Development of CPATT Database and Quality Control Checks for Concrete Pavement Field Data

by

Abid Tabassum

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Civil Engineering

Waterloo, Ontario, Canada, 2022

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner
Professor Dr. Chen-Ping (Kevin) Wang
School of Civil and Environmental Engineering, Oklahoma State University

Supervisor
Professor Susan L. Tighe
Civil and Environmental Engineering, University of Waterloo

Supervisor
Professor Shunde Yin
Civil and Environmental Engineering, University of Waterloo

Internal Member
Professor Alex Penlidis
Civil and Environmental Engineering, University of Waterloo

Internal Member
Professor Wei-Chau Xie
Civil and Environmental Engineering, University of Waterloo

Internal-external Member
Professor Vimy Henderson

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public.

Abstract

The Centre for Pavement and Transportation Technology (CPATT) located at the University of Waterloo, Canada, has robust research capabilities. This is largely related to a team of researchers who have conducted a number of research(s) related to transportation and pavement engineering in collaboration with various public and private sector partners. Numerous research has involved collection, acquisition and development of new data and information involving design, construction, maintenance, economic, rehabilitation, safety and impact of climatic changes on the pavement.

The CPATT researchers collected data has a magnitude of multi-gigabytes. This required that an appropriate repository is provided and maintained for future students, researchers, and research partners. The repository qualifications and its maintenance process are detailed in this study.

At the start of this research, there was a detached data repository and a framework for collecting, storing and maintaining the database. This research has provided solutions to form the basis of a robust, meaningful and useful database, by researching and developing a repository, creation of numerous standard formats for datasets, inter-relationship models and quality control checks. The research has evolved so that concrete pavement field data can be stored safely and accessed by students, and researchers for analysis and its utilization in the future.

In developing a database framework, the literature indicates that an appropriate consultation with experts and rigorous evaluation of database framework (before its implementation) is to be carried out to meet the objectives and goals of the program. This objective was achieved by consulting CPATT management, IT experts (both internal and external to the University of Waterloo) and end-users, such as current and past CPATT students, research associates, and UW staff through a well-articulated "CPATT Database Survey".

Data quality control and datasets format consistency of existing CPATT data were of a major concern, addressed by this research. This concern is addressed by providing numerous standard datasets formats and quality control checks; for dataset utilization to be more feasible and valuable for future researchers.

Acknowledgements

My gratefulness to the Almighty Allah, the creator and beneficent of the universe, without who's countless bounties and blessings, I would not have been able to be where I am today, and My Prophet Muhammad (Peace and salutations upon him).

I wish to express my deepest gratitude for Professor Dr. Susan L. Tighe. My relationship with Dr. Tighe goes back to 2012, when I first studied at the University of Waterloo as an M.Eng. student. Thank you so much for your many kindnesses and providing me the opportunity to work with your auspicious direction. And for keeping the path illuminated by extending support, guidance, empathy and inspiration for the past many years. Dr. Tighe provided a positive environment, where I was encouraged and challenged to do my best at all times.

Special thanks to Dr. Daniel Pickel, who played an instrumental role in helping towards CPATT database implementation, refining data uploading procedures, and provided valuable feedback on different data standard formats.

Thanks for the support of Civil and Environmental Engineering Department Staff at the University of Waterloo such as, Dr. Luke Zhao, Jessica Rossi, Rahnuma Wafa, Lorrain Quast, Victoria Tolton, Beverly Seibel and numerous others for providing their continuous support.

I am also thankful to numerous external organization officials, such as from Ministry of Transportation Ontario, Canada, Highway Administration USA, and Dr. Richard Korczak of Stantec Consulting Inc.

Furthermore, I am thankful to my CPATT colleagues for their provision of data, its clarification and support during my research.

I am grateful to my spouse, children, siblings, in-laws and relatives for their support and inspiration throughout my studies.

Last, but not least, I am thankful to all those who have extended their support during this research.

Dedication

In the name of Almighty Allah (Jalla JalalaHu) the most beneficent and merciful – Ever-existing

Prophet Muhammad ibn e Abdullah (S.A.W. - Peace and Salutations upon him)

And

My Late parents M. Sharif Khan and Qazi Mumlikat

And

My spouse, Sabeen; my children Muhammad, Ayesha, A-Rahman and Imaan

My siblings, my parent-in-laws especially Malik Ahmad Hussain

Table of Contents

Examining Committee Membership ii
Author's Declaration iii
Abstract iv
Acknowledgements vi
Dedication viii
List of Figures xii
List of Tables xiii
List of Abbreviations xv

1. INTRODUCTION 1

1.1 Statement of the Problem 1
1.2 Research Hypothesis 2
1.3 Research Objectives 3
1.4 Methodology of the Study 4
1.4.1. Initial parameters setup process 4
1.4.2. Database Development 5
1.4.3. Data sets standard format 5
1.4.4. Data Inter-relationship Models 5
1.4.5. Datasets Quality Acceptance Checks 6
1.5 Summary and Research Gap 6

2. LITERATURE REVIEW 7

2.1 Current Practices 7
2.1.1 Database Scop 7
2.1.2 Big Data 12
2.2 Cloud based databases 18
2.3 Development of Dataset Standard Format 22
2.4 Dataset Models (Data Analytics and its link with Pavement Engineering/ Management) 23
2.5 Quality Acceptance Checks 27

2.6 Selection of CPATT Database Software and Repository ... 29
2.7 Selection of CPATT Database Creation ... 30
2.8 Chapter Summary ... 31
3. RESEARCH METHODOLOGY ... **32**
3.1 Selection of Methodology ... 32
3.2 Database Framework Development Process ... 33
3.3 Database Normalization Process ... 37
3.4 Chapter Summary ... 41
4. DATABASE FRAMEWORK ... **42**
4.1 Database ... 42
4.1.1 Database Definition ... 42
4.1.2 Database advantages and disadvantages ... 43
4.2 Selection Criterion for CPATT Database ... 44
4.3 Database Management System (DBMS) – Features and Characteristics ... 48
4.4 CPATT Database Management System (DBMS) ... 50
4.4.1 Microsoft Teams ... 50
4.4.2 Is SharePoint a Database? ... 51
4.4.3 Microsoft Teams + SharePoint: Collaboration Tools ... 52
4.5 CPATT Database Survey Results ... 53
4.5.1 General ... 53
4.5.2 Data collection and storing efforts ... 54
4.5.3 Data sort and analysis efforts ... 54
4.5.4 Current practices ... 55
4.5.5 CPATT Preferred software ... 56
4.5.6 Familiarity with different software ... 57
4.5.7 Critical Success Factors Criterion ... 59
4.5.8 CPATT Survey Summary ... 60
4.6 Documents Collaboration and Co-authoring ... 61
4.7 Process to Upload Datasets to CPATT Database ... 61
4.8 Process to maintain/safety of CPATT Database ... 67

4.9 Chapter Summary 68

5. INTER-RELATION DATA MODELS 70

5.1 Current Data collection and developed Models 70

5.2 Data format, structure and requirements 76

5.3 British Pendulum Tester and T2GO 80

5.4 Regression Analysis Model for BPN and T2GO 86

5.4 Regression Analysis Model for FWD and LWD 92

5.5 Chapter Summary 101

6. DATASETS QUALITY CHECKS 102

6.1 Existing Falling Weight Deflectometer (FWD) Data 102

6.2 Statistical Methods 115

6.3 Percent Within Limits (PWL) 115

6.4 Chapter Summary 124

7. CPATT DATABASE 125

7.1 Layout of CPATT Database 125

7.2 Salient datasets and standards included 128

7.3 Chapter Summary 129

8. CONCLUSIONS AND RECOMMENDATIONS 130

8.1 Conclusion 130

8.2 Recommendations 130

References 133

Appendices 139

List of Figures

FIGURE 1: TYPICAL LAYOUT OF 9 SENSORS (LTPP) FALLING WEIGHT DEFLECTOMETER (FWD) 9
FIGURE 2: A DEPICTION OF CPATT LIGHT WEIGHT DEFLECTOMETER (LWD) EQUIPMENT 11
FIGURE 3: DATA ANALYTICS APPROACH 13
FIGURE 4: BIG DATA'S 3V'S 14
FIGURE 5: DIFFERENT TYPES OF VELOCITY FOR BIG DATA 15
FIGURE 6: PERCENTAGE OF CURRENT AND FUTURE USAGE OF CLOUD SERVERS 19
FIGURE 7: PERCENTAGE OF WORKLOAD IN THE CLOUD 19
FIGURE 8: CPATT CURRENT DATA COLLECTION AND DATABASE DEVELOPMENT 24
FIGURE 9: CPATT DATABASE MODEL 36
FIGURE 10: DIAGRAM FOR FIELDS, RECORDS, FILES AND DATABASE 44
FIGURE 11: CPATT DATABASE SELECTION CRITERION 46
FIGURE 12: CPATT DATABASE - UPLOAD PROCESS FLOW CHART 65
FIGURE 13: FWD VS LWD DATASETS 72
FIGURE 14: STRAIN GAUGES INTER-RELATIONSHIP MODEL 75
FIGURE 15: A RDBMS LAYOUT OF USER INTERFACE 80
FIGURE 16: BRITISH PENDULUM TEST APPARATUS 81
FIGURE 17: BRITISH PENDULUM TESTING ON JAMESTON AVENUE, HAMILTON, ONTARIO 84
FIGURE 18: T2GO TESTING ON JAMESTON AVE., HAMILTON CONCRETE OVERLAY 85
FIGURE 19: CHART OF REGRESSION MODEL FOR BPN AND T2GO 90
FIGURE 20: REGRESSION MODEL (ACTUAL AND PREDICTED – VS- OBSERVATION #) 91
FIGURE 21: NORMAL PROBABILITY PLOT 93
FIGURE 22: BPN/T2GO RESIDUALS GRAPH 93
FIGURE 23: REGRESSION MODEL FOR FWD AND LWD (DEFLECTIONS) 95
FIGURE 24: LWD (DEFLECTIONS) - ACTUAL AND PREDICTED 96
FIGURE 25: REGRESSION MODEL FOR FWD AND LWD (MODULUS) 99
FIGURE 26: LWD (MODULUS) - ACTUAL AND PREDICTED 100
FIGURE 27: D1 AND D2 SENSOR READINGS SPC CHART 109
FIGURE 28: D3 AND D4 SENSOR READINGS SPC CHART 110
FIGURE 29: D5 AND D6 SENSOR READINGS SPC CHART 111
FIGURE 30: D7 AND D8 SENSOR READINGS SPC CHART 112
FIGURE 31: D9 SENSOR READINGS SPC CHART 113
FIGURE 32: MACRO VIEW OF TYPICAL STATISTICALLY BASED QC/QA ACCEPTANCE SPECIFICATION 118
FIGURE 33: A SCREEN SHOT FROM DEVELOPED CPATT DATABASE 125
FIGURE 34: A SCREEN SHOT FROM DEVELOPED CPATT DATABASE 126
FIGURE 35: A SCREEN SHOT FROM DEVELOPED CPATT DATABASE 126
FIGURE 36: A SCREEN SHOT FROM DEVELOPED CPATT DATABASE 127

List of Tables

TABLE 1: CHALLENGES FOR BIG DATA MANAGEMENT 15
TABLE 2: SALIENT BENEFITS OF CLOUD (SHAREPOINT) 20
TABLE 3: PREVIOUS STUDY - REGRESSION ANALYSIS BETWEEN LWD AND FWD MODULUS 25
TABLE 4: METHODS (PERCENT) OF TRANSPORTATION AGENCIES (USA) FOR DATA ACCEPTANCE CHECKS 28
TABLE 5: A SCREEN-SHOT FORM ACTUAL CPATT DATA SET FOR LWD 38
TABLE 6: ADJUSTMENTS MADE TO GET THE 1NF 38
TABLE 7: ID LOCATION- SELECTED AS PRIMARY KEY FOR DATA NORMALIZATION 39
TABLE 8: GENERAL NORMALIZATION PROCESS 40
TABLE 9: THE ADVANTAGES OF A MICROSOFT TEAMS/SHAREPOINT INTEGRATION 53
TABLE 10: CPATT SURVEY RESULTS (A) 54
TABLE 11: CPATT SURVEY RESULTS (B) 54
TABLE 12: CPATT SURVEY RESULTS (C) 55
TABLE 13: CPATT SURVEY RESULTS (D) 56
TABLE 14: CPATT SURVEY RESULTS (E) 57
TABLE 15: CPATT SURVEY RESULTS (F) 58
TABLE 16: CPATT SURVEY RESULTS (G) 59
TABLE 17: LIST OF CPATT DATABASE ABBREVIATIONS (FOR EQUIPMENT/TEST) 64
TABLE 18: LIST OF CPATT DATABASE FOUR (4) DIGIT CODES 66
TABLE 19: STRAIN GAUGE DATA DETAILS CONCRETE OVERLAYS RESEARCH 73
TABLE 20: DATA REQUIREMENTS/REDUNDANCY EXAMPLE 78
TABLE 21: FILE TYPES / FORMATS FOR CPATT DATABASE 79
TABLE 22: BPN/T2GO – COUNT OF SECTIONS AND READINGS 86
TABLE 23: BPN/T2GO DATA COMPILED AND USED FOR REGRESSION MODEL 87
TABLE 24: DATASET FOR BPN (RAW) [2016 & 2017] 88
TABLE 25: DATASET FOR T2GO (RAW) [2016 & 2017] 89
TABLE 26: REGRESSION MODEL (CRITICAL INDICATORS OF CORRELATION) BPN & T2GO 91
TABLE 27: REGRESSION MODEL DATA FOR FWD AND LWD (USING NORMALIZED DEFLECTION DATA) 94
TABLE 28: REGRESSION MODEL DATA FOR FWD AND LWD (USING NORMALIZED DEFLECTION AND MODULUS DATA) 98
TABLE 29: MODULUS EQUATION FROM REGRESSION ANALYSIS FOR LWD AND FWD 99
TABLE 30: REGRESSION MODEL (CRITICAL INDICATORS OF CORRELATION) LWD & FWD MODULUS OF ELASTICITY 100
TABLE 31: DATA CHECKS FOR FWD RESULTS FOR JAMESTON AVENUE (NON-DECREASING DEFLECTIONS) 104
TABLE 32: TEMPERATURE BOREHOLES SIZE 107
TABLE 33: SENSORS, NOTED ERRORS DURING SPC AND ITS INTERPRETATIONS 114
TABLE 34: SALIENT FORMULAE FOR PERCENT WITHIN LIMITS 117
TABLE 35: BPN 2017 DATA USED FOR PWL PROCESS (SOURCE: (WAFA R. , 2018) 119
TABLE 36: F-TEST RESULTS FOR PWL PROCESS 121

TABLE 37: T-TEST RESULTS FOR PWL PROCESS 122
TABLE 38: BPN2017 DATA - PWL (STATISTICAL RESULTS) 123

List of Abbreviations

Abbreviation	Details
S.A.W	Sallal-Laho Alaihey Wasallam (Salutation and Peace be Upon Prophet Muhammad – S.A.W)
3V	Volume, Velocity and Variety
4DIG	Four Digit (CPATT Codes)
AASHTO	American Association of State Highway and Transportation Officials
ABBR	Abbreviations
ACID	Accuracy, Completeness, and Isolation and Durability
AN01	Analysis # 1
ANOVA	Statistical Analysis
ASCII	Computer file format
ASFT	Airport Surface Friction Tester
ASTM	American Standards and Testing Methods
BP	British Pendulum
BPN	British Pendulum Number
BPT	British Pendulum Test
CA	Canada
CAPTT	Centre for Pavement and Transportation Technology
COF	Coefficient of Friction
CSF	Critical Success Factors
CSV	Comma-separated values
D1	Deflection at Point 1 (FWD)
DB	Database
DBA	Database Analysis
DBMS	Data base Management System
DMT	Data Management Tool
DOCX	MS Word file extension
DOT	Department of Transportation
E_{LWD}	Modulus of Elasticity for Light Weight Deflectometer
EPCC	Elastic Modulus for Concrete
EQP	Equipment
FHWA	Federal Highway Administration, United States
FWD	Falling Weight Deflectometer
GIS	Geographical Information System

HMA	Hot Mix Asphalt
HWY	Highway
INFO	Information
IRI	International Roughness Index
ISO	International Standards Organization
IWP	Inner Wheel Path
JPEG	Photo file extension
JPG	Photo file extension
K	Kilo
KN	Kilo-Newton
LCCA	Life Cycle Cost Analysis
LCL	Lower Control Limits
LSL	Lower Specification Limit
LTE	Load Transfer Efficiency
LTPP	Long Term Pavement Performance
LWD	Light Weight Deflectometer
MATLAB	Computer Software name
MEPDG	Mechanical-Empirical Pavement Design Guide
MM	Millimeter
MOI	Video file extension
MOV	Video file extension
MP3	Audio file extension
MP4	Audio file extension
MPF	Master_Project_File.xlsx
MTO	Ministry of Transportation Ontario
NF	Normal Form (Database Normalization)
OWP	Outer Wheel Path
PCC	Plain Cement Concrete
PDF	File format
PF	Pay Factor
PNG	Photo file extension
PWL	Percent Within Limit
QA	Quality Assurance
QC	Quality Control
RDBMS	Relational Database Management System
SME	Subject Matter Experts
SPSS	Statistical software
SQL	Database software

SRT	Skid resistant tester
SURPRO	International Surface Index Equipment
T2GO	Skid resistant tester equipment
TAC	Transportation Association Canada
TB	Tera-Byte
UCL	Upper Control Limits
UK	United Kingdom
US	United States
USA	United States of America
UW	University of Waterloo, Waterloo, Ontario, Canada
XML	File format extension
YR	Year
YYYY	Year
YYYYMMDD	Year Month Date

1. INTRODUCTION

1.1 Statement of the Problem

Over the past 15 years, the Centre for Pavement and Transportation Technology (CPATT) located at the University of Waterloo, has been working on various research that involves collection and acquisition of pavement data; for example, data comes from various research(s) which involve design, construction, maintenance, rehabilitation, safety and impact of climatic changes on the pavement. Currently, CPATT team members collect various types of data including pavement condition data, pavement materials data, construction data, economic, cost and time data in laboratory and in the field. At the present time, there is no structured data repository or a framework for collecting, storing and maintaining the database. This research provides basis for data collection for concrete pavement projects, how data could be stored and accessed by students, researchers or the general public for future analysis. The purpose of this research is to provide a stream-lined process for developing a framework, datasets standardization, in-place quality checks criteria, storing data and accessing the collected data in such a way so the data can be stored efficiently and safely such that it can be used in the future along with ancillary information.

Data quality and datasets format consistency of existing CPATT data is a concern identified during the initial evaluation of data. This concern is addressed during the course of this research for dataset use to be more feasible for future researchers.

Since CPATT students and researchers have collected data for more than a decade the amount of data collected is in the order of multi-gigabytes. This requires that an appropriate repository is not only provided but also maintained for future students, researchers and research partners. An apposite repository is identified and will be maintained in future by CPATT.

In this context, the data collected has been efficiently preserved for future analysis. It would thus be desirable to develop, a specific database framework and synergized strategy so that the data collection, suitable storing efforts can be streamlined. When developing database frameworks, the literature indicates an appropriate consultation with experts and rigorous evaluation of database framework (before its implementation) be carried out to meet the objectives and goals of the program. The structure of the data should be consistent to enable efficient and safe data storage for future analyses. Subject Matter Experts and end users (CPATT affiliates) of the database were consulted and recommendations were streamlined and implemented through this research.

The primary research objective is to design and create a useful, robust and meaningful CPATT database; as well as to implement a standardized database framework for CPATT concrete pavement data. Secondary objectives include standardizing data compilation processes for Falling Weight Deflectometer (FWD), Light Weight Deflectometer (LWD), International Roughness Index (IRI), T2GO, British Pendulum Number (BPN), etc. As well standard quality control checks for data; and development of few data-permitting inter-relationship models for different types of datasets. Data-permitting datasets are those data sets which has been collected at the same location during close time proximity.

1.2 Research Hypothesis

The main hypothesis for this research is as below:

- CPATT data acquisition is expected to increase in future and will have significant issues of technical data management, and data being not stored properly.
- The development of an apposite database framework for pavements will streamline efforts of the researchers thus reducing costs.
- Development of a new database framework, standardization of datasets and quality assurance advances the data analysis, accuracy and precision of collected data.
- The developed database framework will significantly enhance data quality, prolong the usable life for data analytics and store CPATT data under one common database repository.

1.3 Research Objectives

This research is directed at the development of a useful and robust database framework for CPATT data, with the final goal of providing access to CPATT concrete data under one portal. This is now available to researchers under one portal where data quality is enhanced thorough rigorous checks, and data investigation avenues are improved. This research will only initially include the CPATT concrete pavement field data, with the possibility of extending these frameworks to other types of data such as asphalt pavement data, and various types of pavement management data.

This research investigated and selected a database structure for CPATT data. The CPATT management is eager to have a more robust database and offer access to researchers to overcome data inconsistencies and further enhance data analysis.

- To develop a database framework for CPATT data.
- To build a robust, useful and meaningful database.
- Combine all available CPATT concrete pavement data under one portal, this will create consistent data and enhance pavement data analyses in the future.
- To investigate appropriate database structure type and database software for CPATT data.
- To create data-permitting inter-relationship models for available data.
- To progress quality check for specific data-permitting datasets
- To create standard formats for concrete pavement datasets, such as FWD, LWD, IRI, BPN, T2GO etc.

1.4 Methodology of the Study

1.4.1. Initial parameters setup process

The first part of the literature review and data collection procedure ran in parallel. Initially, collection of available CPATT concrete field pavement data and selection of initial parameters for the database involved an evaluation which included critical decisions regarding:

- Database structure type,
- File format,
- Selection of apposite Data Management Tool (DMT),
- Selection of a repository for new database,
- Database resources, and
- Development of data set quality acceptance checks.

Initial parameter will include consideration of factors such as:

1. Cost of initial purchase of the data management tool/software for the institution (UW).
2. Ease of data management tool/software and database setup.
3. Database maintenance.
4. Costs for researcher to acquire the data management tool/software and its use.
5. Ease of use of the database software for researcher.
6. Selection of repository for the proposed database.
7. Evaluation of providing access to researchers for the new database.
8. Evaluation of database maintenance personnel.
9. Which data management tool/software among the few chosen based on literature review will be considered for CPATT database.
10. Database structure type for pavement data
11. Which file format will be most efficient?

In the evaluation, it was determined that a few of these selected parameters needed to be considered as critical success factors (CSF) in developing the CPATT database. These CSF were selected and evaluated through CPATT stakeholders survey (discussed later in this research).

1.4.2. Database Development

After the evaluation of critical success factors (CSF), the database will run through the following process:

1. Database Design
2. Database implementation
3. Database conversion and loading
4. Database testing
5. Final database

The results of this process will provide a robust, meaningful and useful CPATT database.

1.4.3. Data sets standard format

Standards were developed for different type of data sets of which were included in the CPATT Database to ensure common understanding and consistency of data sets is achieved. These datasets include standard formats for, FWD, LWD, IRI, BPN, T2GO etc.

1.4.4. Data Inter-relationship Models

A primary objective of this research is to develop inter-relationship models, so that future researchers can wisely use the data and have few guidelines to do so. Parallel to developing the database process, the data-permitting inter-relationship models will be developed for CPATT concrete pavement data. As an example few of the inter-relationship models are developed and presented later in this research.

1.4.5. Datasets Quality Acceptance Checks

Ensuring datasets quality is a major concern. In order to address this appropriately data quality assurance plans/checks are developed and presented in this research. These checks includes range limits, three sigma outliers, comparing means and lower/upper limits; as well several others quality checks.

1.5 Summary and Research Gap

As discussed earlier in this chapter, the current CPATT database requires an overhaul due to the following reasons, such as:

Lack of one database for all the datasets.

No common repository and standardization of formats.

Quality assurance checks to ensure only high-quality datasets are included in the CPATT database.

Missing ancillary information for the datasets of the collected data.

Maintenance of the CPATT database.

The above objectives is addressed through following salient activities, such as:

Provide a robust and meaningful database for CPATT datasets, ensuring high-quality datasets intake and conducting appropriate survey of the CPATT stakeholders to ensure it serves the purpose of researchers involved.

Development of inter-relational models for the data-permitting-datasets, such as LWD and FWD as well between T2GO and BPN

Standardization of the dataset forms for better data analytics in the near and far future.

2. LITERATURE REVIEW

The purpose of this chapter is to review in detail the current practices through extensive literature review; both inside and outside of CPATT for development, standardization of datasets, datasets quality checks, data analytics and inter-relational model developments.

Identify gaps between current CPATT database development and maintenance practices through both literature review and extensive consultation with IT experts.

2.1 Current Practices

2.1.1 Database Scop

With more than 1.42 million km of road networks (Transportation Association of Canada, 2014) and over 2.5 million m^2 of bridge area in Canada (Canadian National Committee of the World Road Association in Partnership with Ontario Ministry of Transportation, 2015). Several organizations (both private and government) are working to collect, sort and analyze data related to pavement and transportation. This includes pavement condition data such as roughness, rutting depth, pavement cracking etc.; as well climate data such as meteorological data, and other types of transportation data such as economic and technical data.

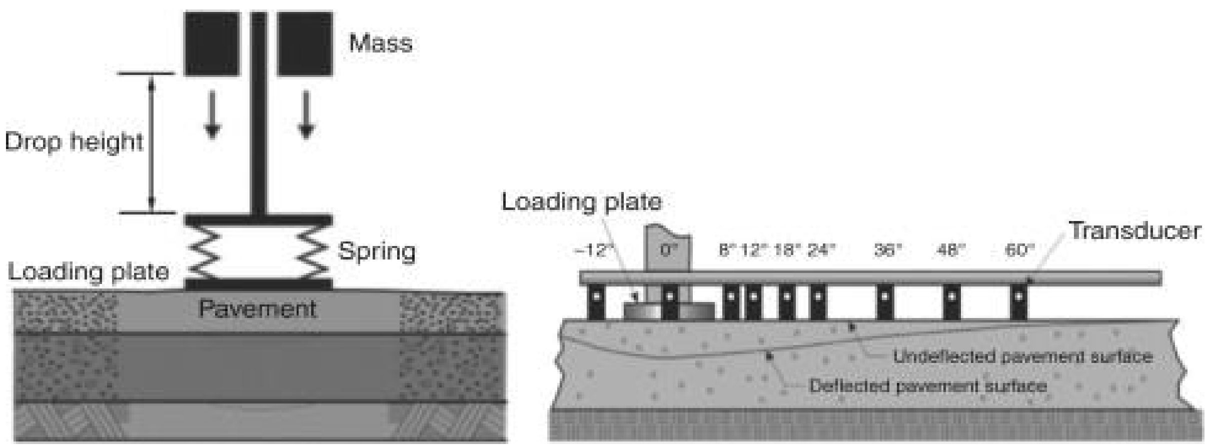
Iowa Department of Transportation (DOT), has determined that the cost to collect data for city streets and country road varies between US$ 47/km to US$ 37/km respectively in the USA (Iowa DOT , 2009). These costs are based on 2009 and should be much higher, when meeting current market rates. The 2018 costs were calculated based on bank of Canada inflation and conversion rates. Data collection cost are CA$ 64/km for city streets and CA$ 55/km for country road in 2018 for Canada.

The data collected is either structured such as categorical and numerical or unstructured using figures, videos and text. Cost-effectiveness of this data is a major concern as a major portion of money allocated for the pavement management system is spent on data collection and the velocity at which the data is acquired ranges from small data (few kilo-bytes) to big-data (multi-thousand-gigabytes) (Frost & Sullivan, 2012).

The Centre for Pavement and Transportation Technology (CPATT) has been involved in many laboratory and field projects since its inception in 2002. As a result of these several and varied research(s), there is a huge amount of data that has been generated based on the various research studies. These data relate to new materials test results and design related to various concrete, asphalt and alternative pavement materials.

An extensive concrete pavement CPATT research program has evolved several projects. Concrete Field Data (such as for precast concrete inlays, concrete overlays, etc.) and Concrete Laboratory Data (such as for recycled aggregate concrete, new aggregate concrete etc.) has been collected for different projects and continues to be collected in some cases. Concrete data is divided into the categories as depicted in Figure 88. In this research the available concrete field datasets is used and combined with data analytics to create a database that is functional and meaningful.

A typical layout of the 9 sensors Dyantest (LTPP) Falling Weight Deflectometer (FWD) is shown below:



Source: ScienceDirect.com

Figure 1: Typical layout of 9 sensors (LTPP) Falling Weight Deflectometer (FWD)

The following salient datasets are considered for in this research:

1. Three datasets of Jameston Avenue, Hamilton, Ontario – Falling Weight Deflectometer (FWD) before construction (2016), FWD after construction (2017) for concrete overlay, and Light Weight Deflectometer (LWD) during construction (2017)
2. Strain gauges data (2016-todate), of Jameston Avenue, Hamilton, Ontario for concrete overlay
3. Strain gauges data (2015-todate), of Spragues Road, Region of Waterloo, Ontario for concrete overlay
4. FWD datasets (2013) for test section at MTO series 400 highways over pre-cast concrete inlay panels (CPATT Database - precast concrete inlays, 2013).
5. International Roughness Index (IRI), friction, moisture and pressure datasets at MTO series 400 highways over pre-cast concrete inlay panels (CPATT Database - precast concrete inlays, 2013).

6. LTPP’s FWD datasets (for numerous States in the USA) (Long Term Pavement Performance (LTPP)).
7. Two datasets for British Pendulum Tester (BPN), 2016 and 2017 for Jameston Avenue, Hamilton, Ontario, Canada.
8. Two datasets for T2GO Test, 2016 and 2017 for Jameston Avenue, Hamilton, Ontario, Canada.

Beside these, there are other CPATT research datasets available for analytics, such as weather station data, surface texture indices, etc., for which inter-relational models can be developed as part of future studies/research(s). It is pertinent to mention here, that initially some CPATT datasets beside mentioned above were also considered for development of inter-relationship models, quality control checks and development of standard formats. However, after detailed analysis of both, the datasets available, its existing quality and need for those datasets to be closely related as data-permitting for development of inter-relationship models, quality control checks and development of standard formats; some of the options are not considered although these are well-thought-out and discussed briefly in numerous occasions in this study.

To meet its robust research requirements and to enhance the capabilities to extend in-house research analysis, CPATT procured a Dynatest Light Weight Deflectometer (LWD) 3031. The following paragraphs highlight some salient features of this equipment as provided by (Dynatest, 2006).

The equipment is precision-engineered, using stainless or anodized material for all metal parts. The system is powered by a pack of four AA alkaline or rechargeable batteries, providing approximately 2000 measurements or the equivalent to more than 12 hours of continuous operation. With the additional (optional) 2 x 5kg weights added, the Dynatest LWD can produce up to 15kN peak loads. The LWD weighs about 22 kg (with the standard 10kg drop weight), and it is highly portable and easily carried around a construction site. This LWD can be mounted on a specially designed trolley for easy usage during the testing process.

