Managing Data and Information Quality in Construction Engineering: a System Design Approach

Soffi Westin

Managing Data and Information Quality in Construction Engineering: a System Design Approach

Doctoral Dissertation

in partial fulfillment of the requirements for the Ph.D. degree in Information Systems

Department of Information Systems Faculty of Economics and Social Sciences University of Agder 2014

Doctoral Dissertations at the University of Agder 79

ISSN: 1504-9272 ISBN: 978-82-7117-758-4

© Soffi Westin, 2014

Printed by the Printing Office, University of Agder Kristiansand

Acknowledgement

I would like to thank my three excellent supervisors, Tero Päivärinta, Maung K. Sein, and Kristin Wallevik, who have guided me through this work and taught me (with a great deal of patience!) a lot about what research is. In addition I would like to thank Olav Birger Holen for fruitful and interesting discussions and for co-authoring one of my papers.

I am also grateful to Peter Axel Nielsen (University of Aalborg, Denmark) for his critical but insightful comments on an earlier draft of this thesis.

I would like to thank the company in which I am employed for allowing me to embark on this journey. I am grateful to my colleagues, who are great sources of knowledge, ideas, and inspiration, and I thank them for their willingness to share it all with me. Without you, this study would never have been completed.

The Norwegian Research Council has contributed by running the Industrial PhD program which I was allowed to attend. The program not only includes partial financing and interesting seminars, it also includes help from two very solution-oriented and nice people: Annette L. Vestlund and Ingeborg Guldal. Thank you to both of you, and also to Jacob Alsted who was in charge of the seminars.

To the faculty and students in the Information System Department at the University of Agder, I am very grateful. Especially, I want to thank Bjørn Erik Munkvold, Director for the PhD program, for designing a program that encourages discussion and collaboration between students and faculty in a most giving way. My gratitude also goes to my fellow students, especially to Karen Stendal, Marius Rohde Johannessen, Fathul Wahid and Christoph Merschbrock for great discussions and friendship.

I would also like to thank my husband, Hans-Petter Westnes, and my three children; Fredrikke, Andreas, and Øivind, for their confidence in me. This journey has been both interesting and challenging and they have supported me all the way. Finally, I would like to thank my mother, Bodil Westin, for her everlasting love and support.

Abstract

Incorrect and missing information in drawings and specifications has recently been identified as a significant factor in causing delays and cost overruns in construction engineering projects. Since the profit margins for construction engineering projects have been narrow for decades, it is important to avoid delays and cost overruns. Hence, the information presented in drawings needs to be as complete and error free as possible. Construction engineering drawings are commonly extracted from the databases and information systems used by construction engineering designers. Hence, the level of data quality and information quality (DQ/IQ) in those sources needs to be sufficient for producing error free drawings.

This thesis focuses on how DQ/IQ can be managed in construction engineering projects. The specific problem under investigation was inspired by the situation in a European construction engineering company. The company was suffering from delays and cost overruns in their construction engineering projects, and their assembly site workers were complaining about missing information in drawings. An investigation of the company's data sources and its existing DQ/IQ assessment tool revealed missing and incorrect information in the sources as well as several shortcomings in the tool. Hence, a new tool was needed.

While the research field of DQ/IQ has produced several assessment frameworks that provide useful information and guidance for developing DQ/IQ assessment tools, few if any of these frameworks address the unique challenges of construction engineering. However, they do provide good insights to the basic considerations needed for such tools. Hence, I chose two of these frameworks - Total Information Quality Management (TIQM) and Data Quality Assessment (DQA) - to guide the development of a new tool.

To support the joint collaborative efforts needed to design and implement a DQ/IQ assessment tool for this construction engineering company, I applied an action design research (ADR) approach. ADR is a method for performing IT artifact design that emerges from interaction with an organizational context, and it considers organizational intervention a requirement for development. For this study, ADR provided guidance for the design process due to its demand for relevance and intervention as well as rigor.

Following the ADR method made it possible not only to develop a tool, but also to formulate design principles and abstract the findings to a generalizable level.

This study contributes to knowledge and practice in several ways. First, I offer five design principles for DQ/IQ assessment tools. These design principles are specifically aimed at mitigating the unavoidable challenges and their consequences in construction engineering projects. By accepting these unavoidable challenges and consequences and subsequently providing means for managing the results in a controlled manner, these principles makes it possible to avoid project delays and still reach a sufficient level of DQ/IQ in the end.

Second, the development and implementation of a tool in which these design principles are embedded demonstrates the effectiveness of the design principles. A formal evaluation performed by comparing a project that used the tool with two projects that did not, showed a significantly better level of DQ/IQ in the project using the tool.

Third, as a result of implementing the tool in a total of 12 construction engineering projects, it was possible to determine three needed and sufficient quality dimensions for rule-based assessment. This finding offers valuable information to theory as well as to practitioners aiming at assessing DQ/IQ in their projects.

