Moving towards data-driven decision-making in maintenance

Manik Madhikermi

Moving towards data-driven decision-making in maintenance

Manik Madhikermi

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Science, at a public examination held at the lecture hall TU2 of the school on 10 June 2019 at 12 noon.

**Aalto University
School of Science
Department of Computer Science
Adaptive Systems of Intelligent Agents**

Supervising professor

Professor Kary Främling

Thesis advisor

Dr Sylvain Kubler

Preliminary examiners

Professor Dante Augusto Couto Barone, Institute of Informatics, Federal University of Rio Grande do Sul Porto Alegre, Brazil

Professor Samir Lamouri, ENSAM, Arts et métiers ParisTech. 151, boulevard de l'Hôpital - 75013 Paris, France

Opponents

Professor Dante Augusto Couto Barone, Institute of Informatics, Federal University of Rio Grande do Sul Porto Alegre, Brazil

Professor Samir Lamouri, ENSAM, Arts et métiers ParisTech. 151, boulevard de l'Hôpital - 75013 Paris, France

Aalto University publication series

DOCTORAL DISSERTATIONS 104/2019

© 2019 Manik Madhikermi

ISBN 978-952-60-8581-4 (printed)

ISBN 978-952-60-8582-1 (pdf)

ISSN 1799-4934 (printed)

ISSN 1799-4942 (pdf)

<http://urn.fi/URN:ISBN:978-952-60-8582-1>

Unigrafia Oy

Helsinki 2019

Finland

Publication orders (printed book):

manik.madhikermi@aalto.fi

Author

Manik Madhikermi

Name of the doctoral dissertation

Moving towards data-driven decision-making in maintenance

Publisher School of Science

Unit Department of Computer Science

Series Aalto University publication series DOCTORAL DISSERTATIONS 104/2019

Field of research Information Technology

Manuscript submitted 8 May 2019

Date of the defence 10 June 2019

Permission for public defence granted (date) 29 April 2019

Language English

☐ **Monograph**

☒ **Article dissertation**

☐ **Essay dissertation**

Abstract

Traditionally, in the maintenance industry, maintenance efficiency is limited by the capability of the experts making the decision. However, the advancement of digital technologies made it possible to improve the effectiveness and efficiency of maintenance activities by adding insight from the data to expert assessment. The opportunity provided by data for decision making made the companies to shift towards a new type of maintenance strategy called data-driven maintenance. Despite of opportunities, data and analytical tools' companies are still struggling to fully harness data asset to improve maintenance activities because of data-centric challenges. Hence, the main objective of this dissertation is to identify and mitigate those challenges that limit organizational decision-making capabilities to improve maintenance effectiveness.

In this dissertation, firstly, quantitative and descriptive analyses of case studies in Finnish Multinational Manufacturing Companies have been carried out to identify key data-centric challenges. The study identified **Data Quality, Interoperability, and Data extraction** as key challenges. Furthermore, each of the identified challenges have been investigated through one or more original publications. The main results achieved in this dissertation are methods and frameworks to i) assess and compare data quality of maintenance reporting procedure ii) two-level interoperability framework for inter-system interoperability iii) data discovery methodology to extract data for Extract, Transform and Load process.

The applicability and validity of each of the proposed methodologies and framework has been validated through one or multiple use cases. For validation, three different tools namely, MRQA Dashboard, Open-messaging Middleware, and Data Model Logger have been developed to tackle each of the identified data-centric challenges.

Keywords Data-driven Decision-making, IoT, Maintenance, Data Quality, Interoperability, Data Extraction

ISBN (printed) 978-952-60-8581-4

ISBN (pdf) 978-952-60-8582-1

ISSN (printed) 1799-4934

ISSN (pdf) 1799-4942

Location of publisher Helsinki

Location of printing Helsinki **Year** 2019

Pages 186

urn <http://urn.fi/URN:ISBN:978-952-60-8582-1>

Preface

The research work presented in this thesis was carried out within Adaptive Systems of Intelligent Agents research group of Department of Computer Science, Aalto University. First and foremost, I would like to express my sincere gratitude to my supervisor Professor Kary Främling for the continuous support to my Ph.D research, for his patience, motivation, and immense knowledge. I am grateful for the opportunity he provided me to join the team as Master Thesis student which later extended to this research work. Without his support and guidance, it would not have been possible to carry out this research.

Besides my supervisor, I would like to thank my Ph.D advisor Dr. Sylvain Kubler for his active role and insight. In addition, I would like to extend my gratitude to Professor Mario Di Francesco for his guidance which has been extremely valuable throughout my Ph.D research.

Furthermore, I want to thank my fellow labmates for the stimulating discussions, workshops, code day, and for all the fun we have had during my research period. Special mention to Mr. Andrea Buda and Dr. Avleen Kaur Malhi who had helped me in various stages of my research.

I would also like to thank all of my friends here in Finland. They have all supported me in one way or other. I would like to have special mention of Dr. Prajwal Osti and Dr. Sandeep Tamrakar for their time, support and help.

Last but not the least, I would like to thank my family, specially to my mother, for supporting me spiritually throughout my Ph.D and my life in general. I thank her for always being on my side through thick and thin. I

also want to thank my beloved furry friend Bantu, who has passed away,
for wonderful memories.

Espoo, May 15, 2019,

Manik Madhikermi

Contents

Preface	1
Contents	3
List of Publications	7
Author's Contribution	9
1. Introduction	15
1.1 Overview	15
1.2 Research design	16
1.3 Research methodology	22
1.3.1 Literature review	22
1.3.2 Comparative study	23
1.3.3 Case study	23
1.3.4 Design science	24
1.4 The structure of the thesis	24
2. Data-driven decision making in maintenance management	25
2.1 Maintenance	25
2.2 Maintenance management	27
2.3 Data-driven decision making in maintenance	27
2.4 Information technology systems for maintenance management	28
2.5 Data-centric challenges towards data-driven decision making	29
2.5.1 Data Quality	30
2.5.2 System interoperability	31
2.5.3 Data extraction	32
2.6 Case study	33

3. Data quality assessment of maintenance reporting procedures	37
3.1 Data quality in the maintenance management system . .	37
3.2 AHP-based MRQA framework	39
3.2.1 Criteria and sub-criteria identification based on Krogstie data quality framework	42
3.2.2 MRQA framework steps	43
3.3 Results and discussion	45
4. Data exchange interoperability with an open messaging standard	49
4.1 Interoperability challenge in maintenance management .	49
4.2 System interoperability	50
4.3 Data exchange interoperability challenge	52
4.3.1 Communication level interoperability	52
4.3.2 Data level interoperability	53
4.4 Building Lifecycle Management (BLM)	54
4.5 Messaging standards for data exchange interoperability for BLM	55
4.6 Open messaging standards	58
4.7 Open messaging standard for data-driven maintenance .	59
4.7.1 Use case: Remote monitoring of air handling unit	59
4.7.2 Use case scenario	62
4.8 Results and discussion	66
5. Data extraction methodology	69
5.1 Data extraction in Extract, Transform and Load	69
5.2 Proposed data discovery methodology for data extraction	71
5.2.1 Step 1: Deploy database trigger	72
5.2.2 Step 2: Record the changes	72
5.2.3 Step 3: Data discovery for data extraction	72
5.3 Results and discussion	73
6. Conclusion	81
6.1 Summary	81
6.2 Contribution	82
6.3 Limitation and avenue for future research	85
References	87

Errata	101
Publications	103

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

- I Manik Madhikermi**, Andrea Buda, Bhargav Dave, Kary Främling. Key Data Quality Pitfalls for Condition Based Maintenance. In *2017 2nd International Conference on System Reliability and Safety (ICSRS)*, Milan, Italy, pp. 474-480, December 2017.
- II Manik Madhikermi**, Sylvain Kubler, Jérémy Robert, Andrea Buda, Kary Främling. Data quality assessment of maintenance reporting procedures. *Expert Systems with Applications*, Vol. 63, pp. 145-164, November 2016.
- III** Sylvain Kubler, **Manik Madhikermi**, Andrea Buda, Kary Främling, William Derigent, André Thomas. Towards data exchange interoperability in building lifecycle management. In *2014 IEEE Emerging Technology and Factory Automation conference*, Barcelona, Spain, pp. 1-8, September 2014.
- IV Manik Madhikermi**, Narges Yousefnezhad, Kary Främling. Data Exchange Standard for Industrial Internet of Things. In *2018 3rd International Conference on System Reliability and Safety*, Barcelona, Spain, pp. 53-61, November 2018.
- V Manik Madhikermi**, Narges Yousefnezhad, Kary Främling. Heat Recovery Unit Failure Detection in Air Handling Unit. In *IFIP International Conference on Advances in Production Management Systems*, Seoul, Korea, pp. 343-350, August 2018.
- VI Manik Madhikermi**, Narges Yousefnezhad, Kary Främling. Data discovery method for Extract-Transform-Load. *2019 5th Interna-*

tional Conference on Information Management and Industrial Engineering, Cape Town, South Africa, pp. 174-181, February 2019.

Author's Contribution

Publication I: "Key Data Quality Pitfalls for Condition Based Maintenance"

In this paper, quantitative and descriptive analysis of the case studies in Finnish Multinational Manufacturing Companies has been carried out to identify key data-centric challenges that hinder companies to carry out effective data-driven maintenance. Key challenges: Data Quality, Interoperability and Data extraction have been identified as primary bottleneck for effective utilization of data for decision-making in maintenance.

Manik Madhikermi is the main author of paper. Co-authors helped in discussing ideas and reviewed written text.

Publication II: "Data quality assessment of maintenance reporting procedures"

In this paper, AHP-based MRQA framework has been proposed to measure and compare the quality of maintenance reports. The proposed framework tries to mitigate the "data quality" challenge identified in Publication I. This framework is based on three data quality dimensions namely: "Believability", "Completeness" and "Timeliness" where each of the dimension has its own one or multiple criteria. The applicability of proposed method has been tested with use case of Finnish Original Equipment Manufacturing Company. In addition, the proposed framework has been implemented in MRQA Dashboard which aids company to measure and compare maintenance reporting quality of different branch offices.

Manik Madhikermi is primary author of the paper and is responsible for developing and testing the proposed framework. Sylvain Kubler and

Jérémy Roberts are primarily responsible for developing MRQA dashboard. In addition, Sylvain Kubler and co-authors helped to write text as well as contributed in discussion and commenting on the paper.

Publication III: “Towards data exchange interoperability in building lifecycle management”

In this paper, application of Open Messaging Interface (O-MI) (previously known as QLM-MI) for interoperability among different entities/systems of product life cycle is presented for overcoming Interoperability challenge identified in Publication I . Here, detailed comparative study of different messaging standards has been carried out and the result of comparative study showed that O-MI is best suited for information exchange within the product lifecycle. In addition, use case of O-MI in building lifecycle management for interoperability is being presented.

Manik Madhikermi carried out comparative study of five messaging standards/systems in this paper. In addition, he implemented O-MI messaging standard for use case of building lifecycle management. In this paper, Sylvain Kubler is the primary author, however Manik Madhikermi contributed in writing messaging comparison framework.

Publication IV: “Data Exchange Standard for Industrial Internet of Things”

In this paper, the Open Messaging Interface (O-MI) and Open Data Format (O-DF) are presented as flexible messaging data exchange interoperability standards that enable seamless integration of heterogenous systems. Here, three use-case scenarios of O-MI messaging standard for condition-based maintenance are presented by utilizing time-centric, event-centric and rate-centric modes of data exchange that are required in maintenance management.

Manik Madhikermi is the primary author of the paper. Here, he utilizes in-house developed message-oriented middleware to highlight applicability of open messaging standard for industrial context such as remote monitoring. He is also responsible for developing agents required for information exchange in the use cases. In this paper, co-authors contributed in the form of discussions and reviews.

Publication V: “Heat Recovery Unit Failure Detection in Air Handling Unit”

In this paper, the method for heat recovery failure detection of Air Handling Unit (AHU) has been proposed. The proposed method utilizes nominal efficiency of air handling unit using Principle Component Analysis (PCA) in combination with logistic regression. The proposed method was able to detect failure of heat recovery unit with 97% of accuracy.

Manik Madhikermi is the primary author of paper and is primarily responsible for developing and implementing the proposed model. Co-authors contributed by reviewing the written text and providing the suggestion for implementation of the method.

Publication VI: “Data discovery method for Extract-Transform-Load”

In this paper, general data discovery methodology has been proposed by leveraging the endusers' knowledge and understanding of the specific information system. The proposed method maps processes' supported information system with its underlying database schema which helps to locate the right data for data extraction. The proposed method was designed to address data extraction challenge identified in publication I. The proposed method has been tested in four information system with different databases.

Manik Madhikermi is the main author of paper and independently identified the problem domain and implemented proposed solution. Co-authors helped to setup use cases and reviewed the written text.

Clarification of subcontracting

Laura Mendoza has checked the language of my dissertation. I have reviewed each and every suggestion she made to make sure that it has not altered the meaning of the scientific content of my dissertation.

Guaranteed by
Manik Madhikermi

List of Abbreviations

AHP	Analytic Hierarchy Process
AHU	Air Handling Unit
BLM	Building Lifecycle Management
BoL	Beginning of Life
CL2M	Closed Loop Lifecycle Management
CBM	Condition Based Maintenance
CM	Condition Monitoring
CMS	Content Management System
CRM	Customer Relation Management
DBMS	Database Management System
DW	Data Warehouse
ebXML	Electronic Business XML
EoL	End of Life
ER	Entity Relation
ERP	Enterprise Resource Planning
ETL	Extract-Transform-Load
GUI	graphical user interface
HRU	Heat Recovery Unit
HTTP	Hyper Text Transfer protocol
ICT	Information and Communication Technology
IFC	Industrial Foundation Class
IoT	Internet of Things
IS	Information System
JMS	Java Message Service
LCIM	Level of conceptual interoperability Model
LISI	Levels of Information Systems Interoperability
MES	Manufacturing Execution Systems
MM	Maintenance Management

MMIS	Maintenance Management Information System
MoL	Middle of Life
MRQA	Maintenance Reporting Quality Assessment
O-DF	Open Data Format
O-MI	Open Messaging Interface
O-MI/O-DF	Open messaging standard
O&M	Operation & Maintenance
oBIX	Open Building Information Exchange
OEM	Original Equipment Manufacturer
PLM	Product Lifecycle Management
PMI	PROMISE Messaging Interface
PUCC	Peer-to-peer Universal Computing Consortium
QLM	Quantum Lifecycle Management
QLM-MF	QLM Messaging Format
QLM-MI	QLM Messaging Interface
SOM	System Object Model
TCP/IP	Transmission Control Protocol/Internet Protocol
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution
ttl	Time to Live
UML	Unified Modeling Language
WSDL/SOAP	Web Service Definition Language / Simple Object Access Protocol
XML	eXtensible Markup Language
XMPP	eXtensible Messaging and Presence Protocol

1. Introduction

1.1 Overview

In today's competitive market space, major companies in the equipment manufacturing space are trying hard to make their products stand out with respect to their competitors in order to derive tangible benefits in terms of increased sales, improved profits, and market share. Pricing and product features are considered as important factors to achieve such goals. However, product reliability and after-sales services play an equal role to improve the odds. Product reliability depends upon the decisions made during design, manufacturing and production process; nonetheless, they tend to deteriorate over time and also occasionally fail due to real-world operating conditions under various degrees of stress (Jardine et al. 2006). Hence, offering after-sales maintenance service is a key tool to assure its customers about the sustained service life of their product and can be a key differentiator (Chase et al. 1992). However, these services may result in additional costs to the manufacturer. Within many large-scale plant-based industries, maintenance costs can account for as much as 40% of the operational budget (Dunn 1998). The surveys of Maintenance Management (MM) effectiveness by Mobley (2002) indicate that one third or 33 cents out of every dollar of all maintenance costs is wasted as a result of unnecessary or improperly carried out maintenance. Understandably, a significant amount of costs is associated with maintenance services and therefore improving maintenance effectiveness is a potential source of financial savings for service providers (Eti et al. 2006).

A successful after-sales maintenance service requires detailed planning, scheduling, controlling as well as deploying maintenance resources to perform the necessary maintenance actions. This is all true given that the

maintenance process involves managing a set of complex tasks to reduce downtime and maximize operational availability. Various stakeholders with different responsibilities are involved in this management, such as (i) maintenance planners who are responsible for scheduling planned maintenance activities; (ii) plant managers who are responsible for cost reporting and savings; (iii) maintenance managers who are responsible for the execution of planned/unplanned maintenance activities. Traditionally, the success of such a program relies on expert assessments and judgments. Currently, this practice has shifted to different direction due to organizational capability to collect and utilize data to improve the effectiveness and efficiency of maintenance activities by adding value from data to their expert assessments. The shift towards data-driven decision making in maintenance and the emerging concept of Industrial Internet of Things encourage companies to collect product data throughout the product lifecycle (Fast et al. 2008, Jun et al. 2007) such as:

- Beginning of life data: Design data, Simulation data, Testing data
- Middle of life data: Failure data, performance data, age of product, operating environment, usage intensity, maintenance report, refund, replacement etc

Effective and efficient maintenance services depend upon companies' ability to leverage their data assets for decision making, planning, and executing maintenance activities. Despite availability of data, as well as advanced statistical and analytical tools, companies are still struggling to exploit it to increase profit, improve customer service and build long-lasting relationships with customers (Lavallo et al. 2011). Thereby, the objective of this dissertation is as follows:

Research objective

To identify and mitigate data-centric challenges in equipment manufacturing companies which obstruct the effectiveness of data-driven decision-making in maintenance.

1.2 Research design

To investigate the overall research objective, the broader research goal is organized into two stages: Identification and Mitigation. The idea behind

the identification step (formulated as research question 1) is to investigate data-centric challenges in equipment manufacturing companies that obstruct effective decision making. The finding of research question 1 resulted in the formulation of three other research questions: 2, 3 and 4 which attempts to mitigate the challenges identified by research question 1 respectively. Primarily, research question 1 focuses on pinpointing data-centric challenges that modern Original Equipment Manufacturer (OEM) are facing while research questions 2, 3, and 4 perform critical analysis on each of the identified shortcomings (depicted in Figure 1.1) and propose the solutions. Each research questions has been investigated in one or more of original publications. The identified shortcomings and the proposed solutions highlighted in this research work are particularly relevant to the companies that are moving towards data driven decision-making. These companies should utilize opportunities provided by industrial internet to improve organizational decision-making capability at different organizational levels by harnessing the available data assets. The tools and methods proposed in this thesis are particularly useful for strategic level decision-makers especially data quality manager, data warehouse planner, and system architect to plan, organize and utilize data assets in order to achieve their organizational goal. Despite that the research focuses on the improvement in maintenance decision-making, the generic nature of the identified problems and proposed solution make this research useful in other domains as well with minor modifications. Details of each research question and its relation with the original publication are detailed below.

Figure 1.1. Overview of Dissertation

Research question 1:

What are key data-centric challenges that hinder companies to carry out effective decision making for MM?