The Dynatest LWD needs no reference measurements and provides an easy, cost-effective alternative to time-consuming and expensive static plate bearing testing. This LWD is ideal for Quality Assurance / Quality Control on subgrade, subbase and thin flexible pavement constructions to verify that specifications are met. It can also be used to identify weaknesses, leading to further tests using FWDs and other material analysis techniques.

This Dynatest LWD was used by a CPATT researcher(s) in year of 2016 and 2017 at Jameston Avenue, in Hamilton, Ontario to record deflection readings off a concrete pavement (Wafa R. , 2018). Figure 2, depicts a typical Light Weight Deflectometer (LWD), owned and utilized by CPATT research team for recording and analyses of deflections.

Source: (Dynatest, 2006)

Figure 2: A depiction of CPATT Light Weight Deflectometer (LWD) equipment

2.1.2 Big Data

In 2019, Cisco projected that for 2020 the total internet annual traffic will raise to more than 2 zettabytes (1000^7 bytes) (Chojecki, 2019). In 1999, our total was at 1.5 exabytes of data and 1 gigabyte was thought of as big data. In 2006, total data was at 160 exabytes - 1000% more in 7 years. In the Zettabyte Era, 1 gigabyte is considered no longer to be big data really, and its more sensible to measure big data starting with 1 terabyte (1000^4 bytes). In mathematical terms, it makes more sense to consider Big Data with regard to datasets which exceed total data created in the world divided by 1000^3 (Chojecki, 2019). Figure 3, shows four sections, based on the current evaluation, CPATT data falls under the category Basic Analytics, as a result of this research the CPATT data will qualify and fall under Big-data Analytics.

This move from Basic Analytics to Big-data Analytics will be achieved through a systematic approach of analyzing existing available data, development of a new framework and quality acceptance checks.

Big-data is the data which can combine different large amounts of data types (gigabytes to petabytes) on a large scale which can provide predictive and real-time analyses. Big-data is a combination of three variables, namely; volume, variety and velocity or 3V's. Big data means there is a large amount of data (Volume), with masses of different types of data (Variety), and the capacity to extract valuable information from the data after processing it in real time (Velocity). Big-data is vast not only because of its volume but also due to the different types of data as well the speed in which it is to be managed. (Frost & Sullivan, 2012)

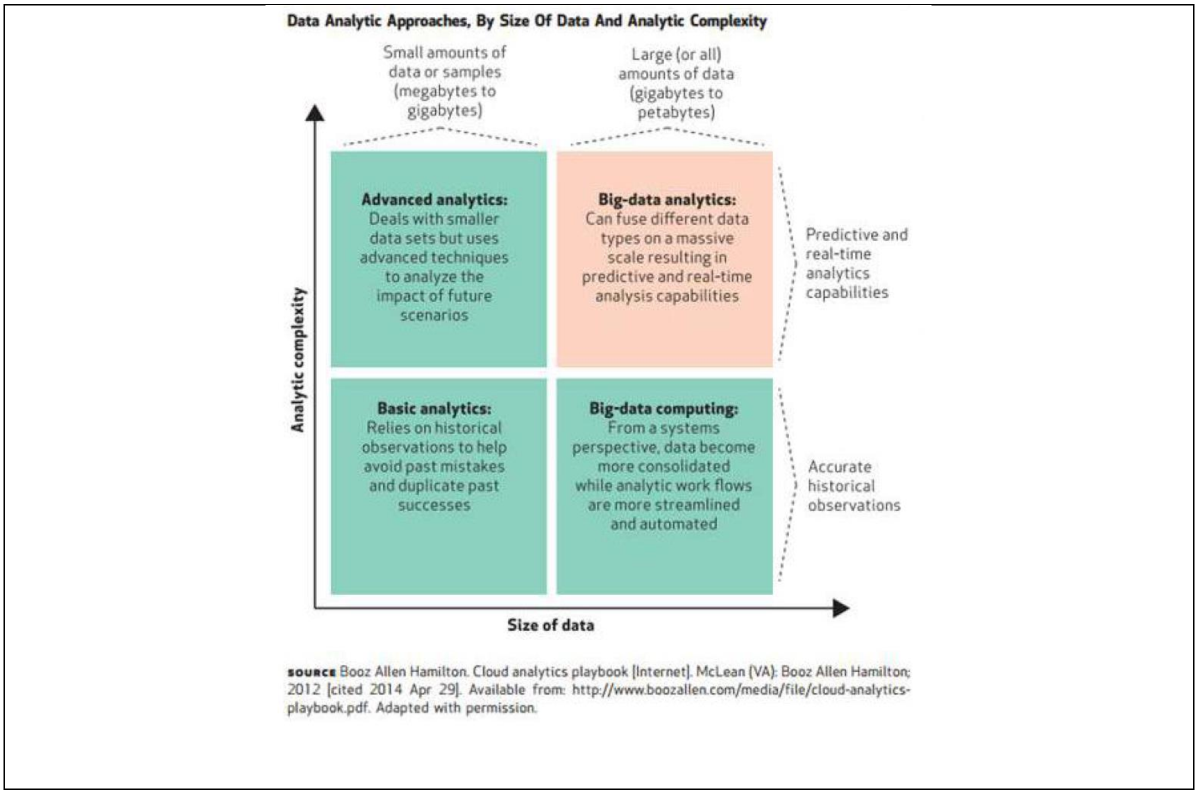
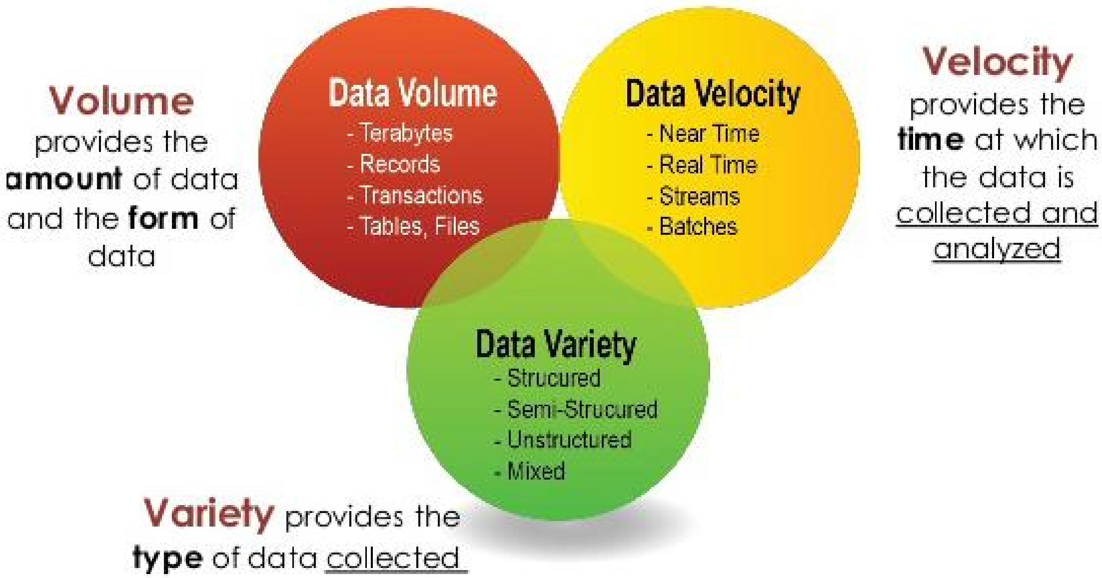


Figure 3: Data Analytics Approach

Among 3V's, Volume is the sheer size of the data measured in bytes (such as gigabytes, terabytes, petabytes etc.). The second among 3V's is Variety, which reflects that a wide range of data is included, for example, a lot of data is structured that can be enquired by data identifiers; while some data is free-text that requires data manipulation before combined with structured and semi-structured data sets (Jee, 2013). The third attribute Velocity depicts that there are three factors affecting this attribute; i) data collection rate, ii) data stored and iii) necessary data analytics speed for the whole data management process effort to be effective (Moore, 2014) .



Source: (Tehseen, 2016)

Figure 4: Big Data's 3V's

As seen in Figure 4, velocity of the big data can be real-time, near real-time, streams and batches. The CPATT database analytics falls under the category of batches. In Figure 3 an expert in the field of Big Data explains the difference between different velocity of Big Data.

The most time sensitive is real time analytics. Batch processing has comparatively smaller time-sensitivity; even less than near real-time. The batch processing jobs can take longer time, hours, or even days. Batch processing includes three different processes. Firstly, over a period of time the data is collected. Secondly, a separate program processes data. Third, is the data output. Data entered for analysis may include historical and archived data, operational data, etc. (Wilson, 2020). Based on the above CPATT database falls under the category of batch processing, as different researches collet data over a course of time (sometimes even years) and then analytics are performed which may include other data such as historical or archived data.

Type of data processing	When do you need it?
Real-time	When you need information processed immediately (such as at a bank ATM)
Near real-time	When speed is important, but you don't need it immediately (such as producing operational intelligence)
Batch	When you can wait for days (or longer) for processing (Payroll is a good example.)

Source: (Wilson, 2020)

Figure 5: Different types of Velocity for Big Data

One of the major benefits of joining Big-data is cost savings, while a major barrier is cost and time required to create Big-data. (Chad Schaeffer, 2016). There are some major challenges which are listed below in Table 1:

Table 1: Challenges for Big Data Management

Challenge	Citation
Lack of standardization of data sets is costly	(Raghupathi, 2014)
Cost of Integration is high	(Jee, 2013)
Quality assurance will be a difficult ongoing task	(Raghupathi, 2014)
Data compliance and security programs	(Jee, 2013)

The new framework have benefits, based on the Long-Term Pavement Performance (LTPP), such as (Federal Highway Administration, 2009):

1. Evaluating existing design methodologies,
2. Producing enhanced design methodologies,
3. Strategizing the rehabilitation and maintenance of existing pavement,
4. Evolvement of better and enhanced design equations for new built pavements and its maintenance over a period of time
5. Identification of test sections which are evaluated to evaluate the impact of environment, loading, material properties, construction quality and required maintenance levels due to pavement distress and performance,
6. Determine effects of specific design features on pavement performance

Another significant benefit is more in-depth analysis and insight into the Life Cycle Cost Analysis (LCCA) and its comparison with the real data from the field.

This benefits CPATT in the following manner:

1. All CPATT data (concrete pavement field data) is available under one portal to both internal (University of Waterloo) and external (with in Canada and international- if permitted under copyrights) students, researchers, academicians, professors and others.
2. Historical data with is available for both laboratory and field data for better comparison with current and future data,
3. The database will welcome data deposits from external sources (subject to copyrights) after rigorous data validation and quality checks and can become the largest pavement database in Canada,
4. After implementation of this research, external source funding is expected to further improve.

Continued support of LTPP activities provides a return on investment from 100:1 to 1000:1 for every dollar spent on pavement in United States (Federal Highway Administration, 2016). This research is expected to return investment which varies between 20:1 to 500:1 for appropriate utilization of CPATT data.

The LTPP data utilizes to calibrate Mechanical-Empirical Pavement Design Guide (MEPDG) adopted by American Association of State Highway and Transportation Officials (AASHTO) (Federal Highway Administration, 2016), related to which CPATT data can be utilized to calibrate and revise the Canadian Design Standards on municipal, provincial and federal levels.

The CPATT data collection program is able to lead developing certain tools and software for Canadian Highways, which will be implemented with modification internationally. The research carried out by CPATT is expected to generate new methods, technologies and practices for further enhancing the manner pavements are design and build in Canada.

Some of the other salient impacts of this research includes; sustainability into pavement design, construction, maintenance and management. A better understanding of effects on long-life infrastructure when exposed to climate change through analyzing data. This ultimately impacts environment, reduction in adverse effects of climate change on infrastructure, cost optimization, better project delivery and ensure extensive research options for subject matter experts (SME), students, researchers, government agencies and academicians. Among other benefits, the findings of this research integrate all CPATT data (concrete pavements) and is available under one portal; enabling researchers to streamline their efforts in advancing the studies in the field of transportation engineering and developing new technologies.

There are two major gaps identified in LTPP data such as materials properties and traffic loading data. These gaps were mainly overcome by development of special software such as LTPP-PLUG (traffic loading gaps identification program) (Federal Highway Administration, 2013) (Federal Highway Administration , 2015).

The benefit of CPATT new database framework will come from its utilization in development of new pavement design, construction, maintenance and rehabilitation methodologies.

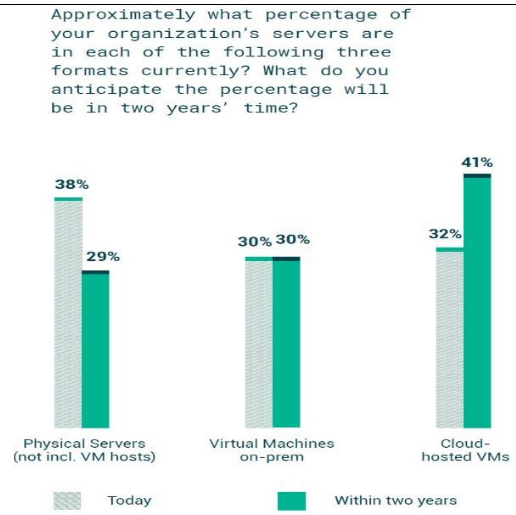
CPATT team members and associates are collecting data at a high velocity. These data are currently located in different databases and requires to be in one database. This research is to develop a database framework and as a result provide a meaningful CPATT database, inter-relationship models and data quality checks.

In order to be a sturdy, meaningful and robust database, these gaps are identified for CPATT database:

1. CPATT data collected through previous research needs to be at one location (one repository)
2. Datasets standard formats are developed for better data integration and reusability and understanding (such as codes of different equipment readings, as different type of equipment uses different codes)
3. Quality Acceptance checks are developed to ensure high quality datasets usage for analytics and are consistent with the high-quality control measures required for accurate and meaningful analytics.
4. Development of datasets inter-relationship models
5. Easily accessible and database can accommodate various formats of collected datasets

2.2 Cloud based databases

An independent research firm with help of VeeAm.com during January 2020, carried out a survey of 1,550 enterprises (all above 1,000 workers) in about 18 countries on what their goals and challenges were for data protection, regardless of what were their present or future data protection vendor choices (VeeAM, 2020). The survey asked partakers a question about their current and future usage of Cloud hosted VMs and results were compiled as depicted in the Figure 6.

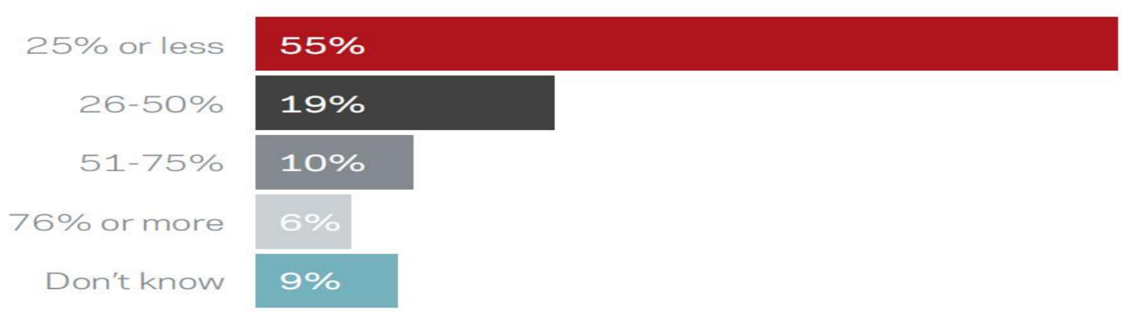


Source: (VeeAM, 2020)

Figure 6: Percentage of current and future usage of cloud servers

As seen in Figure 6; by 2022, over 40% of all servers are expected be cloud-hosted.

In 2019, a survey conducted by veratis.com regarding the state of data protection and cloud was published. The survey asked "What percentage of your workloads is in the cloud?". The survey results are presented in Figure 7. Over 55% said that 25 percent of their workloads is already in the cloud. (Veritas.com, 2019)



Source: (Veritas.com, 2019)

Figure 7: Percentage of workload in the Cloud

Modern organizations are spreading their on-premises environment by moving its data and workloads to the cloud (Veritas.com, 2019).

"A cloud workload is a specific application, service, capability or a specific amount of work that can be run on a cloud resource. Virtual machines, databases, containers, Hadoop nodes and applications are all considered cloud workloads" (Dell Technologies, 2021).

Based on a survey in 2019; transferring workloads to the cloud-based environment is a trend the organizations felt would continue. A huge 86 percent forecasted that cloud workloads would increase in the next two years. This further strengthens the need to ensure that data is also protected. Its best to assume that your data is not being backed up by the cloud provider. In fact, many cloud services require their clients to back up its cloud data (Veritas.com, 2019) .

(Innovative Architects, 2019) reported that Microsoft has two cloud computing services; SharePoint online and MS Azure. They further elaborated that SharePoint online automatically updates and is one of the easiest, lightest, and most accessible approaches to shared cloud hosting. Office 365 is the same Microsoft Office used on a daily basis, apart from it that it is powered by the Cloud. This gives access to manage applications, documents, and files from virtually anywhere and on any device—laptop, tablet, smartphone, etc. The Salient benefits of Cloud (SharePoint) are detailed in Table 2.

Table 2: Salient benefits of Cloud (SharePoint)	
Minimal launch time	Rather than taking hours—if not days—to launch or update, cloud applications are typically up and running in seconds or minutes, and easy to learn.
Immediate global workforce	Cloud can enable a team to access information anywhere, anytime, and on any mobile device — so long as they have an internet connection.

Intelligent automation	Self-provisioning tools give users the ability to spend more time responding to business needs, and less time tinkering with manual intervention.
Security & confidentiality	The primary concern with cloud computing is security. Serious companies need to ensure that private data in the Cloud stays confidential. Using one of two SharePoint cloud hosting platforms, we can make sure migration to the Cloud is quick, easy, and secure.
Greater scalability	Users can effortlessly scale their compute or storage capacity up or down depending on what's needed, keeping your infrastructure simple and efficient.
Reduce or eliminate infrastructure maintenance	Cloud systems can automatically sync with the main server to get the latest updates and patches, which drastically cuts back on time spent doing administrative tasks.
Low startup and capital costs	Maintain easy access to vital information with minimal upfront investment. With the Cloud model, simply pay as you go and based on how much storage space you are using.
Big data	In addition to helping store data, cloud computing services gives the ability to sift through vast amounts of unstructured data to find meaningful business intelligence—a must-have tool for making informed decisions about organization's future goals.
Source: (Innovative Architects, 2019)	

As mentioned in methodology of study, an objective was to select an apposite Data Management Tool (DMT). Data Management would be as successful as the tools used to store, analyze, process, and realize value in an institution's data. These tools are varied multi-platform management systems that complement data (Svitla Team, 2019).

The frequently used DMTs are produced by the industry's biggest software manufacturers whose experience assures a high level of performance, security, effectiveness, data redundancy elimination, and confidentiality (Svitla Team, 2019). PowerBI have been listed at number 4; as an integral part of Microsoft Master Data Services in the top 18 list of the DMTs (Svitla Team, 2019). (Microsoft Corporation, 2019) argues that after the July 2018 version of Power BI Desktop, different tables direct relationship is possible. They further argued that setup of different types of relationship cardinality including many-to-many is conceivable.

(HingePoint, 2020) mentioned that SharePoint is an excellent data management tool that can find on-demand data and organize it for an organization. One of the most valuable things SharePoint can do is that users have data readily accessible and organized in one place, instead of chasing for it.

Constructed on the above facts and to develop a modern and robust CPATT database, cloud options were explored for database.

2.3 Development of Dataset Standard Format

Olson (2013), mentioned that better data integration, enhancement of data exchange process among associates, increased data quality, and reusability are some of the prime benefits of standardizing datasets. When pre-determined standards are used, it can elude the need of post-data collection conversion; from one format to another. Standardization of datasets can help better understand the terminology used in a particular dataset. Although standards do not ensure high quality data, it shall be used to the greatest extent possible.

(Nahm, 2013) reveals that one of the prime functions of dataset is the reuse of data by other than those who collected the data. If the data is not well defined and other user encounter issues with a previously collected data, then the original researcher has not done their job well. Data reuse require standard and specificity for both humans and computational analysis.

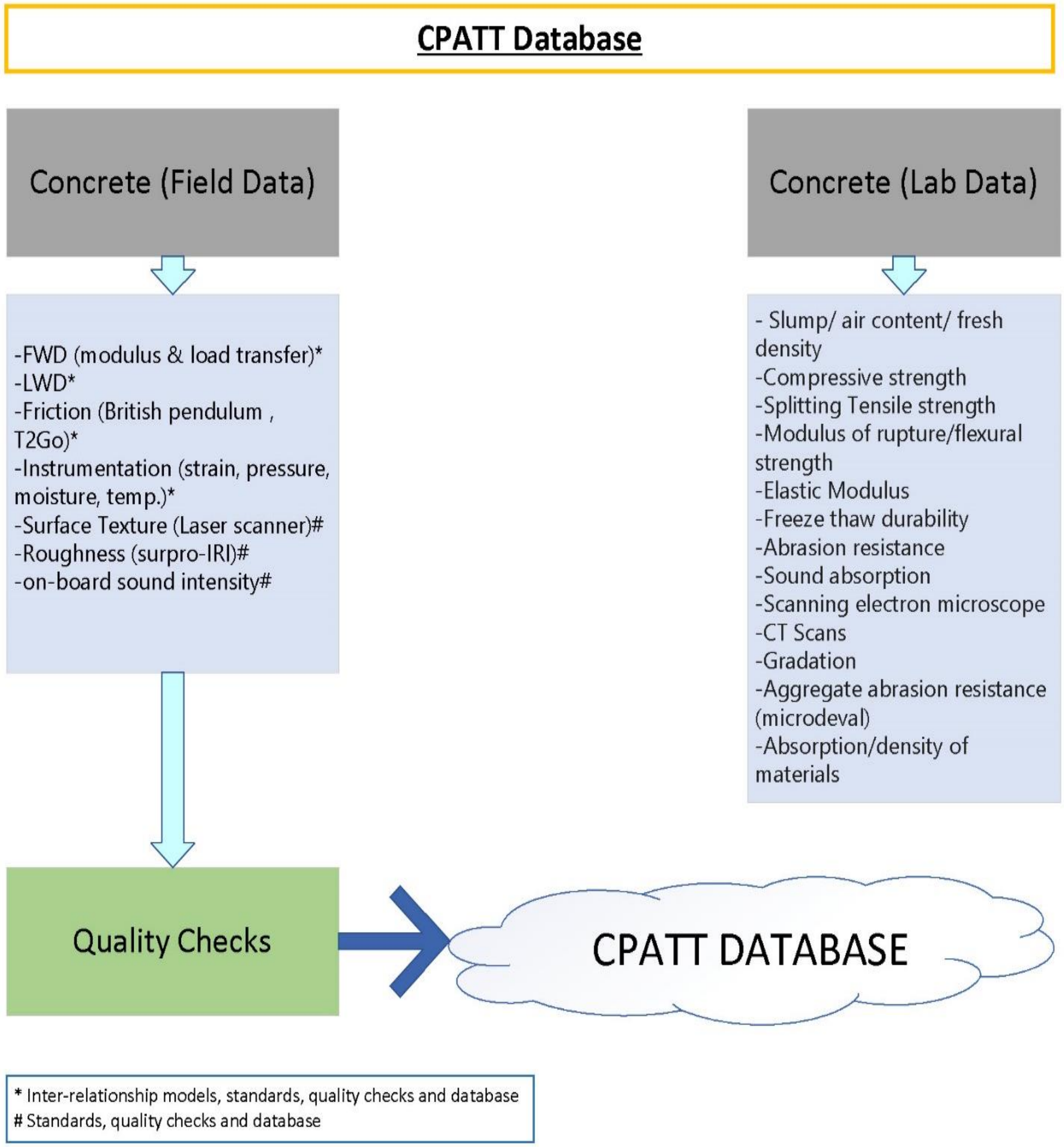
Current CPATT datasets have numerous formats for different dataset of same or similar type of tests/datasets. Most of these are not aligned with each other. This research has developed standard formats for tests of concrete pavements for enhancing consistency, and reusability.

2.4 Dataset Models (Data Analytics and its link with Pavement Engineering/ Management)

Watson (2014), defines data analytics as data "must be analyzed and results used by decision makers and organizational processes in order to generate value". For the purpose of this research decision makers will be the researchers as guided by CPATT management.

Members of the CPATT team, have collected data (in past several decades) using both Light Weight Deflectometer (LWD) and Falling Weight Deflectometer (FWD) testing methodologies. These data are not usually provided in a standard format which presents numerous challenges not only for analysis but also for checking whether it is correct. These corrections relate to standardization, identification of missing data, categorization of data, data structure and development of inter-relation models between FWD and LWD data. In the following sections this will be discussed in detail.

Figure 8, depicts types of CPATT Data Collection and Database development process for different concrete pavement data field and laboratory datasets.



Source: (Pickel, CPATT current data collection for concrete field and laboratory data, 2019)

Figure 8: CPATT Current Data Collection and Database development

Steinert (2006) developed a regression analysis for the relationship between LWD and FWD modulus testing for thin and thick asphalt layers, as seen in Table 3:

Table 3: Previous Study - Regression Analysis between LWD and FWD modulus

Equation (Modulus)	**Layer Description**	**R-Square (R^2) Value**	**LWD Model**
LWD(MPa) = 1.33FWD (MPa)	Thin asphalt layer	0.87	Prima 100
LWD(MPa) = 0.75FWD(MPa)	Thicker asphalt layer	0.56	Prima 100
Source: (Steinert, 2006)			

Fleming (2000), has also developed numerous regression equations for the relationship between LWD and FWD modulus ratio for various layered pavement structures such as, 450-mm granular capping over silt and clay, 260-mm lime-cement treated clay subgrade and 225-mm well-graded crushed stone granular subgrade using Prima 100 LWD equipment.

In 2016, CPATT collected, LWD and FWD data on Jameston Avenue, Hamilton, Ontario on existing pavement structure as part of a major research on concrete overlays. LWD testing was performed by members of CPATT while FWD testing was contracted out to Stantec Consulting Inc. Wafa (2019) stated that the top asphalt layer was removed from the existing pavement structure before the FWD testing was conducted (on before construction concrete panels) by Stantec Consulting. A CPATT team was following the FWD testing team with a CPATT-owned LWD equipment and measured deflections using LWD. Both FWD and LWD were used simultaneously (with a time difference of few minutes) to measure deflection of the rigid concrete pavement. (Wafa R. , 2019) also mentioned that the LWD measured deflection were observed at the same locations where the FWD was used. This data provides basis for developing of a correlation model between FWD and LWD deflections and modulus. Wafa (2019) also advised that LWD testing was conducted only for the Eastbound Lane with a total of 37 stations compared to FWD testing of 44 stations for each eastbound and Westbound Lane. However, after careful evaluation of the LWD data, there were some outliers' readings for seven stations (locations), which were subsequently removed for regression purposes from FWD data as well for apposite comparable data as seen in Chapter 5.

Korczak (2019) commented that the FWD moduli such as EPCC (Elastic Modulus for Concrete) and K_{Static} (Modulus of subgrade reaction) for Jameston Avenue are calculated using AASHTO 1993 Design Guide.

Fleming (2007) and Mallick (2017) provided an equation 1 (Boussineq's theory) for elastic modulus as below:

$$\mathbf{E_{LWD} = \frac{A.P.r.(1 - v^2)}{\mathit{D}}} \qquad \textbf{Equation 1}$$

Where:

E_{LWD} = surface modulus (MPa)

A = shape (plate rigidity) factor, default = 2 for a flexible plate, π/2 for a rigid plate.

P = applied stress (maximum contact pressure) (kPa)

r = plate radius (mm)

v = Poisson's ratio (usually in the range of 0.15-0.25) (but is used as 0.2-0.21 for concrete (Mallick, 2017))

D = surface (peak) deflection (μm)

Based on AASHTO 1993 Design Guide for pavement structure calculations and Fleming (2007), there is a correlation between FWD and LWD modulus calculations. This correlation can be further explored to perform regression analysis and develop a relationship between LWD and FWD for concrete pavements.

There is a relationship between back calculated modulus (LWD and FWD); and Friction Test (British Pendulum and T2GO) at different sites. Specifically strain gauges data and how it responds to loading and temperature changes is also important area to examine, and its effects will be evaluated for curling and warping. Standards and guidelines will be developed for the collection of data from these testing methods, such as FWD & LWD and BP and T2GO; including type of equipment used, its coordinates, calculated fields (formulae are missing from the data) etc. Quality checks will be developed for all of these testing methods, such as statistical analysis (e.g., Nelson's Tests of Statistical Charts), determining ranges, outliers, zero tests etc.

2.5 Quality Acceptance Checks

Many agencies require to apply software routines to check for inconsistencies in data for quality acceptance and control. Although these checks are more inclined towards the quality acceptance than quality control. Different methods are used by different agencies for quality control, among these; 61% examines data for missing data elements or segments; 71%, tests data for out of expected range and 50% utilize statistical analysis for evaluating data inconsistencies (Flintsch, 2009). These data investigations may include data and video checks when data is received in the office, on-vehicle data validations, and/or final database validation after it is entered into the respective pavement management database (Flintsch, 2009).

On-vehicle data tests are run in real-time, while data is being collected. According to Flintsch (2009), these validation checks require a visual display of the collected data to alert the investigator of any out-of-range data; some periodic checks are run for diagnostics or data to validate correct functioning of the equipment during data collection process breaks. The final database checks are conducted to verify the data formatting and that different data has been properly entered in the final database; these data tests include, time-history comparisons, completeness, format, plots on GIS etc. (Flintsch, 2009).

The quality acceptance plan includes procedures which include acceptance plan for the pavement condition data collected by both the agency and service provider (vendor/students etc.). These tests, checks that the data met all the requirements with for proper data collection; before utilization of pavement data. Quality management process usually used for this purpose include sampling and re-rating, complete database checks, time history, GIS-based quality acceptance tests etc. (Flintsch, 2009)

Quality acceptance process requires that some or all the data is checked before analyzing data to ensure that re-collection of data or re-survey is not required. Table 4, below shows the findings of a survey carried out to determine percentage of agencies that do have data acceptance checks in place per the most common data acceptance tests it performs (Flintsch, 2009).