Fourth, by revealing the relationship between unavoidable challenges and their consequences in construction engineering, this thesis offers unique insights into the nature of projects in that field, which is highly needed when performing DQ/IQ assessment. These insights will help DQ/IQ researchers enhance their understanding of a very complex and under-researched context.

Fifth, by providing a ranked list of DQ/IQ problems experienced at EUMEC, this thesis offers a more detailed explanation of DQ/IQ problems causing delays and cost overruns than is the case in previous research.

All in all, this research reduces a gap in the existing literature, namely the scarce amount of DQ/IQ research on construction engineering. The complexity of this industry makes it difficult and time consuming for an information systems researcher to fully understand the nature of construction engineering. This complexity might explain the scarce amount of research in this cross-disciplinary field, and this thesis helps reduce that gap.

Abbreviations

ADR	Action Design Research	
AR	Action Research	
BIE	Stage in ADR (Building, Intervention, and Evaluation)	
CDQ	Comprehensive methodology for Data Quality management	
COLDQ	Loshin Methodology (Cost-effect Of Low Data Quality)	
DP	Design Principle	
DQ	Data Quality	
DQ/IQ	Data and Information Quality	
DQA	Data Quality Assessment	
DR	Design Research	
EAM	Engineering Asset Management	
EUMEC	Artificial acronym used to anonymize the target organization	
IQ	Information Quality	
IQS	Information Quality System	
IS	Information Systems	
IT	Information Technology	
TestP	The construction engineering project during which IQS emerged	
TIQM	Total Information Quality Management	

Definition of terms

Rules	Rules are formally defined on the basis of the requirements for a	
	construction engineering project and used to assess compliance of data	
	to those requirements	
Straw values	Straw values represent various preliminary values in the data records,	
	such as "NA" or "-", which are inserted by the engineers while they	
	are waiting for correct values	

Table of Contents

Part A: Problem	1
1. Introduction	3
1.1 Research Questions	5
1.2 Structure of the Thesis	
2 Related Research	7
2.1 Construction Engineering	7
2.2 Data and Information Quality (DQ/IQ)	9
2.2.1 Data versus Information	10
2.2.2 Overview of DQ/IQ Management	11
2.3 DQ/IQ Assessment Challenges in Construction Engineering	20
2.4 Summary	
Part B: Solution	27
3 Research Approach	29
3.1 Positioning the Research Approach	
3.2 Action Design Research	
3.2.2 Data Collection and Analysis	
4 The Design and Emergence of IQS	
4.1 Target Organization	
4.2 Conceptual Basis	
4.3 Designing IQS through Action Design Research	40
4.3.1 Problem Formulation	
4.3.2 BIE Cycle 1: Designing for Error Detection and Reporting	44
4.3.3 Description of IQS	47
4.3.4 Reflections and Learning from BIE Cycle 1	48
4.3.5 Problem Re-Definition	50
4.3.6 BIE Cycle 2: Designing for Error Handling	51
4.3.7 Reflections and Learning from BIE Cycle 2	
4.3.8 Formalization of Learning	
4.3.9 Summary	
4.4 Use of Guiding Frameworks	
4.4.1 TIQM and IQS	

4.4.2 DQA and IQS	
4.5 Formal Evaluation of IQS	57
Part C: Contribution	61
5 Research Publications	63
5.1 Paper 1: A Delphi Study	64
5.1.1 Findings	64
5.2 Paper 2: A Literature Review	65
5.2.1 Findings	65
5.3 Paper 3: The Design and Emergence of IQS	66
5.3.1 Findings	66
5.4 Paper 4: Identifying Quality Dimensions	66
5.4.1 Findings	67
5.5 Paper 5: Evaluating IQS	67
5.5.1 Findings	
5.5 Overall Summary of the Papers	
6. Contributions	71
6.1 Theoretical Contributions	71
6.1.1 Design Principles	71
6.1.2 Quality Dimensions	75
6.1.3 DQ/IQ Assessment Challenges and Their Consequences	
6.2 Practical Contributions	
6.2.1 Assessment Tool	81
6.2.2 Identified and Ranked Problems	
7. Conclusion	85
7.1 Summary	
7.2 Limitations	
7.3 Future Research	
Part D: Reflection	91
8. Reflections	93
8.1 DQ/IQ Assessment Frameworks in Construction Engineering	93
8.2 Design Principles: Why Did They Emerge?	94
8.3 Method	
8.4 What Did the Practitioners Learn From the ADR Project?	98

8.5 Being an Insider	
References	
Appendix A – Data Collection Activities	111
Appendix B – Rule Description	
Appendix C – Example of Syntax Code	117
Appendix D – Research Publications	

