒	☑	☐	☐	☐
☑	☑	☑	☐	DI	☐	☑	☒	☒	☒	☒	☒	☐	☐
☑	☑	☑	☐	MB	☐	☒	☒	☒	☒	☑	☐	☐	☐
☑	☑	☑	☐	MML	☐	☒	☒	☒	☒	☑	☐	☐	☐
☑	☑	☑	☐	MMM	☐	☒	☒	☒	☒	☑	☐	☐	☐
☑	☑	☑	☐	MMS	☐	☒	☒	☒	☒	☑	☐	☐	☐

Appendix A.2: PACP 6 Metadata

Conditions										
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
ConditionID	AutoNumber	This field is automatically populated when any condition information is entered.	LI	In					Yes (ND)
InspectionID	Number	Software provided designation for this inspection (THIS FIELD USED TO JOIN TABLES)	LI		Au	0		No	Yes (DO)
Distance	Number	Distance is measured to one decimal place to feature location whether it is in feet or meters	S		1			yes	No
Counter	Number	Time into the video of the identified condition, in seconds	S		1			No	No
PACP_Code	Short Text	Combination of Group/Descriptor and Modifier/Severity in a single data field	5				Yes	Yes	Yes (DO)
Continuous	Short Text	Continuous defect number with start (S) and finish (F) matching to denote beginning and ending of defect	3				Yes	No	No

Conditions										
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
Value_1st_Dimension	Number	Dimensions of defects to nearest Inch or mm	S		3			No	No
Value_2nd_Dimension	Number	Used for the intrusion of tap or width of non-circular connecting pipe to nearest inch or mm	S		3			No	No
Value_Percent	Number	Used to express percentage value of defects	LI		0			No	No
Joint	Yes/No	Indicates a defect located near a joint	T/F						No
Clock_At_From	Number	Clock At/From Position of defect/observation	I		0			No	No
Clock_To	Number	Clock To Position of defect/observation	I		0			No	No
Remarks	Short Text	Additional info to describe defect/coding	255				Yes	No	No

Conditions									
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
VCR_Time	Short Text	Time into the video of the identified condition in HHMMSS format with 0 used as a space holder.	6				Yes	No	No

* LI: Long Integer * S: Single * T/F: True/False * I: Integer * In: Increment * Au: Auto * ND: No Duplicates * DO: Duplicates OK

Valid_Codes									
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
Code	Short Text		50				Yes	Yes	Yes (ND)
Description	Short Text		255				Yes	No	No

* LI: Long Integer * S: Single * T/F: True/False * I: Integer * In: Increment * Au: Auto * ND: No Duplicates * DO: Duplicates O

Media_Conditions										
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
MediaCondID	AutoNumber	This field is automatically populated when any media (picture or movie file) is saved.	LI	In					Yes (ND)
ConditionID	Number	Software provided designation for this inspection (THIS FIELD USED TO JOIN TABLES)	LI		Au	0		No	Yes (DO)
Image_Reference	Short Text	If digital snapshots are taken, the name or number of the image file.	255				Yes	No	No
Image_Path	Short Text	Path to digital image reference file	255				Yes	No	No
Video_File	Short Text	For digital recordings, the name of the video file associated with this condition relative to the data file	255				Yes	No	No
Video_File_Path	Short Text	For digital recordings, the path of the video file associated with this condition relative to the data file	255				Yes	No	No

* LI: Long Integer * S: Single * T/F: True/False * I: Integer * In: Increment * Au: Auto * ND: No Duplicates * DO: Duplicates OK

Valid_Materials									
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
Value	Short Text		36				Yes	Yes	Yes (ND)
Abbreviation	Short Text		4				Yes	No	No

* LI: Long Integer * S: Single * T/F: True/False * I: Integer * In: Increment * Au: Auto * ND: No Duplicates * DO: Duplicates OK

Valid_Lining_Methods									
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
Value	Short Text		36				Yes	Yes	Yes (ND)
Abbreviation	Short Text		2				Yes	No	No

* LI: Long Integer * S: Single * T/F: True/False * I: Integer * In: Increment * Au: Auto * ND: No Duplicates * DO: Duplicates OK

Valid_Shapes										
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
Value	Short Text		24				Yes	Yes	Yes (ND)
Abbreviation	Short Text		1				Yes	No	No
Description	Short Text		24				Yes	No	No