Table 4: Methods (Percent) of Transportation Agencies (USA) for data acceptance checks

The most common methods/tools used for quality acceptance	Percentage(s) of methods
1. Calibration of equipment and/or analysis criteria before the data collection	[80%]
2. Testing of known control segments before data collection	[73%]
3. Periodic testing of known control segments during production	[71%]
4. Software routines that check if the data are within the expected ranges	[71%]
5. Software routines that check for missing road segments or data elements	[61%]
6. Statistical/software routines that check for inconsistencies in the data	[50%]
7. Comparison with existing time-series data	[50%]

Source: (Flintsch, 2009)

The New Mexico Department of Transportation (DOT) validates the data quality for the pavement condition collected by a service providers (universities) for completeness, consistency, reasonableness and checks that all values for data collected falls with-in a reasonable range as well carry out a comparison with the previous year's data to evaluate; if large data changes are noted when compared to previous years, the New Mexico DOT carries out further data consistency and reasonableness checks (Flintsch, 2009).

"The Colorado Department of Transportation (DOT) has been a strong supporter of LTPP by being proactive in examining performance at its LTPP test sections and making improvements based on these evaluations. A 2006 report19 documented the implementation of improved PCC pavement practices based on LTPP test sections. The LTPP data confirmed that widening a slab from 12 ft to 14 ft (3.7 m to 4.3 m) provides the structural equivalent of increasing slab thickness by 1 in. (2.5 cm); and that a single 1/8-in. (0.32-cm) cut is as effective as Colorado DOT's previous standard 3/8-in. (0.95-cm) double cut for PCC joints, thereby providing a savings of $0.57 per linear foot of joint ($1.87/m). These results were derived from the SPS-2 and SPS-4 projects, respectively (Federal Highway Administration, 2016)."

2.6 Selection of CPATT Database Software and Repository

Initial literature review and discussions with CPATT management reflects that a low initial cost, low long-term maintenance costs and ease of learning of the software is required of the CPATT database. Based on these discussions, it is also revealed that the initial and maintenance costs for database software shall also be kept at lower side so that every CPATT student and researcher can access the data and can use advanced analytics utilizing CPATT database. The details of CPATT management meetings were discussed with Kevin Rampersad (IT Manager, Civil and Environmental Engineering, University of Waterloo). Following high level summary is provided below:

1. Currently, over 99% of users are active University of Waterloo people, but only a select few.
2. Needs filters to see relevant information.
3. This is need to present data not collect data.
4. Be able to store raw, pdf, docx, jpeg, xlsx etc. files from the data and other miscellaneous types of data file(s).
5. The database will primarily to do modelling after obtaining it from the site(s).
6. No personal data is allowed, this database is for pavement data of concrete pavement field data and expansion capabilities for asphalt pavement data and other pavement related data.

2.7 Selection of CPATT Database Creation

Watson (2014), commented that it’s important to ensure that data “analytics are performed in conjunction with the business strategy. This is why instead of IT, it should be business people to drive the IT projects (Watson, 2014)”. Since CPATT database is required for a pavement engineering and management business/research, a researcher with civil engineering and database background will be well suited for performing the creation of a meaningful and well-designed CPATT database. A civil engineer can better group transportation data, determine datasets inter-relationships to reduce inefficiencies, develop ideal flow of civil engineering data and ultimately enhance the reusability of the datasets.

2.8 Chapter Summary

CPATT Database falls under the category of Big Data based on the fact; that the criterion of 3V's (Volume, Velocity and Variety) is met.

Some salient challenges of Big Data include; such as lack of standardization of datasets as it is costly and datasets quality assurance.

By 2022, over 40% of all servers are expected to be cloud-based. Some salient features of cloud-based Database include Big Data, low startup and confidentiality, etc.

A robust and meaningful Database requires development of data set standard format

Development of quality acceptance checks are of utmost importance to check inconsistency in data for quality assurance and control.

Setting up Critical Success Factors (CSF) for selection of CPATT database Software and Repository.

3. RESEARCH METHODOLOGY

The purpose of this chapter is to include detailed discussion of the selection criteria for the progressing a robust and meaningful CPATT Database. It also includes identification process of Database framework development and its salient features, based on extensive literature review, as well consultations CPATT management and IT experts.

3.1 Selection of Methodology

In order to provide a state-of-the-art solution to the issues identified in Chapter 2, a methodology is developed as depicted in this section. The methodology consists of five broad phases, namely;

1. Development of selection criterion for CPATT database,
2. Database Structure and Design,
3. Database testing (including Datasets standard formats and Datasets Quality Acceptance checks), and
4. Datasets inter-relationship models.
5. Datasets Quality Control checks.

The criterion for a successful CPATT database framework and database itself is developed based on several meetings with CPATT management, its affiliates (University of Waterloo IT staff) and interviews (both verbal and written) of the end users (such as current and past students) were conducted. These meetings and interviews provided guidelines for the selection criterion of database framework critical success factors (CSF) such as database repository, database software, ease of learning, security of database etc.

Thirteen major and one optional CSF were identified as per literature review, meeting with University of Waterloo members and interviews with students. These CSF are detailed later in this research.

The second phase comprised of literature review, a survey of end users (CPATT Team), interviews with CPATT management to determine the specifics of CPATT database design. The survey from end users includes questions about their comfortableness with the use of three CPATT database candidate software.

In the third phase, the database was developed and uploaded for testing to the repository. Researcher of this study and a research associate from CPATT ((Pickel, CPATT Research Associate, 2018-2021) was involved in prototyping, testing and implementation of CPATT database. An appropriate naming convention for the datasets was created based on the literature review and consultation with subject matter experts in this phase. The datasets are then, included into the database according to a set criterion and uploaded. Access to repository and security of the datasets was evaluated as discussed in next sections. Both, adequate quality acceptance checks and newly developed standard datasets formats were applied before uploading to the datasets to CPATT database.

The fourth phase is the development of interrelationships model based on careful review of the datasets and relating the appropriate data entity from one dataset to another dataset. Few data-permitting regression analyses were carried out (as an a demonstration) to find a suitable relationship between two different types of test(s) were developed. In this regard, the data-permitting interrelationships were modeled between FWD and LWD data (for Deflections and Modulus of Elasticity). The other data used to develop a relationship between is T2GO and BPN. Data-permitting datasets are those which are collected from same site at exactly same or similar locations during the same time frame (same hour or so).

According to Merriam-Webster's (Franklin Thesaurus); "Interrelationship" is synonymous to "Correlation".

3.2 Database Framework Development Process

Database Design Development Process (Microsoft Incorporation, 2019)

A good database design contains two major principles as follows:

1. It eludes any errors, inconsistencies and optimize the space (data storage), duplicate information are removed.
2. Correctness and completeness of information is of primary importance as the results based on incorrect and incomplete information is inaccurate; hence, causing the analysis to be of no significant value.

Based on the above a good database design shall have the following four aspects included

1. Subject-based tables are used to divide the data in-order to reduce redundant data
2. Provides information to join tables together when data-permitting
3. Ensure integrity and accuracy of the datasets.
4. Facilitate data processing and analysis.

The design process comprises of the following steps:

1. Purpose of the database: identify the need of database.
2. Discover and organize the information required
3. Divide the information into tables into main entities or subjects (each table becomes a table)
4. Convert information items into column: Ensure that right information in each table. Each item then is known as field and displayed in table as a column
5. Identify primary keys for each table there is a primary key (column which uniquely define each row.
6. Develop table relationships decide how column in one table relates to a Column in another table
7. Evaluate design: Evaluate the database design and identify any errors and fix the errors if the results are not as desired.
8. Apply the normalization rules: to see if the structure is correct.

The CPATT Database development includes steps as outlined below. The newly developed database design has been prototyped using limited portion of existing CPATT data. After the limited implementation of the newly designed database, it is rigorously tested and evaluated by several CPATT members including author of this research. After validation against the set criterion, the mass conversion and loading of the CPATT existing data unto the new database is processed. The end result of this research includes a well-defined, state-of-the-art CPATT framework and concrete pavement field database.

As a first step, the standard format(s) are developed for numerous existing datasets, including, FWD, LWD, IRI, BPT, etc. Some of these standard formats were provided to Daniel Pickel (Pickel, CPATT Research Associate, 2018-2021).

Daniel Pickel, then after several discussions (to improve the standard formats), uploaded the data to the newly created CPATT repository. It was ensured that the process becomes as smooth and user-friendly as possible. Which was achieved through several iterations during the process of data conversion and loading and database testing (Pickel, CPATT Research Associate, 2018-2021).

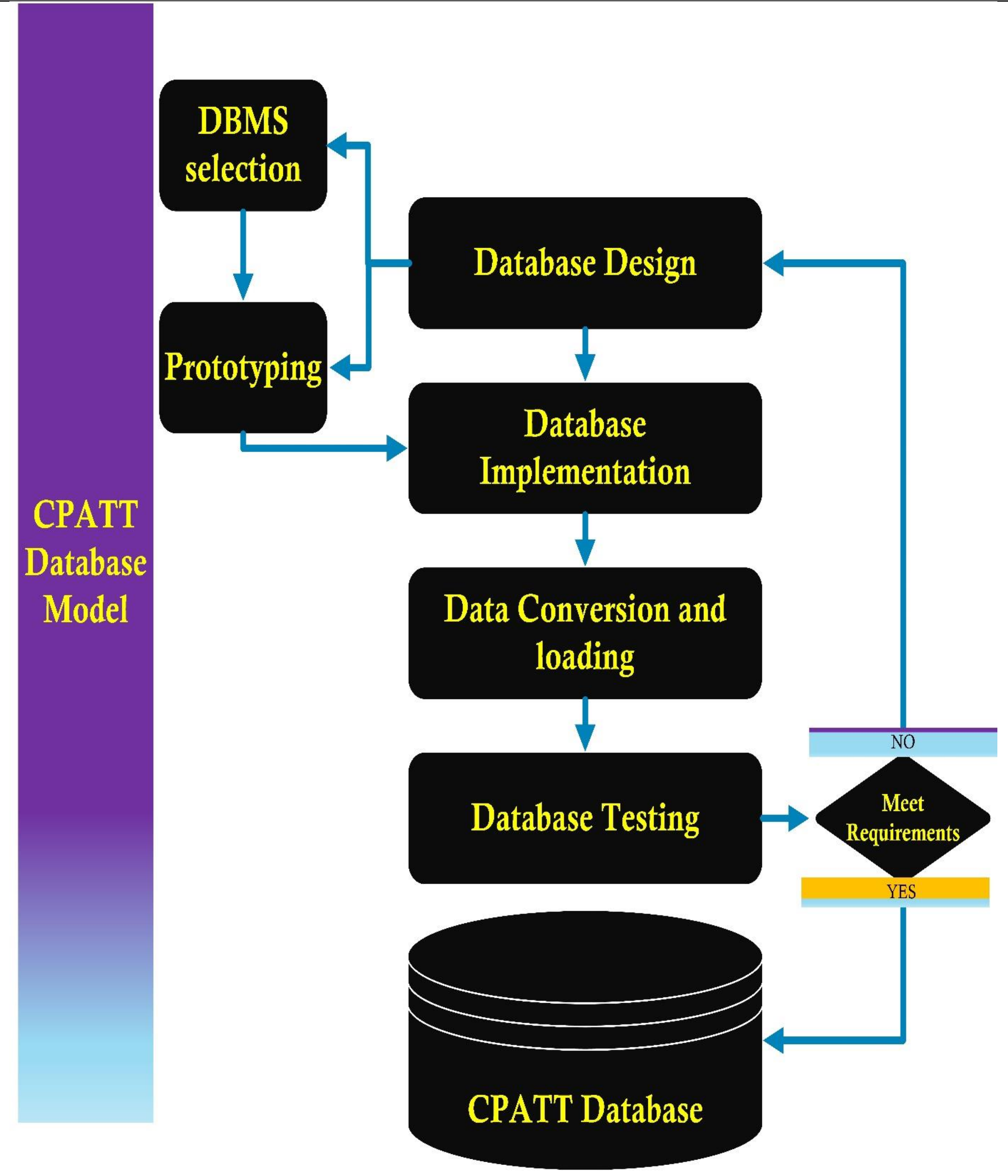


Figure 9: CPATT Database Model

3.3 Database Normalization Process

(Finkel, 2019) is an expert trainer with CBT Nuggets and has explained the normalization of database (3NF) as described below:

1. Database Normalization is structuring a relational database in accordance with a series of "Normal Forms" in order to reduce data redundancy and improve data integrity.
2. There are about 11 Normal Forms and those are progressive, such as; database in third normal form is also in second and first. [See Table 5 (screen shot form real data set); is currently in CPATT Database for a LWD test; as seen there are few columns for which few of the rows have been merged; which doesn't serve the mapping purpose, as values in other columns are not assigned to only one value in average deflection, average (MPa), etc.].
3. Third Normal Form (3NF) is usually considered as a "Normalized Database".
4. Each attribute has "Atomic" data, i-e. no multi-valued fields and each table has a unique identifier called primary key. Which means that every field has one and only one value. [Table 5; reflects that every field has only one value; however, for few of the columns few rows are merged together (for analysis purposes); this does not follow 1NF; few adjustments were made to get the 1NF as seen in Table 6]
5. For the 2NF: every non-prime attribute is dependent exclusively on the primary key. [for Table 6, few column headings including "file name", "location" are repetitive, which cause confusion. To comply with the requirements of 2NF, few steps were taken, such as; a) Rearranging tables; b) taking average of the deflections; c) extracting station from FWD testing and mapping it to the LWD table; d) adding location ID; etc. as depicted in Table 7. ID location was selected as Primary Key for data normalization.

Table 5: A screen-shot form actual CPATT data set for LWD

FileName	Location	Direction	Point No.	Drop No.	Time	Radius (mm)	Load (kN)	Deflection (um)	Avg_Def_Leave&Approach	Avg_Deflection (um)	Stress (kpa)	Modulus (MPa)	Average (Mpa)
testing_august9_EC	L1	approach	1	1	09-Aug-16	150	13.8	177			195	722	
testing_august9_EC	L1	approach	1	2	09-Aug-16	150	14.3	152			202	251	372
testing_august9_EC	L1	approach	1	3	09-Aug-16	150	14.2	142			201	266	
testing_august9_EC	L1	approach	1	4	09-Aug-16	150	14.3	153		156	202	249	
testing_august9_EC	L1	approach	1	5	09-Aug-16	150	14.1	1648			200		
testing_august9_EC	L1	leave	1	6	09-Aug-16	150	14.1	359			199	105	
testing_august9_EC	L1	leave	1	7	09-Aug-16	150	13.2	953			187		
testing_august9_EC	L1	leave	1	8	09-Aug-16	150	12.6	3598			178		
testing_august9_EC	L1	leave	1	9	09-Aug-16	150	14.0	367			199	102	104
testing_august9_EC	L1	leave	1	10	09-Aug-16	150	12.7	1612	259		179		
testing_august9_EC	L1	leave	1	11	09-Aug-16	150	1.7	0			24		
testing_august9_EC	L1	leave	1	12	09-Aug-16	150	14.1	361		362	200	104	

Source: (Wafa R. , 2019)

Table 6: Adjustments made to get the 1NF

FileName	Location	Direction	Point No.	Drop No.	Time	Radius (mm)	Load (kN)	Deflection (um)	Avg_Def	Stress (kpa)	Modulus (MPa)	Average (Mpa)
testing_august9_EC	L1	approach	1	1	09-Aug-16	150	13.8	177	156	195	249	291
testing_august9_EC	L1	approach	1	2	09-Aug-16	150	14.3	152	156	202	300	291
testing_august9_EC	L1	approach	1	3	09-Aug-16	150	14.2	142	156	201	317	291
testing_august9_EC	L1	approach	1	4	09-Aug-16	150	14.3	153	156	202	298	291

Table 7: ID location- selected as Primary Key for data normalization

Station_ID	Station: _FWD16-EB	Avg Def LWD Nor (40kN)
1	0+004	110
2	0+007	96
3	0+013	100
4	0+021	81
5	0+047	278
6	0+054	151
7	0+061	113
8	0+070	233
9	0+090	110
10	0+111	127
11	0+128	169
12	0+135	132
13	0+141	31
14	0+150	155
15	0+160	458
16	0+170	359
17	0+190	78
18	0+211	61
19	0+221	93
20	0+230	63
21	0+240	86
22	0+269	73
23	0+271	71
24	0+280	108
25	0+287	64
26	0+293	76
27	0+302	109
28	0+310	100
29	0+320	157
30	0+328	58

Station ID	Station:_FWD16-EB	D1 Def FWD16
1	0+004	67
2	0+007	97
3	0+013	186
4	0+021	145
5	0+047	216
6	0+054	208
7	0+061	159
8	0+070	236
9	0+090	349
10	0+111	332
11	0+128	221
12	0+135	182
13	0+141	190
14	0+150	358
15	0+160	150
16	0+170	371
17	0+190	221
18	0+211	481
19	0+221	183
20	0+230	158
21	0+240	196
22	0+269	145
23	0+271	208
24	0+280	319
25	0+287	156
26	0+293	214
27	0+302	168
28	0+310	161
29	0+320	97
30	0+328	227

6. In 3NF attributes are determined only by the Primary Key. Setting the Station_ID as the Primary Key can help develop relationship between the Avg Def LWD Nor(40kN) and D1 Def FWD16.

Table 8: General Normalization Process

Normalization	
Normal Form	Characteristics
First Normal Form (1NF)	Table format, no repeating groups, and PK identified
Second Normal Form (2NF)	1NF and no partial dependencies
Third Normal Form (3NF)	2NF and no transitive dependencies

3.4 Chapter Summary

To provide a state-of-the-art solution for the gaps identified in Chapter 2, a methodology is developed for achieving goals of a robust and meaningful CPATT

Database which includes 5 major steps.

Identified the Database Framework development process for salient features such as optimization of the space (data storage), removal of redundant information and correctness of information.

A CPATT Database Model developed for systematic progress of this research through extensive literature review, CPATT management and numerous IT expert(s) consultations

An example of CPATT Database Normalization process.

4. DATABASE FRAMEWORK

This chapter includes salient advantages of a meaningful Database, and some of is disadvantages. It also contains the detailed process of development of selection criterion (Model)for CPATT Database.

It also includes the survey results, analytics and verification of selection criteria; through CPATT stakeholders (25 survey partakers); as well depicts the detailed criteria and procedure for uploading the datasets to the new CPATT Database.

4.1 Database

4.1.1 Database Definition

There are many definitions of a database. Whatisdbs.com (2107)stated few of them as follows:

- A database is the group of inter-related files such that data among files can be combined to get accurate results.
- “A properly ordered collection of interlinked data that are related in a meaningful way which can be accessed in different logical orders but are stored only once” (Whatisdbms.com, 2107).
- A Database is a systematic collection of data where information is easily drawn. A database consists of fields, records and files.

Some basic definitions are as follows (Whatisdbms.com, 2107):

A field is a single part of information.

Where several fields make record. Or (Fields + Fields + Fields + …. = Record)

A file comprises of several records. Or (Record + Records + Records + …. = Files)

In Figure 10, the graphical representation is depicted of what is a database, file, record and field.

Based on the above our definition of database has converged to "Database is a collection of fields, records and files, which are easily retrievable. Where files are related in a meaningful manner and stored only once to elude redundancy of data".

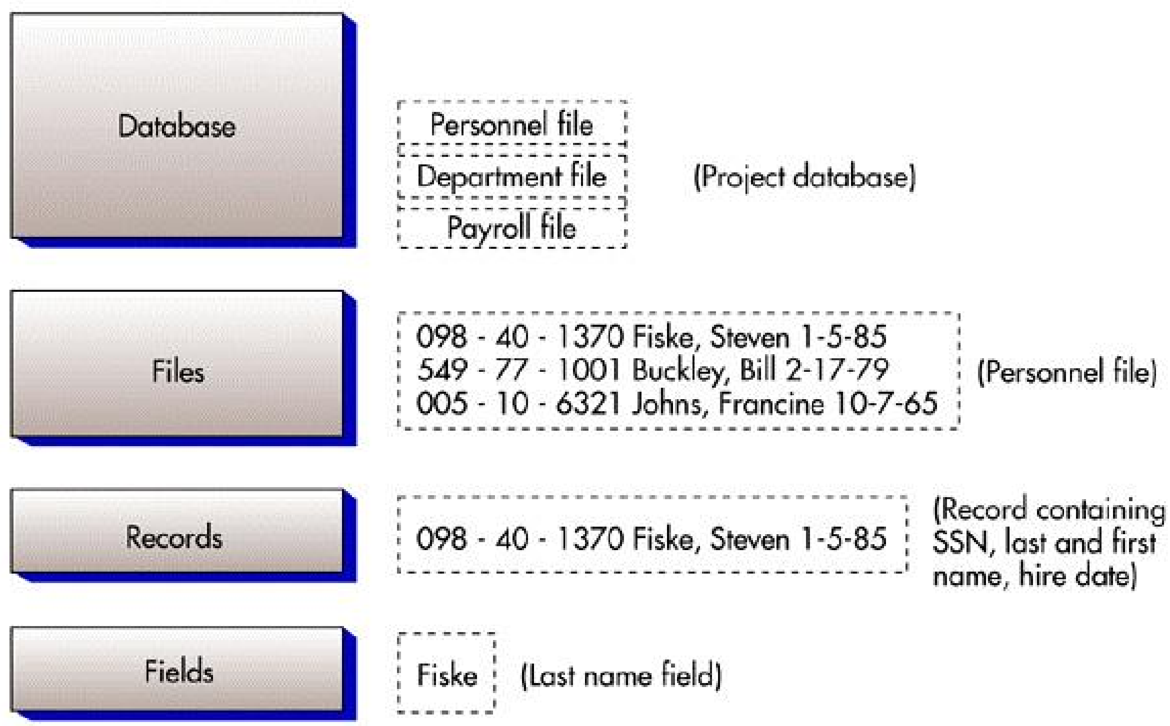
4.1.2 Database advantages and disadvantages

According to Whatisdbms.com (2107) mentioned the following database advantages:

- Minimal Redundancy
- Privacy and Security
- Easy Searching
- Integrity
- Consistency
- Simplicity
- Full Control

Some of the major disadvantages are as below (Whatisdbms.com, 2107):

- Very Costly
- High Hardware Cost.
- Sometime Complex
- Increased Vulnerability



Source: (Whatisdbms.com, 2107)

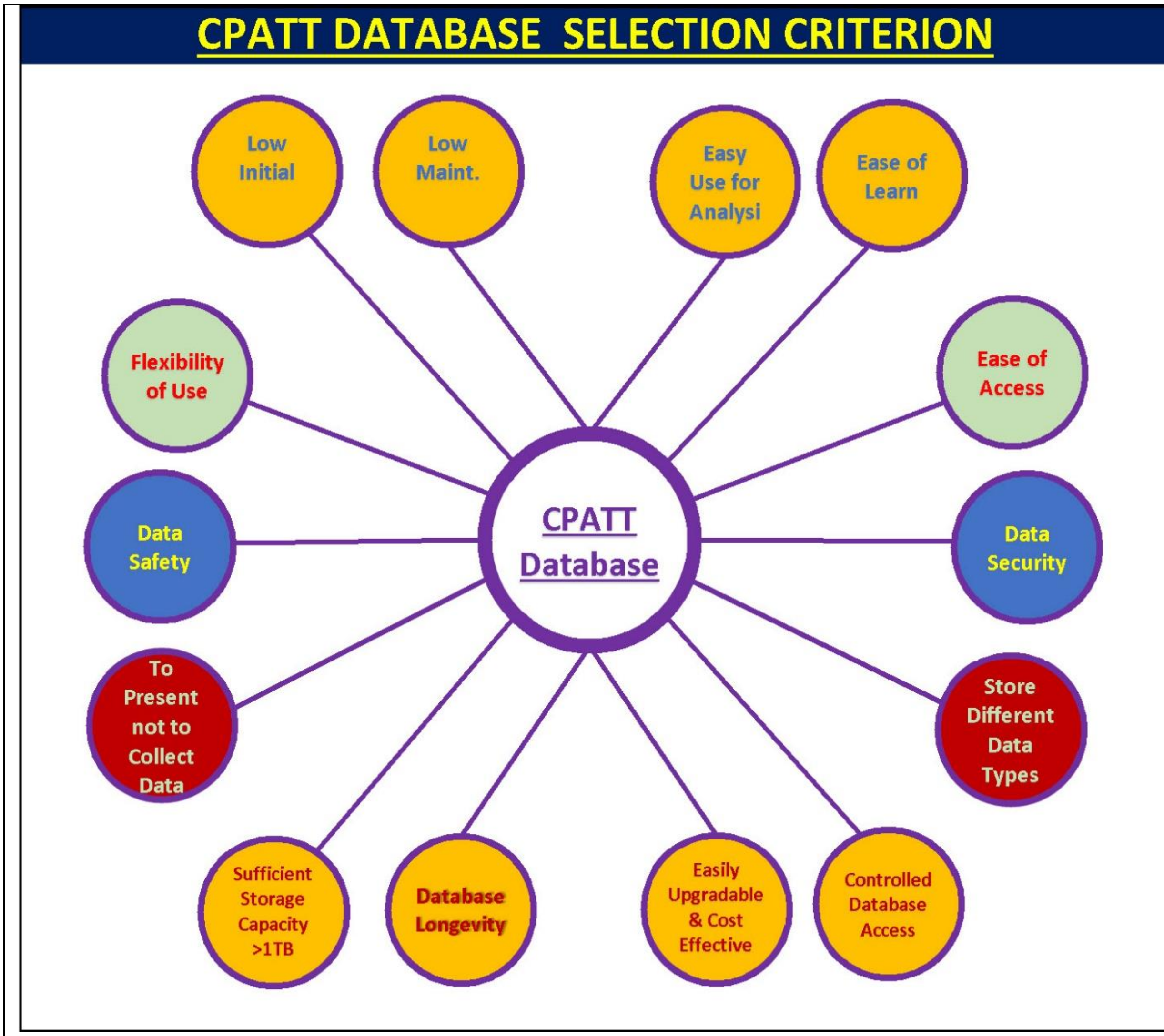
Figure 10: Diagram for Fields, Records, Files and Database

4.2 Selection Criterion for CPATT Database

In order to be a feasible CPATT database solution; database software and repository selection is based on the following required and one optional criterion:

1. Low initial costs
2. Low maintenance costs
3. Longevity
4. Ease of learn (short learning curve)

5. Flexibility of use (Different researchers with different software skills could use the database)
6. Ease of access (readily accessible to one or all allowed researchers)
7. Safety of data is assured through measures
8. It is to present data and not collect data
9. Can store different type of data such as numerical, text, pictures, video; with different formats pdf, xlsx, docx, jpeg, etc.
10. Have sufficient storage capacity (> 1 tera bytes-TB), and easily upgradable to a higher storage capacity in the distant future
11. Cost effective and easily upgradable to the latest version of the software
12. Access to database is controlled
13. Database software can be easily utilized for analytics of datasets
14. Capacity of using with mobile devices (optional)



Source: CPATT Database Survey

Figure 11: CPATT Database Selection Criterion

These criterions were presented in form of a survey to CPATT students and other affiliates for ranking. These rankings are discussed in other section.

These selection criteria were discussed with CPATT management and after initial approval was further evaluated by Kevin Rampersad (IT Manager @ University of Waterloo, Civil & Environmental Engineering Department). Following few points were noted from these Subject

Matter Expert discussions revealed further insight into development of CPATT database framework and are noted as below:

Format of numerical datasets will be that of relational database.

For the purpose of this study a comparison between most popular software such as Microsoft Access, Microsoft Excel in combination with Power BI, and MySQL was carried out. The primary purpose of the CPATT database is to store. The future researchers can perform analysis and extract information from different datasets, simply said "Usability for Analytics of Datasets". In this regard the following is noted (Noah, 2017):

Microsoft Excel is a spreadsheet tool used for the following tasks:

- Data analysis
- Calculations, statistical comparisons
- Chart management
- Generate shareable output

Microsoft Access is a database tool used for the following tasks:

- Data management
- Display data subsets, complex queries
- Reports for data summarization
- Automation for common events
- Database used by multiple users

Microsoft Power BI Desktop is a companion desktop application to Power BI. The following tasks can be performed with Power BI Desktop (Microsoft, 2019):

- Get data: The Power BI Desktop makes discovering data easy. Data can be imported from a numerous data source. After data source is connected, data can be shaped to match analytics and reporting needs.

- Create relationships and new data formats: After importing two or more tables, inter-relationships can be developed between two or more tables. The Power BI Desktop through features such as Manage Relationships and Relationships view, can perform Autodetect create any relationships, or manual inter-relationships can be developed. Power BI Desktop can easily create measures and calculations or customize data formats and categories for additional insights.
- Create reports: The Power BI Desktop can create powerful reports including features such as select the needed fields, numerous visualizations, add filters, etc. The Report View in Power BI Desktop provides same excellent report and visualizations tools as that on PowerBI.com.
- Save your reports: The Power BI Desktop, helps you save your work (reports) as Power BI Desktop file (with .pbix extension).
- Upload or publish your reports: The reports created can be easily saved to Microsoft Excel and Microsoft SharePoint website.

MS Excel + SharePoint + Power BI provides one solution to storing tables, sorting, creating queries and analytics. This is a more versatile solution as most of the graduate students are familiar or advance knowledge of MS Excel.

4.3 Database Management System (DBMS) – Features and Characteristics

Thakur (2016) has mentioned following salient features of the Database Management System (DBMS):

1. Reduced Duplication and Redundancy: More users of database mean higher chances of data duplicity. All data in a database management system is listed only once so that the duplicity can be eluded.
2. Keeps storage space and costs at minimal: All database management system usually has lots of data to handle and provide safe storage. However, integrated data provides saves much more space for database management

system. Organizations usually pays large sums of money for safe data storage for their data. If managed efficiently they can save costs of data storage and data entry.

3. Anyone can work on database: Its usually easy to understand and work with. Th insertion, update, searching and deletion of records is easy for anyone with appropriate level of access.
4. Large Database Maintenance: Large database of organization requires security of data, backup, recovery etc. Database management system includes these features such as appropriate level of access to keep the database recovery and backup system intact.
5. Keeps the data highly secure: Security and appropriate level of access to data is of major concern for organizations. Database management system does not provide full access to anyone except the Database Administrator (DBA) or head of department. All other users have restricted access to database
6. Permanent Storage of Data: Database management system stores keep all database files safe to elude any chance of loss by saving it permanently. DBMS keeps the backup and have a recovery method as well.

There are numerous characteristics of a solid Database management system (DBMS), some of which are included in the features above, while other salient characteristics are included below (Thakur, 2016)

1. Stores all types of Data: A good database management system shall be able to store any type of data including different formats, as it is required to work all sorts of data types that are present around us.
2. A DBMS shall convey accuracy, completeness, Isolation and Durability simply said ACID. It shall include properties such as no mismatch and elude duplicate data.
3. Deals with complex data relationship: the DBMS shall allow to create connections between different datasets. To make accurate and efficient use of datasets a DBMS shall be able to present complex connection effectively.

4. Database backup and recovery: (As discussed above)
5. Structures and described data: A DBMS include not only the data but also structure, format and type of data as well whenever possible the relationships or connections between datasets.
6. Database Integrity: One of the most salient characteristics of the DBMS. Database integrity ensures reliability and quality of the database system. Ensures security of the data and elude unauthorized database access.
7. Concurrent database utilization: The more authorized users of the database; higher likelihood that there will be more users working simultaneously on database. A good DBMS will be providing access to multiple users simultaneously.