The first research question regarding data centric challenges that many modern OEM faces is studied in the form of quantitative and descriptive case studies of two globally operating Finnish companies which provide maintenance as an after-sales service. These OEM provide value-added maintenance service to its customers along with their products. Despite their intention, technology and availability of data, these case companies are struggling to provide effective data-driven maintenance service such as Condition Based Maintenance (CBM) to reduce maintenance cost and downtime. In order to tackle such issues, the study in **Publication I** aims to identify key data-centric challenges that hinder companies to carry out effective decision making. In this research, several statistical and analytical tools are applied to the case companies' data to pinpoint shortcomings that hinder decision making. Furthermore, different kinds of data have been collected through interviews, contextual inquiry and group discussions with representatives from the involved companies to get an insight as well as to validate the results from the quantitative analysis. The finding of this research question gives rise to the rest of the research questions

which are tackled in this dissertation. The relation between each of the research question is highlighted in Figure 1.1.

Research question 2:

How can data quality assessment of maintenance reports be carried out?

The second research question is the study of data quality assessment of maintenance reports. The need for data quality assessment for maintenance report is presented in **Publication II** which is based on the findings of the first research question. Research question 2 is explored through quantitative and descriptive case studies of a Finnish OEM which started to servitize its product. Economic success and failure of such servitization depends upon their ability to plan, schedule and execute maintenance activity to keep their product up and running to maximize its throughput. In this respect, they make use of maintenance reports as decision support tools, which contain useful recorded information such as technical maintenance logs, asset location, description of the defect, location codes, scheduled maintenance date, man-hour etc. One of the major challenges the case company facing is the quality of maintenance reports. Since maintenance decision making relies on maintenance reports, bad reporting results in wrong decision making and financial losses. Hence, data quality assessment of maintenance reports is vital to companies for decision making. For this purpose, **Publication II** proposed the Maintenance Reporting Quality Assessment (MRQA) framework which has been modeled into MRQA dashboard, where MRQA dashboard is to be seen as an artifact. A test of the validity of the proposed artifact is carried out in the use case of a globally operating Finnish OEM with more than 50 maintenance sites distributed throughout the globe. The result of the proposed solution is presented, communicated and validated through the experts of the case company.

Research question 3:

How does an open messaging standard enable data exchange interoperability for data-driven maintenance?

The third research question studies the data exchange interoperability challenge that many OEM companies face due to the heterogeneity of Infor-

mation System (IS) they use. Data-driven decision making in maintenance required data that needs to be collected from different sources such as remote monitoring, maintenance data. The problem becomes even more challenging mostly in the larger plants, where most data in the corporate IT systems are often distributed over several databases and IT systems and even geographically separated. These data are stored in various types making it technically hard to integrate them; the same data set can be interpreted in different ways depending on the user which leads to confusion between different departments or organizational units because data are aggregated in a level that suits one organizational unit. This research question explores the opportunity provided by the open messaging standard to integrate such system enabling data exchange interoperability that is suitable for data-driven maintenance such as CBM. In the given context, **Publication III** presents data exchange interoperability challenges which need to be overcome to enable interoperability. Since messaging standards are ideal to enabling data exchange interoperability, normative style of comparative study among the identified messaging standards has been carried out in **Publication III**. This study also follows the design science research methodology where the identified messaging standard has been implemented as artifacts that enable data exchange interoperability. A test of the validity of the proposed solution is demonstrated through use case scenarios of the smart home in **Publication IV** and **Publication V** in the context of data-driven maintenance.

Research question 4:

How to extract data from complex information systems for data warehouse(DW) by leveraging the end users' knowledge?

Modern enterprises are normally made up of multiple units and sub-units to carry out a specific task and they rely on the specific IS to streamline their part of organizational processes. Lack or limited interaction between organizational units result in multiple data silos. Hence, developing a DW is pre-requisite for any organization to move towards data-driven decision making. However, extracting the right set of data (for data warehousing) from modern IS is a complicated task. This is especially true for commercial ISs consisting of a large number of tables and relationships. In some cases, data extraction becomes almost impossible as some vendors deliberately obfuscate their data model in order to prevent reverse engi-

neering (MarcL 2010). In addition, relationships between tables are not always explicit as they are often encapsulated within the application logic (Pieterse & Olivier 2012). In the given context, **Publication VI** proposes data discovery methodology which maps IS process with subset of database schema where data resides. The study has been carried out following the design science research methodology where artifact named “Data Model Logger” has been developed that implements proposed methodology. A test of the validity of the artifact has been carried out in four different IS where the data model is implemented in four well known databases.

Relation between research questions and publications

Table 1.1 presents the list of original publications used in this dissertation. Table 1.2 presents the relation between each research question and the associated publications. In Table 1.2, “**X**” refers to a strong effort on the publication to the related research question and “**x**” refers to a smaller effort.

Table 1.1. List of publication included in dissertation

Id	Name of publication
I	Key Data Quality Pitfalls for Condition Based Maintenance
II	Data quality assessment of maintenance reporting procedures
III	Towards data exchange interoperability in building lifecycle management
IV	Data exchange standard for industrial Internet of Things
V	Heat Recovery Unit Failure Detection in Air handling Unit
VI	Data Model Logger - Data Discovery for Extract-Transform-Load

Table 1.2. Relation between research questions and publications (**X**= strong effort, x = smaller effort)

Id	Research questions	Publications					
		I	II	III	IV	V	VI
RQ1	<i>What are key data-centric challenges that hinder companies to carry out effective decision making for MM?</i>	X	x	x	x		
RQ2	<i>How can data quality assessment of maintenance reports be carried out?</i>		X				
RQ3	<i>How does an open messaging standard enable data exchange interoperability for data-driven maintenance?</i>			X	X	x	
RQ4	<i>How to extract data from complex information systems for data warehouse(DW) by leveraging the end users' knowledges?</i>						X

1.3 Research methodology

Research methodology is a way to systematically solve the research problem. It may be understood as a science of studying how research is done scientifically (Kothari 2004). The choice of which research method to employ depends upon the nature of the research problem (Morgan & Smircich 1980). To meet the objective of the study, a combination of methods need to be used rather than applying one. In this dissertation, one or more research methods have been employed to answer each of the research questions. The research methods used are detailed below.

1.3.1 Literature review

Hart (2018) has defined literature review as “the analysis, critical evaluation and synthesis of existing knowledge relevant to your research problem help to establish what theories already exist, the relationships between them, to what degree the existing theories have been investigated, and to develop new hypotheses to be tested”. As we know, any research work requires a deep understanding of the problem domain. One of the key sources of such knowledge is through literature review which allows examining existing theories that have accumulated in regard to the particular problem and provide a deeper insight or understanding. As all research

should be based on, or take into account, previous research that has been undertaken within the same subject, this dissertation has carried out a broad literature review that is narrowed down and more specific at later stages in order to highlight the importance or relevance of research that is being carried out. Initially, comprehensive literature regarding data driven maintenance has been carried out, thereafter, narrowed down to different studies performed in order to answer each of the research questions in their respective chapters.

1.3.2 Comparative study

The comparative method has been often used in the early stages of the development of a branch of science. It can help the researcher to ascend from the initial level of exploratory case studies to a more advanced level of general theoretical models, invariances, such as causality or evolution (Pentti 2007). In the comparative study, research work is being carried based on the difference between two or more objects being compared. The nature of the comparative study depends upon the objective of the study. The descriptive style highlights the differences between the cases, describes and possibly explains them. The normative style uses principal criteria to evaluate and point out the best (in this respect) among the alternatives that are being studied. In this dissertation, the normative style of comparative study has been carried out in Chapter 4.

1.3.3 Case study

A case study is an empirical inquiry that investigates a contemporary phenomenon with its real-life context especially when the boundaries between the phenomenon and context are not clearly evident (Yin 1994). There are different types of case studies which are appropriate in different situations and contexts. The descriptive case study is to assess a sample in detail and in depth, based on an articulation of a descriptive theory. This theory must respect the depth and scope of the case under study, which is conveyed through robust propositions and questions. Descriptive studies seek to reveal patterns and connections, in relation to theoretical constructs, in order to advance theory development (Tobin 2010). One of the key criticisms of the case study research method is that it is impossible to get generalizable results; however, Yin (1994) stated that it cannot be generalized to population or the universe but it can be generalized to

theoretical propositions. In this dissertation, quantitative and descriptive style of case study has been carried out in Chapter 1.

1.3.4 Design science

Design science is a research paradigm in which a designer answers questions relevant to human problems via creation of innovative artifacts, thereby contributing new knowledge to the body of scientific evidence. The designed artifacts are both useful and fundamental in understanding that problem (Hevner & Chatterjee 2010). It involves the analysis of use and performance of the designed artefact such as models, diagrams, system designs, algorithms and artificial languages in order to understand, explain, and improve designed systems (Crnkovic 2010). Similarly, constructive research is construction based on existing knowledge used in novel ways, with the possibility to add few missing links (Crnkovic 2010). Constructive work on constructive research starts with well understood ground of the problem which might be the result of a previous empirical investigation where quantitative or qualitative research methods have been used. The design science research presented in this dissertation is constructive research where the constructs are software implementations, method, model, and algorithms, which are detailed in their respective chapters.

1.4 The structure of the thesis

This dissertation is divided into six chapters. In this chapter, a brief introduction to the research topics, objectives, scientific contributions and methodology is presented. Chapter 2 provides an overview of prior literature that is relevant to this work along with finding of the case study of Publication I. Chapter 3, 4, and 5 present reviews of the finding, contribution and discussion of the six original research papers. Finally, in Chapter 6, the summary of the dissertation is discussed and it concludes with some suggestions for future research. The six publications are annexed as appendices.

2. Data-driven decision making in maintenance management

This chapter aims to describe and analyze the existing theories that can be found within the research topic in order to place the research into a context as well as to give the reader an overview of related research. When performing research, it is also necessary to build one's results on both theories and empirics in order to reach reliable conclusions.

2.1 Maintenance

Maintenance is a complex process which requires planning, scheduling and executing various activities with an objective to extend equipment life, improve equipment availability and to retain an equipment in proper condition. BS EN 13306:2010 (BSI 2010) defines maintenance as “*combination of all technical, administrative and managerial actions during the life cycle of an item intended to retain it in, or restore it to, a state in which it can perform the required function*”. Researchers working in the field of maintenance have categorized maintenance based on maintenance strategies. These strategies are developed based on their needs to improve maintenance performance in terms of cost and efficiency. One of the initial categorization is carried out by Bateman (1995) which consists of three types of maintenance strategies; reactive, preventive and predictive maintenance. Here, preventive and predictive maintenance are proactive maintenance strategies (Swanson 2001). Later, another maintenance strategy namely total productive maintenance was introduced by Weil (1998). However, if we look at the definition of maintenance from BS EN 13306:2010 (BSI 2010), it focuses on two key terms “retain” and “restore”, based on which maintenance strategy is primarily classified in two categories as shown in Figure 2.1.

Figure 2.1. Maintenance Strategies

The preventive maintenance is a maintenance strategy which is often referred to as use-based maintenance. The objective of this type of maintenance is to reduce the probability of failure or the degradation of the functioning of an equipment (Swanson 2001). This strategy of maintenance is either predetermined, in which maintenance activities are scheduled based on time or number of uses but without previous condition investigation or are based on the condition which is the combination of and/or inspection and/or testing, analysis to issue maintenance action. The other type of preventive maintenance strategy is called CBM, which can be either continuous or on demand or is based on prediction from the data collected from the monitoring system.

The corrective maintenance is the traditional way of maintaining equipment. This type of maintenance is carried out after fault recognition and intended to put an item into a state in which it can perform a required function (BSI 2010). The corrective maintenance strategy is further classified as deferred or immediate. In immediate, the maintenance action is carried out without any delay once the fault has been noticed or identified in order to avoid unacceptable consequences. However, in deferred maintenance, the maintenance action is scheduled in the next planned maintenance. In general, corrective maintenance strategy is not suitable where the reliability of equipment is critical and failure of equipment results in catastrophic loss in terms of cost and safety.

2.2 Maintenance management

BS EN 13306:2010 (BSI 2010) defined MMs as “*all activities of the management that determine the maintenance objectives, strategies and responsibilities, and implementation of them by such means as maintenance planning, maintenance control, and the improvement of maintenance activities and economics*”. It includes all activities required to restore equipment to, or keep it in a specified operating condition with maximum availability with ensured required operational quality in cost effective and in accordance with safety and environmental legislation (Pintelon et al. 1999). The objective of MM is to improve maintenance decision making by providing the decision-making tool which allows subjective decision making to be replaced by the objective one (Pintelon & Gelders 1992). In order to organize MM, different frameworks have been developed to aid decision making at various organizational levels, one of the earliest being put forward by Pintelon & Gelders (1992) who pointed out three important business levels in the decision-making process: (i) operational level: decision regarding marketing and finance; (ii) planning & control level: decisions regarding resource and scheduling management, and performance reporting; (iii) managerial level: decisions regarding how to optimize actions and policies to be performed on-site. Similarly, Levrat et al. (2008) proposed a similar three business level based MM framework: (i) Strategic level: strategic axis is expressed in quantitative and qualitative terms, and organizational maintenance strategies are defined such as corrective and preventive maintenance, risk-based or condition-based maintenance, etc. (ii) Tactical level: maintenance actions such as scheduling and resource planning are planned; (iii) Operational Level: actual work is carried out in addition to access performance and future equipment conditions.

2.3 Data-driven decision making in maintenance

Data is considered one of the most valuable assets in modern day organizations. Many companies utilize their data asset to improve their decision making capability, which in turn improves their business activities and performance. Decision making in maintenance is a complex process, which requires data from different sources and is affected by the surroundings and context where data is generated (Jonsson 1999, Waeyenbergh & Pintelon 2002). A successful maintenance program requires detailed planning,

scheduling, controlling as well as deploying maintenance resources to perform the necessary maintenance actions. This is all true given that the maintenance process involves managing a set of complex tasks to reduce downtime and maximize operational availability. Various stakeholders with different responsibilities are involved in this management, such as (i) maintenance planners who are responsible for scheduling planned maintenance activities; (ii) plant managers who are responsible for cost reporting and savings; (iii) maintenance managers who are responsible for the execution of planned/unplanned maintenance activities, and so on. All these stakeholders have a common goal: maximize equipment availability in an operating condition permitting the desired output quantity and quality. Several researchers working in the domain of maintenance have viewed decision making from a different perspective: Bertolini & Bevilacqua (2006), Bevilacqua & Braglia (2000), Goossens & Basten (2015), Labib et al. (1998), Pramod et al. (2007), Shyji et al. (2008), Tan et al. (2011), Zaim et al. (2012) worked on aiding maintenance policy selection while Alarcon et al. (2007), Thor et al. (2013), Kumar & Maiti (2012) and Coulter et al. (2006), Eslami et al. (2014), Van den Bergh et al. (2013), Almeida (2012) worked on maintenance action selection and maintenance scheduling respectively. Despite the differences in the aspect of the decision making in maintenance, these researchers utilize available data asset to support decision making at different business levels.

2.4 Information technology systems for maintenance management

Information technology impacts all parts of the product lifecycle, services, and businesses. It has been of strategic importance for many organizations as it has a positive correlation on the profitability and competitiveness of companies (Kini 2002, Dedrick et al. 2003). Most of the modern companies use ISs to streamline their organizational processes and to aid decision making to achieve their business goal which includes MM. IS for MM has been in practice since the early 1970s where maintenance planning system was deployed in mainframe (Kelly 1984). Early Maintenance Management Information System (MMIS) was focused to provide functionality for scheduling, plant inventory, stock control, cost and budgeting and maintenance history (Kans 2009). Progressively, MMIS acquired more decision support capabilities such as work-order planning. More and more specific maintenance software focused on the needs of the maintenance manager

becoming available. It contains modules with technical and historical data on the equipment, work order planning, personnel management inventory control and reporting facilities (Pintelon et al. 1999). Moreover, this MMIS helps to aid decision making at tactical and operational level through a different kind of maintenance performance report which helps them to reduce the number of failures, failure cost and cost associated with repair (Al-Najjar 1996).

Traditionally, maintenance practices were mostly reactive maintenance (Wilkinson et al. 2006, McMillan & Ault 2007). More recently, with the advancement of information and communication technology and increasing need to improve maintenance effectiveness to reduce financial losses, companies are moving towards advanced maintenance strategy called preventive maintenance. This type of maintenance policy requires continuous monitoring system, which gives real-time status of an equipment. Several research studies have been carried out to aid decision using condition monitoring system to support decision making such as Sato et al. (1991) and Itaka et al. (1991) for monitoring and maintenance communication for power transport lines, McMillan & Ault (2007) for monitoring offshore wind turbines etc. In principle, these monitoring programs should be part of a MMIS; however, in reality, these programs are, in general, not integrated with programs for production/operation, quality control, environmental condition or cost accountancy (Al-Najjar 1996) despite its advantages as highlighted by Pearce & Hall (1994). Similarly, Wichers (1996) also showed the applicability of integrated condition monitoring system for reliability-centered maintenance.

2.5 Data-centric challenges towards data-driven decision making

The concept of utilization of data for decision making in MM requires companies to collect large amounts of data that have been generated in different stages of product lifecycle, such as design data, simulation data, failure data, performance data, condition monitoring data, usage intensity, maintenance report etc (Jun et al. 2007) as source of information. Hence, the effectiveness of MM depends upon companies' ability to leverage those data assets. Despite the availability of data, as well as advanced statistical and analytical tools, there are still challenges that limit the organization's capability for effective decision making. Such challenges need to be identified and resolved with an objective to enhance organizational

decision-making capability. With this objective in focus, this dissertation tries to identify these challenges through a literature review and validate it through a case study for its relevance. Finally, the potential solution for each identified challenge is presented in the following chapters.

2.5.1 Data Quality

The term “data quality” can be best defined as “fitness for use” which implies that the concept of data quality is relative. Thus, data with quality considered appropriate for one use may not possess sufficient quality for another use (Tayi & Ballou 1998). Despite major developments in data management and big data, data quality problems are increasingly evident in organizations of all sizes (Strong et al. 1997). Surveys over the years have pointed out that as many as 60% of organizations have problems with data quality (Wang & Strong 1996). Data quality does not just mean accuracy of data but also relates to aspects such as completeness, accessibility, and context. Two major factors – data consistency and data completeness are central to data quality (Kwon et al. 2014). Both of these aspects - data consistency and data completeness are critical from the maintenance perspective. Strong et al. (1997) provide data quality dimensions and common problems affecting them. Table 2.1 shows categories of data and data quality dimensions under each category.

Table 2.1. Data Quality Dimensions (Strong et al. 1997)

Data quality category	Data quality dimensions
Intrinsic code	Accuracy, Objectivity, Believability, Reputation
Accessibility	Accessibility, Access security
Contextual	Relevancy, Value-Added, Timeliness, Completeness, Amount of data
Representational	Interpretability, Ease of understanding, Concise representation, Consistent representation

Among the problems described by the authors, some of the critical ones are, i) multiple sources of same data; ii) poor data accessibility; iii) incomplete data and poor relevancy; iv) and inadequately defined or measured data.

When discussing data quality, it is important to put data consumers first, as quality is subjective to the data use and the context. To each consumer, the parameters that determine data quality may be different and how they consume even the same data can vary greatly (Wang & Strong 1996).