* LI: Long Integer * S: Single * T/F: True/False * I: Integer * In: Increment * Au: Auto * ND: No Duplicates * DO: Duplicates OK

Valid_Flow_Controls									
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
Value	Short Text		24				Yes	Yes	Yes (ND)
Abbreviation	Short Text		1				Yes	No	No

* LI: Long Integer * S: Single * T/F: True/False * I: Integer * In: Increment * Au: Auto * ND: No Duplicates * DO: Duplicates OK

Valid_Locations									
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
Value	Short Text		50				Yes	Yes	Yes (ND)
Abbreviation	Short Text		1				Yes	No	No

* LI: Long Integer * S: Single * T/F: True/False * I: Integer * In: Increment * Au: Auto * ND: No Duplicates * DO: Duplicates OK

Media_Inspections										
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
MediaID	AutoNumber		LI	In					Yes (ND)
InspectionID	Number		LI		Au	0		Yes	Yes (DO)
Video_Name	Short Text		255				No	Yes	No
Viideo_Location	Short Text		255				No	Yes	No

* LI: Long Integer * S: Single * T/F: True/False * I: Integer * In: Increment * Au: Auto * ND: No Duplicates * DO: Duplicates OK

Valid_Sewer_Uses									
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
Value	Short Text		12				Yes	Yes	Yes (ND)
Abbreviation	Short Text		2				Yes	No	No

* LI: Long Integer * S: Single * T/F: True/False * I: Integer * In: Increment * Au: Auto * ND: No Duplicates * DO: Duplicates OK

Valid_Survey_Directions									
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
Value	Short Text		12				Yes	Yes	Yes (ND)
Abbreviation	Short Text		1				Yes	No	No

* LI: Long Integer * S: Single * T/F: True/False * I: Integer * In: Increment * Au: Auto * ND: No Duplicates * DO: Duplicates OK

Valid_Weather									
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
Value	Short Text		50				Yes	Yes	Yes (ND)
Abbreviation	Number		Byte		Au			No	No

* LI: Long Integer * S: Single * T/F: True/False * I: Integer * In: Increment * Au: Auto * ND: No Duplicates * DO: Duplicates OK

Valid_Purposes									
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
Value	Short Text		50				Yes	Yes	Yes (ND)
Abbreviation	Short Text		1				Yes	No	No

* LI: Long Integer * S: Single * T/F: True/False * I: Integer * In: Increment * Au: Auto * ND: No Duplicates * DO: Duplicates OK

Valid_PreCleaning									
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
Value	Short Text		24				Yes	Yes	Yes (ND)
Abbreviation	Short Text		1				Yes	No	No

* LI: Long Integer * S: Single * T/F: True/False * I: Integer * In: Increment * Au: Auto * ND: No Duplicates * DO: Duplicates OK

Valid_AccuracyOfGPS									
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
Value	Short Text		50				Yes	Yes	Yes (ND)
Abbreviation	Short Text		1				Yes	No	Yes (ND)

* LI: Long Integer * S: Single * T/F: True/False * I: Integer * In: Increment * Au: Auto * ND: No Duplicates * DO: Duplicates OK

Inspections										
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
InspectionID	AutoNumber	This field is automatically populated when any inspection information is entered.	LI	In					Yes (ND)
Surveyed_By	Short Text	Name of individual conducting survey	25				No	Yes	No
Certificate_Number	Short Text	NASSCO PACP # of Surveyor	15				No	Yes	No
Owner	Short Text	Owner of the collection system surveyed	30				Yes	No	No
Customer	Short Text	Entity commissioning the survey	30				Yes	No	No
Drainage_Area	Short Text	Common name of the drainage area	15				Yes	No	No
PO_Number	Short Text	Customer's Purchase Order Number	15				Yes	No	No

Inspections									
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
Pipe_Segment_Reference	Short Text	Client provided segment number	25				Yes	No	No
Date	Date/Time	Inspection Date	YYYYMMDD					Yes	No
Time	Date/Time	Time of inspection	Short Time					No	No
Street	Short Text	Street Number and Name	64				No	Yes	No
City	Short Text	City name where sewer located	64				No	Yes	No
Location_Details	Short Text	Descriptive explanation of sewer location	255				Yes	No	No
Upstream_MH	Short Text	Client provided designation for upstream manhole	25				No	Yes	No
Up_Rim_to_Invert	Number	Distance (ft and tenths of ft) or (meters to 2 decimal places max) from rim to invert of upstream manhole	S		2			No	No