List of Figures

Figure 1. Typical phases of an engineering project	8
Figure 2. Connections between identified challenges and level of DQ/IQ (adapted from Paper	: 2)
	24
Figure 3. Positioning ADR (adapted from Mathiassen 2002, with permission from Emerald	
Group Publishing Limited 2013)	31
Figure 4. ADR method: stages and principles (from Sein et al. 2011, Copyright © 2011, Rege	ents
of the University of Minnesota. Used with permission.)	33
Figure 5. Generic schema for IT-dominant BIE (adapted from Sein et al., 2011, Copyright ©	
2011, Regents of the University of Minnesota. Used with permission.)	34
Figure 6. Project timeline (from Paper 3)	41
Figure 7. A schematic model of IQS	47
Figure 8. BIE cycles and contributions (from Paper 3)	54
Figure 9. Comparison of projects (from Paper 5)	59
Figure 10. Handling quality dimensions in construction engineering	77
Figure 11. Challenges and how they are mitigated by IQS	79
Figure 12. Amount of generated information versus amount of processed information	94
Figure 13. Revised ADR model	97
Figure 14. The sliding nature of ADR	98

List of Tables

Table 1. Assessment frameworks and concepts (adapted from Batini et al. 2009 © 2013	
Association for Computing Machinery, Inc. Reprinted by permission)	16
Table 2. The six processes of TIQM (from English, 2003)	18
Table 3. Assessment types in DQA (adapted from Pipino et al., 2002 © 2013 Association for	
Computing Machinery, Inc. Reprinted by permission)	19
Table 4. Engineering disciplines in EUMEC's construction projects (Paper 1)	38
Table 5. DQ/IQ problems identified in the target organization (from Paper 1)	42
Table 6. Implications of BIE cycle 1 (adapted from Paper 3)	49
Table 7. Revised set of design principles (adapted from Paper 3)	53
Table 8. Mapping IQS to the premises of TIQM processes by English (1999, 2003)	55
Table 9. Mapping IQS to the premises of subjective and objective data quality assessment	
proposed by Pipino et al. (2002)	57
Table 10. Research publications	63
Table 11. Contributions of the papers to this thesis	69
Table 12. Summary of the theoretical contributions	71
Table 13. Design principles for proceeding with partial information	72
Table 14. Design principle providing a manageable number of reported errors	73
Table 15. Design principle handling exceptions	74
Table 16. Summary of the practical contributions	81
Table 17. IQS compared with EUMEC's old tool	82
Table 18. DQ/IQ problems in EUMEC and the impact of IQS	83
Table 19. An IQS report displaying identified errors for a work package	.114
Table 20. IQS detail report	

Part A: Problem

"If I had an hour to solve a problem I'd spend 55 minutes thinking about the problem and 5 minutes thinking about solutions."

- Albert Einstein

1. Introduction

Delays and cost overruns are common phenomena in construction engineering projects (Long et al. 2004; Sambasivan and Soon 2007; Toor and Ogunlana 2008). This is particularly true for large construction engineering projects which are characterized by a large number – and broad variety – of participants, geographically spread locations, huge costs, and a vast quantity of data. Several researchers have reported delays as the most costly problem plaguing this industry (Ling et al. 2009; Memon et al. 2011). Since the profit margins for construction engineering projects have been narrow for decades (Agapiou et al. 1998; J. H. Chen and Chen 2012), finding the cause of these delays is important.

Several studies have contributed to this research by identifying some causes for delays such as waiting for information (Frimpong et al. 2003), design complexity (Lim and Mohamed 2000), delays in design information (Assaf and Al-Hejji 2006), and changes in scope (Sambasivan and Soon 2007). One important aspect of construction engineering design is drawings, and recent research has specifically pointed to issues related to drawings as one of the most significant problem areas for the industry (Dai et al. 2009). For example, poor drawings and specifications have been identified as a significant factor that causes delays and cost overruns (Rivas et al. 2010). Variations between the planned and actual task start time and duration have also been traced to errors in design and/or drawings (Wambeke et al. 2011).

Improving drawings by reducing errors is thus vital to reduce delays and cost overruns. Since drawings are based on data, e.g., specifications, errors can be traced to poor data and information quality (DQ/IQ) in the data sources and systems used in construction engineering. Hence, there is a need to improve DQ/IQ in those sources and systems.

Before proceeding I would like to clarify an important issue; the use of the terms *data* quality (DQ) and *information* quality (IQ). Since there is little consensus in the literature on distinguishing the two, there is a need to introduce how these terms are used in this thesis. In short, DQ often refers to technical issues, while IQ is often related to non-technical issues (Madnick et al. 2009). In this thesis I do not distinguish between the two terms; rather, I use the term DQ/IQ to cover both meanings. Furthermore, this study adopts the "fitness of use" perspective (Wang and Strong 1996), which is based on the

intended purposes of information use. In section 2.2.1 I explain some additional background on why I do not distinguish between DQ and IQ.