2.5.2 System interoperability

Advancement in Information and Communications Technology (ICT), emergence of Internet of Things (IoT) and intelligent product evolve the concept of Product Lifecycle Management (PLM) to Closed Loop Lifecycle Management (CL2M) (Kiritsis 2011) with the aim of moving beyond engineering aspects of a product. The extended concept provides a shared platform for the creation, organization and dissemination of product-related information across the enterprise throughout product lifecycle (Ameri & Dutta 2005). This novel concept expanded OEM maintenance capability beyond organizational boundaries by introducing new data-driven maintenance strategy which utilizes a sophisticated analytics engine to identify systems and equipment that are not performing as expected and therefore require, or will require, maintenance (Ecosave 2017). Utilizing PLM data to drive maintenance activities is considered as a core element for such maintenance program. Several researchers (Jonsson 1999, Waeyenbergh & Pintelon 2002) working in the domain of maintenance have shown that maintenance decision-making is affected by several closely related working areas. In considering the role of maintenance in CL2M, close relationships between maintenance activities and those in other phases of product life cycle, such as the design, production, operation, after-sales service requires exchange and management of information between several systems or entities (Främling et al. 2013). This means that data from different IT systems such as Enterprise Resource Planning (ERP), Condition Monitoring (CM), Customer Relation Management (CRM), Manufacturing Execution Systems (MES), MMIS etc needs to be exchanged (Al-Najjar et al. 2001). The major concern with CL2M is how to access the information that is stored in product or heterogeneous systems with variation in technology and standards resulting in interoperability challenge. This challenge becomes even more complicated when the information is not stored only locally in the product or at only one server but is distributed in two or more places.

2.5.3 Data extraction

Data-driven decision making has emerged as a decision-making tool in many organizations, which provides factual insight into organizational processes and performances. Developing such a tool requires data to be collected through multiple heterogeneous systems that are being used in an organization. Several researchers (Jonsson 1999, Waeyenbergh & Pintelon 2002) working in the domain of maintenance have shown that maintenance decision making is affected by several closely related working areas. The standard IEC (ISA et al. 2000) illustrates information overlap between three main areas: production capability information, capacity scheduling and maintenance information as shown in Figure 2.2. This

Figure 2.2. Dependencies between production, scheduling and maintenance (ISA et al. 2000)

means that data from administrative IT systems, industrial IT and other sources must be collected to reach technical, financial and organizational maintenance data (Kans 2008). In larger plants, especially in product systems, data are often distributed over several databases and IT systems and even geographically separated. The diversity of IT systems also leads to data interoperability challenge, making it technically hard to integrate them resulting in data silos. In order to cope with such an issue, data are collected through multiple heterogeneous systems that are being used in an organization into a single coherent system called DW. Getting data out of such systems requires a detailed understanding of the structure and semantics of data. In other words, understanding the data model of an IS practically means to grasp part of its database schema and application logic. However, this task is often challenging because DW designers are usually not experts in the specific domain and processes supported by the

given IS (Viaene 2013) and it is quite easy to get lost in hundreds of tables with hundreds of columns. In some cases, understanding data model of a complex IS becomes almost impossible as some vendors deliberately obfuscate their data model (MarcL 2010), and relationships between tables of data-model are often encapsulated within the application logic (Pieterse & Olivier 2012).

2.6 Case study

In Publication I, studies have been carried out in collaboration with two globally operating original equipment manufacturing companies with the primary objective to identify data-centric challenges that hinder data-driven decision making.

Data Quality in Maintenance report:

Maintenance reports are a summary of maintenance activities that have been carried out by maintenance personnel and are reported through a back reporting tool. It contains details of maintenance activities that include information on workorderID, date, maintenance type, start date and time, end date and time, maintenance personnel, component having problem (Component code), reason of problem (Reason Code), action (Action code) taken to resolve the underlying issue. These reports are one of the main sources of information where several performance metrics are computed for decision support. The study of 45375 maintenance reports collected over 18 months of case companies highlighted that data from a significant number of fields were either missing or were incorrect as shown in Table 2.2. Around 70% of records have empty codes and few invalid codes in all major fields – component code, reason code, and action code. These codes provide vital information regarding commonly failing components, frequency of failures, status of equipment, reasons of failure. In addition, one of the observations in these case companies included lack of data validation techniques to validate the correctness of report. Table 2.2 shows that there exists several empty or invalid codes entered by the technicians during service reporting of their maintenance work. Further investigation through an interview with the maintenance personnel revealed that neither of the case companies practiced any form of data validation process. Once the data is entered, neither the

back-reporting tool performs any validation of the inputted data, nor there exists an external process to identify those errors.

Table 2.2. Total number of Empty and Invalid Codes

Service report fields	Empty count	Invalid count
Action code	31562 (69.55%)	26
Component code	33562 (73.96%)	114
Failure Description	28591 (63.01%)	0
Reason code	33570 (73.98%)	162

Interoperability challenges:

As stated in Section 2.5.2, ISs are fundamental to streamline operations and support processes of any modern equipment manufacturing companies. These processes rely heavily on specialized ISs such as PLM, ERP, MMIS, CM etc. In our studied use cases, the interoperability challenge does exist which results in a data integration problem. There are different workflows and systems dedicated to carrying out maintenance activities depicted in Figure 2.3.

Figure 2.3. Data collection processes

In MM workflow, the maintenance planner creates and assigns work orders to technicians, and the technicians use back reporting tool to log service report data after completion of maintenance tasks in MMIS. While in the remote monitoring process, the embedded software installed in the machine reports errors and alarms only. During the field visit for the study, it was observed that several sensors were installed on the equipments. However, the remote monitoring process (as depicted in Figure 2.3) only stores alarm and

error information from these machines into the database while the associated sensor data have been completely ignored resulting in loss of vital data that is essential for data-driven maintenance such as CBM. This is due to fact that only legacy embedded software in equipment have access to the sensors data which it uses for raising alarms or errors based prior understanding of failure condition. There is no interface in the embedded system that can be used by other systems for collecting sensor data. Moreover, during the study, there was no automatic mechanism to associate these errors and alarm data with the maintenance report data. Due to the lack of interoperability between these two systems, monitoring data and service report data were very hard to integrate. Our analysis on the data of two systems supporting these workflow shows that “Timestamp” is the only key to integrate these data. Due to the occurrence of multiple errors approximately at the same time (shown in Table 2.3), it is nearly impossible to associate the maintenance report with particular errors or alarms.

Table 2.3. Number of errors at same time

Assembly Id	Device Id	Timestamp	Error count
1	215	2013-12-23 06:29:45.673	24
1	215	2012-11-08 15:33:31.517	23
1	215	2012-11-08 15:33:31.517	22
1	215	2012-11-29 11:19:39.780	22

Data Extraction from data silos:

As stated in Section 2.6, we observe three separate ISs supporting monitoring, product design and development, and MM workflow as shown in Figure 2.3. Although these systems support organizational workflow, multiple systems used across the organization and the internal legacy systems developed to support proprietary methods of working have created the landscape of isolated islands of information (Madenas et al. 2014) called data silos. In order to cope with this

issue, our case company attempted to collect data from each of these systems to create DW as depicted in Figure 2.3. However, getting data out of such systems requires detailed understanding of the structure and semantics of data. As detailed in Section 2.5.3, extracting data from these ISs for DW is a huge issue for our case companies as well.

3. Data quality assessment of maintenance reporting procedures

Publication II answers the second research question “*How can data quality assessment of maintenance reports be carried out?*”. This research was motivated by the notion of bad data quality in maintenance reports resulting in financial losses and wrong decision making. As highlighted by many researchers (Tretten & Karim 2014, Mahlamäki et al. 2016), the major source of data quality issue in maintenance reports is due to human involvement in the reporting process after the completion of the maintenance task. Hence, the importance of assessment of such reporting procedures is vital for ensuring the quality of reported data.

The main contribution of Publication II is the proposed Maintenance Reporting Quality Assessment (MRQA) framework which assesses and ranks maintenance reporting procedure based on maintenance report quality. From a theoretical standpoint, the proposed framework utilizes Analytic Hierarchy Process (AHP) to integrate various data quality criteria that have been identified based on Krogstie’s framework (Krogstie et al. 1995). In addition, the application of the proposed MRQA framework to assess and rank branch offices of globally operating Finnish manufacturing company is also presented.

3.1 Data quality in the maintenance management system

Data is considered as one of the most valuable assets in modern day organizations. Many companies utilize their data asset to improve their decision making capability which in turn improves their business activities and performance. In this regards, data and information quality give competitive advantages to the company (Da Xu et al. 2014, Chen et al. 2014) by providing factual insight into organizational processes and performance. This notion of data quality and competitive advantage is also true for man-

ufacturing companies providing after-sales maintenance services. These companies rely on available data as a source of information to improve their maintenance activities. Dunn (1998) highlighted that in large-scale plant-based industries, maintenance costs can account for as much as 40% of the operational budget. Understandably, a significant amount of cost is associated with maintenance services and therefore improving maintenance effectiveness is a potential source for financial savings (Eti et al. 2006). However, as indicated by Mobley (2002), one-third of all the maintenance costs is wasted as the result of unnecessary or improper maintenance practices. More recent studies have confirmed that maintenance is a major cost issue, with a ratio between maintenance costs and added-value higher than 0.25 (Sophie et al. 2014). The dominant reason for ineffective management is the lack of factual data to quantify the actual need for maintenance (Mobley 2002).

A successful maintenance program requires detailed planning, scheduling, controlling as well as deploying maintenance resources to perform the necessary maintenance actions. This is all true given that the maintenance process involves managing a set of complex tasks to reduce downtime and maximize operational availability. Various stakeholders with different responsibilities are involved in this management, such as (i) maintenance planners who are responsible for scheduling planned maintenance activities; (ii) plant managers who are responsible for cost reporting and savings; (iii) maintenance managers who are responsible for the execution of planned/unplanned maintenance activities. All these stakeholders have a common goal: reducing maintenance downtime to increase productivity. In this respect, they usually make use of maintenance reports as decision support tools, which contain useful record information such as technical maintenance logs, asset location, description of defect location codes, scheduled maintenance date. Hence, data reporting practices –including maintenance reports– have a considerable impact on maintenance activities, risks and business performance since poor data quality results in losses across a number of fronts, and reciprocally, high data quality fosters enhanced business activities and decision making.

Several MM frameworks have been developed and applied to aid decision making at various organizational levels as detailed in Section 2.2. Making decisions at each of these three levels implies dealing with multiple, conflicting, and incommensurate criteria and/or objectives, as well as human judgements. Research on human judgements and decision making show

that the human brain is able to consider only a limited amount of information at any one time (Simpson 1996), which makes it unreliable to take decisions while facing complex problems. Multiple Criteria Decision Making (MCDM) techniques, such as AHP, TOPSIS, ELECTRE, PROMETHEE, Fuzzy MCDM, etc., have been proven to be of great value in supporting decision makers at each MM level, as summarized in Table 3.1

Given the significant number of papers studied in Publication II that have been classified in Table 3.1 (40 papers at the strategic level, 19 at the tactical level, and 20 at the operational level) show that existing studies did not take into account data quality-related criteria for their decision making. The most commonly used criteria at each MM level is given below:

- Strategic level: (i) Cost (22.7%), (ii) resource availability & utilization (10.1%), added value (7.6%);
- Tactical level: Cost (21.9%), environmental/operational conditions (8.7%) and safety (9.4%);
- Operational level: Cost (25.8%), resource availability & utilization (19.4%) and added value (4.8%).

These results clearly show that data quality is hardly considered in the reviewed papers despite having a major impact on decision making. However, the only paper that integrates a data quality criterion is Van Horenbeek & Pintelon (2014a), where the authors measure the “accuracy” of maintenance report records with no consideration given to other data quality dimensions. With such finding in studied literature, Publication II attempts to fulfill such gaps by proposing the MRQA framework which enables stakeholders to measure and compare maintenance reporting in real-time and also allows to set preferences for each of the dimensions. The proposed MRQA framework is detailed in Section 3.2.

3.2 AHP-based MRQA framework

The proposed MRQA framework is a three-criterion maintenance reporting assessment scheme where each of the criteria or dimension is defined by one or more sub-criteria (detailed in Section 3.2.1) that have been identified with close coordination with industry experts. From a theoretical standpoint, AHP, developed by Saaty (1980), is used to integrate the identified data quality criteria and sub-criteria to assess the quality of maintenance

Table 3.1. MCDM Techniques Applied in Maintenance Industry

Business levels		AHP	FAHP	ANP	DEA	VIKOR	ELECTRE	PROMOTHEE	MAUT
Strategic Level	Maintenance Policy Selection	(Bertolini & Bevilacqua 2006, Bevilacqua & Braglia 2000, Goossens & Basten 2015, Labib et al. 1998, Pramod et al. 2007, Shyith et al. 2008, Tan et al. 2011, Zaim et al. 2012)	(Azizi & Fathi 2014, Ilangkumaran & Kumaran 2009, Hosseini Frouz & Ghadimi 2015, Ferdousmakan et al. 2014, Wang et al. 2007, Fouladgar et al. 2012)	(Kumar & Maiti 2012, Shahin et al. 2012, Pourjavad et al. 2013, Zaim et al. 2012)	(Azadeh et al. (2014), Sheikhalishahi (2014)	(Ilangkumaran & Kumaran 2012, Ahmadi et al. 2010)	(Zhangqiong & Guozheng 1999, Li et al. 2007)	(Emovon et al. 2015, de Almeida et al. 2015c, da Silva Monte & de Almeida Filho 2015)	
	Tools and companies Selection	(Bertolini et al. 2004, García-Cascales & Lamata 2009, Ha & Krishnan 2008, Triantaphyllou et al. 1997)	(Durán 2011)		(Ha & Krishnan 2008)		(Gonçalves et al. 2014, Gomez & Carnero 2011)		(Kuo & Wang 2012, de Almeida et al. 2015d)
	Cost Estimation	(Chou 2009)					(Chen et al. 2005)		
	Maintenance prioritization	(Farhan & Fwa 2009, Mozazani et al. 2011, Taghipour et al. 2011)	(Ouma et al. 2015)		(Wakchaure & Jha 2011)		(Liu et al. 2012, Cañiso et al. 2002, Hanbachi & Lepert 2011, Trojan & Morris 2012)	(Cavalcante & Costa 2010, Almeida-Filho et al. 2013)	(Monte & de Almeida-Filho 2016, de Almeida et al. 2015b)
	Resource Planning				(Azadeh et al. 2013)				(Garmabaki et al. 2016, de Almeida & Françopol 2006)
Tactical Level	Maintenance Scheduling	(Coulter et al. 2006, Esfami et al. 2014)			(Van den Bergh et al. 2013)		(Certa et al. 2013)		(Almeida 2012)

Table 3.1. (continue...)

Business levels	AHP	FAHP	ANP	DEA	VIKOR	ELECTRE	PROMOTHEE	MAUT
Operational Level	Critical Component Identification	(Dehghanian et al. 2012)					(Cavalcante & De Almeida 2007, Cavalcante et al. 2010)	
	Measuring/Assessment Efficiency	(Wang et al. 2010)	(Mudhiri et al. 2011, Vujanović et al. 2012, Van Horenbeek & Pintelon 2014b)	Sun (2004), Peck Jr et al (1998), Ozbek et al. (2010a,b), Hjalmarsson & Odeck (1996), Liu & Yu (2004), Rouse et al. (2002), Falah-Fini et al. (2015), Jeon et al. (2011), Roll et al. (1989), Charnes et al. (1984)			(Sevilla Juárez & Escobar Toledo 2008)	(e Costa et al. 2012)
	Maintenance Action Selection	(Kumar & Maiti 2012)					(Alarcon et al. 2007, Thor et al. 2013)	

report of an organization. Finally, Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) method has been integrated to rank different branches of the company based on the reporting quality. The proposed MRQA framework comprises of four steps which are discussed in Section 3.2.2.

3.2.1 Criteria and sub-criteria identification based on Krogstie data quality framework

Data quality has been proved as a multi-dimensional concept (Ballou & Pazer 1985, Huang et al. 1999). Hence, data quality should be defined as catering to subjective perceptions of the individuals utilizing the data and objective measurements based on the dataset. Subjective data quality assessments should cater to the needs of the data consumer while objective measurement should reflect the states of the data without contextual knowledge and organizational business rules (Pipino et al. 2002). Given the above requirement and based on the company's requirements, three key quality dimensions of the Krogstie's framework (Krogstie et al. 1995) form the foundation of our framework.

- **Semantic quality:** Semantic quality defines criteria to what extent maintenance report data reported by maintenance personnel can be trusted, or more exactly can be considered as "true", "real" and "credible". This is referred to as the "believability" criterion (C_B), whose various facets are formalized in the form of sub-criteria, i.e. as believability quality indicators denoted by $\{C_{B1}, C_{B2}, C_{B3}\}$ detailed in Table 3.2.
- **Language quality:** Language quality defines criteria to what extent the service data reported by each operator is complete, or is of sufficient depth and breadth for the task at hand. To put it in another way, this criterion, referred to as "completeness" (C_C), reflects the level of details reported by each operator with respect to each report/form field that needs to be entered (in accordance with the company's business logic). Similarly to C_B , several completeness quality indicators are defined, which are denoted by $\{C_{C1}, C_{C2}, ..., C_{C8}\}$ as detailed in Table 3.2.
- **Knowledge quality:** Knowledge quality defines criteria to what extent the service data reported by each operator is sufficiently "up to date", which depends on the time difference between the maintenance com-

pletion and its reporting. This criterion, referred to as “timeliness” (C_T), is based on the assumption that the longer the delay to submit the report, the less the quality of the reporting (operators are likely to forget details over time). As emphasized in Table 3.2, no sub-criterion is defined for this dimension.

Table 3.2. Criteria and its sub-criteria description related to the data quality dimensions

Criteria	Sub-criteria	Description
Believability (C_B)	Length of Work Description (C_{B1})	Length of the work description related to a work order.
	Work Log Conflict (C_{B2})	Work description conflict among different form fields related to a same report.
	Technician Log Variation (C_{B3})	Technical log variation among a set of reports related to a same maintenance work order.
Completeness (C_C)	Asset Location reported (C_{C1})	Asset location (in the product) where maintenance was performed.
	Description reported (C_{C2})	Description of work to be done in particular maintenance work.
	Start & Finish Date reported (C_{C3})	Actual Start and Finish dates and times of work completed.
	Target Start Date reported (C_{C4})	Targeted start date of the maintenance work.
	Target Finish Date reported (C_{C5})	Targeted finish date of the maintenance work.
	DLC Code reported (C_{C6})	Actual location of the defect within product (DLC standing for “Defect Location Code”).
Timeliness (C_T)	Schedule Start Date reported (C_{C7})	Scheduled start date of the maintenance work.
	Schedule Finish Date reported (C_{C8})	Scheduled finish date of the maintenance work.
		This is average delay of reporting on individual site

3.2.2 MRQA framework steps

The proposed MRQA framework is a four-step process starting from setting up AHP hierarchy to ranking maintenance reporting procedure based on reporting quality. The detail of each step is discussed below.

Step 1: MRQA hierarchy

Figure 3.4 shows the hierarchy of the proposed framework. The hierarchy has been identified based on Krogstie framework (Krogstie et al. 1995) and industrial experts as detailed in Section 3.2.1. The first level hierarchy is the goal to measure data quality of maintenance reporting procedure and ranking different branch offices based on their reporting quality. The second level and third level in the hierarchy are constituted of criteria and sub-criteria which will be used to measure reporting quality. The fourth level consists of a list of company’s branch offices for which reporting quality is assessed and ranked relative to each other.