Watt (2018) mentioned another salient feature of a robust database management system is sharing of data. Multiple users can use the same data simultaneously, which is provided through concurrency control strategies. Data accessed in a DBMS are always correct and data integrity is maintained; ensured thorough these concurrency control.

4.4 CPATT Database Management System (DBMS)

4.4.1 Microsoft Teams

Microsoft included a new tool to its Office 365 in November 2016 - Microsoft Teams. It is a chat-based teamwork tool that provides ability to work collaboratively and share data globally, remotely, and among the dispersed teams in a common space. Some of the salient features include one-on-one chat, document collaboration, team chat, etc. Microsoft Teams is also entirely integrated with many other Office 365 including SharePoint, etc. (Sherweb.com, 2020). Microsoft Teams as part of Office 365 suite is used by over 500,000 organizations – including 91 of the Fortune 100 (Unify Square, 2020).

As mentioned by (Sherweb.com, 2020), Microsoft Teams stands apart from other similar platforms due to the data Security. It provides top-of-the-line security and compliance capabilities with Office 365's platform services. The data is encrypted during transfer and at rest. MS-Teams and all other Office 365 services meet compliance standards including ISO 27001, HIPPA. SOC 2 and the EU Model Clauses. (Sherweb.com, 2020)

(Unify Square, 2020) argues that while it's expected that Microsoft will change some of its legacy products (such as Skype) in the future that is not going to be the case with SharePoint. SharePoint supports an advanced set of features to be integrated into Microsoft Teams, and Microsoft Teams adds a user-friendly UI and other collaboration functionality to an existing familiar concept of knowledge and document sharing.

(Unify Square, 2020) commented that there are six key areas that Microsoft Teams excels at that SharePoint does not:

1. "User-Friendly UI – less end-user training required to get them up and running
2. Logical team membership – manage permissions at a team level instead of at a file/folder level
3. Everything is truly in one place – no switching windows and programs for files, emails, meetings, and conversations
4. Work transparently – no need for extra emails about when files have been edited, when a project's status has changed or been completed, etc.
5. Channel conversations – prevent duplicate file sharing, ability to share more than just files (such as links)
6. Persistent private and group chat – eliminates the need to use Skype for Business too"

4.4.2 Is SharePoint a Database?

Golubenko (2019) mentioned that SharePoint is a database as it sits on top of the one of the most robust and well-known relational database management system, SQL. However, SharePoint itself is not a database, although its user will experience relational database looks. SharePoint lists information exactly like SQL server by providing columns and listing data types stored on it.

SharePoint is not a relational database substitute as it is unable to address relationships needs for complex data (Golubenko, 2019).

For simple data structures SharePoint is ideal but for complex relationship building through creating connection between datasets, other software such as PowerBI is needed.

4.4.3 Microsoft Teams + SharePoint: Collaboration Tools

(SharePoint, 2019) said in the post that, SharePoint has been around over a decade and is an excellent collaboration tool. It is a repository/one-stop-shop for content sharing and collaboration within an organization. It is primarily used for document storage and has remarkable document management capacities. A plus point of SharePoint is smooth integration with other Office 365 apps. SharePoint can be conveniently used as repository for all the content within an organization.

(SharePoint, 2019) mentioned that SharePoint is an amazing collaboration tool, but it lacks in the social/communication aspect; to overcome this Microsoft Teams can be utilized.

Microsoft Teams is a chat-based communication tool which is united with SharePoint through Office 365 Group. Any time a new Team is initiated, a separate SharePoint Site along with Office 365 Group and all its other assets like Calendar, Planner are created. When Files Tab within Microsoft Teams is accessed; the files seen are stored not within a Team, but rather in a document library that resides on a SharePoint site that got provisioned, when you created a Team for a chat (SharePoint, 2019).

Microsoft Teams collaboration settings combined with SharePoint's strong document management systems can provide an integral system that effortlessly offer a well-developed collaboration solution (Oliveira, 2019).

Table 9: The advantages of a Microsoft Teams/SharePoint integration	
Advantage	**Description**
One Consolidated Platform	Ability to access, evaluate and sync files for CPATT Team
Automatic Previews (through SharePoint)	Conveniently find the needed documents
Data Protection and Security	Observance of data loss and document retention
Fused searching system	Instant search for searching and discovering files
Linking Conversation	Keeps flow of discussion and topic clear
Ability through MS Teams	To review concise statistical and other crucial information
Ability to preview documents	Through MS Teams easily access and modify their files
Source: (Oliveira, 2019)	

4.5 CPATT Database Survey Results

To arrive at a feasible solution, current practices at CPATT were evaluated through an internal survey. A group of 25 CPATT affiliates (including current and past students, Research associates, post-doc fellows etc.) took the survey, on September 19, 2019. Results of the survey are provided in this section. Survey participants were asked about their usage of different software under different categories and to rank pre-selected factors that they consider to be most important for development of framework for a robust and meaningful CPATT database. Some salient results of this survey are presented in this section.

4.5.1 General

60% of the survey participants have been affiliated with CPATT for one to three years. Doctorate and master's students comprise 60% of the CPATT affiliates surveyed, 16% are staff and 24% are others such as postdoc fellows and co-op students. The highest percentage (40%) of these affiliates work 31-40 hours and 32% work more than 40 hours on CPATT related tasks every week.

4.5.2 Data collection and storing efforts

The survey participants were asked about their collection and storing needs of CPATT related data.

Table 10: CPATT Survey Results (A)	
Survey participants collect and store data each term	Percentage
0-4 times	50%
5-8 times	21%
9-12 times	17%
13-16 times	0%
>16 times	13%

The 0-4 times per term reflects that the maximum @ 50% of CPATT affiliates collect and store data once per month. Greater than 16 times a term reflects that 13% of CPATT affiliates collect and store data on average more than once per week.

4.5.3 Data sort and analysis efforts

The highest percentage as seen above is 33% for 5-8 times per term translating into over one to two times a month.

Table 11: CPATT Survey Results (B)	
Survey participants sort and analyze data each term	Percentage
0-4 times	25%
5-8 times	33%
9-12 times	25%

13-16 times	4%
>16 times	13%

The collection and storing efforts are in line with sort and analysis effort. As seen in Table 10, 50% participants are collecting and storing data once a month while only 25% of the total percentage are sorting and analyzing data 0-4 times per term. Among the same group of greater than 16 times per term (more than once a week) data collectors (13% in Table 10); 8% have reported to sort and analyze data.

4.5.4 Current practices

The survey partakers were asked about their current use of software for storing and analysis. An eighty percent are currently using the MS Excel as their data storing software; and from the same 80% of MS Excel users, 48% of these users also use online storing services such as SharePoint and OneDrive for their data storing needs.

A huge 88% of CPATT affiliates reported that they mainly use MS Excel as their analysis software; with 32% reporting that they also use MATLAB for analysis.

84% participants reported that MS Excel is in their current use for charting and graphing needs. Among these 84%, the users also reported that they use other charting software as shown below in Table 12.

Table 12: CPATT Survey Results (C)	
Survey participants (chart / graph)	Percentage
MATLAB	28%
Minitab	30%
Others (such as Kaliedograph, Origin etc.)	36%

Fourteen responses were received for question regarding participants other uses of spreadsheets. 10 out of 14 among these respondents reported that their other use includes carrying out statistical analysis using spreadsheets, which translates into over 71%.

Among 25 participants only four responded to question if a data specific query is being used by them; for which different software are used including MS Excel.

78% of respondents stated that they develop their spreadsheets from scratch, while 22% conveyed that they never have to go through the process of developing spreadsheets from scratch, which may include reasons such as spreadsheet received from multiple sources including peers and staff.

Over 95% reported that they validate their data models and check integrity of data. Only 4% stated that they never validate their data or check its integrity. About 43% of the correspondents advised that MS Excel is being used by them for validating their data models.

About 46% of the correspondents reported that they run out of space when they are dealing with large data sets (varying from always to usually). While 54% stated that they never run out of space.

4.5.5 CPATT Preferred software

The survey participants were asked about their preferred (future) use of software for storing and analysis. Their responses were as follows; depicted in Table 13:

Table 13: CPATT Survey Results (D)	
Preferred software for storing data	Percentage
MS Excel as their first choice	56%
Online services (SharePoint, OneDrive etc.)	40%
MS Teams	4%

It is important to note that although, MS Access was among the choices, however received a 0% weightage for storing data sets.

Table 14: CPATT Survey Results (E)	
Preferred software for sorting/analyzing data	Percentage
MS Excel as their first choice	76%
MATLAB	12%
Others (nlogit, SPSS etc.)	4%

MS Access as a second choice received a mere 8% of overall score allotted.

4.5.6 Familiarity with different software

The survey partakers were asked about their input on the familiarity and past use of different well-thought-out list of selected software to evaluate and determine the best options for selection of most popular and widely-used software. The response to these asks is recorded in Table 15.

Table 15: CPATT Survey Results (F)								
Ranking	**Excel**	**Access**	**SQL**	**MATLAB**	**SharePoint**	**Online**	**Teams**	**PowerBI**
Below Average	0	9	10	4	10	2	11	12
Average	3	5	6	4	5	7	4	3
Moderately Above Average	4	1	0	3	2	1	2	0
High Above Average	6	1	1	4	1	5	0	0
Advanced	10	1	0	2	0	6	0	0
Total	23	17	17	17	18	21	17	15

About 70% respondents mentioned that they have a well above average to advanced level of knowledge regarding MS Excel use. While 30% reported that their knowledge is average or moderately above average in MS Excel. 84% stated that their knowledge is average or below average for MS Access. 10% reported to have well above average or advanced knowledge in MS Access. For online repositories (such as One Drive, Drop Box etc.) about 52% have well above average to advanced knowledge. SharePoint have ~6% with well above to advanced users; with most users (~83%) with below average to average familiarity with SharePoint.

4.5.7 Critical Success Factors Criterion

As mentioned earlier in this writeup, a model for the criterion of CPATT database repositories and software was developed based on literature review, interviews with both CPATT management and external (to CPATT but within University of Waterloo) experts. The selection criterion for CPATT database have fourteen different criteria. These criterions were presented to the survey partakers and they were asked to rank in order of what they think is the most important among these factors. Every survey participant provided a ranking to every criterion. These ranks were added to achieve a consolidated score for ranking; such as score for "Initial Cost" was considered to be the least important; as among all participant everyone gave it a score of 12 (which translates into 144). The results are presented in Table 16 below:

Table 16: CPATT Survey Results (G)

Score of Rank	Preference Rank	Critical Success Factor
144	12	Initial Cost
142	11	Maintenance Cost
78	**2**	Easy Use for Analysis
83	**3**	Ease of Learn
76	**1**	Ease of Access
84	**4**	Data Security
115	7	Store Different Data Types
123	10	Controlled Database Access
146	13	Easily Upgradable & Cost Effective
117	9	Database Longevity
116	8	Sufficient Storage Capacity >5TB (for all the data)
162	14	To Present not to Collect Data
95	**5**	Data Safety
114	6	Flexibility of Use

Source: CPATT Survey (September 2019)

The top five criterion selected by survey participants are as shown highlighted in green, which include as the number one criterion to be the ease of access, followed by easy use for analysis and third in rank is easy to learn. Data security and data safety are next. The top two criterion as a result of this survey secured almost the same aggregate score.

4.5.8 CPATT Survey Summary

Based on the interviews with CPATT management and other requirements; CPATT requires a robust and meaningful database. This survey results confirm the findings of this research that the CPATT database framework shall include working with MS Excel (spreadsheet), SharePoint / OneDrive (repository), MS Teams (collaboration among CPATT affiliates) and PowerBI for analysis.

The following software were considered as the main software for datasets; 1) MS Excel, 2) MS Access and 3) SQL. As seen through results of this survey MS Excel is the most widely used for storing (@80%), analyzing (88%) and graphs (84%). MS Access and SQL both are rarely used (4% to 12%) for storing, analyzing and graphing purposes.

MS Excel also has the most advanced or well above average users (69%). This will make it easier to handle the database.

The survey participants were asked if they have any other comments regarding the CPATT factors, only two comments came in, stating they would like datasets to be more secure. This result confirms that the selected 14 criterions are the critical success factors for CAPTT database framework and we have them ranked now; wherein concentrating on the most important top five and giving due consideration to the other critical success factors for CPATT database.

This also requires that some kind of training will be required for SharePoint, MS Teams and PowerBI. Currently, internal University of Waterloo provides training for SharePoint and MS Teams in different formats. PowerBI can be learned through some free online available resources.

Based on the CPATT database selection criterion, meetings with University of Waterloo affiliates, management and CPATT database survey results the following has evolved for the framework.

1. For collaboration and management of database access; Microsoft Teams will be used.
2. For the purpose of storing, University of Waterloo SharePoint will be used to provide safe and secure access to database/datasets.
3. For the dataset development the and immediate analysis Microsoft Excel is the best software for utilization
4. For advanced dataset analytics and development of quarries Microsoft PowerBI will be used
5. Master project file is created to elude duplication of datasets and have a unique number to identify each dataset

4.6 Documents Collaboration and Co-authoring

Microsoft Incorporation (n.d.) mentions that SharePoint can be conveniently utilized to allow multiple users simultaneously access and edit the same file. When everyone is working on the same file at the same time, the process is called co-authoring. SharePoint Online and SharePoint Server along with OneDrive have capabilities to allow co-authoring.

Co-authoring has the following limitations (Microsoft Incorporation, n.d.):

1. **Software Limitations:** “Word and PowerPoint on all devices and versions more recent than Office 2010 support co-authoring. The Excel mobile apps and the latest version of Excel for Office 365 also support co-authoring”.

2. **Co-authoring friendly formats:** Co-authoring is possible only on newer file formats such as: .docx (Word), .pptx (PowerPoint), and .xlsx (Excel).
3. **Edit permissions for co-authors:** All co-authors need permission to access and edit the required documents.

4.7 Process to Upload Datasets to CPATT Database

1. Request access to CPATT Database through Database Administrator or Owner.
2. Database access is to be granted to researchers with a valid University of Waterloo email, although access can be given to anyone with a valid email address.
3. Download/access files such as Master Project File, Metadata Schema, related standards the dataset and abbreviation list.
4. Register your dataset in Master_Project_File.xlsx file (MPF); to get the four-digit CPATT_Data_Code (MPF#). This is a Masterfile which is to be edited online and then saved back to the database. The Master Project File is a MS Excel file which is pre-populated with 10,000 (ten thousand) codes. Follow the sequence and fill in the applicable fields in the file. File is to be checked out and checked in immediately after its use (filling all required fields and obtained MPF# the CPATT_Data_Code). CPATT_Data_Code or MPF# is a four-digit alpha-numeric code; used to identify a specific dataset for research.
5. Utilize the Metadata_Schema_for_CPATT_DATA_Ver01.xlsx file to fill in all the required data providing essential information about your data. [this is a downloadable file and a copy can be retained on user's computer). Metadata Schema file shall be named based on the following convention]. Example for a Metadata file of a LWD file of July 15, 2019 will be 20190715_META_LWD_MPF#. Metadata file on the SharePoint is not to be changed so that others can copy and use the original file. Every dataset will have its own Metadata schema file for each dataset of research's data; such as if there are five datasets of a research; five Metadata files will be created as outlined in this step.
6. Use the following naming convention for your data. YYYYMMDD_4DIG_EQP_MPF#; Where: YYYYMMDD is date (YYYY= 4-digit year; MM = 2 digit for Month; DD = two digit for day; e.g., July 15, 2019 is written as 20190715). 4DIG is four Digit # that will be provided in order of sequence, e.g., if there are more than one raw file for a specific data on same date; first RAW file will be 20190715_RA01_IRI_MPF# and second file is 20190715_RA02_IRI_MPF#. EQP is Equipment or Test which is IRI in this specific case and MPF# is obtained from MPF file as CPATT_Data_Code. The 4DIG codes/abbreviations for providing names to data files are given below. The 4DIG in most cases consist of first two alphabets for the type of code

followed by two digits such as 01 is the first file, next file with same date will be 02, 03, 04, …. and so forth [last two digits increases if there is more than one file]. A separate file (Use notepad such as 20190715_RE01_IRI_MPF#.txt file) can be used to explain how the data has been converted from RAW to Processed (workable such as MS Excel) or another format.

7. Use the following files within the Data file (such as, 20190705_DA01_IRI_MPF# (for file of IRI data collected on July 5, 2019) will contain the following sheets
8. README: Important Initial and general information to understand data and analysis of a specific research process. General information (this sheet is used to provide any general information about the data not mentioned anywhere else); description/information that is necessary to understand the data background, data itself, understanding of data or any other useful information.
9. INFO: Specific information (sample files)
10. DATA: Data ready for analysis
11. DRAW: drawing, sketch of layout, location, etc. (any drawing sketches of the site from where the data has been collected or related drawings/sketches)

Table 17: List of CPATT Database Abbreviations (for Equipment/Test)	
Abbreviation (Three Alphabets)	**Name of Equipment / Test**
LWD	Light Weight Deflectometer
FWD	Falling Weight Deflectometer
IRI	International Roughness Index /SURPRO
TGO	T2GO
STR	Strain Gauge
BPN	British Pendulum Number
TEM	Temperature Gauge
PRE	Pressure Gauge
MOI	Moisture Gauge
ARA	ARAN
BOR	Boreholes Data

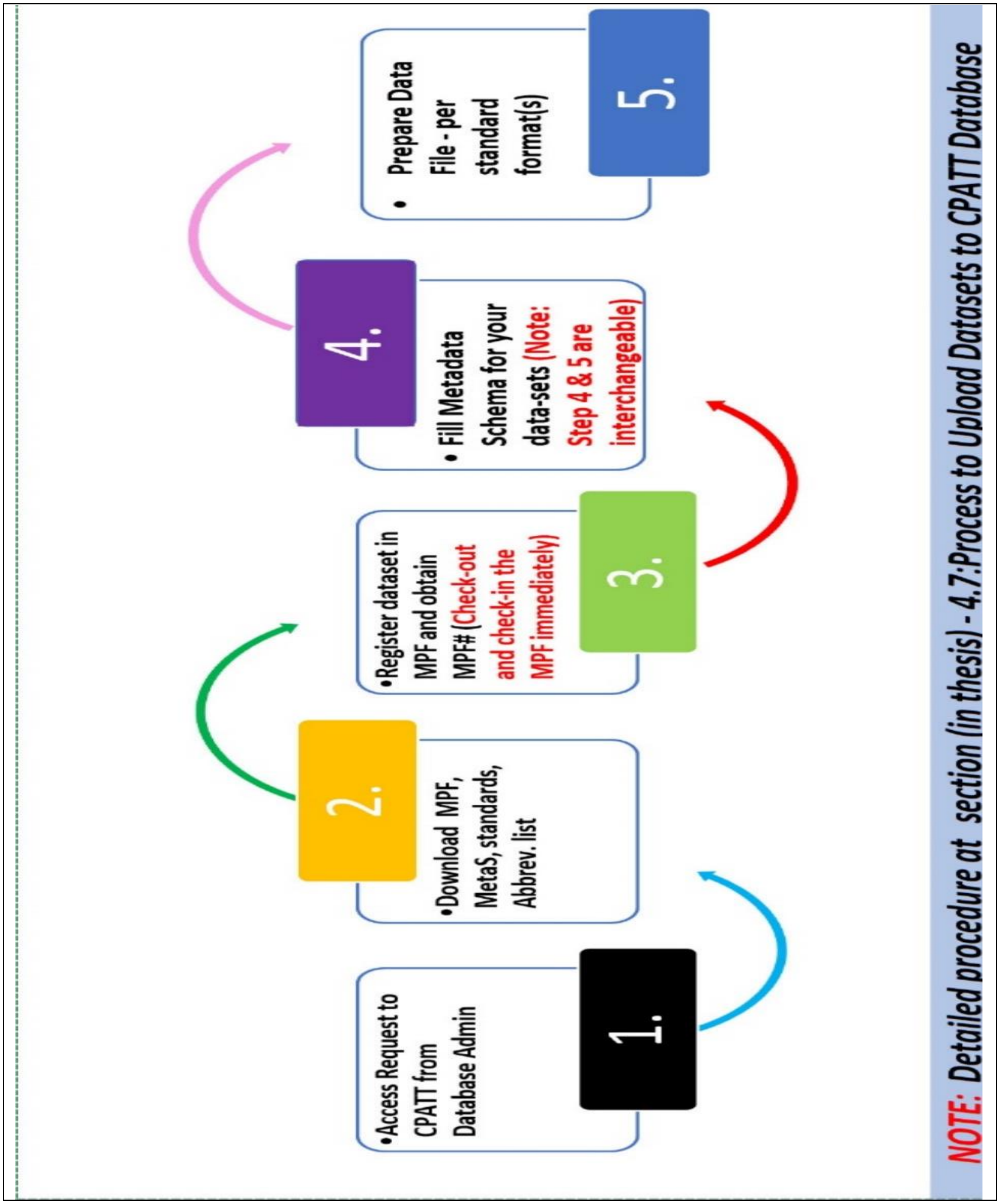


Figure 12: CPATT Database - Upload Process Flow Chart

Details of 4DIG codes are provided as below:

Table 18: List of CPATT Database Four (4) Digit Codes

4DIG	Description
ABBR	Abbreviations (One standard Abbreviation's file will be created for use with all research, its responsibility of the Data Collector to update this file with all their data related abbreviations they have used in their research) 20190821_CPATT_ABBR
AN01	Analysis (Analysis of the data in Excel format with first sheet explaining the logic, reasoning, steps behind analysis, use this to mention any formulas with reference used in analysis)
DA01	Data (Excel; use this file to submit your data in a standard format – data submitted must meet the standard data format requirements)
MS01	Manual / Standard (Use this 4DIG code for manual of equipment, used process, applicable standards etc.)
META	Meta Data file (Basic information about the data, separate format is provided in Metadata_Schema_for_CPATT_DATA_Ver01.xlsx file to fill in all the required data)
PR01	Procedure (Explain here details of the procedures / processes used to collect, and sort data. AS well for info about converting data from raw format to workable (Excel/CSV) format
PH01	Photos [Use this as a Word file with tables to insert numerous pictures in jpeg/jpg/png (Bitmap) format, with Caption of photo, Name of photographer, date and other basic data. Maximum number of pictures shall be no more than 100 in one file].
RA01	Raw data (such as obtained from equipment such as LWD)
RE01	README File for explanation of Data processing from raw to an excel or another format.
VI01	Videos (Acceptable formats includes what is playable on Windows platform without requiring any external software)

4.8 Process to maintain/safety of CPATT Database

The process of maintaining as discussed with the CPATT management is as follows:

1. Researcher requiring access to CPATT Database request the database administrator (which could be a co-op student, research associate or a staff member-including IT team). This depends on the final decision of the CPATT management, considering critical factors such as time, cost and expertise required to properly maintain the CPATT Database.
2. The process of selection for CPATT Database has established that keeping data safe and secure is of the utmost importance. In this regard, the University of Waterloo has a policy of Duo-factor security which will is required to access the online portals of UW.
3. Per Kevin Rampersad (IT Manager, Civil and Environmental Engineering, University of Waterloo), the selected repository has an automated backup system, however, the CPATT management will be provided a soft copy on a storage media as another backup.

4.9 Chapter Summary

Identified the appropriate and most relevant CPATT Database definition.

Salient Database advantages and disadvantages such as minimal redundancy, privacy and security as advantages. Disadvantages are that very costly and high hardware costs.

Development of selection criterion (model) for CPATT Database which includes salient factors such as:

Low initial cost,

Low maintenance costs,

Longevity,

Short learning curves,

Security of data, etc.

Development of CPATT Database Management System (DBMS) based on the modeled selection criterion.

Verification of selection criterion through conducting of a survey for 25 CPATT stakeholders based on questions such as:

CPATT data collection and storing efforts

Number of times CPATT data is sorted and analyzed each term

Preferred software for storing data

Preferred software for analyzing data

Summarizing of CPATT survey results

Discussion on CPATT survey results

Detailed process of uploading datasets to CPATT database.

5. INTER-RELATION DATA MODELS

This chapter shows the detailed discussions on the development of inter-relation models for data-permitting-datasets.

Identification of salient gaps between the existing and revised datasets. Inter-relation models between FWD and LWD datasets as well between BPN and T2GO datasets., with a regression equation for relationship between the BPN and T2GO 2017 datasets.

5.1 Current Data collection and developed Models

Figure 13, depicts a model developed for this research based on the datasets collected through CPATT for past research, Long Term Pavement Performance (LTPP) and evaluation of ASTM standards. Two datasets for FWD and LWD collected on Jameston Avenue, Hamilton, Ontario by CPATT researchers were evaluated for missing data information from both datasets. The elastic moduli for FWD (both Ep (Effective pavement modulus) and Epcc (Elastic modulus for concrete)) are calculated per AASHTO 1993 Design Guide for pavement structures (Korczak, 2019). The elastic modulus of FWD data has a correlation with LWD modulus. The datasets provided had different results for modulus calculations which were further explored and verification of datasets was carried out required in conjunction with each other.

In Figure 13Figure 13, the yellow highlighted fields are the missing data based on extensive literature review and datasets received from different databases for both LWD and FWD data. An effective FWD and LWD data-set should include these fields for better data analytics. Data type is also classified as measured, standard, or calculated. Standard refers to a data point or value which will be same or similar throughout the data-set. Measured is a value actually measured in the field and calculated field is for values calculated by use of the respective formulae.

The three-double-arrows in Figure 13; depicts the inter-relations that can be developed between FWD and LWD existing data. LWD data collected generally provides back calculated resilient modulus M_R of subgrade. However, since CPATT has collected deflection data for concrete pavement using both FWD and LWD on the same day, this data can be utilized to develop an inter-relational model between FWD and LWD for equivalency equation between the two different datasets.

Both FWD and LWD deflections vary over a wide range and to determine the range of data-set, outliers shall be identified and removed from the data. As presented later in this research; for FWD and LWD there could be outlier deflection data points which can be caused by the pavement distortions. To effectively identify the outliers for FWD, Nelson's eight rules of statistical quality chart can be utilized. For LWD either Nelson's eight rules can be utilized or identified by the equipment operator as deflection data points have a range of about 300 micrometers recorded and few data points are well above 1000 micrometers (observation made from LWD data-set for Jameston Avenue, Hamilton) (CPATT Database - concrete overlays strain data, 2016-2019).

FWD vs LWD (Standards data and missing data)

Data Type	FWD Data
Standard	Equipment Operator
Standard	Equipment Manufacturer/Model
Standard	Equipment calibration report
Standard	Date
Standard	Street
Standard	Direction
Measured	Station
Measured	Joint Location
Standard	Wheel Path
Standard	Lane
Measured	LoadSize
Measured	D_1
Measured	D_2
Measured	D_3
Measured	D_4
Measured	D_5
Measured	D_6
Measured	D_7
Measured	D_8
Measured	D_9
Calculated	MR
Calculated	EP
Calculated	SN_{Eff}
Calculated	K_{Static}
Calculated	E_{pcc}
Calculated	D_{Eff}
Measured	AirTemperature C
Measured	SurfaceTemperature C
Measured	Latitude
Measured	Longitude
Measured	Construction Number
Measured	Boreholes (if any)
Measured	Load Transfer Efficiency
Measured	Notes on Pavement condition
Measured	Test Material
Measured	Void Detection Intercept
Measured	Layer Structure
Standard	Map
Standard	Contract information
Measured	Deflection sensors offsets

DEFLECTION

LWD Data	Data Type
Equipment Operator	Standard
Equipment Manufacturer/Model	Standard
Equipment calibration report	Standard
Date	Standard
Radius (mm)	Standard
FileName	Standard
Street	Standard
Direction	Standard
Location/Station	Measured
LoadSize (kN)	Measured
Point No.	Measured
Drop No.	Measured
Time	Measured
Deflection (um)	Measured
Avg_Deflection (um)	Calculated
Stress (Mpa)	Measured
Modulus (MPa) *	Calculated
Joint Location	Measured
Wheel Path	Standard
Modulus (MPa) *	Calculated
Avg. Modulus (Mpa)	Calculated
AirTemperature C	Measured
SurfaceTemperature C	Measured
Latitude	Measured
Longitude	Measured
Construction Number	Measured
Boreholes (if any)	Measured
Load Transfer Efficiency	Measured
Notes on Pavement condition	Measured
Test Material	Measured
Layer Structure	Measured
Map	Standard
Contract information	Standard

Legend:		
	FWD Data set	LWD Data set
	Missing FWD & LWD Data	*= Same values

Figure 13: FWD vs LWD datasets

The other inter-relational model presented in this research relates to two different sites where CPATT collects deformation data through strain gauges installed at Jameston Avenue, Hamilton and Spragues Road in the Region of Waterloo. Both are recently constructed concrete overlay pavements. Jameston Avenue has strain gauges which were installed in 2016 and Spragues Road strain gauges (CPATT Database - concrete overlays; strain and temperature data, 2015-2019) were installed in 2015.

Table 19: Strain Gauge Data Details Concrete overlays research

Data Type	Jameston Avenue (City of Hamilton)	Spragues Road (Region of Waterloo)
Pavement	Concrete overlay	Concrete overlay
Concrete depth	100mm	160mm
Separation Layer	HMA HL 3/Geotextile	HMA SP 12.5
Separation Layer depth	25mm	25mm
Total sensors	Six (east and west bound)	Ten (east bound)
Location in pavement	Outer Wheel Path (OWP)	Four OWP/ Four Inner Wheel Path (IWP) / Two middle
Temperature recording	No	Yes
Source: (CPATT Database - concrete overlays strain data, 2016-2019) (CPATT Database - concrete overlays; strain and temperature data, 2015-2019)		

Figure 14. depicts the potential data inter-relationship model based on the CPATT data collection efforts and literature review. Strain (1 to10) are strain measurements for Spragues Road, while VW_uE (1 to 4) are strain measurements for Jameston Avenue. It should be noted that even the different naming conventions indicate a need for CPATT data collection guidelines to result in consistent results which can be compared efficiently.

All data is recorded in units of micro strain for strain and Celsius (ºC) for temperature where applicable. Among ten strain gauges at Spragues Road, five are placed in longitudinal and five are placed in transverse direction. For Jameston Avenue, three are in longitudinal and three are in transverse direction. Two of six strain gauges at Jameston Avenue (east bound) are not considered in comparison with Spragues Road as they are over geotextile membrane instead of HMA.

The inter-relationship can be developed between longitudinal to longitudinal and transverse to transverse, such as, longitudinal Strain (1) readings will be related to VW_uE (1) and VW_uE(3) which are both longitudinal as well, as depicted in Figure 14.