DQ/IQ management has, over the years, received increasing attention in the field of information systems (IS) (Madnick et al. 2009). Due to the high costs of poor DQ/IQ (Ramaswamy 2006), it is important for organizations to be able to assess, manage, and maintain a sufficient level of DQ/IQ in data sources and data deliverables (i.e. drawings and other documentation). Research on DQ/IQ has been performed in a variety of contexts such as data warehousing, health care, and retailing. Researchers agree that DQ/IQ assessment is context-dependent (Strong et al. 1997), and several assessment frameworks and tools have been developed to fit these various contexts (Batini et al. 2009). Although construction engineering is a field suffering from poor DQ/IQ, unfortunately existing frameworks and tools cannot be used for sufficient assessment since they are based on assumptions that are not valid in this context (Neely et al. 2006). Further, while research has identified challenges related to DQ/IQ assessment in construction engineering, few, if any, solutions have been suggested.

My research is also rooted in a practical problem experienced by an organization where I am employed (anonymized as a European multi-discipline construction engineering company – EUMEC). It operates in the oil and gas sector and delivers engineering design and equipment mostly related to the offshore oil and gas industry (e.g., the design of and equipment for oil rigs and drill ships). EUMEC was experiencing delays and cost overruns in their construction engineering projects, and they wanted to find a way to mitigate this problem. During the early stages of this research, I identified poor DQ/IQ in data sources and drawings as the most important factor causing these problems.

I am an industrial PhD candidate, which requires me to perform research on a real problem experienced by the organization in which I am employed. I found that a possible solution to EUMEC's problem would contribute to both theory and practice. Thus, the importance of relevant DQ/IQ assessment frameworks and tools for the construction engineering context in general and for EUMEC in particular, combined with the challenging task of developing tools for such a complex environment, constitute the main motivation for this research.

The empirical basis for this research was the development and implementation of a DQ/IQ assessment tool named Information Quality System (IQS). I applied an action design research (ADR) approach, through which IQS gradually emerged. By the completion of this research, IQS had been implemented in the test project (TestP) and 15 other construction engineering projects at EUMEC. I discuss the DQ/IQ assessment challenges described in the literature and examine the very interesting relationships among these challenges. I examine how the various actions carried out by the ADR team affected the work of the engineers, and how the level of DQ/IQ for TestP was proven better than for projects not using IQS. The design principles embedded in IQS contribute to existing theory and provide useful advice to practitioners.

1.1 Research Questions

It is important to stress that the main research question was not formulated from the outset of this study. When I began this study, the problem was framed as delays and cost overruns experienced in relation to construction engineering projects within EUMEC. I conducted a Delphi study to identify and rank the problems causing the delays and cost overruns. Among 18 ranked problems, 8 were related to DQ/IQ, including the top 6. My focus then narrowed from problems causing delays and cost overruns to those problems specifically related to DQ/IQ. Hence, the main research question became:

RQ: How can *DQ/IQ* be managed in construction engineering?

Two sub-questions were formulated to answer the main question. As a starting point, I aimed to learn the breadth and state of the field. I especially wanted to know if any known challenges related to DQ/IQ assessment existed in construction engineering projects. The first sub-question was formulated as follows:

RQ1: What are the main challenges for managing DQ/IQ assessment in construction engineering projects?

When designing an information system in a unique and challenging environment, the challenges need to be embodied in the design process. Hence, challenges related to DQ/IQ management in construction engineering projects needed to be embodied in the

system design process of the artifact to be developed. A second sub-question was framed to handle this issue:

RQ2: How can the challenges identified in RQ1 be addressed through system design? (i.e., which design principles are needed for a DQ/IQ assessment tool to be relevant for construction engineering projects?)

The results from this research have been presented in five research publications (see Section 5). In this thesis I integrate these publications and present the research findings as a coherent whole.

1.2 Structure of the Thesis

This thesis integrates the findings from my five research articles (Appendix D). Some of the text, figures, and tables from the articles are brought into the thesis to make the story flow and to make it easier for the reader to understand the main issues without having to consult the articles. However, the articles include deeper elaboration on the issues in focus.

The thesis is divided into four parts: A) Problem, B) Solution, C) Contribution, and D) Reflection.

In part A, I present my motivation for performing this research, I introduce the research questions, and then I discuss related research.

In Part B, I briefly present the research approach before I describe in detail how the approach guided the emergent design, development, and implementation of IQS.

In Part C, I summarize the five publications accompanying this thesis before I discuss the theoretical and practical contributions. Then I draw conclusions, discuss limitations, and present suggestions for future research.

In Part D, I provide my reflections on several aspects of this research.

2 Related Research

This section provides an overview of the literature related to the research described in this thesis. First, I discuss pertinent literature within the areas of a) construction engineering, and b) DQ/IQ, and then narrow the focus to literature concerning c) DQ/IQ challenges in construction engineering.