Figure 3.1. MRQA Framework

Step 2: Pairwise comparison based preference measurement

The second step in the MRQA framework is to derive the weights for the criteria and sub-criteria based on user preferences. Not all the criteria will have the same importance, here a user can define its own preference of each criterion with respect to others using 1 to 9 point scale for comparison developed by Saaty (Saaty 1977). This is done by using pairwise comparisons as a ratio (detailed in Section 4.1 of Publication II). Finally, simple additive weight method (Hwang & Yoon 1981) is used to compute the preference weight of each criterion and its sub-criteria. In general, preference weight of each criterion and sub-criteria at level 2 and level 3 is computed at this step.

Step 3: Pairwise comparison based ratio measurement

The third step in the MRQA framework is to compute the weight of each branch to compare with other branch office with respect to every criterion. This is done based on the computed value of sub-criteria. For this purpose, the computation of each sub-criteria based on maintenance report is carried out for all branch offices. The calculation of these values is based on business rules detailed in Section 4.2 of Publication II. Once an actual value of each sub-criteria is computed for every branch office, pairwise comparison as a ratio of measured value is used to compute the weight of each branch offices. In general, comparative weight of each branch office with respect to other branch office at level 4 is computed at this step.

Step 4: Alternative ranking using TOPSIS

The fourth or final step in the MRQA framework is to compute the global score for every branch office based on all identified criteria and sub-criteria. This global score is a numeric value between 0 to 1 (1 being best), which is used to assess and rank quality of maintenance reporting procedure. For this purpose, firstly, global weight of each criterion for every branch is

computed. This global weight of each criterion is used by TOPSIS method to compute the final global score. The computation of the global weight and global score is detailed in Section 4.3 of Publication II.

3.3 Results and discussion

In this study, the proposed MRQA framework has been applied to assess and compare maintenance reporting of 54 branch offices of a globally operating Finnish multinational company through the MRQA dashboard that has been developed as part of this study. The MRQA dashboard thereby enables to assess and rank the reporting quality of the 54 branch offices of the case company based on his/her preferences. The MRQA dashboard enables the stakeholder to

- Visualize in a user-friendly way both the locations of the OEM sites (refer to "Map of Companies Sites" dashboard element in Figure 3.2) and their corresponding quality/ranking (refer to "Final Assessment and ranking" dashboard element in Figure 3.2);
- Deeper investigate the maintenance reporting quality of one or more sites with respect to one or more quality dimensions (refer to "disintegrated quality view" dashboard element in Figure 3.2);
- Adjust the assessment period, e.g., to assess/compare sites over the last few weeks, months or years (refer to "Date range input" dashboard element in Figure 3.2);
- Modify his/her preferences related to the criteria importance, e.g. if he/she wants to give –at a specific point of time – further importance to one or more dimensions or sub-dimensions (as depicted in Figure 3.3)

Figure 3.2. MRQA dashboard

Figure 3.3. MRQA dashboard/user interface to adjust the pairwise comparison based preference measurement

Figure 3.4 present results of MRQA framework which successfully ranked branch offices of the case company based on the maintenance reporting quality. In Figure 3.4(a), where equal preferences is given to all its criteria and sub-criteria, it can be observed that three sites 46, 12 and 18 stand out as having the highest quality scores 0.94, 0.81 and 0.62 respectively whereas sites 6 and 54 have the poorest quality both having quality scores

Figure 3.5. Overall site ranking based on preferences

of 0.08. Further investigation on site 6 and 54 (through the disintegrated view of MRQA dashboard) as depicted in Figure 3.5 shows that site 6 has a particularly poor ranking regarding both the “timeliness” and “completeness” dimensions, while site 54 has a poor ranking regarding “timeliness” and “believability”.

Figure 3.4. Overall site ranking based on preferences

Similarly, in Figure 3.4(b), where highest preferences are given to Timeliness with respect to Believability and Completeness (depicted in Figure 3.3) it is observed that sites 6,12,18,21 and 42 have the highest quality scores. Here, we showed that our MRQA framework is able to compute rank maintenance reporting based on user preferences. It should be noted that the user preference has a huge impact on the ranking and these preferences should be selected based on the purpose of measuring the quality of data. In the first case where equal preference is given to all dimensions, our case companies used these results to identify the branches with

the lowest reporting quality in order to take necessary steps to improve reporting quality on identified sites. However in the second case, where “timeliness” has the highest preference with respect to “completeness” and “believability”, case companies used these results to identify the branches with the highest reporting delay which is one of the sources of financial losses as detailed in Section 5.2 of Publication II.

This research presents three main theoretical implications. First, it contributes to the literature on MM by proposing a thorough state-of-the-art on the use of MCDM techniques at each MM level (strategic, tactical, operational), which helps in identifying criteria at each of these levels (cf. Table 5 in Appendix A of Publication II). This list of criteria can be of potential value to future researchers working in MM. Secondly, this research contains an approach to identify relevant data quality dimensions (based on existing data quality frameworks) and to turn them into a hierarchical AHP structure. A theoretical framework is then proposed, enabling the assessment and ranking of different company branch offices in terms of maintenance reporting quality.

4. Data exchange interoperability with an open messaging standard

The fourth research question *“How does an open messaging standard enable data exchange interoperability for data-driven maintenance?”* has been answered using the comparative study carried out in Publication III in conjunction with the use-case scenario detailed in Publication IV and fault detection method presented in Publication V. This research question was motivated by the data exchange interoperability challenges many companies face due to variation in standards and technology in systems they use to carry out data-driven maintenance.

4.1 Interoperability challenge in maintenance management

Advancement in information technology, the emergence of IoT and intelligent products evolve the concept of PLM to CL2M with the aim to move beyond engineering aspects of a product (Kiritsis 2011) . The extended concept provides a shared platform for the creation, organization and dissemination of product-related information across the enterprise throughout a product lifecycle (Ameri & Dutta 2005). This novel concept expanded OEM maintenance capability beyond organizational boundaries by introducing new data-driven maintenance concept called E-maintenance which utilizes a sophisticated analytics engine to identify systems and equipment that are not performing as expected and therefore require, or will require, maintenance (Ecosave 2017). Utilizing PLM data to drive maintenance activities is considered as a core element for such maintenance program. In considering the role of maintenance in CL2M, close relationships between maintenance activities and those in other phases of product life cycle, such as the design, production, operation, after-sales service require exchange and management of information. The information needs to be exchanged between different systems such as ERP, CM, CRM,

MES, MMIS etc. The major concern with CL2M is to exchange or access the information that is stored in products or systems with variation in technology and standards. Hence interoperability becomes a key challenge for data-driven maintenance. This challenge becomes even more complicated when the information is not stored only locally in the product or at only one server but is distributed in two or more places. Hence interoperability among different systems is an integral part of CL2M which enables flow of information between the different phases of a product lifecycle.

4.2 System interoperability

System interoperability is a growing area of interest in the industrial domain due to emerging need to integrate new, legacy systems to expand their business capability by utilizing opportunity presented by novel digital information infrastructures. Enterprises today face challenges such as “Data integration” and “Data silos” (as highlighted in the case study of Publication I) due to lack of interoperability between different systems. These systems need to be interoperable in order to achieve seamless business integration across organizational boundaries (Berre et al. 2007) and beyond. In the given context, researchers working on the domain of interoperability have defined it based on their need and context. Geraci et al. (1991) define interoperability as

“the ability of two or more systems or components to exchange information and to use the information that has been exchanged”

Similarly, other researchers/institutes have also defined interoperability as listed in Table 4.1. Much research on interoperability has been carried out in past to define interoperability between two systems or enterprises. A widely recognized model for IS interoperability is “Levels of Information Systems Interoperability” (LISI) (Group et al. 1998) describing increasing levels (Isolated, Connected, Functional, Domain, Enterprise) of sophistication regarding the ability of systems to exchange information with each other. LISI interoperability framework only focuses on technical interoperability ignoring the importance of organizational interoperability. Acknowledging this limitation, several researchers developed conceptual frameworks to cater to the need of interoperability at the organizational level. IDEAS interoperability framework (Chen et al. 2008) reflects the view that interoperability is achieved on multiple levels: inter-

Table 4.1. Definition of interoperability

Definition	source
<i>The ability of a set of communicating entities to (i) exchange specified state data and (ii) operate on that state data according to specified, agreed-upon, operational semantics.</i>	Morris et al. (2004)
<i>The ability of systems, units, or forces to provide services to and accept services from other systems, units, or forces, and to use the services so exchanged to enable them to operate effectively together.</i>	Department of defense Staff (2013)

enterprise coordination, business process integration, semantic application integration, syntactical application integration, and physical integration. Chen (2006) proposes enterprise interoperability framework with two basic dimensions (i) Enterprise dimension representing enterprise levels (interoperability aspects), (ii) Interoperability dimension representing interoperability barriers. ATHENA Interoperability Framework (AIF) (Berre et al. 2007) proposed interoperability framework at three levels: Business, Knowledge and ICT Systems. Turnitsa (2005) presents Level of Conceptual Interoperability Model (LCIM) consisting of six levels: technical, syntactic, semantic pragmatic, dynamic and conceptual interoperability. These conceptual frameworks incorporate both technical and business aspects of interoperability as defined by European Interoperability Framework (EIF 2004), which defines interoperability at three levels: organizational, semantic, and technical. Organizational interoperability focuses on modeling processes in a way to enable information exchange between different organizational units. Semantic interoperability is concerned with the meaning of data exchanged whereas technical interoperability deals with technical issues of linking computer systems and services between the administration and public. Despite the difference in definition of interoperability which is based on requirement and context, one can agree that interoperability enables flow on information fulfilling the need of information exchange.

4.3 Data exchange interoperability challenge

Interoperability is a comprehensive paradigm, that can be seen on many levels and in different domains base. In general, data exchange interoperability can be defined as the ability to communicate, exchange and process information whenever needed, whoever needed, whatever needed between two or more entities involved to manage, optimize the objective. Within the given context, Publication III presented data exchange interoperability as two-level challenge as shown in Figure 4.1.

Figure 4.1. Data Exchange Interoperability Challenge

4.3.1 Communication level interoperability

The communication interoperability provides path, infrastructures, and mechanisms for data exchange between two or more distinct organizations, networks, applications and IT solutions. In order to achieve communication level interoperability, technological and interface level interoperability need to be achieved.

Technology level interoperability

The technological level interoperability is the foundation of data exchange interoperability which enables two or more systems or entities to communicate with each other. The basic technology level interoperability can be achieved by a common agreement of the data transmission medium, low-level data encoding and rules of accessing the medium (Perumal et al. 2008). Basically, it combines hardware and software which unifies the physical communication channel and the low-level protocol structure (Järvinen et al. 2015). Since Transmission Control Protocol/Internet Protocol (TCP/IP) is

widely accepted as a solution for technology level interoperability; It can be concluded that technology level interoperability is fully developed (Kubicek et al. 2011).

Interface level interoperability

The interface level interoperability should provide the necessary communication interface which allows systems to communicate with each other for exchange of data or information whenever needed, whoever needed and whatever needed. These communication interfaces need to be general enough to be able to cater to the data exchange requirement of each of the systems. To gain time and efficiency and to avoid re-defining cooperation rules and software supporting it each time, these references must be based on business standards or norms. The business standards must be independent and weakly coupled with the technology so as to support openness.

4.3.2 Data level interoperability

While communication level interoperability enables data exchange capabilities between two or more heterogeneous systems, it does not dictate the format or interpretation of the data that has been exchanged. Communication level interoperable systems can be compared as two or more people from different countries speaking their own languages to each other where messages are received but not understood. Hence, Data level interoperability should deal with the structure and semantics of data to make systems interoperable.

Syntactic level interoperability

Syntactic interoperability refers to the format and structure of data that has been exchanged between heterogeneous systems. In order to achieve syntactic interoperability, the data exchanged between systems, which are communication level interoperable, should have a well-defined syntax, encoding and structure. These data formats should be generic enough that enable representing any data or information that will be exchanged. The syntactic level interoperability allows the heterogeneous systems to understand the structure or format of the data. However, it does not enable the involved entities to understand the meaning of exchanged data.

Semantic level interoperability

While syntactic interoperability refers to the shared format and structure of data that has been exchanged, semantic interoperability deals with the shared meaning of it between the heterogeneous systems that are designed for different purposes. The main objective of semantic interoperability is to ensure that the meaning of data that has been exchanged is understood precisely by communicating entities with the purpose to create value out of it. Generally, semantic interoperability can be achieved via ontologies which are used to define concepts and relations between those concepts on a certain domain. The ontologies can be linked with the data for logical inference for the meaning.

4.4 Building Lifecycle Management (BLM)

Wikipedia¹ defines building lifecycle management as

“the adaptation of product lifecycle management-like techniques to the design, construction, and management of buildings. ”

The main aim of BLM is to design and develop a flexible system that enables various aspects of building management such as energy management, facility management, maintenance management, traceability management. The system should be able to integrate and reuse building information and domain knowledge throughout the building lifecycle (Lee et al. 2006). BLM often begins with a physical analysis of the building to obtain a numerical representation (e.g. CAD documents). In this context, it is more than necessary to develop an appropriate BLM system at the beginning of its lifecycle in order to avoid information loss acquired during the building construction, during its use/maintenance and disposal (Bonandrini et al. 2005). A key ingredient to a successful BLM system rests upon a flexible communication infrastructure that provides sufficiently generic interfaces for exchanging all types of data consumed or provided by the actors, corporations or information systems in general, involved in the building lifecycle (Vanlande et al. 2008). Industry Foundation Classes (IFCs) provide a useful structure for data sharing among applications (Eastman et al. 2011). However, IFCs only focus on the information structure. They do not specify how to capture, exchange and synchronize information between different information systems. The current integrated IT tools also appear limited

¹Wikipedia: https://en.wikipedia.org/wiki/Building_lifecycle_management

in providing such generic communication interfaces, thus hindering the extension of existing monitoring and control systems, even when considering IFCs. In order to achieve data exchange interoperability among multiple systems or entities in BLM, interoperability needs to be achieved at two levels: (i) Communication level (ii) Data level.

4.5 Messaging standards for data exchange interoperability for BLM

Messaging standard enables interoperability, consistency, and clarity among two or more heterogeneous systems by defining methods, rules and format of data exchange. Solving interoperability challenge requires standardization at different levels. Some standards are specific to technology level interoperability such as Peer-to-peer Universal Computing Consortium (PUCC) (Sumino et al. 2007) which focuses on designing a general peer-to-peer networking platform that enhances the communication capabilities of various devices/applications via various networks. While others such as IFC provide structure for data sharing among applications (Eastman et al. 2011) and did not specify how to capture, exchange and synchronize information between different systems. It is a daunting task to develop solutions to support and enable both (communication and data level interoperability; described in Section 4.3) levels of interoperability. Some standard has nevertheless emerged for supporting both levels of interoperability which has been identified in Publication III which are oBIX (Open Building Information Exchange), ebXML (Electronic Business XML), XMPP (eXtensible Messaging and Presence Protocol), Open messaging standard (O-MI/O-DF) or JMS (Java Message Service)(detailed in Section II-B of Publication III). These messaging standards are application level standards for data exchange interoperability that rely on underlying widely accepted lower layer protocol for technology level interoperability. Adopting the effective standard is a challenging and daunting task as it depends on the nature of the system and its messaging requirements. In the context of BLM, Publication III presented three important requirements listed below.

- I to support the complete building lifecycle and related activities, from the Beginning of Life (BoL) including design, analysis, production, and construction of the building, through the Middle of Life (MoL) including its operation and maintenance, up to the End of Life (EoL)

including its recycling and disposal;

II to be able to communicate with any types of intelligent items/systems involved in the building lifecycle (sensors, actuators, RFID, databases). A key element of the communication lies in the fact that the messaging interface standard should be protocol and network agnostic to support all types of technologies and infrastructures, should be able to convey building lifecycle information (related to the building itself or activities that surround it) to relevant data systems and users either synchronously or asynchronously, and should be able to support a wide range of notification mechanisms;

III to support multiple payload formats and structures to handle many changes in data media and formats throughout the building lifetime.

Based on these requirements, Publication III presented messaging standard comparison framework (detailed in Section III of Publication III) for comparative study of identified standard for its applicability in BLM. The result of the study is presented in Table 4.2.

Given these requirements, an aspect of prime importance considering complex product lifecycles, such as building lifecycle, is support of multiple payloads and formats within a message (i.e., to meet in part requirement I). The nonsupport of this functionality is particularly critical when many actors, corporations and systems are involved in the building lifecycle. Indeed, the higher the product complexity, the higher the number of formats used to store information. In this regard, all protocols support this functionality. Considering now requirement II, some protocols are limited in some respects. First, some of them do not support synchronous and asynchronous communications (refer property 9 in Table 4.2), which is a hurdle for supporting a wide range of activities in the building lifecycle (e.g., when communicating with devices that are frequently disconnected from the network). Second, some of these protocols are not protocol-agnostic (refer property 8, e.g. oBIX is only specified for Hyper Text Transfer protocol (HTTP) and Web Service Definition Language / Simple Object Access Protocol (WSDL/SOAP protocols)) that could prevent building stakeholders from communicating using specific technologies. Third, none of these protocols, except O-MI/O-DF, have support for piggy-backing (refer property 11) that is of utmost importance to deal properly with products and systems that are mobile or located behind firewall systems (Kubler et al. 2013). Fourth,

²Previously known as Quantum Lifecycle Management (QLM)

Table 4.2. Messaging protocol comparison based messaging comparison framework

n	Property	Sub-property	oBix	ebXML	XMPP	JMS	O-MI/O-DF ²
1	Representation	Object				✓	
		Data element	✓	✓	✓		✓
2	Messaging API	Application-specific	✓		✓		
		Application-independent		✓		✓	✓
3	Initiation	Push		✓	✓	✓	✓
		Pull	✓	✓		✓	✓
4	Intermediation		✓	✓	✓	✓	✓
5	Persistence	Transient				✓	
		Persistent	✓			✓	✓
6	Subscription	Interval-based	✓			✓	✓
		Event-based	✓	✓	✓	✓	✓
7	Self-contained		✓	✓		✓	✓
8	Protocol agnostic			✓		✓	✓
9	Synchronicity	Synchronous		✓	✓		✓
		Asynchronous	✓	✓	✓	✓	✓
10	Delivery-guarantee		✓			✓	✓
11	Piggy backing						✓
12	Multiple payloads		✓	✓	✓	✓	✓
13	Processing result	Single return value		✓	✓	✓	✓
		Single integrated return value	✓				✓
		Set of individual return value	✓				✓
14	Communication	Separate message	✓	✓	✓	✓	✓
		Callback address					✓
15	Failure notification	Timeout of acknowledgement				✓	✓
		Reply with error message	✓	✓	✓	✓	✓
		Exception				✓	

only O-MI/O-DF provides the possibility of making subscription with call-back address (refer property 14) that is an important feature for real-time notifications. The lack of one of these properties has a significant negative impact on meeting requirement II. Given these observations, O-MI/O-DF seems to closely meet the major BLM requirements. Another determining factor in our decision for using O-MI/O-DF for building lifecycle management is that O-MI/O-DF can use different payload formats including, for instance, oBIX objects. Thus, oBIX can be used as the payload format when communicating with building-related information systems that already have existing support for handling oBIX object structures. Therefore, our choice is to use the O-MI/O-DF standards which are introduced in greater detail in Section 4.6.