These relationships will help develop models for curling in PCC pavements. Curling and warping is a process that occurs in PCC pavement due to changes in environmental conditions that can result in upward (concave) or downward (convex) deformation. Temperature and moisture are the two most significant environmental factors that can influence volumetric changes in PCC. The deformation induced by a non-uniform temperature gradient is referred to as curling, while the deformation induced by a non-uniform moisture gradient is referred to as warping. When the top of a PCC slab has a higher temperature or greater moisture content than the bottom, a positive gradient will be induced, and the top part of the PCC slab will experience more expansion than the bottom, resulting in downward slab curling or warping. Conversely, if the bottom of a PCC slab has a higher temperature or greater moisture content than the top, a negative gradient will occur, and the bottom part of the slab will experience more expansion than top, resulting in upward

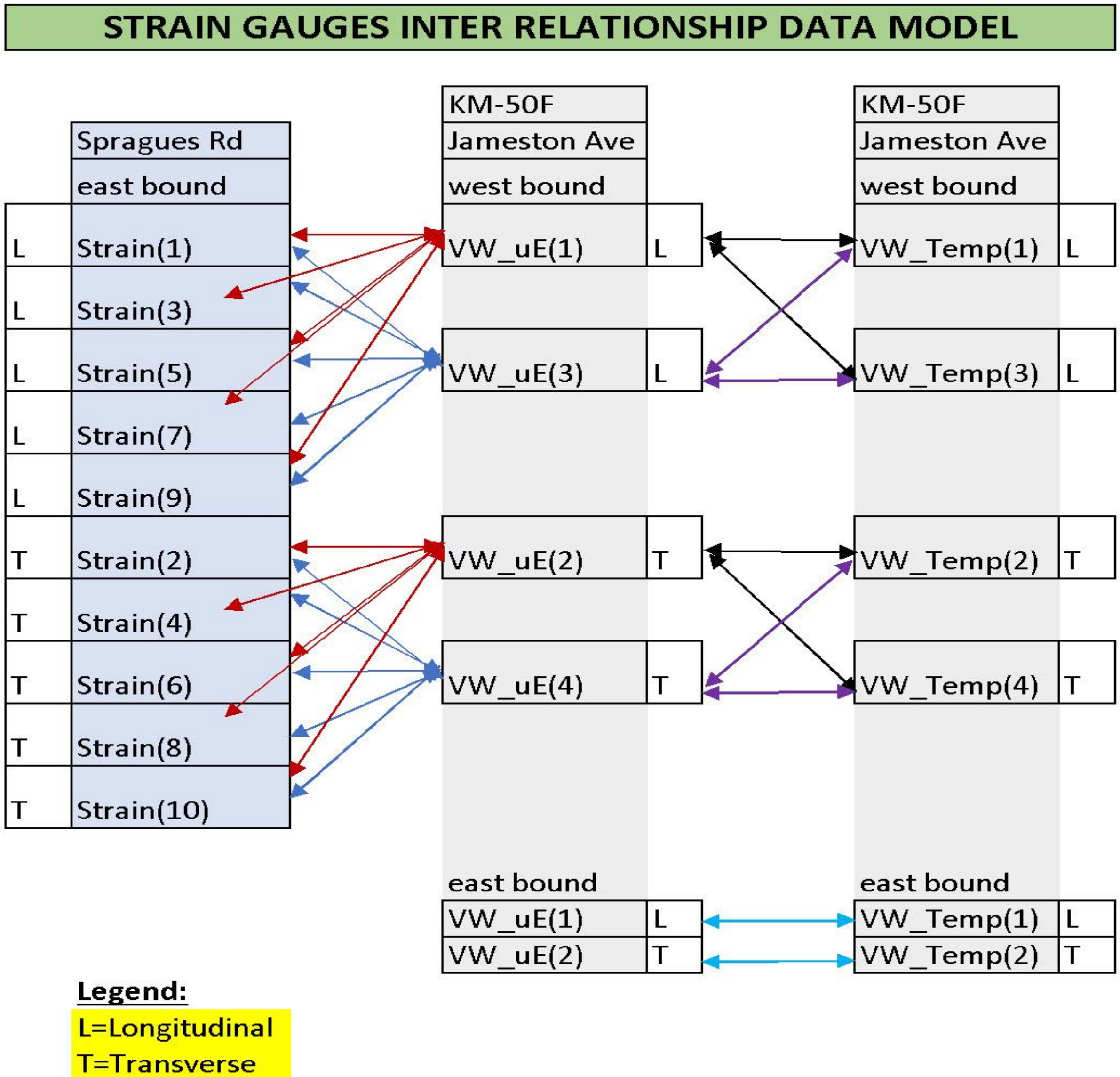


Figure 14: Strain Gauges inter-relationship Model

curling or warping of the slab. A positive temperature gradient usually occurs during daytime and a negative temperature gradient usually occurs during nighttime. Conversely, a positive moisture gradient usually occurs during nighttime and a negative moisture gradient usually occurs during daytime (Ceylan, 2016).

CPATT currently collects temperature and moisture data from several concrete pavement sites, which can be utilized for exploring and developing new methodologies, relationships and design inputs. An inter-relationship model can be developed for temperature, strain and moisture data to better reflect on curling and warping. The range for the strain gauge measurements would be determined based on the manufacturer's recommendation.

There are some datasets which are from the same site or if from different sites, have the potential for development of inter-relationship data models. However, this may not be possible for all the concrete pavement field data. In cases where the inter-relationship data models can't be developed within a reasonable time frame (during this study), a data standardization and quality check will be created for data inclusion into the CPATT database, such as statistical quality control checks developed for data-permitting datasets, etc.

5.2 Data format, structure and requirements

Initially the option of CSV (Comma Separated Values) file format was evaluated as file format for the CPATT Database. It was proposed that the CSV format will be used for the numeric datasets, however, after careful evaluation of the options, based on interviews with CPATT management and affiliates; and a survey conducted of CPATT members, MS Excel (.xlsx) is considered to be the best option for CPATT database file format. The CPATT database will be available on the University of Waterloo website (online CPATT database through MS Teams and Repository such as SharePoint/OneDrive) with secure access to datasets. In Table 20, the data is imported into Excel from the strain gauges native file *.DAT format (only part of data seen here).

As depicted in Table 20, row 1 to 10 contains data, where top four rows are related to general information about the research and column headings. Among these four top rows, Rows 1, 3 and 4 reflect the information such as research name, site name, some general codes (specific to the strain gauges used). Column-2 is a record number which is not required for analysis and along with Row 1, 3 and 4 will be omitted before the dataset is added to the CPATT database. This will reduce the data redundancy meeting the requirements of normalization forms (explained later in this study).

Based on the evolvement of this study it is considered reasonable to limit the development of inter-relationship models to 2-3 as this will lead to meeting the time constraints set for the completion of this study. Recommendations will be made for specific data-permitting-sets for further studies in the future research(s).

CSV data to be included in CPATT database will be measured and calculated data. Standardized data, which is common among almost all the fields will be provided in a different file and will be reflective in the data-set name, which will be incorporated before uploading data to online CPATT database.

CPATT database will consist mainly of the following types of files as seen in Table 21.

Table 20: Data requirements/redundancy example

	Column-1	Column-2	Column-3	Column-4	Column-5	Column-6	Column-7	Column-8	Column-9	Column-10	Column-11	Column-12
Row-1	TOA5	CR1000_caledon	CR1000	6916	CR1000 .Std.28	CPU:Spragues Road Jan287.CR1	4896	Max				
Row-2	TIMESTAMP	RECORD	Strain(1)	Strain(2)	Strain(3)	Strain(4)	Strain(5)	Strain(6)	Strain(7)	Strain(8)	Strain(9)	Strain(10)
Row-3	TS	RN										
Row-4			Smp	Smp	Smp	Smp	Smp	Smp	Smp	Smp	Smp	Smp
Row-5	9/16/2016 16:00	5545	256.84	980.93	117.92	-710.16	315.68	219.77	778.23	-450.72	-218.10	124.92
Row-6	9/16/2016 17:00	5546	256.31	981.62	115.90	-711.12	314.14	221.51	775.93	-448.93	-220.33	121.68
Row-7	9/16/2016 18:00	5547	258.28	980.93	111.35	-708.70	314.33	221.97	778.27	-448.60	-219.54	116.44
Row-8	9/16/2016 19:00	5548	261.39	979.03	107.01	-704.42	317.15	220.32	785.72	-451.80	-212.17	113.64
Row-9	9/16/2016 20:00	5549	268.08	975.61	102.95	-697.29	322.05	217.72	797.34	-457.60	-201.39	110.62
Row-10	9/16/2016 21:00	5550	275.03	971.49	99.90	-689.68	327.21	214.37	808.00	-463.67	-191.61	108.59

Source: Strain gauges data collected from Spragues Road

Table 21: File types / formats for CPATT database

Data Type	File Type
Databases:	XML, CSV, XLSX, others
E-Books:	EPUB, others
Images:	JPG, JPEG, PNG, PDF, BMP, others
Documents:	DOCX, others
Sound:	MP3, FLAC, others
Text:	TXT, CSV, PDF/A, ASCII, UTF-8, others
Video:	MPG, MOV, AVI, MP4, others
Spreadsheets:	CSV, XLSX, others

Microsoft Excel and its tools such as Power Pivot, Power Query and Power BI can be utilized for data query and mining hence developing multi-faceted data exploration methodologies. The benefit of using MS Excel can be converted to CSV which can be further exported to MS Access and linking together of different datasets is easier. This will give the future researcher the capability of using either MS Excel or MS Access for their data analytics requirements.

Every data site will be given an alpha-numeric code for easy identification of the datasets. If someone would like to search for all the FWD data collected for CPATT in Ontario or any other specific area, they would be able to find the general information about the site, datasets, videos, images, PDF file, drawings etc. easily. A master project list will be developed for CPATT database, where more research(s) will be added as they occur.

The RDBMS is a set of tables which are interrelated for accessing data easily. Other data such as videos, images, drawings, PDF etc. will be stored in online CPATT database. Data in relational databases is stored in different access control tables, each having a key field that mainly identifies each row. Relational databases require that each table has a key field that uniquely indicates each row, and that these key fields can be used to connect one table of data to another. Relational database is preferred over the other available database structure types for the following reasons:

- Ease of use, maintenance and learning.
- In RDBMS, DB can be changed without changing the whole database.
- Provides a wide-range of availability in terms of DB software.

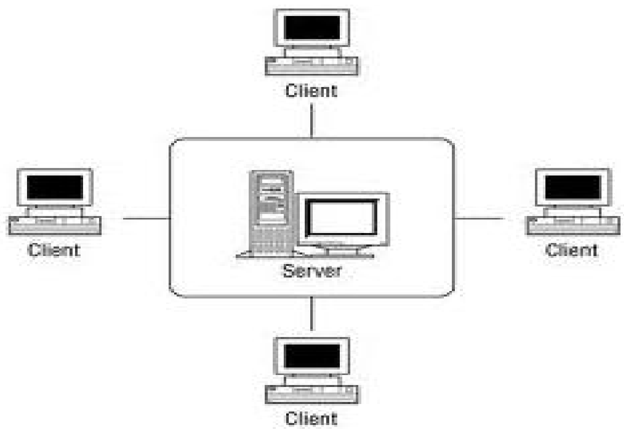


Figure 15: A RDBMS layout of user interface

5.3 British Pendulum Tester and T2GO

British Pendulum Tester (BPT) is an equipment used to test friction or skid resistance of the pavement surface. Its units are British Pendulum Number (BPN). The tests are conducted yearly to determine change in skid resistance over time (ASTM E303-93., 2013). T2GO is rolling device (similar to the BPT), which is used to find the coefficient of friction (COF) of a surface, and can be monitored at numerous times of the year.

Skid resistance is used to define tire-pavement interaction. The two terms (skid resistance and friction) are used interchangeably, as they are in the literature (Corsello, 1993).

The British Pendulum Tester is a dynamic pendulum impact-type tester used to measure the energy loss when a rubber slider edge is propelled over a test surface, Figure 16. The BPN values represent the frictional properties obtained with the apparatus.

A MASc student from CPATT carried out testing at Jameston Avenue, Hamilton, Ontario, Canada by using both BPT and T2GO to determine friction. These set of tests (both BPT and T2GO) were carried out one year apart, one right after construction of the newly built overlay concrete pavement in August 2016 and the other set of tests were conducted one year after construction in August of 2017.

Wafa (2018) reports that six locations on each section Jameston Avenue were selected, three in each direction, eastbound and westbound. A BPN value of >54 indicates good slip resistance. To determine the yearly performance of the overlay, tests should be conducted at various times of the year.

The pendulum has swung 4 times in each location to get a good overall average of the skid resistance values of Jameston Avenue, as seen in Figure 16.

Figure 16: British Pendulum Test Apparatus

Source: (Wafa R. , 2018)

Each direction (east bound and west bound) was divided into three sections. Test locations 1-3 is Section 1, 4-6 is Section 2 and 7-9 is Section 3. Over the one year apart testing period there were no significant change observed in the BPN. Most of the BPN values did decrease non-significantly which was expected due to normal wear and tear of the road owing to traffic, however these BPN values recorded are well above the threshold for a road with good slip resistance (Wafa R. , 2018).

The coefficient of friction of the Jameston Avenue pavement surface was measured using the device called the T2GO. T2GO is made by ASFT (Airport Surface Friction Tester (ASFT), 2017). T2GO is a mechanism similar to a wheelbarrow that runs over the section being tested Figure 18. It collects friction values and is a good indicator of pavement performance. Like the British Pendulum test, T2Go was conducted on 6 locations on each section, a friction value greater than 0.5 indicates good braking 96 and skid resistance. The testing for T2GO was conducted right after construction (before the road opened) and a year the road has been open to daily traffic (Wafa R. , 2018).

Similar to the BPN the T2GO showed great success as well a year after the road has been opened to daily traffic. The road has gone through one freeze thaw cycle and has not experience a significant decrease in skid resistance. Initially after construction the T2GO friction values ranged from a high of 0.84 to a low of 0.5. A year after construction the recorded range of values remained the same. Both BP test and T2GO test found section three (locations 8, 9 and 10) to have the lowest skid resistance.

Figure 17 depicts the CPATT research team in-action for British Pendulum testing on Jameston Avenue, Hamilton, Ontario.

In a guideline developed for Road Administrations around the world, (Airport Surface Friction Tester (ASFT), 2017) mentions that T2GO is developed as a quick, reliable, and easy-to-use skid resistant tester (SRT).

As both BPT and T2GO are skid resistant tester (SRT), an inter-relationship model among the measured values of these two SRT is developed. There are four set of data collected by CPATT researchers at a time apart by on year; two datasets of BPT in 2016 and 2017 each; and two datasets for T2GO in 2016 and 2017 each. Since these readings have been taken at the same location or very close to each other, the datasets are comparable and the inter-relationship model can be developed for these four datasets (CPATT Database, concrete overlay BPT & T2GO Data, 2016-2017).

Source: (Wafa R. , 2018)

Figure 17: British Pendulum testing on Jameston Avenue, Hamilton, Ontario

Source: (Wafa R. , 2018)

Figure 18: T2GO testing on Jameston Ave., Hamilton concrete overlay

These four data sets are spread over two years as follow:

Table 22: BPN/T2GO – count of sections and readings

Equipment	No. of Sections (2016 & 2017)	Reading each section	Section divided into no. of points
BPT	Three	Four	Three
T2GO	Three	Three	Three

There were total of 36 readings measured for each east and west bound lane for BPT. For T2GO the respective data readings collected for each east bound and west bound were 27 each. In order to better analyze the T2GO and BPN datasets were grouped in three different sets; 1) Dataset for 2016, 2) Dataset for 2017 and lastly 3) Dataset for combination of both 2016 and 2017 for getting a larger dataset.

5.4 Regression Analysis Model for BPN and T2GO

Using Regressit (A Microsoft Excel Add-in); Regression analysis were carried out for the datasets after combining BPN and T2GO for both 2016 and 2017 as shown in Table 23. Table 24 and Table 25, respectively reflects the raw data that was compiled to get the datasets as seen in Table 23. The average of all the readings in 2016 and 2017 for both BPN and T2GO were calculated. This total no of readings or n=36. BPN was selected as dependent variable while T2GO was the predicted variable.

In his explanation, (Minitab Blog Editor, 2013) say that in linear regression analysis, the constant term is known as the y intercept, it’s a value at which the fitted line crosses the y-axis. He further elaborated that the value of the constant term is almost always pointless as the true value of a regression analysis is to comprehend the changes in response variable with changes in values of predictor variables. Due to the above facts, no constant was selected for regression analysis.

Table 23: BPN/T2GO Data compiled and used for Regression Model

BPN (2016&2017)	T2GO (2016&2017)
BPN Units	**Mu-values**
73.75	0.78
70.25	0.74
75.5	0.77
77.75	0.80
75.5	0.78
79.75	0.77
66.5	0.71
53.25	0.62
54.25	0.69
63.5	0.75
57.5	0.75
83	0.75
72.5	0.76
72	0.78
75	0.70
64.5	0.67
64.75	0.62
60	0.59
76	0.78
69	0.75
74.5	0.75
63.25	0.80
74	0.79
73.25	0.77
64.5	0.79
57.5	0.61
56.5	0.68
74.75	0.73
82.25	0.71
70	0.73
70	0.73
74.5	0.77
70.5	0.75
64	0.74
60.25	0.63
57.75	0.58

Table 24: Dataset for BPN (RAW) [2016 & 2017]

BPT: New Construction Aug 2016				BPT: 1 YR after Construction Aug 2017			
Section	Reading #	West	East	Section	Reading #	West	East
Section1	1	72	64	Section1	1	76	75
	1	73	64		1	77	74
	1	74	63		1	76	75
	1	76	63		1	75	75
	2	66	57		2	69	81
	2	71	60		2	71	83
	2	72	56		2	64	81
	2	72	57		2	72	84
	3	73	80		3	76	70
	3	74	84		3	73	70
	3	78	84		3	74	67
	3	77	84		3	75	73
Section2	4	77	72	Section2	4	64	68
	4	75	72		4	64	68
	4	79	72		4	61	70
	4	80	74		4	64	74
	5	75	69		5	71	74
	5	75	72		5	74	74
	5	76	75		5	77	73
	5	76	72		5	74	77
	6	82	74		6	77	75
	6	78	73		6	71	71
	6	82	75		6	74	68
	6	77	78		6	71	68
Section3	7	65	65	Section3	7	64	60
	7	66	64		7	63	62
	7	68	66		7	65	67
	7	67	63		7	66	67
	8	54	64		8	53	65
	8	54	65		8	59	59
	8	53	65		8	57	57
	8	52	65		8	61	60
	9	52	60		9	55	60
	9	55	60		9	54	59
	9	55	60		9	56	56
	9	55	60		9	61	56

Source: (Wafa R. , 2018)

Table 25: Dataset for T2GO (RAW) [2016 & 2017]

T2GO: New Construction Aug 2016				T2GO: 1 YR after Const. Aug 2017			
Section	Reading	West	East	Section	Reading	West	East
Section1	1	0.75	0.76	Section1	1	0.79	0.75
	1	0.79	0.7		1	0.77	0.65
	1	0.8	0.8		1	0.77	0.8
	2	0.64	0.7		2	0.65	0.72
	2	0.79	0.77		2	0.8	0.7
	2	0.8	0.79		2	0.8	0.7
	3	0.77	0.78		3	0.76	0.8
	3	0.8	0.77		3	0.72	0.73
	3	0.75	0.7		3	0.77	0.65
Section2	4	0.77	0.76	Section2	4	0.79	0.77
	4	0.84	0.73		4	0.8	0.65
	4	0.79	0.79		4	0.8	0.78
	5	0.79	0.79		5	0.8	0.77
	5	0.77	0.79		5	0.77	0.75
	5	0.79	0.76		5	0.8	0.8
	6	0.74	0.67		6	0.74	0.67
	6	0.78	0.7		6	0.77	0.8
	6	0.78	0.72		6	0.8	0.77
Section3	7	0.71	0.67	Section3	7	0.79	0.75
	7	0.68	0.68		7	0.77	0.76
	7	0.73	0.66		7	0.8	0.7
	8	0.66	0.63		8	0.62	0.6
	8	0.58	0.59		8	0.65	0.61
	8	0.62	0.65		8	0.56	0.67
	9	0.72	0.6		9	0.68	0.62
	9	0.66	0.52		9	0.7	0.55
	9	0.7	0.65		9	0.67	0.57

Source: (Wafa R. , 2018)

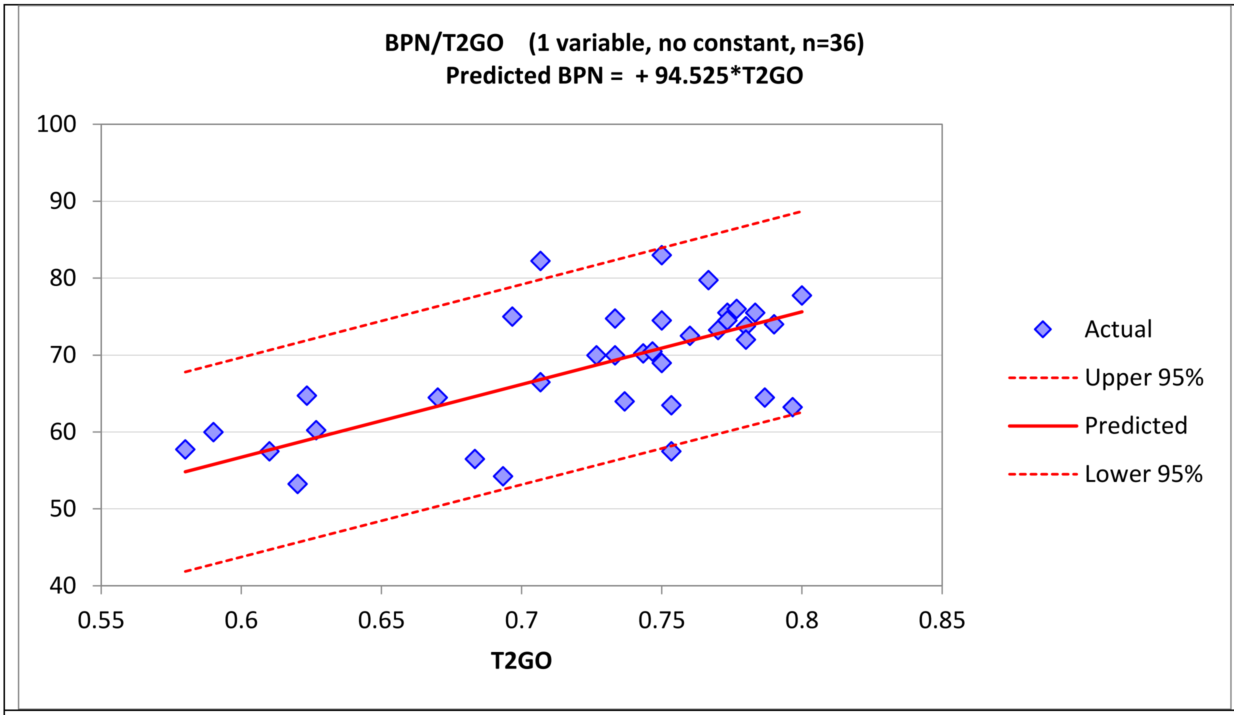


Figure 19: Chart of Regression Model for BPN and T2GO

The regression analysis, derived a relationship equation between BPN and T2GO, given as below:

$$\mathbf{BPN = 94.525 * T2GO}$$

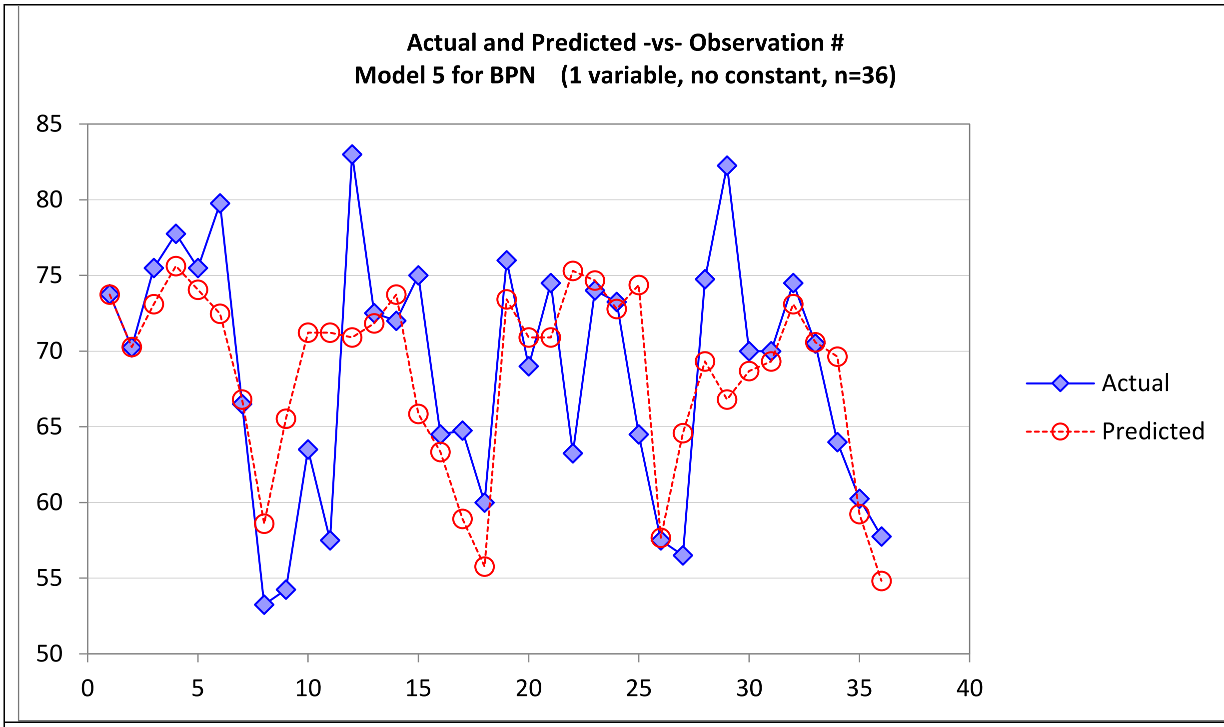


Figure 20: Regression Model (Actual and Predicted – vs- Observation #)

Some salient/critical indicators of the correlation are provided in Table 26. As seen the R-squared is at 99.18% which is very high. (Frost, 2017) reported that High R2 values are neither always good nor always a problem. Under certain circumstances, R2 can be legitimately very high value (in the high 90s); such as a physical process which has very precise and accurate data. As collection of data for BPN and T2GO is considered an accurate process, the R2 value is acceptable.

Table 26: Regression Model (Critical Indicators of Correlation) BPN & T2GO

R-Squared	Adj.R-Sqr.	R-sq(pred)	t-Statistic	P-value	Std.Err.	Durbin-Watson
99.18%	99.16%	99.14%	65.259	0.000	1.448	1.765

5.4 Regression Analysis Model for FWD and LWD

A Regression Model was developed using the Regressit (A Microsoft Excel Add-in) between Falling Weight Deflectometer (FWD) and Light Weight Deflectometer (LWD) deflections and modulus of elasticity. In 2016, CPATT researchers collected, LWD and FWD data on Jameston Avenue, Hamilton, Ontario on existing pavement structure as part of a major research on concrete overlays. LWD testing was performed by members of CPATT while FWD testing was contracted out to Stantec Consulting Inc. Wafa (2019) stated that the top asphalt layer was removed from the existing pavement structure before the FWD testing was conducted (on before construction concrete panels) by Stantec Consulting. A CPATT team was following the FWD testing team with a CPATT-owned Dynatest LWD equipment and measured deflections using LWD. Both FWD and LWD were used simultaneously (with a time difference of few minutes) to measure deflection of the rigid concrete pavement. (Wafa R. , 2019) also mentioned that the LWD measured deflection were observed at the same locations where the FWD was used to measure deflections. This data provides basis for developing of a correlation model between FWD and LWD deflections and modulus. Wafa (2019) mentioned that LWD testing was conducted only for the Eastbound Lane with a total of 37 stations compared to FWD testing of 44 stations for each Eastbound and Westbound Lane. However, after careful evaluation of the LWD data, there were some outliers in the data readings for seven stations (locations), which were subsequently removed for regression analysis purposes from FWD data as well for apposite comparable data, making the data points at n=30 (instead of 37 for LWD and 44 for FWD). This data (n=30) is presented in Table 27.

The original LWD data was analyzed in LWDMod (an accompanied software with Dynatest LWD). The LWD was normalized to 40kN for making it comparable to the stress point (40KN) of FWD data collected by Stantec in 2016. These new deflections (normalized) were then used in the regression analysis as a variable to obtain the correlation between Def_LWD16 and the D1_FWD16. Some of the LWD measurements were taken multiple times (3 to 7 readings), those were converted to get the average deflections for each station.

The process of analyzing the normalization was discussed with (Lund, 2021) of Dynatest to confirm the process followed in LWDMod is as per the standards.