Inspections									
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
Up_Grade_to_Invert	Number	Distance (ft and tenths of ft) or (meters to 2 decimal places max) from average grade to invert of upstream manhole	S		2			No	No
Up_Rim_to_Grade	Number	Distance (ft and tenths of ft) or (meters to 2 decimal places max) from rim to an average grade of upstream manhole	S		2			No	No
Downstream_MH	Short Text	Client provided designation for downstream manhole	25				No	Yes	No
Down_Rim_to_Invert	Number	Distance (ft and tenths of ft) or (meters to 2 decimal places max) from rim to invert of downstream manhole	S		2			No	No
Down_Grade_to_Invert	Number	Distance (ft and tenths of ft) or (meters to 2 decimal places max) from average grade to invert of downstream manhole	S		2			No	No
Down_Rim_to_Grade	Number	Distance (ft and tenths of ft) or (meters to 2 decimal places max) from rim to an average grade of downstream manhole	S		2			No	No
Sewer_Use	Short Text	Purpose of sewer	15				Yes	No	No
Direction	Short Text	Direction of the survey, Upstream or Downstream	10				No	Yes	No

Inspections									
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
Flow_Control	Short Text	Type restriction of the flow used	25				Yes	No	No
Height	Number	Diameter of sewer (or height if non-circular) to the nearest inch(999) or nearest mm(99999)	LI		0			Yes	No
Width	Number	Width of non-circular sewer to nearest inch(999) or nearest mm(99999)	LI		0			No	No
Shape	Short Text	Sewer shape	15				No	Yes	No
Material	Short Text	Type of pipe material	64				Yes	Yes	No
Lining_Method	Short Text	Type of process used to line the host pipe	30				Yes	No	No
Pipe_Joint_Length	Number	Length of pipe joint sections measured to one decimal place whether in feet or meters	S		1			No	No
Total_Length	Number	Distance between the exit of the start manhole and the entrance of the finish is measured to one decimal place whether it is feet or meters	S		1			No	No

Inspections									
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
Length_Surveyed	Number	If the survey is abandoned, enter the actual length surveyed to one decimal place whether it is feet or meters	S		1			No	No
Year_Laid	Number	Year sewer surveyed was constructed	LI		0			No	No
Year_Renewed	Number	Year sewer surveyed was renewed	LI		0			No	No
Media_Label	Short Text	Unique identifier for tape/media	64				Yes	No	No
Purpose	Short Text	Reason for conducting the survey	64				Yes	No	No
Sewer_Category	Short Text	Importance of sewer, to be provided by the client	2				Yes	No	No
Pre-Cleaning	Short Text	Type of preparatory cleaning conducted prior to the survey	15				No	Yes	No
Date_Cleaned	Date/Time	Date when the sewer was cleaned prior to survey						No	No

Inspections										
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
Weather	Short Text	Weather conditions when a survey conducted	12				Yes	No	No
Location_Code	Short Text	General description of ground cover of the surveyed segment	30				Yes	No	Yes (DO)
Additional_Info	Short Text	Supplemental info regarding survey or segment	255				Yes	No	No
Reverse_Setup	Number	Specifies that a second survey has been done on the pipe segment--use inspection ID from matching survey	LI		Au	0		No	No
Sheet_Number	Number	Number used to identify individual surveys done within a group	LI		Au	0		No	No
IsImperial	Yes/No	Used to identify whether units are metric or imperial. Defaults to imperial.	T/F			True			No
PressureValue	Number	Grouting pressure value	S		3	0		No	No
WorkOrder	Short Text	Work order or Project reference for Asset Management	20				Yes	No	No

Inspections										
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
Project	Short Text	Project Title or reference for Asset Management	64				Yes	No	No
Northing	Short Text	Y Coordinate - Latitude at the center point of the Starting Access Point - If value exists, Easting and Coordinate System are also required	50				Yes	No	No
Easting	Short Text	X Coordinate - Longitude at the center point of the Starting Access Point- If value exists, Northing and Coordinate System are also required	50				Yes	No	No
Elevation	Short Text	Z Coordinate - Height at the center point of the Starting Access Point	50				Yes	No	No
Coordinate_System	Short Text	Datum or reference system used for the gps coordinates - If value exists, Northing and Easting are also required	50				Yes	No	No
GPS_Accuracy	Short Text	Describes degree of accuracy obtained from coordinates	50				Yes	No	No