4.6 Open messaging standards

The O-MI/O-DF has emerged out of the PROMISE EU FP6 project³ in which real-life industrial applications required the collection and management of product instance-level information for many domains such as heavy and personal vehicles, BLM, household equipment, phone switches. Information such as sensor readings, alarms, assembly, disassembly, shipping event, and other information related to the entire product lifecycle needs to be exchanged between several organizations. The project consortium is set out to find suitable standards for exchanging such information. PROMISE created two main specifications that fulfilled the necessary requirements: the PROMISE Messaging Interface (PMI) and the PROMISE System Object Model (SOM). At the end of the PROMISE project, the work on these standards' proposals was moved to the QLM workgroup of The Open Group⁴. The QLM messaging protocol is the continuity of PMI and consists of two standards proposals (Främling & Maharjan 2013), namely the QLM Messaging Interface (QLM-MI) that defines what kinds of interactions (refer to Figure 4.2) between entities are possible, and the QLM Messaging Format (QLM-MF) that defines the structure (refer to Figure 4.3) of the information included in a QLM message. Later the evolved messaging standard has renamed and published as Open Messaging Interface (O-MI) and Open Data Format (O-DF) by the The Open Group⁵ and are detailed in Section IV of Publication III.

Figure 4.2. Open messaging interface

³<http://promise-innovation.com>

⁴<http://www.opengroup.org/qlm/>

⁵<http://www.opengroup.org/iot/omi/p1.htm>

Figure 4.3. Open data format

4.7 Open messaging standard for data-driven maintenance

In the context of building lifecycle, Operation & Maintenance (O&M) play a vital role to resolve operational problems, improve comfort, optimize energy use, and recommend retrofits if necessary to its assets. Data-driven maintenance such as CBM has emerged as best maintenance practice that recommends maintenance actions based on the information collected through condition monitoring. CBM attempts to avoid unnecessary maintenance tasks by taking maintenance actions only when there is evidence of abnormal behaviors of a physical asset (Jardine et al. 2006). A CBM program, if properly established and effectively implemented, can significantly reduce maintenance cost and downtime by reducing the number of unnecessary scheduled preventive maintenance operations. However, the collection and processing of data is extremely difficult because buildings are dissimilar, are located in different areas, built at different times, and operated by a variety of owners and their agents (Krstić & Marenjak 2012). The problem becomes even more challenging when each of these agents has their own system resulting in interoperability challenges. The use case in following subsections demonstrates the applicability of O-MI/O-DF in remote monitoring to resolve the problem of data exchange interoperability between various entities involved for data driven maintenance in BLM.

4.7.1 Use case: Remote monitoring of air handling unit

Remote monitoring is the heart of data-driven maintenance, which involves data acquisition, processing, analysis, interpretation and extracting useful information to carry out maintenance task. Monitoring enables access

to the data that is required for the involved entities to carry out their task. For example, an engineering team might be interested in operational sensor data which helps them to develop fault detection, failure prediction methods whereas maintenance unit might be interested in error, alert and diagnosis information which help them to prepare for maintenance action. Hence interoperability between these entities is vital for successful implementation of data driven maintenance. In order to demonstrate how O-MI/O-DF enables data exchange interoperability between these entities, a smart home with Air Handling Unit (AHU) has been selected as a proof of concept. This use-case involves four entities namely smart home, O-MI/O-DF Middleware, OEM-Engineering Unit and OEM-Maintenance Unit which are detailed below.

Smart House: Smart house is equipped with AHU to maintain air quality throughout the house. The smart house and AHU are equipped with several sensors as shown in Table 4.3. In addition, the smart house also implements O-MI node which generates O-MI write request (depicted in Figure 4.4) of sensor values and sends to O-MI middleware every 10 seconds as shown in Figure 4.5.

Table 4.3. Air Handling Unit Sensor Details

Sensor Name	Sensor Description
TE10	Temperature at operator panel 1
TE15	Temperature at operator panel 2
TE20	The current effective supply-side fan speed
TE45	The current effective extract-side fan speed
TE60	Incoming air temperature
HRE_T_FRS	Fresh air temperature
HRE_T_SPLY_LTO	Supply air temperature after heat recovery unit
HRE_T_SPLY	Supply air temperature
HRE_T_WST	Waste air temperature
HRE_T_EXT	Room removed air temperature

O-MI/O-DF Middleware: O-MI Middleware is an open source⁶ reference message-oriented middleware developed at Aalto University

⁶<https://github.com/AaltoAsia/O-MI>

Figure 4.4. O-MI/O-DF write request from smart home to O-MI/O-DF Middleware

based on published O-MI/O-DF standards. This implementation is an agent-based system, which encapsulates the complexity of lower level communication with sensors and AHU of the smart house backend. For this study, all sensor data from smart house (namely SmartHome4456) are published in O-MI node in O-DF structure⁷. These published sensor data are updated every 10 seconds using O-MI write request. The security module (Yousefnezhad et al. 2017) in O-MI middleware handles the access control to prevent unauthorized access to the published data.

OEM-Engineering Unit: The engineering unit is responsible for various activities including design and development of fault detection and diagnosis tool to ease maintenance unit for the diagnosis of different kinds of failure in equipment.

OEM-Maintenance Unit: The maintenance unit is responsible for maintenance activities which includes ensuring smooth operation of AHU by reducing downtime subsequently, improving operations and maintenance. In this study, we focus on three use-case scenarios of remote monitoring of AHU where O-MI messaging standard can be effective.

⁷<http://otaniemi3d.cs.hut.fi/omi/node/html/webclient/index.html>

4.7.2 Use case scenario

Data collection for developing fault detection and diagnosis method:

Figure 4.5. Rate-Centric Subscription

In this use case scenario, the engineering unit is interested in developing heat recovery failure detection method for AHU. For this purpose, the engineering unit requires operational data from AHU, which is done by placing a subscription request of object “AHU4456” by setting “time to live (ttl)” parameter to -1 and interval to 10 seconds along with callback address as shown in Figure 4.5. Then, this subscription request is sent to O-MI middleware using HTTP/POST. Once the O-MI middleware receives the subscription request, it will respond immediately with subscriptionId which can be used by engineering unit in future to cancel the subscription (as depicted in Figure 4.6). Once the subscription is placed, O-MI middleware starts sending operational data of all InfoItems (sensors value) of “AHU4456” every 10 seconds in the provided callback address in O-DF format.

```

1 <omiEnvelope xmlns="http://www.opengroup.org/xsd/omi/1.0/"
  version="1.0" ttl="0">
2   <cancel>
3     <requestID>REQ4456</requestID>
4   </cancel>
5 </omiEnvelope>

```

Figure 4.6. Cancel subscription request

Based on the collected data, Publication V proposes a process-history based method to detect Heat Recovery Unit (HRU) failure of balance AHU using the generic principal of nominal efficiency as shown in Equation 4.1

(detailed in the Section 4 of Publication V).

$$\mu_{nom} = \frac{T_{ext} - T_{wst}}{T_{ext} - T_{frs}} \tag{4.1}$$

(a) Accuracy at different cutt-off (b) Confusion matrix with cut-off at 0.7

Figure 4.7. Results

The proposed method utilizes Principal components analysis (Pearson 1901) in combination of logistic regression (Cox 1958) to detect HRU failure. The technical details of the proposed methodology are presented in Section 4 of Publication V. The test results of the proposed method are shown in Table 4.4 and Figure 4.7a. Figure 4.7a shows that the accuracy of our proposed method for “Train” and “Test” at different cut-off values. It is clearly seen that accuracy of the model is increased from 95% to 97% when the cut-off is changed from 0.3 to 0.8. In addition, Figure 4.7b depicts the tradeoff choosing a reasonable cut-off. If we increase the cut-off value, the number of True Negative increases and the number of True Positive decreases i.e, if we increase the cut-off value, the number of False Positive falls, while the number of False Negative rises. With the achieved results, the proposed method can detect HRU failure with the accuracy of at least 95% with given test and train dataset.

Table 4.4. Performance metrics of the proposed methodology to detect HRU failure

Cut-off value	Accuracy	Sensitivity	Specificity
0.2	0.95	0.95	0.95
0.4	0.96	0.91	0.98
0.6	0.97	0.91	0.99
0.8	0.97	0.91	1.00

Figure 4.8. Event-centric subscription

Use case: remote monitoring of Air handling unit for fault notification:

Failure notification is one of the important activities of the maintenance process, which aids the maintenance planner to plan and execute maintenance activities. Employing remote monitoring technologies for early failure notification helps to ensure smooth operations of equipment and also increases the lifetime of equipment which might be reduced if not dealt immediately. For such notification service, O-MI messaging standard provides event-centric subscription for remote monitoring.

In this use case scenario, the maintenance unit utilizes event-centric subscription for failure notification, by creating an event-centric subscription of InfoItem “ErrorMsg” and “AlarmMsg” for AHU4456. The event-centric subscription request is created by setting the “interval” and “ttl” parameters of subscription request to “-1” along with callback address as shown in Figure 4.8. Then, the subscription request is sent to O-MI middleware using HTTP/POST. Once the O-MI middleware receives the request, it responds immediately with subscriptionId which can be used later by the maintenance unit to cancel the subscription. Once the subscription is placed, O-MI middleware will notify the maintenance unit about only the latest errors (along with subscriptionId) immediately after they are detected. Figure 4.8 shows that there are new errors at time $t=10$ and $t = n$ which are immediately notified to the maintenance unit by O-MI middleware.

Use case: remote monitoring of Air handling unit for reparation monitoring

Reparation monitoring is another important task of the maintenance unit. Applying remote monitoring technologies for reparation monitoring

Figure 4.9. Time-centric subscription

helps to ensure the quality of maintenance by monitoring operational behavior of equipment once maintenance is carried out. For such services, O-MI messaging standard provides a time-centric subscription for remote monitoring.

In this use case scenario of reparation monitoring, the maintenance unit can create a time-centric subscription request of InfoItem AHU4456-TE10 of AHU4456 for 2 hours at the rate of 20 seconds. This subscription request is created by setting “ttl” parameter to 7200 (in seconds) and “interval” to 20 seconds along with callback address as shown in Figure 4.9. This subscription request is then sent to O-MI middleware using HTTP/POST. Once the O-MI middleware receives the subscription request, it will respond immediately with subscription ID which can be used by the maintenance unit to cancel the subscription before “ttl” expires. Once the subscription is placed, O-MI middleware will send the value (both old and new) of AHU4456-TE10 every 20 seconds for the next two hours (as depicted in Figure 4.9) highlighting operational behavior of AHU. Once the provided “ttl” in the subscription is expired, the request will be automatically canceled and maintenance unit will not receive any message from O-MI middleware. This type of time-limited subscription is quite useful for reparation monitoring; once the maintenance is carried out, it helps to monitor how equipment is performing after maintenance to ensure maintenance quality.

4.8 Results and discussion

The use cases above highlight three different data-driven maintenance scenarios where an open messaging standard enables data exchange by solving interoperability challenges between the different entities involved. For the first use-case scenario, the engineering unit utilizes “**read**” (as depicted in Figure 4.10a) messaging interface provided by O-MI to subscribe data from object of interest (“AHU4456” in this case). This subscription request utilizes O-DF structure to specify the object of interest as shown in Figure 4.10a. Similarly, O-MI middleware sends the response every 10 seconds by providing requested data using O-DF structure shown in Figure 4.10b.

Figure 4.10. Data collection for developing fault detection and diagnosis method

In the second use case scenario, the maintenance unit utilizes “**read**” (as depicted in Figure 4.11a) with different parameters (as detailed in Section 4.7.2) to create an event-centric subscription to get a notification with an error message when a failure occurs. The request from maintenance unit to O-MI/O-DF middleware and response from O-MI/O-DF middleware to maintenance unit is shown in Figure 4.11.

Figure 4.11. Remote monitoring of Air handling unit for fault notification

In the third use case scenario, maintenance unit utilizes “read” (as depicted in Figure 4.12a) with different parameters (as detailed in Section 4.7.2) to create time-limited subscription to get sensor data for a predetermined interval of time which is “7200” seconds in this case. The request from the maintenance unit to O-MI/O-DF middleware and response from O-MI/O-DF middleware to maintenance unit is shown in Figure 4.12.

Figure 4.12. Remote monitoring of Air handling unit for reparation monitoring

During the middle of life (operation and maintenance) phase of BLM, several entities such as maintenance unit (planning, scheduling and executing maintenance activities), engineering unit (responsible to develop tools to support data-driven maintenance such as CBM), owner (or agent) of building need to exchange data/information to ensure smooth operation of building and its assets. In such a scenario, the O-MI/O-DF enables data exchange interoperability between these entities by providing two standards;

i) Open messaging interface (O-MI) provides interfaces which enable communication level interoperability between two or more systems. In

the use case scenario detailed above, O-MI provides three modes of “read” messaging interface for creating three types of subscription which are required for data collection (as depicted in Figure 4.10a), event notification (as depicted in Figure 4.11a) and reparation monitoring (depicted in Figure 4.12a). These modes of data exchange are a vital requirement for any data-driven maintenance. In addition, O-MI also provides “write” (depicted in Figure 4.4) messaging interface for writing or updating data as seen in use-case for updating sensor data of smart home to O-MI/O-DF middleware and “cancel” (as depicted in Figure 4.6) messaging interface is used to cancel subscription which is no longer necessary. The O-MI messaging interface is not only useful in data-driven maintenance, but its application can extend the whole product lifecycle as highlighted in Främling et al. (2013).

ii) Open data format (O-DF) provides data level interoperability by defining the structure or format of the message for syntactic interoperability and provision to embed domain specific ontologies to provide semantic interoperability. It is defined as a basic ontology using eXtensible Markup Language (XML) schema that is generic enough for representing “any” object and information needed for information exchange. The “Objects” element can contain any number of “Object” sub-elements. “Object” elements can have any number of properties, referred to as InfoItem(s), as well as “Object” sub-elements. The resulting Object tree may contain various number of levels as shown in Figure 4.10, Figure 4.11, and Figure 4.12 of the aforementioned use-case scenarios. However, one of the limitations of these use-case scenarios is their inability to demonstrate the semantic interoperability between different involved entities. Nevertheless, the capability of O-DF to enable semantic interoperability by semantic enrichment of object has been well documented in (Karpenko et al. 2018) within the context of mobility domain.

5. Data extraction methodology

Publication VI is used to answer the fourth research question *“How to extract data from complex information systems for data warehouse(DW) by leveraging the end users’ knowledge?”*. This research question was motivated by the data extraction challenge many organizations face due to the complex nature of the modern information system. More concretely, complications arise at the stage of mapping of the source schema with the target schema during the Extract, Transform, and Load (ETL) process of data warehousing.

In this regard, Publication VI proposes a data discovery methodology for ETL and its implementation which utilizes end users’ knowledge about the process supported by specific IS. This method maps the processes supported by IS with its underlying database schema, which helps to locate the right set of data that needs to be extracted and mapped for DW.

5.1 Data extraction in Extract, Transform and Load

Data-driven decision making has emerged as a decision-making tool in many organizations, which provides factual insight into organizational processes and performances. Developing such a tool requires data to be collected through multiple heterogeneous systems that are used in an organization into a single coherent system called DW. Building such DW requires pre-processing steps that are commonly referred as ETL. The ETL is an end to end process taking data from one or more system(s), loading it into a single unified system after processing and transformation (Agrawal et al. 2008). The implementation of such a process requires close focus and understanding on three main areas: the source system, the destination system, and the mapping (El-Sappagh et al. 2011).

Given the industrial/commercial relevance of the problem, several re-

searchers have been working on different aspect of ETL process. Golfarelli et al. (1998), Caruccio et al. (2014) proposed a graphical formalism for conceptual modeling of ETL process while Vassiliadis et al. (2002) proposed conceptual modeling dedicated to the design and documentation of the DW refreshment process. Another conceptual modeling framework (Trujillo et al. 2001, Lujan-Mora & Trujillo 2003) was proposed by leveraging on the Unified Modeling Language (UML) as a formalism for their model. Similarly, other work has focused on mapping aspect of ETL process based on schema structure (Madhavan et al. 2001), query over the source and target schema (Fagin et al. 2009, Rifaieh & Benharkat 2002), domain ontologies (Touma et al. 2015, Jiang et al. 2010).

The first step in any ETL scenario is data extraction which is responsible for extracting data from the source systems typically enterprise information systems which consist of two phases: initial extraction and changed data extraction (El-Sappagh et al. 2011). During initial extraction having data from different data sources, the ETL expert should be aware and understand the structure and semantic of data. In other words, understanding the data model of an IS practically means to grasp a part of its database schema and application logic. However, this task is often challenging because DW designers are not usually experts in all IS used in different business domains (Viaene 2013) and it is quite easy to get lost in hundreds of tables with hundreds of columns. In order to understand these source schemas, in some of these techniques, data engineers had to typically rely on UML or Entity Relation (ER) (El-Sappagh et al. 2011). Unfortunately, UML diagrams do not provide a detailed understanding of data models, while ER diagrams have a serious practical limitation in their inability to cope with complexity (Feldman & Miller 1986, Kimball & Ross 1996). In addition, the ER model lacks explicit abstraction mechanisms for managing the size and complexity of real world data models (Weber et al. 1997). With large numbers of entities, complexity quickly becomes overwhelming and as a result, data models become very difficult for people to understand (Moody 1991, Kimball & Ross 1996). In some cases, understanding data model of complex IS becomes almost impossible as some vendors deliberately obfuscate their data model (Marcl 2010) and relationships between tables of data model are often encapsulated within the application logic (Pieterse & Olivier 2012).

In order to ease the data extraction process by leveraging end user knowledge about the understanding of IS, Publication VI proposes a data

discovery methodology for ETL and its implementation which utilizes end users knowledge about process supported by specific IS. This method maps the processes supported by IS with its underlying database schema which helps to understand and locate the right set of data that need to be extracted and mapped. The proposed methodology is detailed in Section 5.2.

5.2 Proposed data discovery methodology for data extraction

The proposed method is a three-step process that helps to guide ETL experts in the identification of the desired data from a pool of tables of an arbitrarily complex database by leveraging the end users' knowledge and understanding of the specific IS. The core concept is based on tracking database changes while users are performing their daily operations in the front-end, typically a GUI of an IS. As most IS uses database management system (DBMS), a natural choice to track events/changes within a DBMS is to use triggers. Triggers are the procedures that are automatically executed in response to certain events such as INSERT, UPDATE, or DELETE on a particular table or view in a database. Triggers are mostly used for maintaining the integrity of the information in the database. However, in the proposed methodology, triggers are used to log changes made by IS to its data model in order to map user action (eg. creating workOrder) in IS with corresponding tables and columns of the underlying data model. Figure 5.1 shows the three-step process of the proposed methodology which is discussed below.

Figure 5.1. Data discovery steps for data extraction

5.2.1 Step 1: Deploy database trigger

The first step of the proposed method is to create and deploy triggers that logs the changes made by IS in its data model. This step is based on Algorithm 1 which is detailed in Publication VI. This algorithm dynamically creates required temporary tables and triggers.