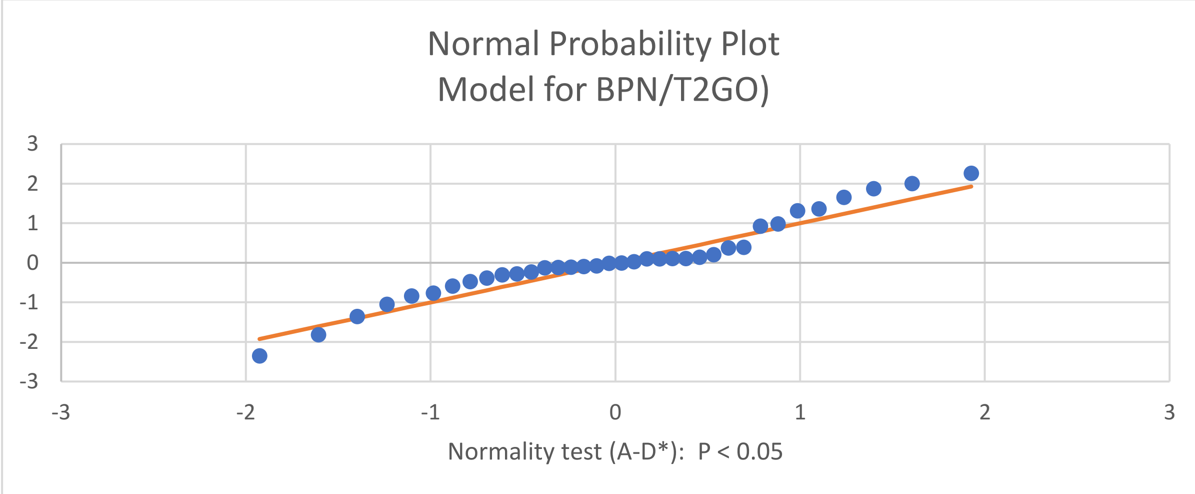


Figure 21: Normal Probability Plot

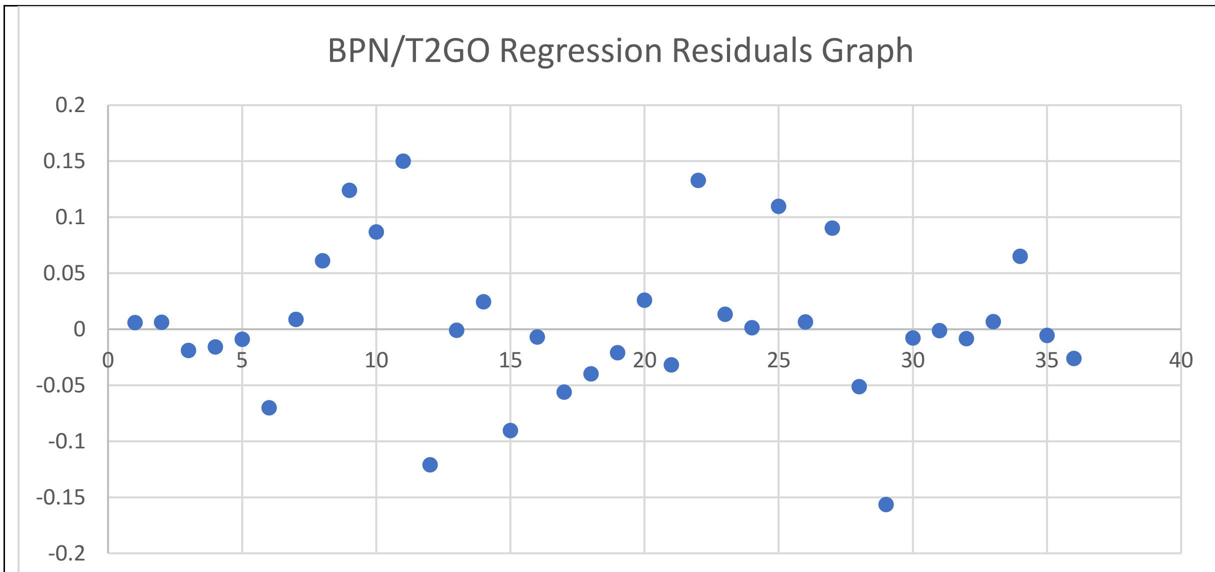


Figure 22: BPN/T2GO Residuals Graph

Table 27: Regression Model Data for FWD and LWD (using normalized deflection data)

Col A	Col B	Col C
Station-FWD2016EB	Def_D1_FWD2016 (µm)	Avg. Def_LWD (D1) Normalized to (40kN) (µm)
0+004	67	110
0+007	97	96
0+013	186	100
0+021	145	81
0+047	216	278
0+054	208	151
0+061	159	113
0+070	236	233
0+090	349	110
0+111	332	127
0+128	221	169
0+135	182	132
0+141	190	31
0+150	358	155
0+160	150	458
0+170	371	359
0+190	221	78
0+211	481	61
0+221	183	93
0+230	158	63
0+240	196	86
0+269	145	73
0+271	208	71
0+280	319	108
0+287	156	64
0+293	214	76
0+302	168	109
0+310	161	100
0+320	97	157
0+328	227	58

Source: (CPATT Database, 2016)

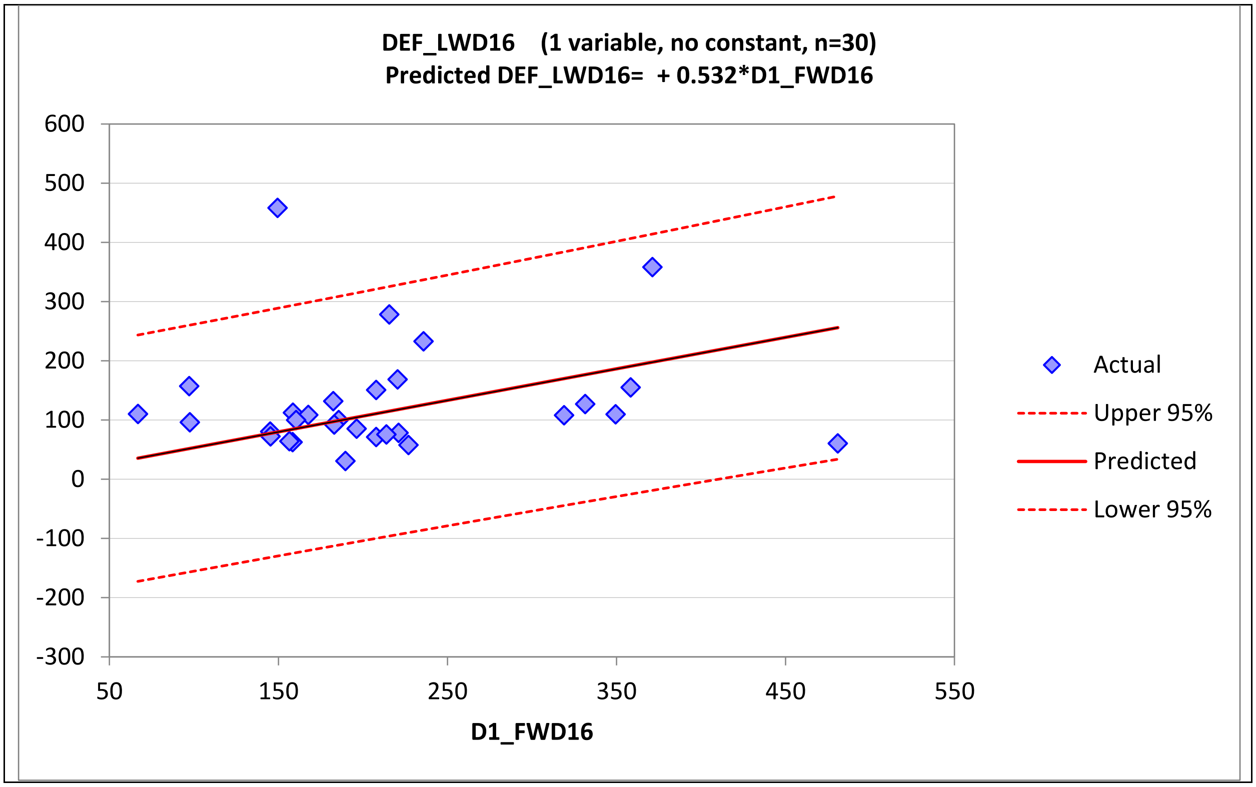


Figure 23: Regression Model for FWD and LWD (deflections)

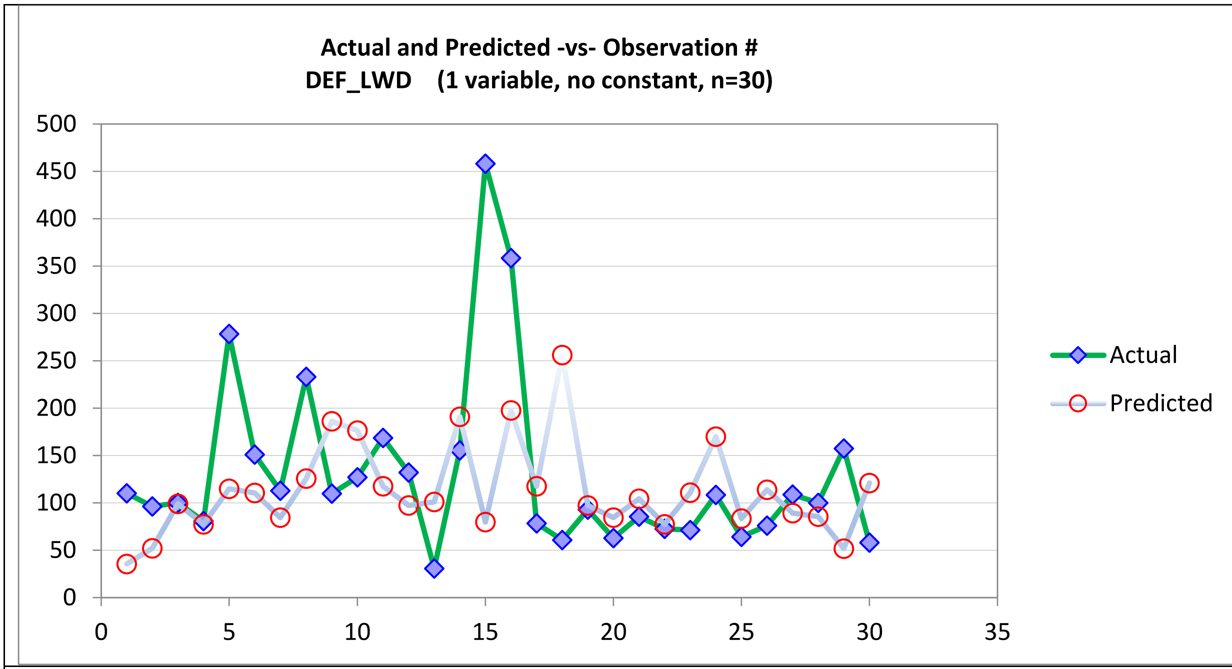


Figure 24: LWD (deflections) - Actual and Predicted

The deflection measured for both FWD16 and normalized LWD16 were used to calculate the surface modulus per the formula provided by Fleming (2007) and Mallick (2017) equation (Boussineq's theory) for elastic modulus as below:

$$\mathbf{E_{LWD}} = \frac{\mathbf{A.P.r.(1 - v^2)}}{\boldsymbol{D}} \qquad \textbf{Equation 2}$$

Where:

E_{LWD} = surface modulus (MPa)

A = shape (plate rigidity) factor, default = 2 for a flexible plate, π/2 for a rigid plate.

P = applied stress (maximum contact pressure) (kPa)

r = plate radius (mm)

v = Poisson's ratio (usually in the range of 0.15-0.25) (but is used as 0.2-0.21 for concrete (Mallick, 2017))

D = surface (peak) deflection (μm)

Table 28: Regression Model Data for FWD and LWD (using normalized deflection and Modulus data)

Station:FWD16 -EB	Deflection - D1 FWD16	Avg Def LWD Normalized (D1) (40kN)	Modulus FWD16	Modulus LWD Normalized
0+004	67	110	1908	1157
0+007	97	96	1308	1326
0+013	186	100	686	1275
0+021	145	81	879	1582
0+047	216	278	591	458
0+054	208	151	614	844
0+061	159	113	804	1131
0+070	236	233	541	547
0+090	349	110	365	1161
0+111	332	127	385	1004
0+128	221	169	578	756
0+135	182	132	699	965
0+141	190	31	672	4140
0+150	358	155	356	821
0+160	150	458	853	278
0+170	371	359	343	356
0+190	221	78	577	1627
0+211	481	61	265	2097
0+221	183	93	697	1374
0+230	158	63	805	2024
0+240	196	86	650	1491
0+269	145	73	878	1756
0+271	208	71	614	1790
0+280	319	108	400	1176
0+287	156	64	815	1986
0+293	214	76	596	1678
0+302	168	109	760	1172
0+310	161	100	794	1275
0+320	97	157	1313	810
0+328	227	58	562	2191

Source: FWD and LWD datasets (2016) for Jameston Road, Hamilton, Ontario, Canada

A regression model was developed based on the above data using Regressit (a MS Excel Ad-in). There are thirty data points (n=30). The results are presented below:

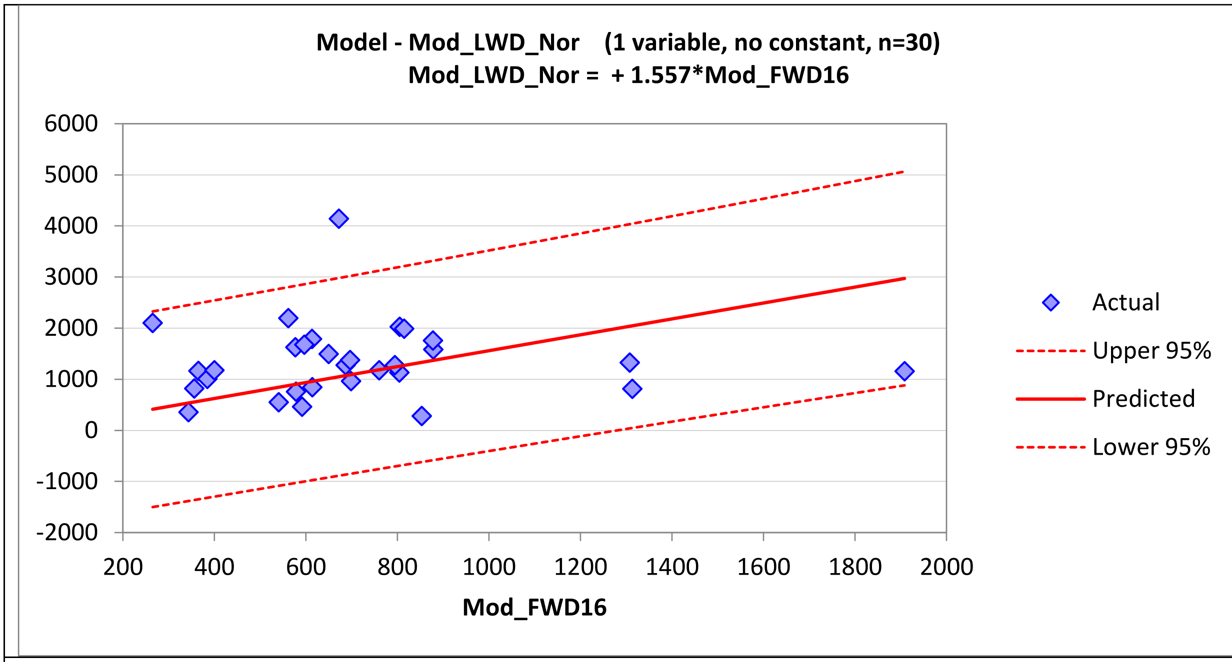


Figure 25: Regression Model for FWD and LWD (modulus)

Based on the regression analysis, Table 29, depicts the equation between Dynatest 3031 LWD and Dynatest FWD equipment(s) modulus of elasticity.

Table 29: Modulus Equation from Regression Analysis for LWD and FWD

Equation (Modulus)	**Layer Description**	**R-Squared (R^2)**	**LWD Model**
LWD(MPa) = 1.557FWD(MPa)	Concrete Overlay Pavement	0.64	DYNATEST 3031 LWD

These regression results are very close to a similar study conducted by (Steinert, 2006), where he has mentioned that LWD(MPa) = 1.33FWD (MPa) for thin asphalt layer.

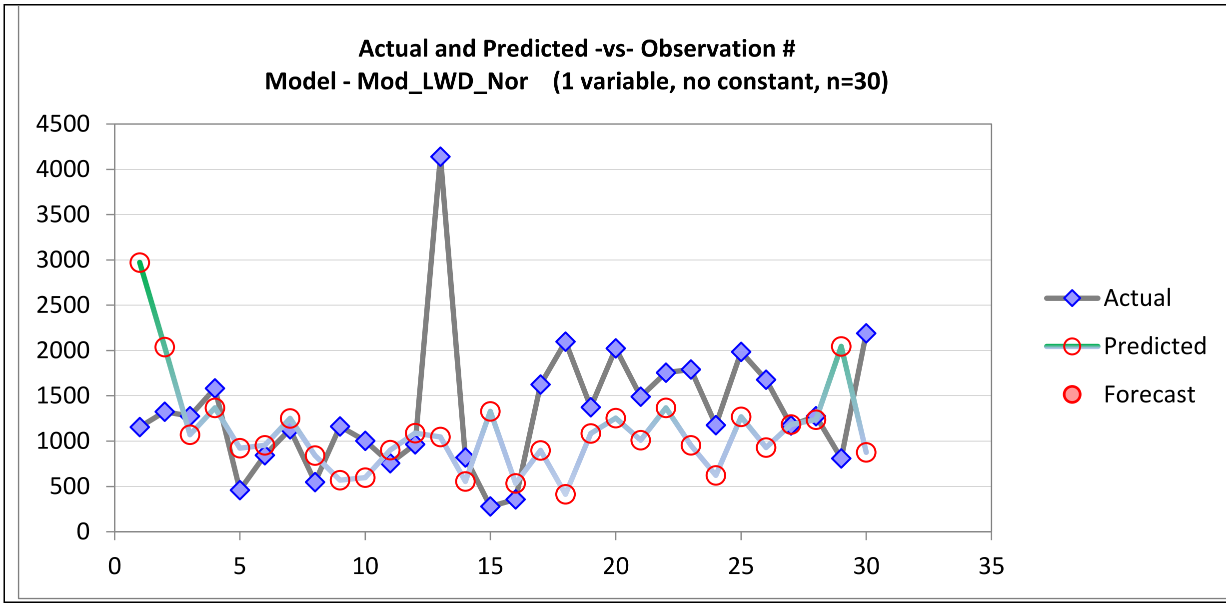


Figure 26: LWD (Modulus) - Actual and Predicted

Table 30: Regression Model (Critical Indicators of Correlation) LWD & FWD Modulus of Elasticity

R-Squared	Adj.R-Sqr.	t-Statistic	P-value
63.74%	62.50%	7.136	0.000

The results of regression analyses provided above confirms that data-permitting datasets can be utilized for creating correlation and basis for future analytics when properly standardized, stored in a meaningful manner in the CPATT database.

5.5 Chapter Summary

Discussion on current data collection and development of inter-relation models such as what are the salient requisites for selection of data sets for inter-relation model, etc.

Evaluation of potential of numerous datasets for selection to develop inter-relation model; such as FWD and LWD, etc.

Evaluation of data format(s) structure and requirements.

Identification of salient gaps from existing/past datasets.

Identification and evaluation of several different types of file formats.

Regression analysis model for BPN and T2GO, based on selection criterion, re-arrangement of data-permitting datasets for BPN and T2GO from 2016 and 2017.

Development of regression equation between BPN and T2GO as

$$\mathbf{BPN = 94.525 * T2GO}$$

Regression analysis model for deflections of FWD and LWD, based on selection criterion, re-arrangement of data-permitting datasets, normalization, modulus of elasticity for FWD and LWD datasets from 2016.

6. DATASETS QUALITY CHECKS

In this chapter the details of salient quality checks for Falling Weight Deflectometer are presented. It also represents the selection of requisite statistical quality checks, its application on sample datasets and interpretations based on the predefined criterion.

Later, in this chapter, Percent Within Limits in discussed in detail with the QC/QA acceptance criteria based on statistical methods and tests.

6.1 Existing Falling Weight Deflectometer (FWD) Data

In the past, there has been some Falling Weight Deflectometer (FWD) data acquired by CPATT team members for a concrete pavement research. From this research, two datasets provided, one for 2016 (before construction- dataset 1) and one for 2017 (one year after construction- dataset 2). These datasets were analyzed for the missing information and developing quality checks. The missing data mentioned here can affect the analysis and integrity of data.

The following errors are most commonly encountered during FWD data collection (Federal Highway Administration, 2006):

1. Roll-Off
2. Non-decreasing deflections
3. Overflow
4. Load variations
5. Deflection variation
6. Zero values

Among these six errors, there are four errors (Roll-off, Overflow, Load variations and deflection variations) which can be verified during the data collection process. The zero values and non-decreasing deflections can be checked both during and after data is acquired. The available FWD data was evaluated for these two errors. There were no values found to be zero in the data set 1 or 2, this error check has a status of “Pass”.

The other error that can be checked after data collection is non-decreasing deflections as offset from sensor at zero increases. Both datasets were evaluated for this error. Dataset 2 was found to be free from this error, however dataset 1 had several errors. There were total of 88 reading for each sensor from D1 to D9; 44 each for East bound and West bound direction. Among these 88, 32 data points failed for non-decreasing between D1 and D2. A list of these non-decreasing checks is provided in Table 31. These 32 data points makes 36% of total datapoints for D1.

These errors can occur if there is an existing discontinuity between two sensors such as a transverse crack (Federal Highway Administration, 2006), however it should be noted under the comments portion which was missing from the provided data and accompanied report. Non-decreasing deflections may occur occasionally between sensor 1 and sensor 2 on extremely weak pavement. However, the specific pavement is not considered to be extremely weak. Its average K_{static} value is at 28 MPa/m which is 13MPa/m more than the K_{static} value of 15MPa/m for extremely weak pavement (Wafa R. , 2018). This is an error which should have been mentioned in the data or accompanied report.

The overflow error ranges in this study are used for development of quality checks. These checks are based on the fact that outside a certain range, such as 2000 microns (.08 inch or 80 mils) is considered to be out of range for overflow values of a LTPP operated FWD (Federal Highway Administration, 2006).

Table 31: Data checks for FWD results for Jameston Avenue (non-decreasing deflections)

Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Fail	Fail	Pass	Pass	Pass	Pass	Pass	Fail	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass	Fail	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass	Fail	Pass	Pass	Pass	Pass	Pass	Pass
Fail	Pass	Pass	Pass	Pass	Pass	Pass	Fail	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass	Fail	Pass	Pass	Pass	Pass	Pass	Pass
Fail	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Fail	Pass	Pass	Pass	Pass	Pass	Pass	Fail	Pass	Pass	Pass	Pass	Pass	Pass
Fail	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Fail	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass	Fail	Pass	Pass	Pass	Pass	Pass	Pass

Fail	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass
Fail	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass
Fail	Pass	Pass	Pass	Pass	Pass	Pass
Fail	Pass	Pass	Pass	Pass	Pass	Pass
Fail	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Fail	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass

Fail	Pass	Pass	Pass	Pass	Pass	Pass
Fail	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass
Fail	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass
Fail	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass
Fail	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass
Fail	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass

Pass	Pass	Pass	Pass	Pass	Pass	Pass	Fail	Pass	Pass	Pass	Pass	Pass	Pass
Fail	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Fail	Pass	Pass	Pass	Pass	Pass	Pass	Fail	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass	Fail	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass	Fail	Pass	Pass	Pass	Pass	Pass	Pass
Fail	Pass	Pass	Pass	Pass	Pass	Pass	Fail	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
East bound data points							West bound data points						

The provided FWD data is missing the following information based on standard nine sensors FWD equipment:

- Not all Longitude/Latitude for data points are provided
- FWD equipment make, model and last calibration and verification (reports)
- Data gathering holes for subsurface temperature per Table 32
- Borehole for pavement structure layers
- Not all surface and air temperatures readings were provided
- Description of data fields was not elaborated
- Distress identification and commenting; other comments such as non-decreasing deflections, excess variations etc.

A list of non-critical missing data information is provided in appendix C.

Table 32: Temperature Boreholes Size

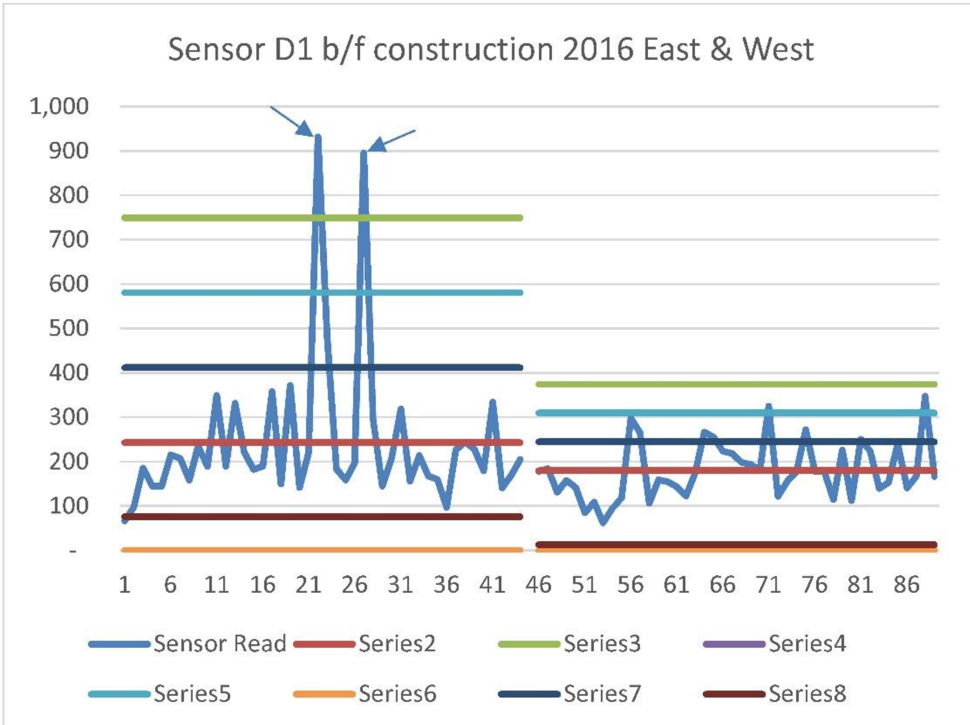
Hole Number	Hole Depth (mm)
1	25
2	50
3	100
4	200
5	300

Source: (Federal Highway Administration, 2006)

Statistical Process Control (SPC) is a methodology that have been widely used in both industrial and non-industrial professions to better understand the process variability (Noskievicova, 2013). These methods can be successfully used for validating quality of processes and datasets. A new data check was developed for the FWD datasets received. In development of this new quality checks for FWD data, the following were devised:

- All the data under one sensor is grouped into two datasets, one is East bound and the other is West bound

- Rational subgroup concept was used to develop subgroups. Rational subgroups are snap shot of a process (Mintab, 2018)
- Eighteen dataset subgroups were developed for all the sensors from D1 to D9
- The mean for each dataset was determined
- Upper Control Limits (UCL) and Lower Control Limits (LCL) are set at Mean + 3σ and Mean-3σ respectively. In some instances, the LCL falls below zero which are adjusted to zero.
- A mean X graph for each subgroup was developed and analyzed against the Nelson's nine rules of Statistical Process Control (SPC) Rules
- The SPC control charts from sensor D1 to D9 are depicted from Figures 27 to 31
- The data sets (1 & 2) pass all the tests, except rule number 1, the results and its interpretation and likely reasons are provided in Table 33
- The pavement related interpretations are provided at the end of this section for further elaboration
- The top most line (colored) in the graph reflects UCL (upper control limit) while the lowest line depicts the LCL (lower control limit)
- In some instances, the LCL was negative which was removed from the graphs



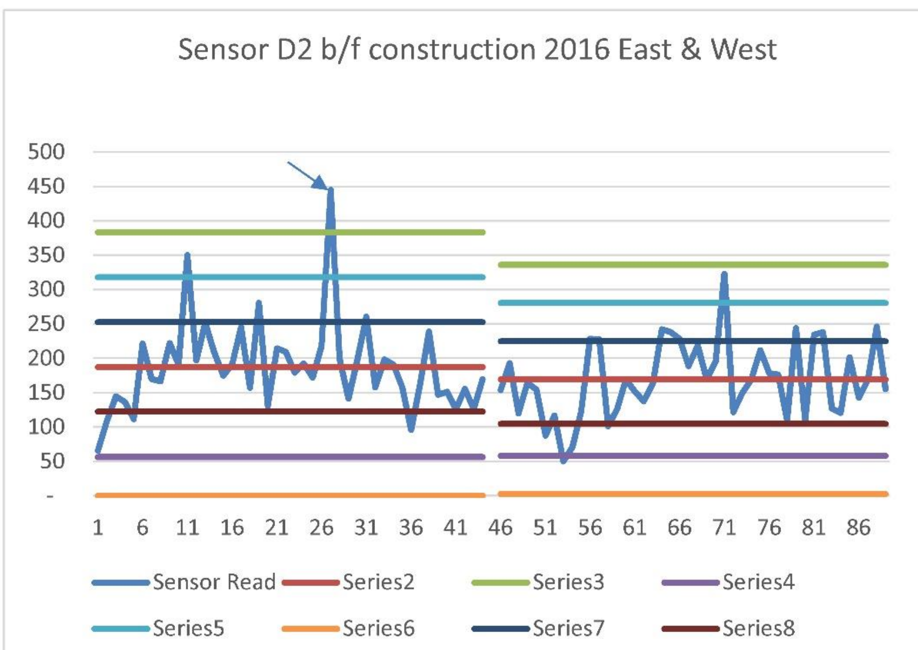
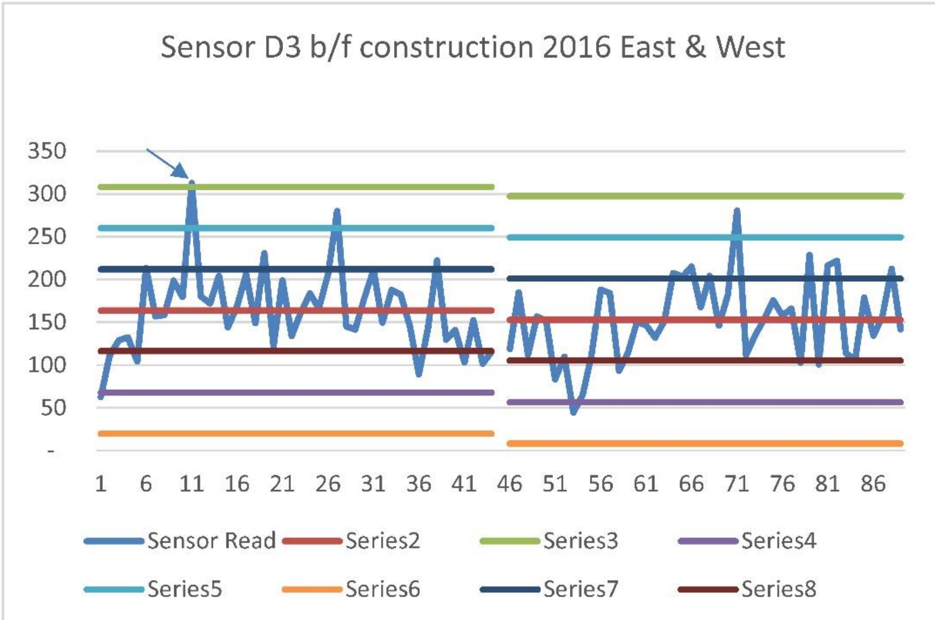