2.1 Construction Engineering

Large construction engineering projects are generally complex and challenging to manage (Miller and Lessard 2001). These projects are characterized by a large number of project participants and a broad variety of stakeholders (Yeo and Ning 2002). In addition, the construction (assembly) sites are usually located far away from where the design is carried out, which raises the continuous challenge of providing adequate project information to construction sites (Braimah and Ndekugri 2009).

The teams carrying out these projects are of a transient nature and deliver their products through multiple, temporary project-based organizations that exist only for a single project (Cherns and Bryant 1984). The projects are usually carried out in alignment with a Project Execution Model (PEM) (Kumar et al. 2006; Sunavala 2008). In the following, I present the PEM phases typical for managing construction engineering projects in the oil and gas industry, which is the context of this thesis. These projects may include designing an oil rig, a drill ship, or the parts of such constructions. The design includes definitions of the complete system ("system" here means the entire construction which makes up a system of assets such as equipment, pipelines, and valves). Every detail is defined down to the smallest assets such as light switches, nuts, and bolts (including their functions, sizes and placement). All of these physical assets are known as engineering assets (Amadi-Echendu et al. 2010). The PEM is divided into main phases with several subphases. Figure 1 depicts the main phases; the circle indicates the detail engineering phase with handover to the assembly phase.

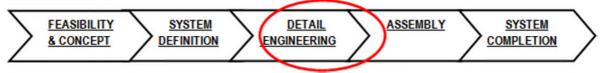


Figure 1. Typical phases of an engineering project

In the feasibility and concept phase, alternative solutions are identified, and one or two design concepts are selected for further work. At the end of this phase, one concept is chosen and some requirements related to this concept are frozen. In the system definition phase, drilling processes and equipment concepts are frozen along with the layout and main structure. At the end of the system definition phase, global design and equipment design are completed based on the chosen concept. In the detail engineering phase, detail design is completed and deliverables (e.g., drawings) are handed over to the assembly phase. During the assembly phase, the physical work is done based on the drawings received from detail engineering. The system completion phase includes system assembly, commissioning, and close-out, and concludes with the project complete.

The outcome of the detail engineering phase is mainly drawings and specification documents, which are later handed over to the construction site for the assembly phase (Ogunlana et al. 1998). It is the level of DQ/IQ in the drawings and their sources that is the focus of this thesis.

Drawings carry the data needed for assembly. A Machinery Arrangement drawing will commonly display the placement of various assets. One example that illustrates how machinery might be displayed is a roughneck, a machine that performs pipe handling for drilling activities. The drawing will illustrate where the roughneck is placed in relation to all other assets in the drawing, and it will display the exact coordinates (X, Y, and, Z) and total weight for every asset. In addition, another drawing is needed for the roughneck, called the General Arrangement drawing. In this drawing, the roughneck is illustrated in detail. The drawing also contains a list of needed information including weight per part making up the total roughneck weight, overall dimensions, rotation angle, drill pipe capacity, pipe spinner elevation, tube connection system, and hydraulic pressure requirement. If the level of DQ/IQ in sources is poor, then these drawings will contain less information than needed or even incorrect information.

The drawings are extracted from various information systems such as 2D or 3D tools and engineering databases. During the detail engineering design phase, the engineers insert

data into these information systems. In this phase, the physical engineering assets may not yet have been produced. Hence, the physical assets may exist only as referenced logical/functional elements in databases and data models. This situation underscores the key issue related to the construction engineering industry: that the correct data is not always known by the engineers at the time they need to insert the data to a data base record.

The typical phases of a construction engineering project displayed in Figure 1 also constitute the initial phases in the life cycle of engineering asset management (EAM). After the assembly and system completion phases, the physical engineering assets continue their life cycle through the phases of operation, maintenance, and finally retiring or disposal (Lin et al. 2006a, 2006b).

Although phase based, engineering today is iterative and concurrent. Concurrent engineering means that tasks previously performed in sequence are now performed in parallel. This break from the traditional linear engineering model is mainly attributed to the globalization of engineering, which requires shorter delivery schedules (Dobson and Martinez 2007). Concurrent engineering is a challenge when it comes to DQ/IQ and is further explained in Section 2.3.

2.2 Data and Information Quality (DQ/IQ)

DQ/IQ has been a long standing issue in Information Systems (IS) research. The negative consequences of low quality data contribute to the high interest in the topic, reflected by the huge number of publications. Given the interdisciplinary nature of such research, it touches a broad range of topics – IS, economics, medicine, information technology, etc. – and both individuals and organization are affected. To some extent one might say that wherever an IS is involved, so is the concern for DQ/IQ.

In the following sections, I discuss several important aspects of DQ/IQ, including concepts such as quality dimensions and assessment frameworks. However, I will first discuss data versus information to clarify why I do not distinguish between the two.