Algorithm 1: Trigger creation procedure

Input : Database Credentials

Output : Successful creation of triggers and temporary tables to track changes in deployed database schema

```

1 TableList  $\leftarrow$  get_TableLists(IsTemp = False)
   /* Return list of tables for creating triggers */
2 foreach Table  $\in$  TableList do
3   ColumnList  $\leftarrow$  ColumnLists(Table Name)
   /* Return list of columns for particular table */
4   CreateTemporaryTables(Table Name, ColumnList)
   /* Creates temporary table to log changes. */
5   foreach Action  $\in$  (Insert, Update, Delete) do
6     Query  $\leftarrow$  generateTriggerQuery(Table, ColumnList, DatabaseType, Action)
     /* Return SQL query for creating trigger based on action and
        databasetype */
7     DeployQuery(Query)
     /* Deploy trigger in the database for logging track changes */
8   end
9 end

```

5.2.2 Step 2: Record the changes

Once the required temporary tables and triggers are deployed in step 1, this step is responsible for logging the changes that IS made to its underlying data. For this purpose, the IS end-user will provide a task name describing the operation (e.g. create maintenance request, create work-order, etc.) and then followed by task on the IS'GUI, involving inputting or reading data. Here the deployed trigger will log every change made by the IS on its database on deployed temporary tables.

5.2.3 Step 3: Data discovery for data extraction

This is the final step of the proposed methodology. In this step, report is generated listing the tasks that had been carried out with the associated list of tables and columns that have been modified in underlying database

during its operations. The ETL expert can review this report to locate and understand where certain type of data related to the task carried out by the end-user is stored in the database. Mapping the process or task supported by IS with its underlying database helps ETL experts to locate the right set of data that need to be extracted for ETL process.

5.3 Results and discussion

In this study, the proposed data discovery methodology had been applied to map processes supported by IS with its underlying database schema to locate the right set of data that needs to be extracted and mapped for DW. The rationale behind the development of the proposed methodology is based on the fact that the most complex task of any ETL process is understanding the structure and semantics of the data to be imported. This step is often given as pre-requisite in literature, but in reality the personnel performing the ETL, in many cases, have a very limited knowledge regarding the data they are manipulating. Grasping even partially the data model of a commercial IS can be very frustrating and it is quite easy to get lost in hundreds of tables with hundreds of columns. The typical setup of an ETL process is often far from ideal and the following conditions generally occur:

- ETL experts usually are not aware which data and processes are supported by the IS from where they are extracting the data.
- Domain experts, such as the IS end-users, can provide insight when it comes to data and processes, but they are seldom technically competent to understand the underlying IS architecture and database schema.
- Some information systems deliberately obfuscate their data model in order to prevent reverse engineering (MarcL 2010).
- The sheer size of the database, in terms of the number of tables and relationships, often makes it difficult to browse and determine where certain data is stored (Kim & Everest 1994, Earl 1993).
- The relationships between tables are not always explicit as they are often encapsulated within the application logic (the controller in the MVC model) (Pieterse & Olivier 2012).

As part of the research, a data model logger tool supporting the proposed methodology (detailed in Section 5.2) has been developed (depicted in

Figure 5.2) to ease ETL experts for data warehousing. Since ETL experts are not always aware which data and processes are supported by the IS from which they are extracting the data, the developed tool leverage on the end-users’ knowledge about the specific IS for guiding the ETL expert in the discovery of the relevant pool of tables from the underlying IS data model.

Figure 5.2. Data Model Logger

The three steps of the proposed methodology are supported by the data model logger tool as shown in Figure 5.2. Figure 5.2a corresponds to deploying database triggers and temporary tables in the DBMS of IS that are used to track the changes IS made in its underlying database. In this step, the user must provide valid database credentials (with permission to create and drop triggers and tables) of the underlying database (refer to Figure 5.2a). Once the user clicks on Deploy Database Logger button, the tool dynamically generates necessary scripts to create temporary tables and triggers on tables and deploys them in the database. For each table, three types of triggers are created, specifically once for each operation of SQL (INSERT, UPDATE, DELETE). After successfully deploying the trigger, the end user will provide a task name describing the operation (e.g. create maintenance request, create work-order, publish a post, etc.) that the IS end user is about to perform on the IS. At this point, the ETL experts will click on the START button (refer to Figure 5.2b) and the IS end user will start performing the planned task by operating on the GUI of an IS. Every change made by the IS Controller on its database is logged in our temporary tables which helps to map the process supported by IS with underlying data-model. The final step is the discovery phase. Here, once the IS end user completes the task, the data engineer can click on the STOP button (refer to Figure 5.2b). At this point, Results tab (refer to Figure 5.2c), shows the list of inserted, deleted or updated table, their columns and the specific row data that has been created or updated. This

result is vital for the initial extraction phase of the extraction process of ETL which consists of two stages: initial extraction and changed data extraction. The initial extraction phase is a one time process which involves locating the data from the source schema that need to be mapped into the target schema for further extraction. Based on initial extraction, further extraction can be carried out, which is called changed data extraction in the ETL process.

In order to validate and test general applicability of the proposed methodology, four different ISs with its database deployed in different DBMSs are selected. The list of selected ISs and their underlying databases is shown in Table 5.1.

Table 5.1. List of IS and their corresponding database

Information system	Database	Number of tables
WordPress	MYSQL	12
openMAINT	PostgreSQL	581
SharePoint	MSSQL	2827
OpenCms	Oracle	40

Wordpress:

WordPress is a free open-source Content Management System (CMS) based on PHP and MySQL. It is one of the most popular CMS, used by more than 27.5% of the top 10 million websites as of February 2017 (W3Techs 2017). Typical tasks in WordPress include create and publish posts, manage users, install plugins etc. The full list of tested operations (Tasks) and the related changes in the underlying database are reported in Table 5.2.

openMAINT:

openMAINT is an asset management application for managing buildings, installations, and maintenance activities. It is a complex system with more than 581 database tables deployed in PostgreSQL. Common openMAINT tasks and corresponding database changes are listed in Table 5.3.

Table 5.2. Changes log for wordpress

Tasks	SQL insert	SQL update	
	Tables name	Tables name	Columns name
Add new user	wp_usermeta	-	-
	wp_users	-	-
		-	-
Add new page	wp_options	-	-
	wp_postmeta	-	-
	wp_posts	-	-
	wp_term_relationships	-	-
Install plugin(Buddy Press)	wp_options	wp_options	option_value
Add new post	wp_options	wp_posts	Guid
	wp_postmeta		post_name
	wp_posts		post_date
	wp_term_relationships		post_title
			post_status
			post_modified
			post_date_gmt
			post_modified_gmt
		wp_term_taxonomy	Count
		wp_options	option_value
		wp_postmeta	meta_value
Post a comment	wp_comments	-	-
Approve comment		wp_comments	comment_approved
		wp_posts	comment_count

SharePoint:

SharePoint is the most widely used enterprise collaboration and document management platform developed by Microsoft. Hence, SharePoint naturally leverage on MSSQL Server database to store SharePoint data. SharePoint is one of the most advanced general-purpose CMS, providing sophisticated managed metadata service for improving filtering while performing searches. The full list of tested operations (Tasks) and the related changes in the underlying database are reported in Table 5.4.

OpenCMS:

OpenCMS is a powerful web-based CMS designed and developed to create and maintain websites quickly and efficiently. OpenCMS can be deployed with different database systems such as MySQL and Oracle. In this case study, OpenCMS was deployed with Oracle 11 Express edition. The full list of tested operations (Tasks) and the related changes in the underlying database are reported in Table 5.5.

The results highlight that the proposed methodology is able to map the process supported by IS with its underlying data model by utilizing end users' knowledge in particular IS which helps ETL experts to understand and locate the right set of data from the database. In many complex IS,

Table 5.3. Change log for openMAINT

Tasks	SQL insert	SQL update	
	Tables name	Tables name	Columns name
Create maintenance activity	maintenanceactivity map_maintenancegroupmaintenanceactivity map_servicecategorymaintenanceactivity map_servicemaintenanceactivity map_suggestedassigneeactivity map_teammaintenanceactivity	maintenanceactivity	BeginDate
Update maintenance activity (ActivityStatus)	maintenanceactivity maintenanceactivity_history map_maintenancegroupmaintenanceactivity map_servicecategorymaintenanceactivity map_servicemaintenanceactivity map_suggestedassigneeactivity map_teammaintenanceactivity	MaintenanceActivity - -	ActivityStatus User
Add rent contract	map_mainpartycontract rent		
Update rent contract (StartingDate)		rent	BeginDate StartingDate
Create maintenance request	shkactivities shkactivitydata shkassignmentstable shkprocessdata shkprocesses shkprocessrequesters	objectid	next
		shkcounters	Objectversion the_number
		shkprocesses	Objectversion State started laststatetime
		shkactivities	Objectversion State laststatetime
		MaintenanceRequest	Description BeginDate Type Details
		shkprocessdata	Objectversion variablevalue

only a small portion of the IS database might contain relevant data to be imported in a DW. For instance, openMAINT, an asset management system, supporting complex workflows for streamlining maintenance activities of building, equipment, heating systems where the database schema counts 581 tables. However, during our tests while performing typical openMAINT tasks, only 20 tables were actually modified. Understanding a pool of 20 tables with a clear relationship between performed operation and modified tables/columns aids better understanding of the semantic and structure of the database schema which in return reduces the complexity of understanding the entire schema. In addition, some ISs provide APIs (for example, SharePoint) for accessing programmatically the data they manage. However, the complexity of these APIs is generally directly proportional with the complexity of the ISs and by reflection its database

Table 5.4. Changes log for SharePoint

Tasks	SQL insert	SQL update	
	Tables name	Tables name	Columns name
Create new group	ECMChangeLog	-	-
	ECMGroup	-	-
		-	-
Create new termset	ECMChangeLog	-	-
	ECMTermSet	-	-
		-	-
Create term	ECMChangeLog	-	-
	ECMTerm	-	-
	ECMTermLabel	-	-
	ECMTermSetMembership	-	-
Create new group	ECMChangeLog	-	-
	ECMGroup	-	-
		-	-
Update group	ECMChangeLog	ECMGroup	Name Description LastModifiedTime
Update term	ECMChangeLog ECMTermLabel	ECMGroup	LastModifiedTime UseCount Owner IsDeprecated

schema. Therefore, using these APIs correctly might be an overwhelming task because it requires an understanding of the entire API for data extraction. In such case, ETL expert needs to spend time and effort to understand API to extract few relevant data from the database. In order to help ETL expert to extract data for ETL process, the proposed method aids to locate the right set of data by providing the context of how these data are generated. The main benefit of the proposed methodology is the ability to leverage the knowledge of the IS end users, for identifying and performing a set of tasks (directly on the Graphical User Interface (GUI) of IS) involving the data items that should be imported in the DW. The proposed method allows ETL experts to pinpoint exact tables and columns of the database that have been updated/inserted, while understanding the context (operation performed and data semantics) in which that data was created/modified. In conclusion, the proposed methodology simplifies the extraction and mapping of complex data sources into a DW. This technique is generally applicable to any IS using a DBMS and it can be easily integrated with any existing ETL process.

Table 5.5. Changes log for openCMS

Tasks	SQL Insert	SQL Update	
	Tables Name	Tables Name	Columns Name
Create a Page	cms_offline_structure cms_offline_resources cms_offline_properties cms_offline_contents_relationships	cms_offline_resources	project_lastmodified resource_type resource_flags resource_size date_content sibling_count date_lastmodified user_lastmodified
			property_value
		cms_offline_structure	resource_id parent_id resource_path rstructure_stat resource_path date_released date_expired
		cms_userdata	data_type
Publish a Page	ccms_online_structure cms_history_projects cms_history_resources cms_online_resources cms_history_properties cms_history_structure cms_online_properties cms_publish_history cms_publish_jobs cms_contents cms_user_publish_list	cms_contents	online_flag
		cms_publish_jobs	project_id project_name user_id publish_locale publish_flags resource_count enqueue_time start_time finish_time
		cms_userdata	data_type
		cms_users	user_firstname user_lastname user_email user_lastlogin user_flags
		cms_online_structure	resource_id parent_id resource_path structure_state date_released date_expired

6. Conclusion

6.1 Summary

The research reported in this thesis focused on challenges that OEM is facing to provide effective data-driven maintenance service. Traditionally, maintenance scheduling, asset availability forecasts, remaining useful life, fault detection and diagnosis, as well as spare-part inventory forecasting have been based on expert assessments and intuition from personal experience. This personal understanding is invaluable; yet, decision making is a complex process especially where downtime is extremely costly and new solutions are demanded to optimize maintenance activities. Most assets have data readily available that have been collected at different stages of product lifecycle. Despite the availability of data as well as advanced statistical and analytical tools, companies are still struggling to fully harness effective data-driven decision making due to several data-related pitfalls. This dissertation aimed to identify those key data-related pitfalls that hinder effective data-driven decision making in maintenance. This study made attempts to systematically study each of the bottlenecks and tried to provide the possible solutions.

The work was done in stages and in a systematic way using the combination of research methodology as detailed in Section 1.3. The first stage focused on the descriptive study of literature which highlights importance and challenges of data in the data driven maintenance. It is supplemented by quantitative and descriptive case studies on two multinational Finnish manufacturers for highlighting key data-centric challenges that hinder companies to carry out effective data-driven decision-making maintenance such as CBM.

The aim of this work is to identify and provide possible solutions for

each of the identified challenges. This was based on the assumption that providing a solution of identified challenges helps OEM companies to improve the effectiveness and efficiency of maintenance decision making by adding value from data to their expert assessments of maintenance activities. Hence, the central research objective of this dissertation is

To identify and mitigate data-centric challenges in equipment manufacturing companies which obstruct the effectiveness of data-driven decision making in maintenance.

The work is studied based on the central research question from four complementary viewpoints with six research articles. research question 1 focuses on pinpointing data centric challenges while research question 2, research question 3, and research question 4 proposed the solutions for the identified challenges. However, the evaluation of the success or applicability of solution was accomplished through validation in each research contribution. The assessment is by theoretical evaluation and through case study example to demonstrate the support. In addition, there was verification by acceptance through publication to the relevant scientific community (Buur 1990).

6.2 Contribution

The main contribution of the research presented in this dissertation is answered in four research questions.

Key data-centric challenges

The first research question defines the problem to identify the data-centric challenges that hinder companies to carry out effective data-driven maintenance such as CBM. The answer to this question is shaped based on the experience and result of case study of two multinational OEM in Publication I, where five data-centric challenges at different business levels have been identified as shown in Table 6.1. The identified challenges are broadly grouped into three categories that are detailed in Section 2.6. Data-driven decision making requires effective utilization of available data assets to add value from data to improve maintenance activities. Hence, I believe these identified challenges can be the key bottleneck for data-driven maintenance and need to be resolved.

Table 6.1. Key data centric pitfalls

Key challenges	Business level	Challenges
Wrong or incomplete data reported	Operational level	Data quality
Lack of data validation technique	Operational level	Data quality
Sensor data not logged for CBM	Strategic level	System inter-operability
Machine data and service reports are hard to integrate	Strategic level	System inter-operability
Data silos	Strategic level	Data extraction

Data Quality assessment of maintenance reporting procedure

The second research question “How can data quality assessment of maintenance reports be carried out?” has been answered by proposing AHP-based MRQA (detailed in Section 3.2.2) framework in Publication II. The proposed framework defines three data quality dimensions namely “Believability”, “Completeness” and “Timeliness” based on Krogstie framework (Krogstie et al. 1995) where each criterion has its own quality indicators. These quality indicators (detailed in Table 3.2) are derived from the maintenance reports generated from maintenance reporting process. The framework utilizes these quality dimensions to quantify the maintenance reporting procedure of an organization. The applicability of the proposed framework via MRQA dashboard has been demonstrated in the use case presented in Section 3.3, which is used to assess and rank maintenance reporting quality of 54 branch offices of the case company. I believe, maintenance reporting quality plays an important role in data driven decision-making and assessing the quality of maintenance report helps decision makers critically think about the decision that will be carried out based on those reports.

Data exchange interoperability challenges in data driven maintenance

The third research question considers how open standards enable inter-system interoperability for data-driven maintenance. The answer to this question is based on experience and results from Publications III, IV and V which are detailed in Chapter 4. I believe the key reason for data-centric challenges “Sensor data not logged for CBM” and “Machine data and service reports are hard to integrate” (discussed in Publication I) are due to the lack of data exchange interoperability between different systems

deployed in case company. “Sensor data not logged for CBM” (detailed in Section 2.6) is due to serious limitations in architecture where access to the data are limited to the embedded software which does not have any interface to share data with other units of organization such as the maintenance unit resulting in other challenges “Machine data and service reports are hard to integrate” (detailed in Section 2.6) and “Data Silos (detailed in Section 2.6)”. In order to address such challenges, Publication IV defines data exchange interoperability as a two level challenge: communication level challenge and data level challenge. Communication level challenge consists of technology level challenge and interface level challenge whereas data level challenge consist of syntactic level and semantic level challenge. For any companies moving towards data-driven maintenance, I believe that there is a strong need of an open messaging standard that enables data exchange interoperability, providing a mechanism to share structured data with its semantic. Publication IV provides a detailed extensive and comparative study of such standards. The result of the comparative study shows that an open messaging standard (O-MI/O-DF) is the best candidate for data exchange interoperability. The application of the identified messaging standard O-MI/O-DF in three different use case scenarios: data collection, fault notification and reparation monitoring in the context BLM is highlighted in Publications III and IV.

Data Extraction for ETL

The answer to the fourth research question “How to extract data from complex information systems for data warehouse (DW) by leveraging the end users’ knowledges?” has been presented by proposing data extraction methodology (detailed in Section 5.2) in Publication VI. The proposed method tracks the changes made by IS to its underlying database while end users use user interface to carry out their task. This method helps to map the process supported by the IS with its underlying tables and columns in the database schema. This mapping will be very helpful for DW designer to understand the structure of the database from where the data need to be extracted for DW. The rationale behind such method is based on the fact that the most complex task of any ETL process is understanding the structure and semantics of the data to be imported. This step is often given as pre-requisite in literature, but in reality, the personnel performing the ETL process have, in many cases, a very limited knowledge regarding the data they import. This is especially true when there are multiple data silos created by heterogeneous ISs deployed in an organization to support

their organizational processes (as detailed in Section 2.6). Understanding the data model for each and every IS for data extraction is a complicated and time consuming task as some ISs deliberately obfuscate their data model in order to prevent reverse engineering (Marcl 2010). Sometimes the relationships between tables are not always explicit as they are often encapsulated within the application logic (Pieterse & Olivier 2012). Hence, I believe, the mapping between IS processes with underlying table and columns in database schema helps DW designer to understand and locate the data that need to be extracted for DW.

6.3 Limitation and avenue for future research

As most academic research, research carried out in this dissertation contains limitations stemming from the chosen methods, theories, and scope. The primary methodological limitation of this research is the use of the case study methodology to identify data-centric challenges that hinder companies to carry out effective data-driven decision making. The case study method has a limited ability to draw generalizations to the population or universe; however, it can be generalized to theoretical propositions (Yin 1994). In addition, the scope of this research was delimited to data-centric challenges that have been identified in Publication I which provides an answer to research question 1. However, the objective of this dissertation is not to claim that the challenges are limited to those identified in Publication I. In the future, further research needs to be carried out to identify other data-centric challenges that hinder effective decision making in MMs. However, this work can be a starting point for companies that are moving towards data-driven decision making for maintenance management. As stated, scope of this dissertation is limited to these identified challenges where each challenge is explored and studied to propose a solution.