Figure 27: D1 and D2 Sensor readings SPC Chart



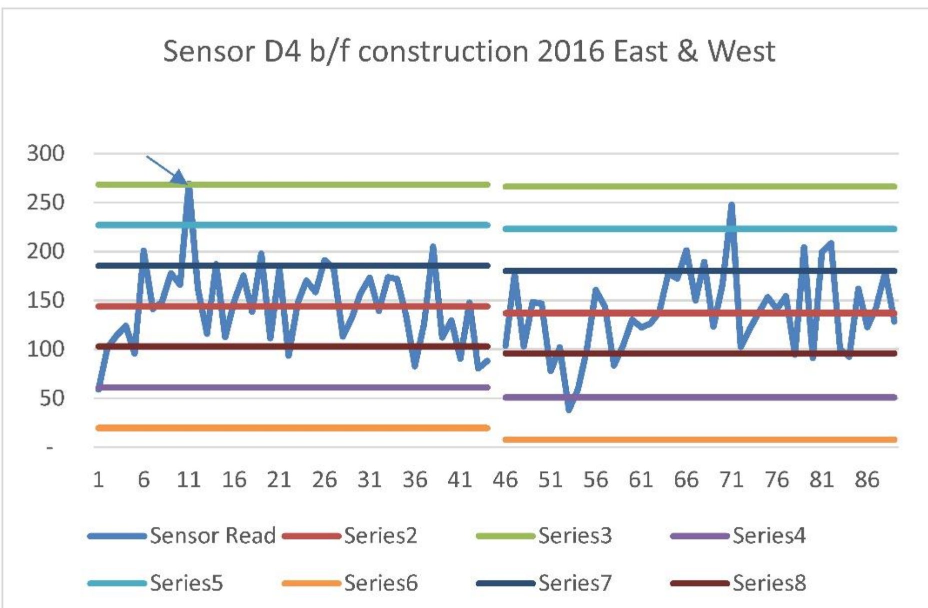


Figure 28: D3 and D4 Sensor readings SPC Chart

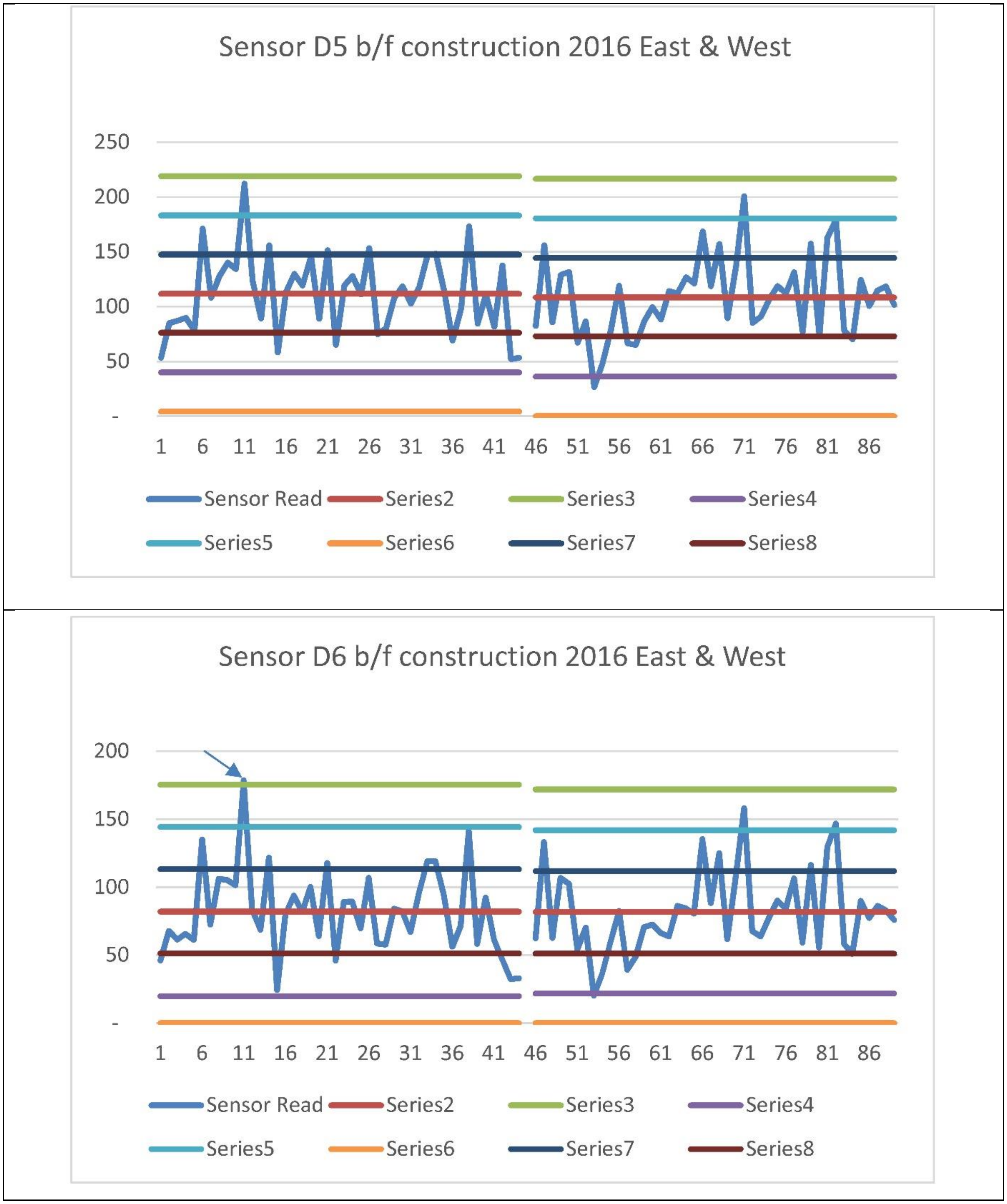


Figure 29: D5 and D6 Sensor readings SPC Chart

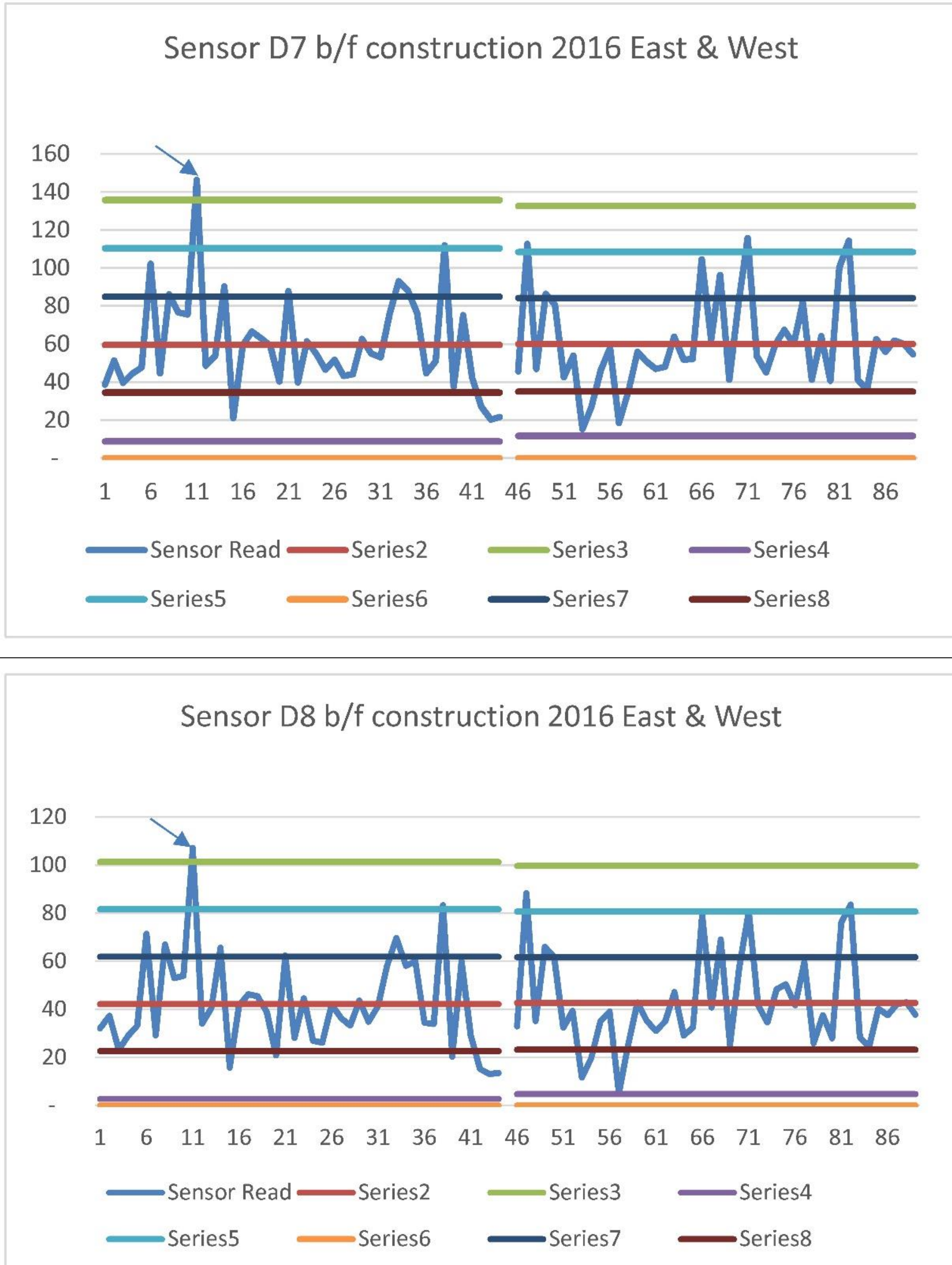


Figure 30: D7 and D8 Sensor readings SPC Chart

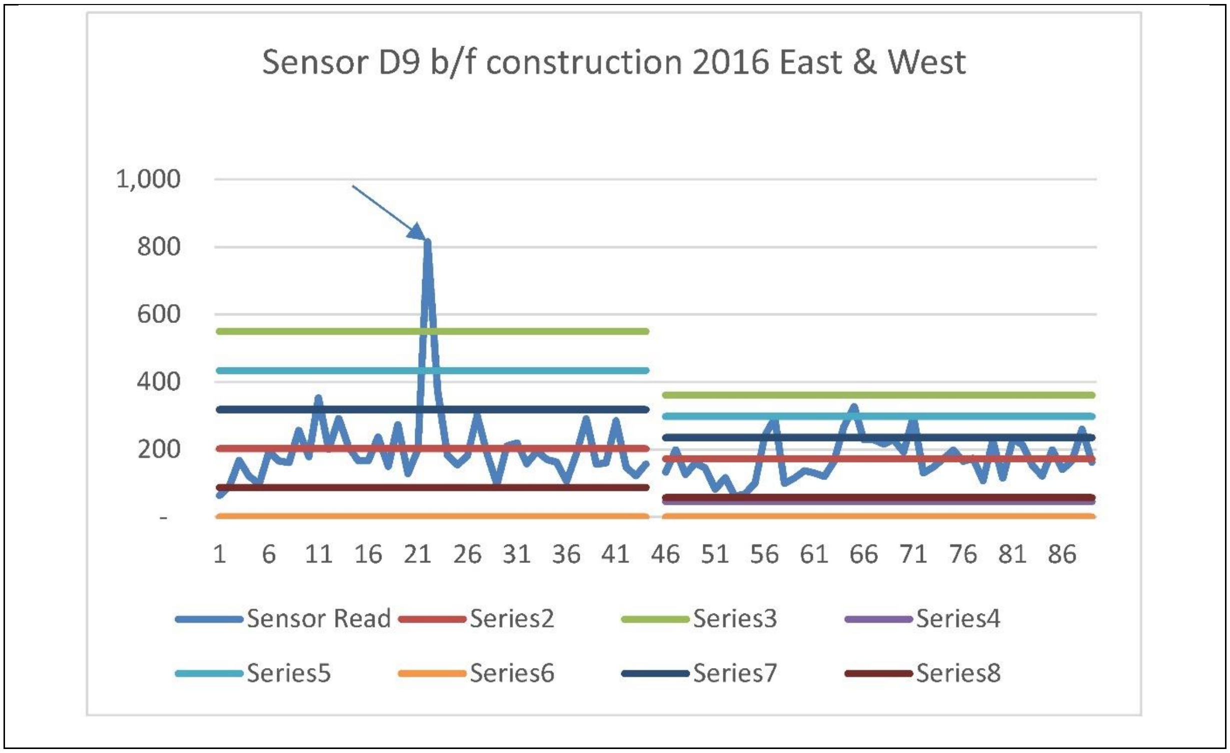


Figure 31: D9 Sensor readings SPC Chart

As seen in Table 33, there are nine outliers' points which indicates that this FWD process is statistically out of control. Although there could be three different reasons, based on the reasons (as mentioned in Table 33) closest possible reason could be "Error in measuring". Since this FWD testing is conducted on an over two decades old pavement, the following causes can be assigned for getting such reading in pavement terminology:

- For D1 and D9 sensors reading at station 0+200, there is a very high likely hood of presence of a longitudinal crack in the pavement
- For D1 and D2 sensors reading there is a possibility of small transverse or longitudinal crack at station 0+249
- At station 0+090, from D3 to D8, there is a strong possibility of existence of a longitudinal crack affecting the readings at these sensors

Table 33: Sensors, noted errors during SPC and its interpretations

Sensor	Direction	Station	Reason	Interpretation*	Possible reason*[1]
D1	E	0+200	Point above 3σ	Process grossly out of control	Wrong Setting, Error in measuring, Incomplete operation, Overcorrection
D1	E	0+249	Point above 3σ	Process grossly out of control	Wrong Setting, Error in measuring, Incomplete operation
D2	E	0+249	Point above 3σ	Process grossly out of control	Wrong Setting, Error in measuring, Incomplete operation
D3	E	0+090	Point above 3σ	Process grossly out of control	Wrong Setting, Error in measuring, Incomplete operation
D4	E	0+090	Point above 3σ	Process grossly out of control	Wrong Setting, Error in measuring, Incomplete operation
D6	E	0+090	Point above 3σ	Process grossly out of control	Wrong Setting, Error in measuring, Incomplete operation
D7	E	0+090	Point above 3σ	Process grossly out of control	Wrong Setting, Error in measuring, Incomplete operation
D8	E	0+090	Point above 3σ	Process grossly out of control	Wrong Setting, Error in measuring, Incomplete operation
D9	E	0+200	Point above 3σ	Process grossly out of control	Wrong Setting, Error in measuring, Incomplete operation

Source of * and *[1]: (Noskievicova, 2013)

Such outliers shall be appropriately reported by the FWD operator during testing and be a part of the data provided as comments. Including appropriate comments can elude the misrepresentation of data for analysis as well such outliers are recommended for elimination from the data to get better representation for the calculations of key factors such as Kstatic and Load Transfer Efficiency (LTE %).

Emphasis is to receive appropriate comments along with the data so that specific concerns can be evaluated and verified during the data quality checks.

6.2 Statistical Methods

Several Departments of Transportation (DOT) are using different methods to evaluate pavement condition data quality, such as the paired t-test, the Cohen's kappa statistic, and percent within limits. Both paired t-test and the Cohen's kappa statistic requires a network level or repetitive data collection. Such as "paired t-test is used to evaluate a sample of matched pairs or similar units, or one group of units that has been tested twice (e.g., comparing the ground truth to the pavement condition survey results, comparing two raters evaluating the same pavement segment)" (Linda M. Pierce, 2014). "Cohen's kappa statistic can be used to measure the level of agreement between raters. A score is calculated that quantifies how much consensus exists among the different raters, as well as the possibility of raters agreeing or disagreeing simply by chance" (Linda M. Pierce, 2014). Percent Within Limits (PWL) is utilized for evaluating construction material quality by several highway agencies in the USA. Ministry of Transportation (MTO) also utilize the PWL method to calculate Pay Factor (PF) for transportation projects.

6.3 Percent Within Limits (PWL)

PWL is defined as the "percentage of the lot falling above the lower specification limit (LSL), beneath the upper specification limit (USL), or between the USL and LSL" (American Association of State Highway and Transportation Officials (AASHTO)., 2011). PWL is estimate of the percentage of the population falling within the specification limits. The Pennsylvania DOT was among the prime agencies to utilize PWL concept for evaluating pavement condition data (Linda M. Pierce, 2014).

Colorado State Transportation Agency, USA, issued a detailed method for calculating PWL (Colorado Procedure 71-01 , 2017). The procedure determines Pay Factor (PF) for the construction of HMA pavement structures depending on the specification and may vary from agency to agency as, this method is used by different transportation agencies throughout the North America. The critical factor is the specifications of for the construction which is followed rigorously through Quality Control (QC) procedures during and after the construction. The Pay Factor is utilized to pay the contractor fairly through complicated calculations which usually ensure both Quality of Hot Mix Asphalt (HMA) and Voids Acceptance of Hot Mix Asphalt.

A detailed procedure of calculations, formulae and tables are provided in this standard practice (Colorado Procedure 71-01 , 2017); some of the salient formulae are as follows:

Table 34: Salient formulae for Percent Within Limits

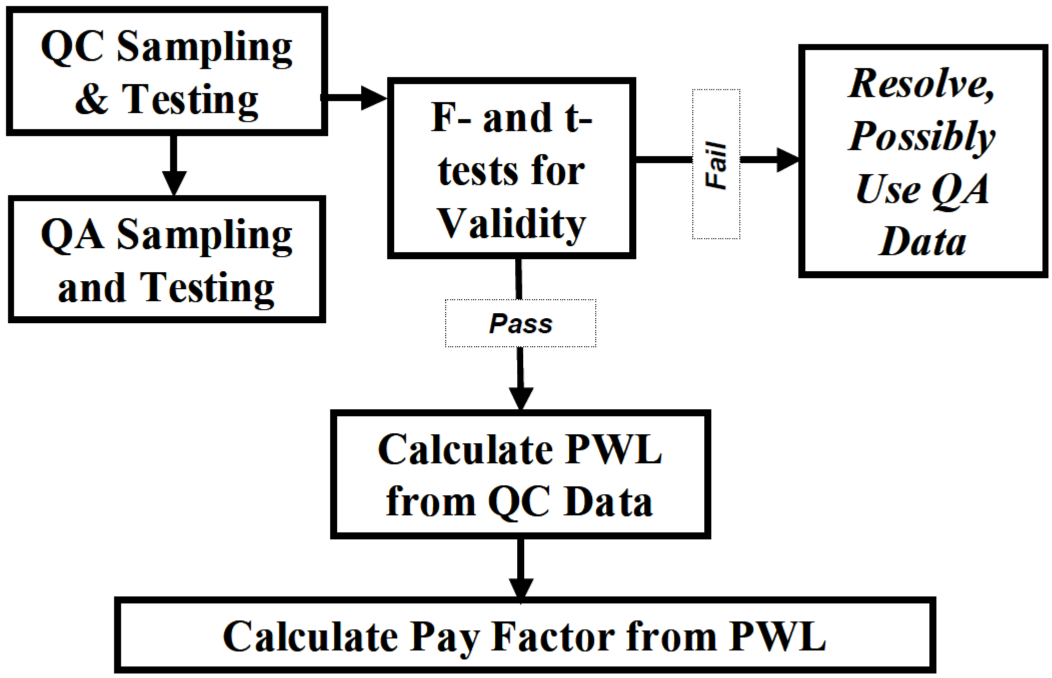
$$Q_U = \frac{T_U - \bar{X}}{s}$$	Where: Q_u = Upper quality index, T_u = Upper specification limits.	Equation A1
$$Q_L = \frac{\bar{X} - T_L}{s}$$	Where: Q_L = Lower quality index, T_L = Lower specification limits.	Equation A2
$$QL = (P_U + P_L) - 100$$	Where: QL = Quality Level Pu = Upper Specification Limit and, Pl = Lower Specification Limit	Equation A3
Source: (Colorado Procedure 71-01 , 2017)		

"The method involves calculating statistical parameters from three or more representative measurements, test results, or values for each specified element in a lot or sample. The arithmetic average (mean) value of the sample is calculated. As a measure of variability, the sample Standard Deviation is calculated. Using these results, the distance from the sample mean to each limit is divided by the standard deviation, which yields the Quality Index."

With those known, use Equations A1 and A2 to solve for the upper quality index (QU) and the lower quality index (QL). These values are then used to find percent within the upper specification limit (PU) and the percent within the lower specification limit (PL).

In a publication, (Transportation Research Board, 2006) has given the procedure of PWL, which is depicted as below:

Figure PWL1X: Macro view of typical statistically based QC/QA acceptance specification



Source: (Transportation Research Board, 2006)

Figure 32: Macro view of typical statistically based QC/QA acceptance specification

The following steps were followed to calculate the PWL from QC data:

1. Existing data sets of two British Pendulum Test conducted by (Wafa R. , 2018), in 2017 for determining BPN were utilized as seen in Table 35.
2. The F-Test was carried out for both the datasets.
3. After obtaining reasonable results from F-test, T-test was carried out.
4. Once it was determined per the flow chart (Figure 30) that the data sets meet the requisites, the PWL was calculated.

For this research, the data used is as given below:

Table 35: BPN 2017 data used for PWL process (Source: (Wafa R. , 2018)

BPN2017	
1 YR after Construction Aug 2017	
West	**East**
76	75
77	74
76	75
75	75
69	81
71	83
64	81
72	84
76	70
73	70
74	67
75	73
64	68
64	68
61	70
64	74
71	74
74	74
77	73
74	77
77	75
71	71
74	68
71	68
64	60
63	62
65	67
66	67
53	65
59	59
57	57
61	60
55	60
54	59
56	56

	61	56	

F-Test

As described by Minitab Blog Editor (Minitab Blog Editor, 2016); F-tests are known after its statistic test, F, which is named after Sir Ronald Fisher. The F-statistic is simply a ratio of two variances. Variances are a measure of dispersion, or how far the data are scattered from the mean. Larger values represent greater dispersion.

According to (Bradburn, 2021); statistical F-test is performed to determine if the variances of two populations are equal. F-test is important to know before carrying out an independent student t-test if the two group have equal variances.

Salient steps to obtain results for F-test (MS Excel) are as follows:

- Calculate the variance of each group using “VAR.S” command in MS Excel.
- For variable selection, it is recommended to select the variable with the highest variance value (calculated in step 1) so that MS Excel can calculate the correct f-value, which in this case for BPN2017 (East) units.
- Alpha is the significance threshold. Alpha is a threshold value used to judge whether a test statistic is statistically significant. Because alpha corresponds to a probability, it can range from 0 to 1. Per Minitab the significance level, also denoted as alpha or α, is the probability of rejecting the null hypothesis when it is true. Such as, a significance level of 0.05 indicates a 5% risk of concluding that a difference exists when there is no actual difference. So, if the p-value for the test is less than 0.05, we will conclude that the test is significantly significant.
- We assume the null hypothesis ($p>0.05$) that there is no difference the BPN2017 values measured between two populations (East and West). The alternative hypothesis ($p<0.05$)

is that BPN2017 (East) have a higher variance for values measured compared with BPN2017(West). As calculated the p value for our F-test is .44, which is greater than p=0.05; so, we fail to reject the null hypothesis and conclude that there is no difference between the BPN2017(East) values when compared to BPN2017(West). We will now perform T-test for the two sets of population.

(Top Tip Bio, 2021) specifies that if the results of F-test reflect that, p>0.05 we would proceed to perform T-test and assume that two populations have equal variance.

The results for F-test are depicted in Table 36:

Table 36: F-Test Results for PWL process

F-Test Two-Sample Results for Variances BPN 2017		
	East	*West*
Mean	69.33	67.61
Variance	58.80	55.73
Observations	36	36
df	35	35
F	1.06	
P(F<=f) one-tail	0.44	
F Critical one-tail	1.76	
T-Test Two-Sample Results for BPN 2017		
P value to T-test (two tail) = 0.338	T-test	2-Tail/Type 2

T-Test

The t test tells you how significant the differences between groups are; In other words it lets you know if those differences (measured in means) could have happened by chance (Statistics How To, 2021).

(Statistics How To, 2021) also add that the t score is a ratio between the difference between two groups and the difference within the groups. The larger the t score, the more difference there is between groups. The smaller the t score, the more similarity there is between groups. A t score of 3 means that the groups are three times as different *from* each other as they are within each other. When you run a t test, the bigger the t-value, the more likely it is that the results are repeatable.

- A large t-score tells you that the groups are different.
- A small t-score tells you that the groups are similar.

Every t-value has a p-value. A p-value is the probability that the results from the data occurred are coincidental. P-values are from 0% to 100%. They are usually written as a decimal, such as 0.05 for 5%. Low p-values are good; which indicates that the data did not occur by chance. Such as, .01 p-value would mean that there is a mere 1% probability that the results from the test happened by chance. (Statistics How To, 2021)

(Tip Top Bio, 2019), mentions that if the standard deviation for both the population is roughly the same, then we shall use type 2 for the t-test. As our datasets have a standard deviation of 7.46 for BPN2017 (West) and 7.66 for BPN2017 (East) with a difference of 0.2 which translates into around 2%; therefore, the type 2 was selected for performing t-test in MS Excel. A two-tailed is utilized as the most common option. (Minitab Blog Editor, 2015), mentions that if $p > 0.05$, the results are not significant.

Results of t test are given below in Table 37:

Table 37: T-Test Results for PWL process

T-Test Two-Sample Results for BPN 2017		
P value to T-test (two tail) = 0.338	T-test	2-Tail/Type 2

The salient statistical results for PWL process per (Colorado Procedure 71-01 , 2017) are given in Table 38:

Table 38: BPN2017 data - PWL (Statistical Results)

BPN2017 Data PWL Results		
Statistic	**West**	**East**
Variance	**55.73**	**58.8**
St. Dev	**7.47**	**7.67**
Mean	**67.61**	**69.33**
Maximum	**77**	**84**
Minimum	**53**	**56**
Qupper	**1.26**	**1.91**
Qlower	**1.96**	**1.74**
Quality Level (QL)	**87.42**	**94.75**

Quality Assurance type specifications where acceptance decisions are based on Quality Level (QL), defined as percent within specification (tolerance) limits. QL is a measure of quality of a lot or process. QL represents the percentage of the population (lot or process) that falls above a single lower limit, below a single upper limit, or between the upper and lower limits of double limit specifications. Since the Quality Level for the datasets for BPN2017 is calculated as 87.42 for West and 94.75 for East; both of which are well above the required minimum Quality Level of 80, we can accept these test results and make it and include it into the CPATT database.

6.4 Chapter Summary

Evaluation and analytics of existing Falling Weight Deflectometer (FWD) Data.

Detailed analysis of appropriate and applicable quality checks for existing FWD data.

Selection of requisite statistical quality data checks, its evaluation, analysis and interpretation based on established criterion.

Discussion, evaluation and calculations of percent within limits (PWL).

Development of PWL statistically based QC/QA acceptance criterion.

F-test and T-test requirements, evaluation and application in relation to PWL for BPN 2017 datasets.

7. CPATT DATABASE

7.1 Layout of CPATT Database

Figure 31, depicts a snap shot from the database created in UW repository.