2.2.1 Data versus Information

Ackoff's (1989) article "From data to wisdom" is often cited when the definitions of and differences between the terms *data*, *information*, *knowledge*, and *wisdom* are discussed (Rowley 2007). M. Chen et al. (2009) have captured Ackoff's definitions of *data* as "symbols," and *information* as "data that are processed to be useful, providing answers to 'who,' 'what,' 'where,' and 'when' questions" (p. 13). This definition is intuitively easy to understand: if, for example, the letter "A" (data) was presented, the reader may not fully understand the meaning unless something more was included, for example, "the grade is A," "the alphabet starts with the letter A," or "the energy class of this refrigerator is A" (information).

The importance of context is reflected in several studies on the following topics: the need for including the context in the conceptualization of data quality (Strong et al. 1997); that data could be meaningless unless placed in some context (Tayi and Ballou 1998); practitioners' ability to view data in its context and by that be able to link otherwise separate data processes (Lee 2004); collecting contextual information as part of a data quality methodology (Batini et al. 2009); and "the notions of 'good' or 'bad' data cannot be separated from the context in which the data is produced or used" (Bertossi et al. 2011, p. 52).

When discussing knowledge and how individuals understand data and/or information, Alavi and Leidner (2001) state that "...for individuals to arrive at the same understanding of data or information, they must share a certain knowledge base" (p. 109). This shared "certain knowledge base" could be a shared context, for example, a group of students in a classroom or a group of engineers in a construction company. The members of each of these groups share a certain knowledge base. This means that data occurring in that specific contextual environment will be perceived as information for those belonging to that environment, but will be perceived simply as symbols (data) with no meaning by those who do not.

Moving on to the area of data and information quality, the discussion continues: is DQ different from IQ, and if so what is the difference? Most researchers agree there is a difference (Talburt 2011), and that difference is often described as technical issues (for

DQ) versus non-technical issues (for IQ) (Madnick et al. 2009). Even if most researchers agree that there is a difference, the terms "data quality" (DQ) and "information quality" (IQ) are often used interchangeably (Talburt 2011). In their book "Introduction to information quality", Fisher et al. (2012) briefly discuss data versus information in the first chapter before concluding that they will treat the terms interchangeably since "the context will make it [the difference] clear" (p. 4).

In the case of this thesis, the level of DQ/IQ is measured based on whether data or information handed over from detail engineering to assembly meets the requirements determined for the individual construction engineering project. The project requirements are contractual, and the IQS tool developed for this thesis assesses the data based on these requirements. Hence, the quality assessment is objective and task-dependent (Pipino et al. 2002), and the level of quality is determined by comparing the specific requirements for any given project with the actual data/information produced by the engineers for that project. It does not really matter whether a detected error relates to a technical or a non-technical factor. The error has to be corrected one way or another, and the main goal of the developed tool is to *identify* the error. Hence, in this thesis it is not important to distinguish between DQ and IQ, and I use the term DQ/IQ to cover both. Where the difference could possibly matter, the context will make it clear.

In the following section, I provide an overview of DQ/IQ management research, including the important concepts mentioned previously, quality dimensions, and assessment frameworks.

2.2.2 Overview of DQ/IQ Management

Over the years, DQ/IQ management has received increasing attention from researchers. The cost of an insufficient level of DQ/IQ has been found to be huge (Ramaswamy 2006; Strong et al. 1997), and problems have been observed across numerous corporations (Ramaswamy 2006; Wand and Wang 1996). Also, organizations have become increasingly aware of the importance of DQ/IQ because of the impact it may have on the success of their business (Otto et al. 2009).

Influenced by the Total Quality Management (TQM) framework expounded inter alia by Deming (1986), which aims at improving quality in manufacturing, DQ/IQ researchers have started to view information as a product. Acknowledging that the concepts and procedures of product quality control could be applied to information quality management (Ballou et al. 1998), researchers have recognized that "information may be treated as a *product* and the steps involved in creating it as a set of *manufacturing* processes" (Shankaranarayanan et al. 2000, p. 1). Wang et al. (1998) explain four principles for treating information as a product and term the application of these principles "the information product (IP) approach" (p. 96). Understanding the information needs of the consumers is one of those principles. This contextual principal is important because different purposes for the same information may indicate varying judgments about the quality of that information (Talburt 2011).

Whereas early DQ/IQ research focused on query techniques for multiple data sources and data warehouses at the end of the 1980s, the research field has since spread to a number of new application areas such as customer resource management, knowledge management, supply chain management, and enterprise resources planning (Madnick et al. 2009), and includes varied contexts such as data warehousing (e.g. Blake and Mangiameli 2011; Wixom and Todd 2005), health care registers (e.g. Pipino and Lee 2007; Vician 2011), and retailing and internet-related systems (e.g. de Corbiere 2009; Helfert and Hossain 2010). Among the issues that have emerged from this body of literature are quality dimensions, which is an important aspect of DQ/IQ assessment.