The second research question regarding the data quality assessment of the maintenance reporting procedure is based on a use case with the objective to assess the reporting quality of 54 maintenance sites of the case company. Publication II proposes an AHP based framework for assessing maintenance reporting procedure, which relies on the Krogstie's data quality framework, where only a few concepts and relationships are considered to match the requirement and relevancy for the case company. Direct utilization of the proposed framework for other companies should be considered with caution. However, the proposed method might have

some transferable knowledge that can be utilized for other companies with the objective to assess reporting quality. In future research work, the proposed AHP framework and underlying criteria should be extended to take into consideration the other concepts/relationships such as language quality (e.g., for domain appropriateness, participant knowledge appropriateness.). Such an extension might potentially require to combine AHP with other tools and techniques for semantic processing and matching purposes, for example, handling uncertainty and vagueness in the expert judgments/preferences (e.g., using fuzzy logic).

The third research question regarding data exchange interoperability challenge is focused on the applicability of an open messaging standard for data exchange interoperability within the product lifecycle. The research carried out was based on use case scenario of maintenance which only considered the middle of life product lifecycle. The use case scenarios are based on requirements that came from experience and learning from case companies. The test and validity of the open messaging standard is proven to enable data exchange interoperability within the use case. In addition, there was verification by acceptance through publication to the relevant scientific community. However, the real test of the validity of standard has not been tested or implemented in an industrial setting due to the limitation of time and scope of the research project. As of future work, the applicability of the open messaging standard needs to be tested and validated in an industrial setting and its applicability needs to be tested considering the whole product lifecycle.

The fourth research question regarding data extraction challenges many organizations face due to data silos created by heterogeneous IS is answered by proposing a methodology in Publication VI. One of the key limitations of the proposed methodology is lack of integration with existing ETL framework. As of future research work, integration of the proposed methodology with the existing ETL framework should be carried out to extend ETL to D-ETL(Discovery-Extract-Transform-Load). As of artifacts developed in Publication VI, currently it supports only four databases, namely, Oracle, MSSQL, PostgreSQL, and MySQL. In the future, additional support for other databases such as DB2 and SQLite need to be added. Reporting is another area that needs to be improved. Now, the report displays a list of changed tables and related affected columns along with activities the user is conducting. In addition, artifact could be extended to suggest SQL query for data extraction from underlying database.

References

- Agrawal, H., Chafle, G., Goyal, S., Mittal, S. & Mukherjea, S. (2008), An enhanced extract-transform-load system for migrating data in telecom billing, *in* 'Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on', IEEE, pp. 1277–1286.
- Ahmadi, A., Gupta, S., Karim, R. & Kumar, U. (2010), 'Selection of maintenance strategy for aircraft systems using multi-criteria decision making methodologies', *International Journal of Reliability, Quality and Safety Engineering* **17**(03), 223–243.
- Al-Najjar, B. (1996), 'Total quality maintenance: an approach for continuous reduction in costs of quality products', *Journal of Quality in Maintenance Engineering* **2**(3), 4–20.
- Al-Najjar, B., Alsyoud, I., Salgado, E., Khoshaba, S. & Faaborg, K. (2001), *The economic importance of maintenance planning when using vibration-based maintenance*.
- Alarcon, M. J., Grau, J. B. & Torres, J. (2007), Application of electre i method to restoration actions in telecommunication network maintenance, *in* 'Industrial Electronics, 2007. ISIE 2007. IEEE International Symposium on', IEEE, pp. 3430–3434.
- Almeida, A. T. (2012), 'Multicriteria model for selection of preventive maintenance intervals', *Quality and Reliability Engineering International* **28**(6), 585–593.
- Almeida-Filho, A., Ferreira, R. J. & Almeida, A. (2013), A dss based on multiple criteria decision making for maintenance planning in an electrical power distributor, *in* 'International Conference on Evolutionary Multi-Criterion Optimization', Springer, pp. 787–795.
- Ameri, F. & Dutta, D. (2005), 'Product lifecycle management: closing the knowledge loops', *Computer-Aided Design and Applications* **2**(5), 577–590.
- Azadeh, A., Sheikhalishahi, M., Firoozi, M. & Khalili, S. (2013), 'An integrated multi-criteria taguchi computer simulation-dea approach for optimum maintenance policy and planning by incorporating learning effects', *International Journal of Production Research* **51**(18), 5374–5385.
- Azadeh, A., Sheikhalishahi, M., Khalili, S. M. & Firoozi, M. (2014), 'An integrated fuzzy simulation–fuzzy data envelopment analysis approach for optimum maintenance planning', *International Journal of Computer Integrated Manufacturing* **27**(2), 181–199.

- Azizi, A. & Fathi, K. (2014), 'Selection of optimum maintenance strategies based on a fuzzy analytic hierarchy process', *Management Science Letters* **4**(5), 893–898.
- Ballou, D. P. & Pazer, H. L. (1985), 'Modeling data and process quality in multi-input, multi-output information systems', *Management science* **31**(2), 150–162.
- Bateman, J. F. (1995), 'Preventive maintenance: Stand alone manufacturing compared with cellular manufacturing', *INDUSTRIAL MANAGEMENT-CHICAGO THEN ATLANTA*- **37**, 19–19.
- Berre, A.-J., Elvesæter, B., Figay, N., Guglielmina, C., Johnsen, S. G., Karlsen, D., Knothe, T. & Lippe, S. (2007), The athena interoperability framework, in 'Enterprise interoperability II', Springer, pp. 569–580.
- Bertolini, M. & Bevilacqua, M. (2006), 'A combined goal programming—ahp approach to maintenance selection problem', *Reliability Engineering & System Safety* **91**(7), 839–848.
- Bertolini, M., Bevilacqua, M., Braglia, M. & Frosolini, M. (2004), 'An analytical method for maintenance outsourcing service selection', *International Journal of Quality & Reliability Management* **21**(7), 772–788.
- Bevilacqua, M. & Braglia, M. (2000), 'The analytic hierarchy process applied to maintenance strategy selection', *Reliability Engineering & System Safety* **70**(1), 71–83.
- Bonandrini, S., Cruz, C. & Nicolle, C. (2005), Building lifecycle management, in 'Proceedings of the International Conference on Product Lifecycle Management, Lyon, France', Vol. 1113, Citeseer.
- BSI (2010), *Maintenance - Maintenance terminology BS EN 13306:2010*, BSI Standard Publication.
- Buur, J. (1990), *A theoretical approach to mechatronics design*, Institute for Engineering Design, Technical University of Denmark Denmark.
- Cafiso, S., Di Graziano, A., Kerali, H. & Odoki, J. (2002), 'Multicriteria analysis method for pavement maintenance management', *Transportation Research Record: Journal of the Transportation Research Board* (1816), 73–84.
- Caruccio, L., Deufemia, V. & Polese, G. (2014), Visual data integration based on description logic reasoning, in 'Proceedings of the 18th International Database Engineering & Applications Symposium', ACM, pp. 19–28.
- Cavalcante, C. A. V. & Costa, A. P. C. S. (2010), 'Multicriteria model of preventive maintenance', *Brazilian Journal of Operations & Production Management* **3**(1), 71–86.
- Cavalcante, C. A. V. & De Almeida, A. (2007), 'A multi-criteria decision-aiding model using promethee iii for preventive maintenance planning under uncertain conditions', *Journal of Quality in Maintenance Engineering* **13**(4), 385–397.
- Cavalcante, V., Alexandre, C., Pires Ferreira, R. J. & de Almeida, A. T. (2010), 'A preventive maintenance decision model based on multicriteria method promethee ii integrated with bayesian approach', *IMA Journal of Management Mathematics* **21**(4), 333–348.

- Certa, A., Enea, M. & Lupo, T. (2013), 'Electre iii to dynamically support the decision maker about the periodic replacements configurations for a multi-component system', *Decision support systems* **55**(1), 126–134.
- Charnes, A., Clark, C. T., Cooper, W. W. & Golany, B. (1984), 'A developmental study of data envelopment analysis in measuring the efficiency of maintenance units in the us air forces', *Annals of Operations Research* **2**(1), 95–112.
- Chase, R. B., Kumar, K. & Youngdahl, W. E. (1992), 'Service-based manufacturing: The service factory', *Production and Operations Management* **1**(2), 175–184.
- Chen, D. (2006), Enterprise interoperability framework., in 'EMOI-INTEROP'.
- Chen, D., Doumeingts, G. & Vernadat, F. (2008), 'Architectures for enterprise integration and interoperability: Past, present and future', *Computers in industry* **59**(7), 647–659.
- Chen, L.-q., Weng, M. & Zhang, G.-z. (2005), 'Utility optimality method for pipeline integrity maintenance costs', *Petroleum Engineering Construction* **2**(004).
- Chen, M., Mao, S. & Liu, Y. (2014), 'Big data: A survey', *Mobile networks and applications* **19**(2), 171–209.
- Chou, J.-S. (2009), 'Web-based cbr system applied to early cost budgeting for pavement maintenance project', *Expert Systems with Applications* **36**(2), 2947–2960.
- Coulter, E. D., Sessions, J. & Wing, M. G. (2006), 'Scheduling forest road maintenance using the analytic hierarchy process and heuristics', *Silva Fennica* **40**(1), 143.
- Cox, D. R. (1958), 'The regression analysis of binary sequences', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 215–242.
- Crnkovic, G. D. (2010), Constructive research and info-computational knowledge generation, in 'Model-Based Reasoning in Science and Technology', Springer, pp. 359–380.
- da Silva Monte, M. B. & de Almeida Filho, A. T. (2015), A mcdm model for preventive maintenance on wells for water distribution, in 'Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on', IEEE, pp. 268–272.
- Da Xu, L., He, W. & Li, S. (2014), 'Internet of things in industries: A survey', *IEEE Transactions on industrial informatics* **10**(4), 2233–2243.
- de Almeida, A. T. (2001), 'Multicriteria decision making on maintenance: spares and contracts planning', *European Journal of Operational Research* **129**(2), 235–241.
- de Almeida, A. T., Cavalcante, C. A. V., Alencar, M. H., Ferreira, R. J. P., de Almeida-Filho, A. T. & Garcez, T. V. (2015a), Decision on maintenance outsourcing, in 'Multicriteria and Multiobjective Models for Risk, Reliability and Maintenance Decision Analysis', Springer, pp. 249–272.
- de Almeida, A. T., Cavalcante, C. A. V., Alencar, M. H., Ferreira, R. J. P., de Almeida-Filho, A. T. & Garcez, T. V. (2015b), Decisions on priority assignment for maintenance planning, in 'Multicriteria and Multiobjective Models for Risk, Reliability and Maintenance Decision Analysis', Springer, pp. 335–349.

- de Almeida, A. T., Cavalcante, C. A. V., Alencar, M. H., Ferreira, R. J. P., de Almeida-Filho, A. T. & Garcez, T. V. (2015c), Preventive maintenance decisions, in 'Multicriteria and Multiobjective Models for Risk, Reliability and Maintenance Decision Analysis', Springer, pp. 215–232.
- Dedrick, J., Gurbaxani, V. & Kraemer, K. L. (2003), 'Information technology and economic performance: A critical review of the empirical evidence', *ACM Computing Surveys (CSUR)* **35**(1), 1–28.
- Dehghanian, P., Fotuhi-Firuzabad, M., Bagheri-Shouraki, S. & Kazemi, A. A. R. (2012), 'Critical component identification in reliability centered asset management of power distribution systems via fuzzy ahp', *IEEE Systems Journal* **6**(4), 593–602.
- Dunn, S. (1998), Reinventing the maintenance process: towards zero downtime, in 'Queensland Maintenance Conference Proceedings, Queensland, Australia'.
- Durán, O. (2011), 'Computer-aided maintenance management systems selection based on a fuzzy ahp approach', *Advances in Engineering Software* **42**(10), 821–829.
- e Costa, C. A. B., Carnero, M. C. & Oliveira, M. D. (2012), 'A multi-criteria model for auditing a predictive maintenance programme', *European Journal of Operational Research* **217**(2), 381–393.
- Earl, M. J. (1993), 'Experiences in strategic information systems planning', *MIS quarterly* pp. 1–24.
- Eastman, C., Teicholz, P., Sacks, R. & Liston, K. (2011), *BIM handbook: A guide to building information modeling for owners, managers, designers, engineers and contractors*, John Wiley & Sons.
- Ecosave (2017), 'What is data driven maintenance?'.
URL: <https://ecosave.com.au/news-insights/data-driven-maintenance/>
- EIF (2004), European interoperability framework for pan-european egovernment services, Technical report, European Commission.
URL: <http://ec.europa.eu/idabc/servlets/Docd552.pdf?id=19529>
- El-Sappagh, S. H. A., Hendawi, A. M. A. & El Bastawissy, A. H. (2011), 'A proposed model for data warehouse etl processes', *Journal of King Saud University-Computer and Information Sciences* **23**(2), 91–104.
- Emovon, I., Norman, R. A. & Murphy, A. J. (2015), 'Hybrid mcdm based methodology for selecting the optimum maintenance strategy for ship machinery systems', *Journal of Intelligent Manufacturing* **29**(3), 519–531.
- Eslami, S., Sajadi, S. M. & Kashan, A. H. (2014), 'Selecting a preventive maintenance scheduling method by using simulation and multi criteria decision making', *International Journal of Logistics Systems and Management* **18**(2), 250–269.
- Eti, M. C., Ogaji, S. & Probert, S. (2006), 'Reducing the cost of preventive maintenance (pm) through adopting a proactive reliability-focused culture', *Applied energy* **83**(11), 1235–1248.

- Fagin, R., Haas, L. M., Hernández, M., Miller, R. J., Popa, L. & Velegrakis, Y. (2009), Clio: Schema mapping creation and data exchange, in 'Conceptual Modeling: Foundations and Applications', Springer, pp. 198–236.
- Fallah-Fini, S., Triantis, K., Rahmandad, H. & Jesus, M. (2015), 'Measuring dynamic efficiency of highway maintenance operations', *Omega* **50**, 18–28.
- Farhan, J. & Fwa, T. (2009), 'Pavement maintenance prioritization using analytic hierarchy process', *Transportation Research Record* **2093**(1), 12–24.
- Fast, M., Assadi, M. & De, S. (2008), Condition based maintenance of gas turbines using simulation data and artificial neural network: A demonstration of feasibility, in 'ASME Turbo Expo 2008: Power for Land, Sea, and Air', American Society of Mechanical Engineers, pp. 153–161.
- Feldman, P. & Miller, D. (1986), 'Entity model clustering: Structuring a data model by abstraction', *The Computer Journal* **29**(4), 348–360.
- Ferdousmakan, M., Vasili, M., Vasili, M., Tang, S. & Lim, N. T. (2014), Selection of appropriate risk-based maintenance strategy by using fuzzy analytical hierarchy process, in '4rd European Seminar on Computing, Pilsen, Czech Republic', p. 77.
- Fouladgar, M. M., Yazdani-Chamzini, A., Lashgari, A., Zavadskas, E. K. & Turskis, Z. (2012), 'Maintenance strategy selection using ahp and copras under fuzzy environment', *International journal of strategic property management* **16**(1), 85–104.
- Främling, K., Holmström, J., Loukkola, J., Nyman, J. & Kaustell, A. (2013), 'Sustainable plm through intelligent products', *Engineering Applications of Artificial Intelligence* **26**(2), 789–799.
- Främling, K. & Maharjan, M. (2013), 'Standardized communication between intelligent products for the iot', *IFAC Proceedings Volumes* **46**(7), 157–162.
- García-Cascales, M. S. & Lamata, M. T. (2009), 'Selection of a cleaning system for engine maintenance based on the analytic hierarchy process', *Computers & Industrial Engineering* **56**(4), 1442–1451.
- Garmabaki, A., Ahmadi, A. & Ahmadi, M. (2016), Maintenance optimization using multi-attribute utility theory, in 'Current Trends in Reliability, Availability, Maintainability and Safety', Springer, pp. 13–25.
- Geraci, A., Katki, F., McMonegal, L., Meyer, B., Lane, J., Wilson, P., Radatz, J., Yee, M., Porteous, H. & Springsteel, F. (1991), *IEEE standard computer dictionary: Compilation of IEEE standard computer glossaries*, IEEE Press.
- Golfarelli, M., Maio, D. & Rizzi, S. (1998), 'The dimensional fact model: A conceptual model for data warehouses', *International Journal of Cooperative Information Systems* **7**(02n03), 215–247.
- Gomez, A. & Carnero, M. C. (2011), 'Selection of a computerised maintenance management system: a case study in a regional health service', *Production Planning and Control* **22**(4), 426–436.

- Gonçalves, C. D. F., Dias, J. A. M. & Cruz-Machado, V. A. (2014), Decision methodology for maintenance kpi selection: Based on electre i, in 'Proceedings of the Eighth International Conference on Management Science and Engineering Management', Springer, pp. 1001–1012.
- Goossens, A. J. & Basten, R. J. (2015), 'Exploring maintenance policy selection using the analytic hierarchy process; an application for naval ships', *Reliability Engineering & System Safety* **142**, 31–41.
- Group, C. A. W. et al. (1998), 'Levels of information systems interoperability (lisi)', *US DoD*.
- Ha, S. H. & Krishnan, R. (2008), 'A hybrid approach to supplier selection for the maintenance of a competitive supply chain', *Expert systems with applications* **34**(2), 1303–1311.
- Hankach, P. & Lepert, P. (2011), 'Multi-criteria decision analysis for prioritizing road network maintenance interventions', *International Journal of Pavements* **10**(1), 12p.
- Hart, C. (2018), *Doing a Literature Review: Releasing the Research Imagination*, Sage.
- Hevner, A. & Chatterjee, S. (2010), Design science research in information systems, in 'Design research in information systems', Springer, pp. 9–22.
- Hjalmarsson, L. & Odeck, J. (1996), 'Efficiency of trucks in road construction and maintenance: an evaluation with data envelopment analysis', *Computers & operations research* **23**(4), 393–404.
- Hosseini Firouz, M. & Ghadimi, N. (2015), 'Optimal preventive maintenance policy for electric power distribution systems based on the fuzzy ahp methods', *Complexity* **21**(6), 70–88.
- Huang, K.-T., Lee, Y. W. & Wang, R. Y. (1999), 'Quality information and knowledge management', *Publisher: Prentice Hall*.
- Hwang, C.-L. & Yoon, K. (1981), Methods for multiple attribute decision making, in 'Multiple attribute decision making', Springer, pp. 58–191.
- Ilankumaran, M. & Kumanan, S. (2009), 'Selection of maintenance policy for textile industry using hybrid multi-criteria decision making approach', *Journal of Manufacturing Technology Management* **20**(7), 1009–1022.
- Ilankumaran, M. & Kumanan, S. (2012), 'Application of hybrid vikor model in selection of maintenance strategy', *International Journal of Information Systems and Supply Chain Management (IJISSCM)* **5**(2), 59–81.
- ISA, A. et al. (2000), 'Enterprise-control system integration. part 1: Models and terminology'.
- Itaka, K., Matsubara, I., Nakano, T., Sakurai, K. & Taga, H. (1991), Advanced maintenance information systems for overhead power transmission lines, in 'Advances in Power System Control, Operation and Management, 1991. APSCOM-91., 1991 International Conference on', IET, pp. 927–932.