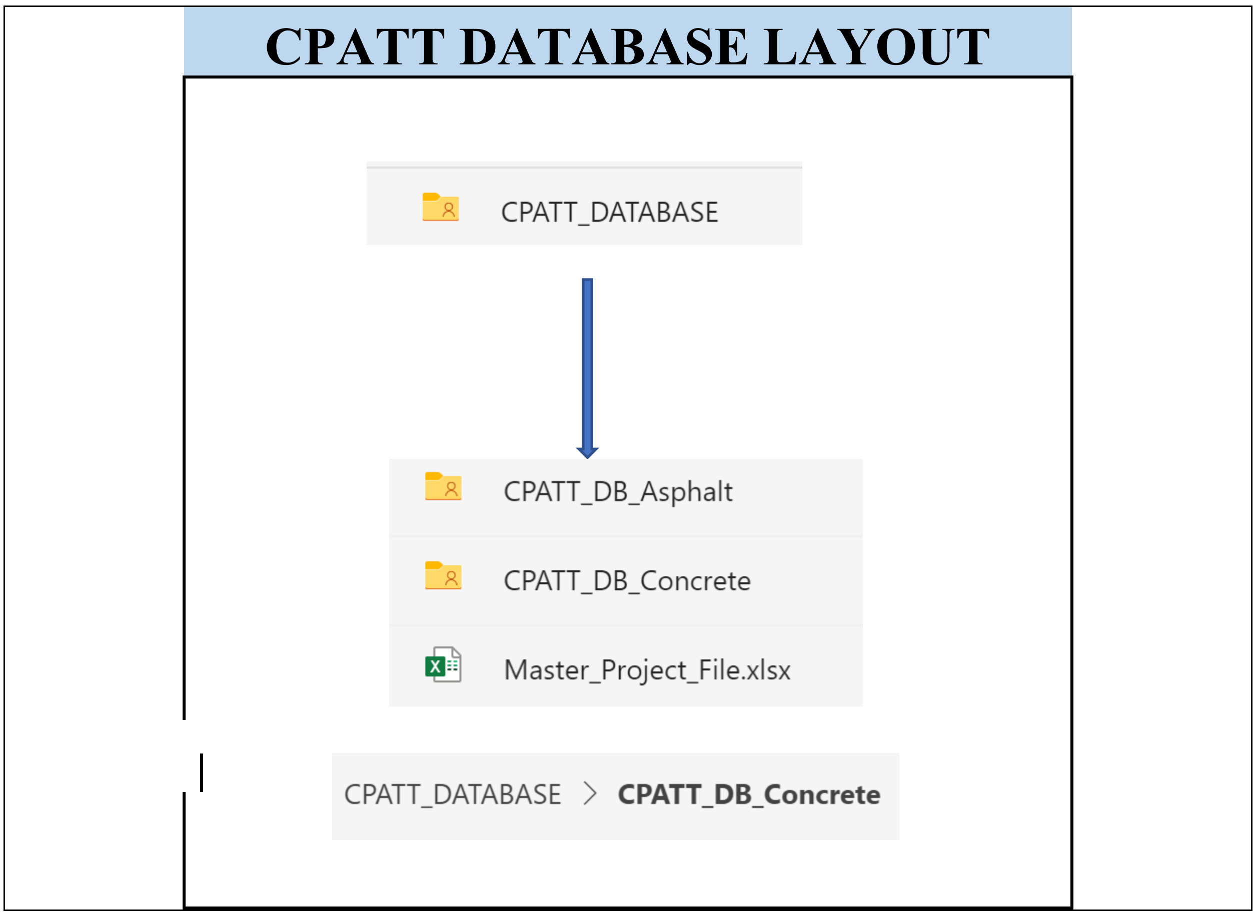


Figure 33: A screen shot from developed CPATT Database

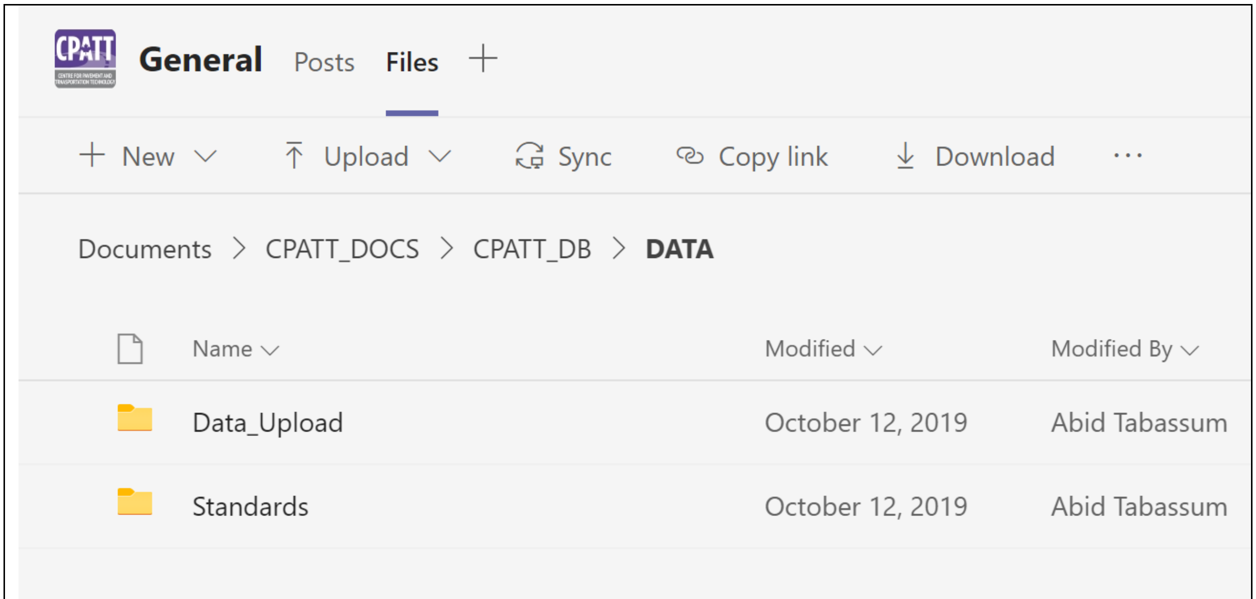


Figure 34: A screen shot from developed CPATT Database

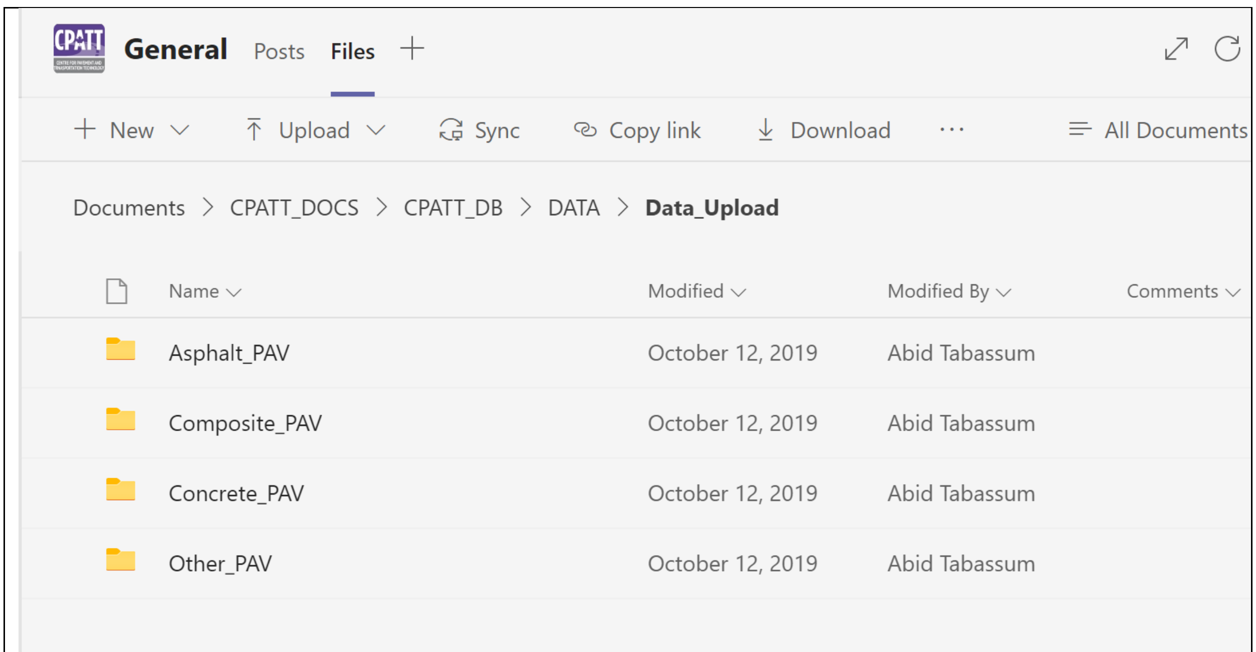


Figure 35: A screen shot from developed CPATT Database

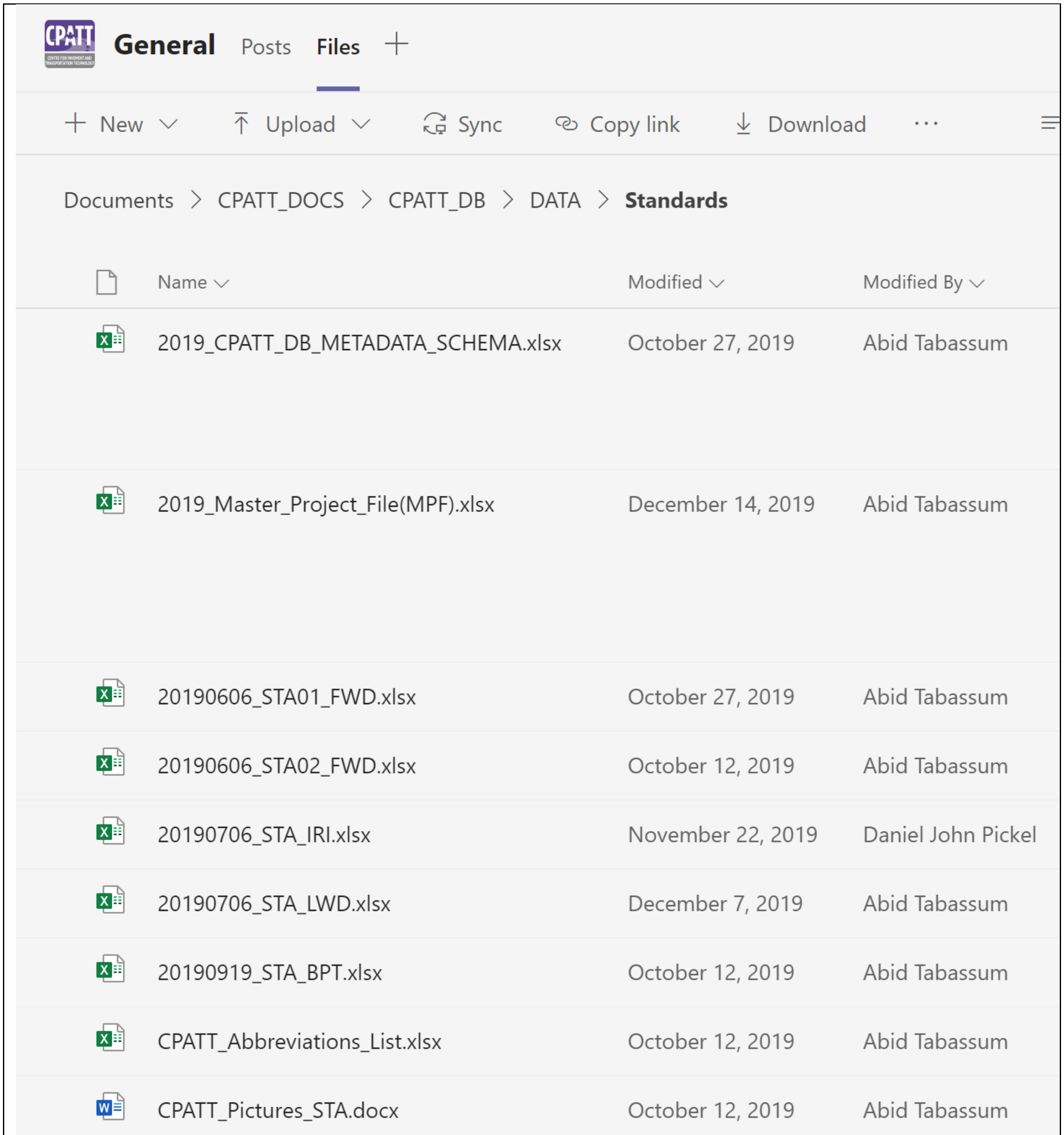


Figure 36: A screen shot from developed CPATT Database

7.2 Salient datasets and standards included

This includes the following:

1. Procedure of uploading the data to the repository,
2. Master Project File,
3. Standards for different test where possible, such as Falling Weight Deflectormeter (two standards FWD), Light Weight Deflectormeter (LWD), British Pendulum Tester (BPT or BPN), International Roughness Index (IRI), T2GO, etc.,
4. Abbreviations list for use with database,
5. Picture taking and standards for CPATT Database.

7.3 Chapter Summary

Development of CPATT Database layout based on selection criterion as explained in chapter 4.

Salient datasets and standards included in CPATT Database.

8. CONCLUSIONS AND RECOMMENDATIONS

8.1 Conclusion

As a result of this research the following salient goals were achieved;

1. Based on the discussions with CPATT management, University of Waterloo IT experts, and CPATT (end users), such as researcher(s), including current and past students, post-doctoral staff and staff; robust and meaningful CPATT database is developed.
2. The CAPTT database requirement(s) gathered through a survey and complied, forming basis for development of an end user friendly CPATT database.
3. Numerous standards developed for CPATT data-permitting datasets.
4. Numerous Quality Control checks developed for data-permitting datasets such as Falling Weight Deflectometer (FWD) and British Pendulum Tester (BPT or BPN).
5. Regression analysis carried out between British Pendulum Tester (BPT or BPN) and T2GO. Which resulted in several regression equations providing a relationship between BPN and T2GO.

Author of the research would like to acknowledge consistent guidance of Professor Susan L. Tighe, and contributions of Dr. Daniel Pickel, Kevin Ramparsad, Dr. Luke Zhao, Rahnuma Wafa, Beverly Seibel from the University of Waterloo as well Dr. Richard Korczak from Stantec consulting for their valuable inputs.

8.2 Recommendations

Since the developed database is for concrete pavement field data, the following research(s) are recommended for future CPATT researcher(s);

1. Develop CPATT database and standards for concrete pavement laboratory data
2. Develop CPATT database and standards for asphalt pavement field data
3. Develop CPATT database and standards for asphalt pavement laboratory data
4. An inter-relationship model can be developed between FWD, and strain gauges data, collected after construction in 2017 for Jameston Avenue, Ontario, Hamilton, as this data exist in the CPATT databases.
5. Other inter-relationship model(s) can be developed between regularly/continuously data gathered by CPATT team for concrete pavement field (for past few years), such as strain-gauges data collected on Spragus Road and Jameston Road. An illustration of the process of data interrelation models (for strain gauges) is provided earlier in the study.
6. The apposite process of uploading the datasets to CPATT and keeping it as per the requisite standards (outlined in this study) are extensive. This may cause for some of the researchers to elude some steps in uploading the data. It is recommended that the CPATT management requires every student/researcher to provide a consent from designated CPATT Database Administrator for graduation.

References

(n.d.).

Airport Surface Friction Tester (ASFT). (2017). T2GO Technical Specification. Köpingebro, Sweden: Airport Surface Friction Tester (ASFT).

American Association of State Highway and Transportation Officials (AASHTO). (2011). Standard Practice for Definition of Terms Related to Quality and Statistics as Used in Highway Construction. . Washington, DC, USA: AASHTO. Retrieved 2020

ASTM E303-93. (2013). *Standard Test Method for Measuring Surface Frictional Properties Using the British Pendulum Tester.* West Conshohocken: : ASTM International.

Bradburn, S. (2021, May). Retrieved 2021, from https://www.youtube.com/watch?v=1KbXEyoCxmQ

Canadian National Committee of the World Road Association in Partnership with Ontario Ministry of Transportation. (2015). *CANADA - NATIONAL REPORT*. Retrieved from TAC-ATC.CA: http://www.tac-atc.ca/sites/tac-atc.ca/files/site/doc/resources/piarc-report.pdf

Ceylan, H. &. (2016, July). Impact of Curling and Warping on Concrete (Program for Sustainable Pavement Engineering and Research, Institute for Transportation, Iowa State University). Ames, Iowa, USA.

Chad Schaeffer, A. H. (2016, April). Big Data Management in United States Hospitals: Benfits and Barriers. *In J. Sanchez (Ed.), Proceedings of the Business and Health Administration Association Annual Conference,.* Chicago, IL. Retrieved from http://mds.marshall.edu/cgi/viewcontent.cgi?article=1154&context=mgmt_faculty

Chojecki, P. (2019, January 31). How Big is Big Data? Retrieved January 5, 2021, from https://towardsdatascience.com/how-big-is-big-data-3fb14d5351ba

Colorado Procedure 71-01 . (2017). Standard Practice for Determining Quality Level (Percent Within Tolerance. Retrieved July 5, 2021, from https://www.codot.gov/business/designsupport/materials-and-geotechnical/manuals/2018-fmm/cps/CP-70s/15-cp-71-18

Corsello, P. (1993). *EVALUATION OF SURFACE FRICTION GUIDELINES FOR WASHINGTON STATE HIGHWAYS.* Seattle, Washinl'ton: Washington State Transportation Commission; Transit, Research, and lntermodal Planning (TRIP) Division.

CPATT Database - concrete overlays strain data. (2016-2019). Jameston Avenue. Hamilton, Ontario, Canada.

CPATT Database - concrete overlays; strain and temperature data. (2015-2019). Spragues Road. Region of Waterloo, Ontario, Canada.

CPATT Database - precast concrete inlays. (2013). MTO 400 series Highway. 400 sereis Highway, Ontario, Canada.

CPATT Database. (2016, August 9). CPATT (FWD and LWD) Data - 2016. Hamilton, Ontario, Canada.

CPATT Database, concrete overlay BPT & T2GO Data. (2016-2017). Jameston Avenue, Hamilton, Ontario, Canada.

Dell Technologies. (2021). Cloud Workloads. Retrieved February 16, 2021, from https://www.delltechnologies.com/en-us/learn/cloud/cloud-workloads.htm

Dynatest. (2006, June 13). Dynatest 3031 LWD. Retrieved March 3, 2019

e, M. B. (n.d.).

Federal Highway Administration . (2015, September 15). *The Long-Term Pavement Performance Program.* US Department of Transportation, Washington, DC. Retrieved November 17, 2017, from https://www.fhwa.dot.gov/publications/research/infrastructure/pavements/ltpp/15049/15049.pdf

Federal Highway Administration. (2006). *Long-Term Pavement Performance for Falling Weight Deflectometer Measurements -FHWA-HRT-06-132.* U.S. Department of Transporatation.

Federal Highway Administration. (2006, December). Long-Term Pavement Performance Program Manual for Falling Weight Defelectometer Measurements. *Publication No. FHWA-HRT-06-132*. McLean, VA 22101-2296, USA. Retrieved January 18, 2020

Federal Highway Administration. (2009). *Long-Term Pavement Performance Program - Accomplishments and Benefits (1989-2009).* Federal Highway Authority. Retrieved February 25, 2018, from https://www.fhwa.dot.gov/publications/research/infrastructure/pavements/ltpp/10072/index.cfm

Federal Highway Administration. (2013, May). Retrieved September 1, 2017, from https://www.fhwa.dot.gov/publications/focus/13may/may2013.pdf

Federal Highway Administration. (2016). *The Long-Term Pavement Performance Program (Summary of Key Benefits 1989-2015).* US Department of Transportation.

Finkel, B. (2019, January 3). MicroNugget: How to Normalize Databases. Retrieved March 8, 2021, from https://www.youtube.com/watch?v=upS2HlUj1gI

Fleming, P. F. (2000). A Comparison of Devices for Measuring Stiffness In-situ. *Unbound Aggregates in Road Construction, Proceedings of the Fifth International Symposium on Unbound Aggregates in Roads*, 193-200. Nottingham,, UK.

Fleming, P. F. (2007). Review of lightweight deflectometer for routine in situ assessment of pavement material stiffness. *Transportation Research Record : Journal of the Transportation Research Board, 2004*, 80-87.

Flintsch, G. M. (2009). *Quality Management of Pavement Condition Data Collection.* Washington, DC: Transportation Research Board.

Frost & Sullivan. (2012). *Drowning in Big Data? Reducing Information Technology Complexities and Costs For Healthcare Organizations*. Retrieved March 18, 2018, from https://www.emc.com/collateral/analyst-reports/frost-sullivan-reducing-information-technology-complexities-ar.pdf

Frost, J. (2017). Five Reasons Why Your R-squared can be Too High. Retrieved February 2, 2021, from https://statisticsbyjim.com/regression/r-squared-too-high/

Golubenko, S. (2019). *SharePoint MythBusters: Top 5 Misconceptions of the Platform*. Retrieved September 2019, from TopSharePoint: http://www.topsharepoint.com/sharepoint-mythbusters-top-5-misconceptions-of-the-platform

HingePoint. (2020). 4 Ways To Use SharePoint For Company Data Management. Retrieved January 19, 2021, from https://www.hingepoint.com/blog/sharepoint/4-ways-to-use-sharepoint-to-manage-company-data/#:~:text=SharePoint%20is%20an%20incredible%20data,instead%20of%20hunting%20for%20it.

Innovative Architects. (2019). Cloud Computing Solutions: Effective, Economical, Secure. Retrieved January 30, 2021, from https://www.innovativearchitects.com/SharePoint-Services/Cloud-Computing-Solutions.aspx

Iowa DOT . (2009). Retrieved October 10, 2017, from http://www.iowadot.gov/local_systems/mailing/2009/march/08_ipmp_update.pdf

Jee, K. K. (2013). Potentiality of Big Data in The Medical Sector: Focus on How to Reshape the Healthcare systems. *Healthcare Informatics Research*, 79-85.

Korczak, R. (2019, February). Calculation methodology of Modulus from FWD data. (A. Tabassum, Interviewer)

Linda M. Pierce, K. A. (2014, August 15). QUALITY MANAGEMENT FOR PAVEMENT CONDITION DATA COLLECTION. Santa Fe, NM 87508, USA.

Long Term Pavement Performance (LTPP). (n.d.). Falling Weight Deflectometer Data. Numerous Regions, Numerous States, USA. Retrieved 2018

Lund, B. (2021, March 2). R & D Engineer (Dynatest, Denmark). (A. Tabassum, Interviewer)

Mallick, R. B.-K. (2017). *Pavement Engineering Principles and Practice.* CRC Press Taylor & Francis Group.

Microsoft. (2019, August 20). *Microsoft Power BI Desktop*. Retrieved from www.microsoft.com: https://www.microsoft.com/en-us/download/details.aspx?id=45331

Microsoft Corporation. (2019, December 19). Apply many-many relationships in Power BI Desktop. Retrieved January 12, 2021, from https://docs.microsoft.com/en-us/power-bi/transform-model/desktop-many-to-many-relationships#:~:text=With%20the%20July%202018%20version,neither%20table%20contains%20unique%20values.

Microsoft Incorporation. (2019, September 1). *Database Design Basics*. Retrieved from https://support.office.com/en-us/article/database-design-basics-eb2159cf-1e30-401a-8084-bd4f9c9ca1f5

Microsoft Incorporation. (n.d.). *Document collaboration and co-authoring*. Retrieved August 12, 2019, from https://support.office.com/en-us/article/document-collaboration-and-co-authoring-ee1509b4-1f6e-401e-b04a-782d26f564a4

Minitab Blog Editor. (2013, July 13). Regression Analysis: How to Interpret the Constant (Y Intercept). Retrieved January 02, 2021, from https://blog.minitab.com/en/adventures-in-statistics-2/regression-analysis-how-to-interpret-the-constant-y-intercept

Minitab Blog Editor. (2015, December 3). What Can You Say When Your P-Value is Greater Than 0.05? Retrieved July 2021, from https://blog.minitab.com/en/understanding-statistics/what-can-you-say-when-your-p-value-is-greater-than-005

Minitab Blog Editor. (2016, May). Understanding analysis of variance ANOVA and the F-test. Retrieved May 2021, from https://blog.minitab.com/en/adventures-in-statistics-2/understanding-analysis-of-variance-anova-and-the-f-test

Mintab. (2018, August 26). *Working with subgroups in variables control charts*. Retrieved from Support.Minitab.com: https://support.minitab.com/en-us/minitab/18/help-and-how-to/quality-and-process-improvement/control-charts/supporting-topics/data/working-with-subgroups-in-variables-control-charts/

Moore, K. D. (2014). The Big Deal about Big Data. *Healthcare Financial*, 60-68.

Nahm, M. (2013). *Standardization to Enhance Data Sharing*. Retrieved from https://www.nap.edu/read/18267/chapter/6

Noah. (2017, January 24). *Excel vs Access: When it is time to switch?* Retrieved from https://www.arkware.com/2017/01/24/excel-vs-access-time-switch/

Noskievicova, D. (2013). Complex Control Chart Interpretation. International Journal of Engineering Business Management.

Oliveira, I. (2019, April 1). *SharePoint vs. Teams*. Retrieved from https://www.cwps.com/blog/sharepoint-vs-teams

Olson, S. a. (2013). Sharing Clinical Research Data: Workshop Summary. *Washington, D.C.* Retrieved from https://www.nap.edu/read/18267/chapter/6

Pickel, D. (2018-2021). CPATT Research Associate. (A. Tabassum, Interviewer)

Pickel, D. (2019, February). CPATT current data collection for concrete field and laboratory data. Waterloo, Ontario, Canada.

Raghupathi, W. &. (2014). Big Data Analytics in Healthcare: Promise and Potential. *Health Information Science and Systems*, 3.

SharePoint. (2019, August 28). Microsoft Teams vs. SharePoint. Retrieved January 30, 2021, from https://sharepointmaven.com/microsoft-teams-vs-sharepoint/

Sherweb.com. (2020, March 10). 7 things you should know about Microsoft Teams. Retrieved February 17, 2021, from https://www.sherweb.com/blog/office-365/microsoft-teams/

Statistics How To. (2021). T Test (Student's T-Test): Definition and Examples . Retrieved June 2021, from https://www.statisticshowto.com/probability-and-statistics/t-test/

Steinert, B. H. (2006). Portable Falling Weight Deflectometers for Tracking Seasonal Stiffness Variations in Asphalt Surfaced Roads. *85th Transportation Research Board annual meeting, Compendium of papers CD-ROM,*. Washington, D.C.

Svitla Team. (2019, AUGUST 16). 18 Best Data Management Tools. Retrieved March 20, 2020, from https://svitla.com/blog/18-best-data-management-tools#:~:text=Data%20Management%20is%20as%20successful,management%20systems%20that%20harmonize%20data.

Tehseen. (2016). Big Data (3V's). Retrieved February 13, 202221, from http://tehseen.dbsdataprojects.com/wp-content/uploads/sites/107/2016/04/3vs1.jpg

Thakur, S. (2016, March 16). *Characteristics of Database Management System*. Retrieved May 15, 2019, from Whatisdbms.com: https://whatisdbms.com/characteristics-of-database-management-system/

Thakur, S. (2016, March 14). *Features of Database Management System (DBMS)*. Retrieved April 20, 2019, from Whatisdbms.com: https://whatisdbms.com/features-of-database-management-system-dbms/

Tip Top Bio. (2019, January 19). How To Perform T-Tests In Microsoft Excel. Retrieved June 2021, from https://www.youtube.com/watch?v=q0ckcKsSPXU

Top Tip Bio. (2021, May). Perform A Two-Sample F Test In Excel. Retrieved June 2021, from https://www.youtube.com/watch?v=1KbXEyoCxmQ

Transportation Association of Canada. (2014). *Primer - Pavement Asset Design and Management.* Transportation Association of Canada.

Transportation Research Board. (2006, September). Factors Affecting Compaction of Asphalt Pavements. Retrieved June 2021, from http://onlinepubs.trb.org/onlinepubs/circulars/ec105.pdf

Unify Square. (2020). Microsoft Teams vs SharePoint: Dueling Devils or a Match Made in the Cloud? Retrieved February 17, 2021, from https://www.unifysquare.com/blog/microsoft-teams-vs-sharepoint/

VeeAM. (2020, January). 2020 Data Protection Trends. Retrieved January 9, 2021, from https://static.cbsileads.com/direct/whitepapers/wp_data_protection_trends.pdf

Veritas.com. (2019). The State of Data Protection and Cloud. Retrieved January 10, 2021, from https://static.cbsileads.com/direct/whitepapers/V0856_SB_State-of-DP-and-

Wafa, R. (2018). *Evaluating Unbonded Concrete Overlay for Usage on Ontario Residential Streets*. Waterloo, Canada.

Wafa, R. (2019, February). LWD and FWD data collection. (A. Tabassum, Interviewer) Toronto, Ontario, Canada.

Watson, H. (2014). Tutorial: Big Data Analytics: Concepts, Technologies, and Applications. *Communications of the Association for Information Systems, 34*.

Watt, A. (2018). *opentextbc.ca*. Retrieved June 25, 2019, from Chapter 3 Characteristics and Benefits of a Database: https://opentextbc.ca/dbdesign01/chapter/chapter-3-characteristics-and-benefits-of-a-database/

Whatisdbms.com. (2107, 17 July). *http://whatisdbms.com/what-is-a-database/*. Retrieved July 19, 2019, from What is a database: http://whatisdbms.com/what-is-a-database/

Wilson, C. (2020, January 24). The Difference Between Real-Time, Near Real-Time, and Batch Processing in Big Data. Retrieved January 10, 2021, from https://www.precisely.com/blog/big-data/difference-between-real-time-near-real-time-batch-processing-big-data

Appendices

APPENDIX 1 (CPATT SCHEMA DATA)

CENTRE FOR PAVEMENT AND TRANSPORTATION TECHNOLOGY

UNIVERSITY OF WATERLOO FACULTY OF ENGINEERING

CPATT Standard Metadata

Sr	All fields with an asterisk are required	
1	Data Title:* (e.g. IRI, FWD, LWD etc.)	
2	Creator/Author:* (e.g. data collector etc.)	
3	Creator/Author Affiliations:* (e.g. CPATT, external vendor, consultant, etc.)	
4	CPATT_Data_Code [Obtained from Master_Project_File (MPF)]*	
5	Pavement Type*	Concrete Pavement
6	Type of Data:*	
7	Data File Title:* (File Name as per CPATT naming convention)	
8	Description of Data*^ **(Some equipment collects numerous types of data, e.g. pressure, moisture, temperature etc.)**	
9	Data set File format:*	
10	Author's Contact Name* (for questions about using the dataset)	
11	Author's Contact Email:*	
12	Data Collection Period (s) Start:*	
13	Data Collection Period (s) End:*	
14	Contact Instructions: (e.g. for data questions the authorized person's response time etc.)	
15	Keywords:*	
16	Geolocation Place:@	
17	Geolocation Point (Latitude Longitude):*	
18	Visible to (Who is allowed to download it):*	
19	Funder:	
20	Funding Award Number:	
21	Funding Award Title:	
22	Contributors (E.g. Name data collector, supervisor, researcher, analyst):	
23	ORCID: Register at https://orcid.org)	
24	File Size: (units)	
	Notes:	
	* = Mandatory fields	
	^ = please provide description of data such as deflection, pressure, temperature, etc..	
	@ = optional as it is a paid service	
	Dublin Core and DataCite Metadata Schema used in FRDR	

APPENDIX 2 - MASTER PROJECT FILE (MPF)

MASTER PROJECT FILE (MPF)

CPATT_Data_Code (MP	Project Name	City, Province, Country	Pavement_OR_Asset_Type	Data_Type (Dropdown lis	YYYYMMDD Date_Star	YYYYMMDD Date_End	Equipment_Type (Dropdown list	Investigator (First and Last Nar	Supervisor	Comments if any
B5XZ	Jameston Road	Hamilton, Ontario, Canada	Concrete Pavement	Friction	20160816	20170816	BritishPT	Rahanuma V	Susan L. Tighe	
7NG2	PCIP, King City	Ontario, Canada	Concrete Pavement	Pressure	20161003	20170803	Pressure Guage	Daniel Pickel	Susan L. Tighe	
1EK2			Concrete Pavement	Friction			BritishPT			
OS22			Concrete Pavement	Friction			BritishPT			
LJ2O			Concrete Pavement	Friction			BritishPT			
OI1K			Concrete Pavement	Friction			BritishPT			
PP9D			Concrete Pavement	Friction			BritishPT			
GUSZ			Concrete Pavement	Friction			BritishPT			
JN6R			Concrete Pavement	Friction			BritishPT			
EFP2			Concrete Pavement	Friction			BritishPT			
3ZCK			Concrete Pavement	Friction			BritishPT			
NNRT			Concrete Pavement	Friction			BritishPT			
W321			Concrete Pavement	Friction			BritishPT			
7S0A			Concrete Pavement	Friction			BritishPT			
LHRA			Concrete Pavement	Friction			BritishPT			
KYYZ			Concrete Pavement	Friction			BritishPT			
U3L5			Concrete Pavement	Friction			BritishPT			
BJ20			Concrete Pavement	Friction			BritishPT			

APPENDIX 3 - (FALLING WEIGHT DEFLECTOMETER (FWD) STANDARD 1)

20210606_STA01_FW
D.xlsx

APPENDIX 4 - (FALLING WEIGHT DEFLECTOMETER (FWD) STANDARD 2)

20210606_STA02_FW
D.xlsx

APPENDIX 5 - (BRITISH PENDULUM TESTER STANDARD - BPN)

20210619_STA_BPT.xl
sx

APPENDIX 6 - (IRI STANDARD)

20210706_STA_IRI.xls
x

APPENDIX 7 - (LIGHT WEIGHT DEFLECTOMETER (LWD) STANDARD)

20210706_STA_LWD.
xlsx

APPENDIX 8 - CPATT Database uploading process VERSION 02

CPATT Database
uploading process_v0

APPENDIX 9 – CPATT Abbreviations List

CPATT CENTRE FOR PAVEMENT AND TRANSPORTATION TECHNOLOGY

UNIVERSITY OF WATERLOO FACULTY OF ENGINEERING

Sr. No.	Abbreviation	Description	Source	Added By	Date Added (YYYYMMDD)
1	FWD	Falling Weight Deflectometer		Abid Tabassum	2020120
2	LWD	Light Weight Deflectometer		Abid Tabassum	2020120
3	IRI	International Roughness Index		Abid Tabassum	2020120
4	STR	Strain Gauge		Abid Tabassum	2020120
5	BPN	British Pendulum Number		Abid Tabassum	2020120
6	TEM	Temperature Gauge		Abid Tabassum	2020120
7	PRE	Pressure Gauge		Abid Tabassum	2020120
8	MOI	Moisture Gauge		Abid Tabassum	2020120
9	ARA	ARAN		Abid Tabassum	2020120
10	BOR	Boreholes Data		Abid Tabassum	2020120
11	BPT	British Pendulum Test or Tester		Abid Tabassum	2020120
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					

APPENDIX 10 – (CPATT PICTURE STANDRAD)

MPF#:	Date and time:
Picture/Photo taker (Name/designation/affiliation)	
Location of photo (include here street/HWY; type and condition of pavement; lane and direction – if applicable)	
Description of photo (general description for better understanding of the picture information such as any specific reason for taking the picture etc.)	
No more than 100 photos in one MS Word file.	