2.2.2.1 Quality Dimensions

It is hard to find a definition of the term "quality dimensions" in the DQ/IQ literature, and authors instead try to explain what they mean when they use the term without really providing a definition. Here are two examples: 1) Wand and Wang (1996, p. 87) state that "Data quality...is a multidimensional concept. Frequently mentioned dimensions are accuracy, completeness, consistency, and timeliness." and 2) Batini et al. (2009, p. 6) state that "quality dimensions can be referred either to the *extension* of data – to data values, or to their *intension* – to their schema," before they move on to describe the classification of the different quality dimensions.

Even if the term's meaning may be rather intuitive when explained like this, a definition that might be suitable in the DQ/IQ area of research comes from the field of cognitive science and defines the "quality dimension" through its primary function:

"The primary function of the quality dimensions is to represent various qualities of objects" (Gärdenfors 2004, p. 6).

Gärdenfors mentions height, width, and depth as examples of spatial quality dimensions. In the case of this thesis, these examples are not directly transferable, but if stated as "The primary function of the quality dimensions is to represent various qualities of data values," then examples of more familiar quality dimensions such as accuracy, timeliness, and consistency make sense.

Quality dimensions make it easier to define and discuss issues related to DQ/IQ without referring to specific data attributes (Abate et al. 1998). Rather, quality dimensions reference groups of data values with the same qualities (e.g., accurate values, consistent values, etc.). Over the years a vast amount of research on quality dimensions has been conducted, with two of the more important studies being Wand and Wang (1996), aimed at defining the various quality dimensions, and Wang and Strong (1996), aimed at identifying quality dimensions important to data consumers. Batini et al. (2009) have identified altogether 28 quality dimensions in their review of data quality methodologies. Many of these dimensions have varying definitions and measures for data quality evaluation due to the contextual nature of quality. Four core dimensions, accuracy, completeness, consistency, and timeliness, were emphasized frequently throughout these methodologies. In another study, Ge et al. (2011) assessed 17 dimensions commonly used for data quality evaluation and tailored them to a set of 11 dimensions they perceived as the most important dimensions: accessibility, security, relevance, accuracy, completeness, consistency, timeliness, ease of understanding, reliability, ease of manipulation, and objectivity. These are the 11 dimensions I used as the basis for assessing quality dimensions in EUMEC's construction engineering projects (Paper 4). One of the findings was that the accuracy dimension at EUMEC was understood as a combination of completeness and consistency, an understanding also pointed out by Ge et al.: "users presume that inaccurate information consists of incomplete and inconsistent information" (p. 7). I want to stress this point here, because the accuracy dimension is perceived to measure if data is "correct, accurate, free of error, precise" (p. 5), and error-free data is of course the ultimate goal in the data sources and drawings at EUMEC. Hence, the importance of the accuracy dimension is indisputable, but in the case of EUMEC this dimension was covered by the completeness and consistency dimensions.

When Neely and Cook (2008) reviewed DQ/IQ literature for the period 1996–2007, they concluded that the quality dimension was one of three over-researched areas (semantics and distribution being the other two). However, due to the contextual nature of quality as mentioned by Batini et al. (2009) and the importance of quality dimensions for the guidance of system designers (Wand and Wang 1996), it seems necessary to determine which dimensions are needed and sufficient for new contexts being researched. Hence, since construction engineering is a relatively new context in DQ/IQ research, quality dimensions are important to investigate in that environment.

Quality dimensions are vital concepts in assessment frameworks, which are discussed in the next section.

2.2.2.2 DQ/IQ Frameworks

DQ/IQ assessment frameworks are meant to guide actions that aim to increase the level of DQ/IQ in various contexts. The assessment tools informed by these frameworks are meant to increase the level of DQ/IQ in various information systems by identifying and automatically correcting data errors. Such tools focus on typical business systems and primarily on the accuracy of the data (Neely et al. 2006). The goal is to analyze data describing real world objects (Batini et al. 2009). Hence, the "correct answer" (as Neely et al. 2006 termed it) has to be known in order to perform the comparison of inserted data with real world data. This is the predominant paradigm for DQ/IQ assessment. However, as will be seen in section 2.3, this paradigm does not hold in construction engineering.