- Jardine, A. K., Lin, D. & Banjevic, D. (2006), 'A review on machinery diagnostics and prognostics implementing condition-based maintenance', *Mechanical systems and signal processing* **20**(7), 1483–1510.
- Järvinen, H. et al. (2015), 'Web technology based smart home interoperability'.
- Jeon, J., Kim, C. & Lee, H. (2011), 'Measuring efficiency of total productive maintenance (tpm): a three-stage data envelopment analysis (dea) approach', *Total Quality Management & Business Excellence* **22**(8), 911–924.
- Jiang, L., Cai, H. & Xu, B. (2010), A domain ontology approach in the etl process of data warehousing, in 'e-Business Engineering (ICEBE), 2010 IEEE 7th International Conference on', IEEE, pp. 30–35.
- Jonsson, P. (1999), *The impact of maintenance on the production process-achieving high performance*, Division of Production Management, Lund Institute of Technology, SE-221 00 Lund.
- Jun, H.-B., Kiritsis, D. & Xirouchakis, P. (2007), 'Research issues on closed-loop plm', *Computers in industry* **58**(8), 855–868.
- Kans, M. (2008), On the utilisation of information technology for the management of profitable maintenance, PhD thesis, Institutionen för teknik och design.
- Kans, M. (2009), 'The advancement of maintenance information technology: A literature review', *Journal of Quality in Maintenance Engineering* **15**(1), 5–16.
- Karpenko, A., Kinnunen, T., Madhikermi, M., Robert, J., Främling, K., Dave, B. & Nurminen, A. (2018), 'Data exchange interoperability in iot ecosystem for smart parking and ev charging', *Sensors* **18**(12), 4404.
- Kelly, A. (1984), *Maintenance planning and control*, Butterworths.
- Kim, Y.-G. & Everest, G. C. (1994), 'Building an is architecture: Collective wisdom from the field', *Information & Management* **26**(1), 1–11.
- Kimball, R. & Ross, M. (1996), 'The data warehousing toolkit', *John Wiley & Sons, New York*.
- Kini, R. B. (2002), 'It in manufacturing for performance: the challenge for thai manufacturers', *Information management & computer security* **10**(1), 41–48.
- Kiritsis, D. (2011), 'Closed-loop plm for intelligent products in the era of the internet of things', *Computer-Aided Design* **43**(5), 479–501.
- Kothari, C. R. (2004), *Research methodology: Methods and techniques*, New Age International.
- Krogstie, J., Lindland, O. I. & Sindre, G. (1995), Defining quality aspects for conceptual models, in 'Information System Concepts', Springer, pp. 216–231.
- Krstić, H. & Marenjak, S. (2012), 'Analysis of buildings operation and maintenance costs', *Građevinar* **64**(04.), 293–303.
- Kubicek, H., Cimander, R. & Scholl, H. J. (2011), Layers of interoperability, in 'Organizational Interoperability in E-Government', Springer, pp. 85–96.

- Kubler, S., Madhikermi, M., Buda, A. & Främling, K. (2013), Two-way communications through firewalls using qlm messaging, in 'International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services', Springer, pp. 743–747.
- Kumar, G. & Maiti, J. (2012), 'Modeling risk based maintenance using fuzzy analytic network process', *Expert Systems with Applications* **39**(11), 9946–9954.
- Kuo, T. C. & Wang, M. L. (2012), 'The optimisation of maintenance service levels to support the product service system', *International Journal of Production Research* **50**(23), 6691–6708.
- Kwon, O., Lee, N. & Shin, B. (2014), 'Data quality management, data usage experience and acquisition intention of big data analytics', *International Journal of Information Management* **34**(3), 387–394.
- Labib, A. W., O'Connor, R. F. & Williams, G. B. (1998), 'An effective maintenance system using the analytic hierarchy process', *Integrated Manufacturing Systems* **9**(2), 87–98.
- Lavalle, S., Lesser, E., Shockley, R., Hopkins, M. S. & Kruschwitz, N. (2011), 'Big data, analytics and the path from insights to value', *MIT sloan management review* **52**(2), 21.
- Lee, G., Sacks, R. & Eastman, C. M. (2006), 'Specifying parametric building object behavior (bob) for a building information modeling system', *Automation in construction* **15**(6), 758–776.
- Levrat, E., Iung, B. & Crespo Marquez, A. (2008), 'E-maintenance: review and conceptual framework', *Production Planning & Control* **19**(4), 408–429.
- Li, C., Xu, M. & Guo, S. (2007), Electre iii based on ranking fuzzy numbers for deterministic and fuzzy maintenance strategy decision problems, in 'Automation and Logistics, 2007 IEEE International Conference on', IEEE, pp. 309–312.
- Liu, J.-j., LU, X.-z. & QU, C.-z. (2012), 'A priority sorting approach of maintenance task during mission based on electre tri', *Fire Control & Command Control* p. S1.
- Liu, J. & Yu, D.-j. (2004), 'Evaluation of plant maintenance based on data envelopment analysis', *Journal of Quality in Maintenance Engineering* **10**(3), 203–209.
- Liu, M. & Frangopol, D. M. (2006), Decision support system for bridge network maintenance planning, in 'Advances in Engineering Structures, Mechanics & Construction', Springer, pp. 833–840.
- Lujan-Mora, S. & Trujillo, J. (2003), A comprehensive method for data warehouse design, in 'Workshop on Design and Management of Data Warehouses', pp. 1–1.
- Madenas, N., Davies, A. & Filton, C. (2014), 'Integrating plm systems with maintenance information across the supply chain', *G T E C H N O L O G Y* p. 41.
- Madhavan, J., Bernstein, P. A. & Rahm, E. (2001), Generic schema matching with cupid, in 'vldb', Vol. 1, pp. 49–58.
- Mahlamäki, K., Niemi, A., Jokinen, J. & Borgman, J. (2016), 'Importance of maintenance data quality in extended warranty simulation', *Intl. J. COMADEM* **19**(1), 3–10.

- MareL (2010), 'What is plm data "obfuscation" and i why should i care?', <http://community.aras.com/en/plm-data-obfuscation-care/>. Accessed on 2016-05-16.
- McMillan, D. & Ault, G. W. (2007), 'Quantification of condition monitoring benefit for offshore wind turbines', *Wind Engineering* **31**(4), 267–285.
- Moazami, D., Behbahani, H. & Muniandy, R. (2011), 'Pavement rehabilitation and maintenance prioritization of urban roads using fuzzy logic', *Expert Systems with Applications* **38**(10), 12869–12879.
- Mobley, R. K. (2002), *An introduction to predictive maintenance*, Butterworth-Heinemann.
- Monte, M. B. & de Almeida-Filho, A. T. (2016), 'A multicriteria approach using maut to assist the maintenance of a water supply system located in a low-income community', *Water Resources Management* **30**(9), 3093–3106.
- Moody, D. (1991), A practical methodology for the representation of large data models, in 'Proceedings of the Australian Database and Information Systems Conference'.
- Morgan, G. & Smircich, L. (1980), 'The case for qualitative research', *Academy of management review* **5**(4), 491–500.
- Morris, E., Levine, L., Meyers, C., Place, P. & Plakosh, D. (2004), System of systems interoperability (sosi), Technical report, CARNEGIE-MELLON UNIV PITTSBURGH PA SOFTWARE ENGINEERING INST.
- Muchiri, P., Pintelon, L., Gelders, L. & Martin, H. (2011), 'Development of maintenance function performance measurement framework and indicators', *International Journal of Production Economics* **131**(1), 295–302.
- Ouma, Y. O., Opudo, J. & Nyambenya, S. (2015), 'Comparison of fuzzy ahp and fuzzy topsis for road pavement maintenance prioritization: methodological exposition and case study', *Advances in Civil Engineering* **2015**.
- Ozbek, M. E., de la Garza, J. M. & Triantis, K. (2010a), 'Data and modeling issues faced during the efficiency measurement of road maintenance using data envelopment analysis', *Journal of Infrastructure Systems* **16**(1), 21–30.
- Ozbek, M. E., de la Garza, J. M. & Triantis, K. (2010b), 'Efficiency measurement of bridge maintenance using data envelopment analysis', *Journal of Infrastructure Systems* **16**(1), 31–39.
- Pearce, D. & Hall, S. (1994), Using vibration monitoring techniques to minimize plant downtime, in 'Advances in Control in the Process Industries: An Exercise in Technology Transfer, 1994. IEE Colloquium on.(Digest No. 1994/081)', IET, pp. 8–1.
- Pearson, K. (1901), 'Principal components analysis', *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **6**(2), 559.
- Peck Jr, M. W., Scheraga, C. A. & Boisjoly, R. P. (1998), 'Assessing the relative efficiency of aircraft maintenance technologies: an application of data envelopment analysis', *Transportation Research Part A: Policy and Practice* **32**(4), 261–269.

- Pentti, R. (2007), 'Comparative study'.
URL: <http://www2.uiah.fi/projekti/metodi/172.htm>
- Perumal, T., Ramli, A. R., Leong, C. Y., Mansor, S. & Samsudin, K. (2008), Interoperability among heterogeneous systems in smart home environment, in 'Signal Image Technology and Internet Based Systems, 2008. SITIS'08. IEEE International Conference on', IEEE, pp. 177–186.
- Pieterse, H. & Olivier, M. (2012), Data hiding techniques for database environments, in 'IFIP International Conference on Digital Forensics', Springer, pp. 289–301.
- Pintelon, L., Du Preez, N. & Van Puyvelde, F. (1999), 'Information technology: opportunities for maintenance management', *Journal of Quality in Maintenance Engineering* **5**(1), 9–24.
- Pintelon, L. M. & Gelders, L. (1992), 'Maintenance management decision making', *European journal of operational research* **58**(3), 301–317.
- Pipino, L. L., Lee, Y. W. & Wang, R. Y. (2002), 'Data quality assessment', *Communications of the ACM* **45**(4), 211–218.
- Pourjavad, E., Shirouyehzad, H. & Shahin, A. (2013), 'Selecting maintenance strategy in mining industry by analytic network process and topsis', *International Journal of Industrial and Systems Engineering* **15**(2), 171–192.
- Pramod, V., Sampath, K., Devadasan, S., Jagathy Raj, V. & Moorthy, G. D. (2007), 'Multicriteria decision making in maintenance quality function deployment through the analytical hierarchy process', *International Journal of Industrial and Systems Engineering* **2**(4), 454–478.
- Rifaieh, R. & Benharkat, N. A. (2002), Query-based data warehousing tool, in 'Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP', ACM, pp. 35–42.
- Roll, Y., Golany, B. & Seroussy, D. (1989), 'Measuring the efficiency of maintenance units in the israeli air force', *European Journal of Operational Research* **43**(2), 136–142.
- Rouse, P., Putterill, M. & Ryan, D. (2002), 'Integrated performance measurement design: insights from an application in aircraft maintenance', *Management Accounting Research* **13**(2), 229–248.
- Saaty, T. L. (1977), 'A scaling method for priorities in hierarchical structures', *Journal of mathematical psychology* **15**(3), 234–281.
- Saaty, T. L. (1980), 'The analytic hierarchy process: planning, priority setting, resources allocation', *New York: McGraw* **281**.
- Sato, K., Atsumi, S., Shibata, A. & Kanemaru, K. (1991), Power transmission line maintenance information system for hokusei line with snow accretion monitoring capability, in 'Transmission and Distribution Conference, 1991., Proceedings of the 1991 IEEE Power Engineering Society', IEEE, pp. 811–816.
- Sevilla Juárez, E. & Escobar Toledo, C. E. (2008), 'The efficiency of preventive maintenance planning and the multicriteria methods: a case study', *Computación y Sistemas* **12**(2), 208–215.

- Shahin, A., Pourjavad, E. & Shirouyehzad, H. (2012), 'Selecting optimum maintenance strategy by analytic network process with a case study in the mining industry', *International Journal of Productivity and Quality Management* **10**(4), 464–483.
- Sheikhalishahi, M. (2014), 'An integrated simulation-data envelopment analysis approach for maintenance activities planning', *International Journal of Computer Integrated Manufacturing* **27**(9), 858–868.
- Shyjith, K., Ilankumaran, M. & Kumanan, S. (2008), 'Multi-criteria decision-making approach to evaluate optimum maintenance strategy in textile industry', *Journal of Quality in Maintenance Engineering* **14**(4), 375–386.
- Simpson, L. (1996), 'Do decision makers know what they prefer?: Mavt and electre ii', *Journal of the Operational Research Society* **47**(7), 919–929.
- Sophie, S., Thomas, A., Dominik, L., Ralf, H., Pierrick, B. & Javier, G. (2014), 'Supreme sustainable predictive maintenance for manufacturing equipment, in 'European Congress', pp. 1–6.
- Staff, J. C. O. (2013), *Department of Defense dictionary of military and associated terms*, MILITARY BOOKSHOP.
- Strong, D. M., Lee, Y. W. & Wang, R. Y. (1997), 'Data quality in context', *Communications of the ACM* **40**(5), 103–110.
- Sumino, H., Ishikawa, N., Murakami, S., Kato, T. & Hjelm, J. (2007), 'Pucc architecture, protocols and applications, in 'Consumer Communications and Networking Conference, 2007. CCNC 2007. 4th IEEE', IEEE, pp. 788–792.
- Sun, S. (2004), 'Assessing joint maintenance shops in the taiwanese army using data envelopment analysis', *Journal of Operations Management* **22**(3), 233–245.
- Swanson, L. (2001), 'Linking maintenance strategies to performance', *International journal of production economics* **70**(3), 237–244.
- Taghipour, S., Banjevic, D. & Jardine, A. K. (2011), 'Prioritization of medical equipment for maintenance decisions', *Journal of the Operational Research Society* **62**(9), 1666–1687.
- Tan, Z., Li, J., Wu, Z., Zheng, J. & He, W. (2011), 'An evaluation of maintenance strategy using risk based inspection', *Safety science* **49**(6), 852–860.
- Tayi, G. K. & Ballou, D. P. (1998), 'Examining data quality', *Communications of the ACM* **41**(2), 54–57.
- Thor, J., Ding, S.-H. & Kamaruddin, S. (2013), 'Comparison of multi criteria decision making methods from the maintenance alternative selection perspective', *The International Journal of Engineering and Science* **2**(6), 27–34.
- Tobin, R. (2010), 'Descriptive case study', *Encyclopedia of case study research* **1**.
- Touma, R., Romero, O. & Jovanovic, P. (2015), 'Supporting data integration tasks with semi-automatic ontology construction, in 'Proceedings of the ACM Eighteenth International Workshop on Data Warehousing and OLAP', ACM, pp. 89–98.

- Tretten, P. & Karim, R. (2014), 'Enhancing the usability of maintenance data management systems', *Journal of Quality in Maintenance Engineering* **20**(3), 290–303.
- Triantaphyllou, E., Kovalerchuk, B., Mann, L. & Knapp, G. M. (1997), 'Determining the most important criteria in maintenance decision making', *Journal of Quality in Maintenance Engineering* **3**(1), 16–28.
- Trojan, F. & Morais, D. C. (2012), 'Prioritising alternatives for maintenance of water distribution networks: A group decision approach', *Water Sa* **38**(4), 555–564.
- Trujillo, J., Palomar, M., Gomez, J. & Song, I.-Y. (2001), 'Designing data warehouses with oo conceptual models', *Computer* **34**(12), 66–75.
- Turnitsa, C. (2005), Extending the levels of conceptual interoperability model, in 'Proceedings IEEE summer computer simulation conference, IEEE CS Press'.
- Van den Bergh, J., De Bruecker, P., Beliën, J., De Boeck, L. & Demeulemeester, E. (2013), 'A three-stage approach for aircraft line maintenance personnel rostering using mip, discrete event simulation and dea', *Expert Systems with Applications* **40**(7), 2659–2668.
- Van Horenbeek, A. & Pintelon, L. (2014a), 'Development of a maintenance performance measurement framework—using the analytic network process (anp) for maintenance performance indicator selection', *Omega* **42**(1), 33–46.
- Van Horenbeek, A. & Pintelon, L. (2014b), 'Development of a maintenance performance measurement framework—using the analytic network process (anp) for maintenance performance indicator selection', *Omega* **42**(1), 33–46.
- Vanlande, R., Nicolle, C. & Cruz, C. (2008), 'Ifc and building lifecycle management', *Automation in construction* **18**(1), 70–78.
- Vassiliadis, P., Simitsis, A. & Skiadopoulos, S. (2002), Conceptual modeling for etl processes, in 'Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP', ACM, pp. 14–21.
- Viaene, S. (2013), 'Data scientists aren't domain experts', *IT Professional* **15**(6), 12–17.
- Vujanović, D., Momčilović, V., Bojović, N. & Papić, V. (2012), 'Evaluation of vehicle fleet maintenance management indicators by application of dematel and anp', *Expert Systems with Applications* **39**(12), 10552–10563.
- W3Techs (2017).
URL: https://w3techs.com/technologies/overview/content_management/all/
- Waeyenbergh, G. & Pintelon, L. (2002), 'A framework for maintenance concept development', *International journal of production economics* **77**(3), 299–313.
- Wakchaure, S. S. & Jha, K. N. (2011), 'Prioritization of bridges for maintenance planning using data envelopment analysis', *Construction Management and Economics* **29**(9), 957–968.

- Wang, J., Fan, K. & Wang, W. (2010), 'Integration of fuzzy ahp and fpp with topsis methodology for aeroengine health assessment', *Expert Systems with Applications* **37**(12), 8516–8526.
- Wang, L., Chu, J. & Wu, J. (2007), 'Selection of optimum maintenance strategies based on a fuzzy analytic hierarchy process', *International journal of production economics* **107**(1), 151–163.
- Wang, R. Y. & Strong, D. M. (1996), 'Beyond accuracy: What data quality means to data consumers', *Journal of management information systems* **12**(4), 5–33.
- Weber, R. et al. (1997), *Ontological foundations of information systems*, Coopers & Lybrand and the Accounting Association of Australia and New Zealand Melbourne.
- Weil, N. (1998), 'Make the most of maintenance', *Manufacturing engineering* **120**(5).
- Wichers, J. (1996), Optimising maintenance functions by ensuring effective management of your computerised maintenance management system, in 'AFRICON, 1996., IEEE AFRICON 4th', Vol. 2, IEEE, pp. 788–794.
- Wilkinson, M., Spianto, F. & Knowles, M. (2006), Towards the zero maintenance wind turbine, in 'Universities Power Engineering Conference, 2006. UPEC'06. Proceedings of the 41st International', Vol. 1, IEEE, pp. 74–78.
- Yin, R. K. (1994), 'Case study: design and methods', *Newbury Park, Sage Publications*.
- Yousefnezhad, N., Filippov, R., Javed, A., Buda, A., Madhikermi, M. & Främling, K. (2017), 'Authentication and access control for open messaging interface standard'.
- Zaim, S., Turkyilmaz, A., Acar, M. F., Al-Turki, U. & Demirel, O. F. (2012), 'Maintenance strategy selection using ahp and anp algorithms: a case study', *Journal of Quality in Maintenance Engineering* **18**(1), 16–29.
- Zhangqiong, W. & Guozheng, S. (1999), 'A study of appraisal and decision-making support system for maintenance scheme of metal structure of crane', *JOURNAL OF WUHAN TRANSPORTATION UNIVERSITY* **5**, 007.

Errata

Publication I:

- The last sentence of the Section II-A should read as: “So, it is vital to identify the key reasons behind poor data quality that hinder effective CBM.”

ISBN 978-952-60-8581-4 (printed)
ISBN 978-952-60-8582-1 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Science
Department of Computer Science
www.aalto.fi

**BUSINESS +
ECONOMY**

**ART +
DESIGN +
ARCHITECTURE**

**SCIENCE +
TECHNOLOGY**

CROSSOVER

**DOCTORAL
DISSERTATIONS**