* LI: Long Integer * S: Single * T/F: True/False * I: Integer * In: Increment * Au: Auto * ND: No Duplicates * DO: Duplicates OK

PACP_Version									
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
Version	Short Text	PACP/LACP export version	50				No	Yes	Yes (ND)
Date	Date/Time		1					No	No
Notes	Long Text						Yes	No	No
SoftwareVendor	Short Text	Software Vendor for the program used to create the export	25				No	Yes	No
SoftwareProgram	Short Text	Software Program used to create the export	25				No	Yes	No
SoftwareVersionNumber	Short Text	Version Number of Software Program	10				No	Yes	No

* LI: Long Integer * S: Single * T/F: True/False * I: Integer * In: Increment * Au: Auto * ND: No Duplicates * DO: Duplicates OK

Code_Substitutions										
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
Code	Short Text	Obsolete Code	10				No	Yes	Yes (ND)
Type	Short Text	PACP, LACP, or Both	255				Yes	No	No
Substitute_Code	Short Text	Code to substitute for Obsolete Code	10				No	Yes	No
Instructions	Short Text		255				Yes	No	No

* LI: Long Integer * S: Single * T/F: True/False * I: Integer * In: Increment * Au: Auto * ND: No Duplicates * DO: Duplicates OK

Custom_Labels									
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
Field_Name	Short Text		18				No	Yes	Yes (ND)
Label	Short Text		255				Yes	No	No
Description	Short Text		255				Yes	No	No

* LI: Long Integer * S: Single * T/F: True/False * I: Integer * In: Increment * Au: Auto * ND: No Duplicates * DO: Duplicates OK

Custom_Fields									
Field Name	**Data Type**	**Description**	**Field Size**	**New Values**	**Decimal Places**	**Default Value**	**Allow zero length**	**Required**	**Indexed**
CustomID	AutoNumber	This field is automatically populated for each custom field.	LI	In					Yes (ND)
InspectionID	Number	Software provided designation for this inspection (THIS FIELD USED TO JOIN TABLES)	Li		Au	0		No	Yes (ND)
Custom_Field_One	Short Text		255				Yes	No	No
Custom_Field_One	Short Text		255				Yes	No	No
Custom_Field_One	Short Text		255				Yes	No	No
Custom_Field_One	Short Text		255				Yes	No	No
Custom_Field_One	Short Text		255				Yes	No	No
Custom_Field_One	Short Text		255				Yes	No	No

* LI: Long Integer * S: Single * T/F: True/False * I: Integer * In: Increment * Au: Auto * ND: No Duplicates * DO: Duplicates OK

VITA

MOHAMAD HOSSEIN KHALEGHIAN

Candidate for the Degree of

Doctor of Philosophy

Thesis: TOTAL QUALITY MANAGEMENT AND INTEGRATION OF HETEROGENEOUS SEWER INSPECTION DATABASES FOR ASSET MANAGEMENT

Major Field: Civil Engineering

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Civil Engineering at Oklahoma State University, Stillwater, Oklahoma in July 2021.

Master of Science in Civil Engineering at Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran in July 2009

Bachelor of Science in Civil Engineering at Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran in July 2006

Experience:

One-Voice Team Lead and Researcher, VentureWell E- Team 2020 - 2021
One-Voice Team Lead and Researcher, OSU Pre-Seed Fund 2019 - 2020
Entrepreneurial Lead, NSF I-Corps Program 2018 - 2019
Graduate Research Assistant, Oklahoma State University 2012 - 2018

Professional Memberships:

- Member of Water Environment Federation (WEF)
- Member of American Society of Civil Engineers (ASCE)
- Member of Chi Epsilon (The Civil Engineering Honor Society)

Grants and Awards:

- E-Team Grant for Student Startups, VentureWell ($25,000) – 2020
- OSU Foundation Pre-Seed Fund,OSU ($25,000) – 2019
- National I-Corps grant, NSF ($50,000) – 2018