Several assessment frameworks exist, and all have the following basic DQ/IQ concepts in common: methodological phases and steps, strategies and techniques, DQ/IQ dimensions, types of data, and types of information systems. Based on these common concepts, Batini et al. (2009) used various assessment and improvement steps to compare several frameworks. From these I chose a set of concepts I found most relevant, and I added one additional concept: context focus. This addition was a consequence of the challenges I

have identified for construction engineering, which I discuss in section 2.3. The concepts were chosen based on these reflections:

- "*Extensible to other dimensions and metrics*": I did not know at the outset which dimensions would be relevant, hence the possibility to extend to needed dimensions was necessary.
- "*DQ requirement analysis*": I knew the DQ requirements would be heavily related to the requirements for each construction engineering project; hence, a DQ/IQ requirement analysis would be necessary for each new project.
- "*Process control*": Given the iterative nature of concurrent engineering, it was necessary to identify in which phases of a construction engineering project the various correct data values should be known. This information could be used to determine *when* it was useful to assess the various data values. From the description of construction engineering projects and their phase-based execution (section 2.1), at least one such check point appeared to be obvious: the handover from the detail engineering phase (e.g. drawings) to the assembly phase. It also seemed necessary to monitor the quality up to and including that point.
- "Measurement of quality" was needed for monitoring progress and for comparison.
- "Identification of causes of errors" was needed to possibly avoid future errors.
- "*Context focus*" was needed since construction engineering is characterized by unique challenges related to DQ/IQ assessment.

Table 1 displays the frameworks and the chosen set of concepts as identified by Batini et al., in addition to the added concept I also perceived as important (last column). However, the table displays an '(X)' for TIQM in the "Process control" column, which differs from Batini et al. According to them, process control means defining "check points in the data production processes, to monitor quality during process execution" (p. 16:4). But, they did not include process control as one of the characteristics for TIQM. I, on the other hand, think process control *is* one on the characteristics for TIQM since one of its process definitions explicitly states "Identify and implement quality controls and assessment schedules as necessary" (English 1999, p. 301).

Framework Acronym	Extended name	Extensible to other dimensions and metrics	DQ requirement analysis	Process control	Measurement of quality	Identification of causes of errors	Context focus
TDQM	Total Data Quality Management	Fixed			х	X	
DWQ	The Datawarehouse Quality Methodology	Open	X		x	X	
TIQM	Total Information Quality Management	Fixed	X	(X)	X	X	
AIMQ	A methodology for information quality assessment	Fixed			Х		
CIHI	Canadian Institute for Health Information methodology	Fixed					
DQA	Data Quality Assessment	Open			X	X	Х
IQM	Information Quality Measurement	Open			Х		
ISTAT	ISTAT-methodology	Fixed			Х	Х	
AMEQ	Activity-based Measuring and Evaluating of product information Quality (AMEQ) methodology	Open			Х	Х	
corpo	Loshin Methodology (Cost-effect Of Low Data Quality)	Fixed	х	х	Х	Х	
DaQuinCIS	Data Quality in Cooperative Information Systems	Open			Х	Х	
QAFD	Methodology for the Quality Assessment of Financial Data	Fixed	Х		Х		
сDQ	Comprehensive methodology for Data Quality management	Open	х	х	х	х	

 Table 1. Assessment frameworks and concepts (adapted from Batini et al. 2009 © 2013 Association for Computing Machinery, Inc. Reprinted by permission)

This comparison provided useful information when selecting the guiding frameworks. As can be seen from Table 1, DQA was the only framework that included specific guidance for contextual issues. More explicitly, this relates to task-dependent measures which are explained below in the discussion on selected frameworks. In addition, DQA allowed for extending to other dimensions and metrics. Three other frameworks, TIQM, COLDQ, and CDQ, fulfilled the needs related to all of the other concepts. However, I was familiar with TIQM from previous experience and trusted it would provide the needed guidance, while I was unfamiliar with COLDQ and CDQ.

Ultimately I chose two frameworks for guiding the research performed in this thesis: Total Information Quality Management (TIQM), proposed by English (1999, 2003), and Data Quality Assessment (DQA), proposed by Pipino et al. (2002).

Total Information Quality Management

The TIQM framework, as presented by English (1999, 2003), was developed on the assumption that DQ/IQ problems can only be solved by addressing "the organization's management systems that set performance measures and influence employee behavior" (p. 1). The framework is based on principles, techniques and processes used for quality management in general, which English considers equally useful for information quality management. The framework establishes six processes (see Table 2). The first three are assessment processes, the fourth is a process of data correction, and the fifth is an improvement process to prevent recurrence of errors. Finally, the sixth process is an overall process of culture transformation adapted from Deming's (1986) 14 Points of Quality for assuring a supportive business environment in terms of enabling continuous improvement.

Table 2. The six	processes (of TIOM	(from]	English, 20)03)
I doit at Inc Six	processes .			- inginoing # 0	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,

Process	Process description
(P1) Assess Data Definition	This process defines how to measure the quality of data definition to
and Information	meet the requirements of the knowledge workers to know what they need
Architecture Quality	to know and to understand the meaning of the information they require.
	This process also defines how to measure the quality (stability, flexibility
	and reusability) of data models and database and data warehouse designs.
(P2) Assess Information	This process defines how to measure the quality of information to meet
Quality	the various quality characteristics, such as accuracy, completeness, non-
	duplication and consistency across multiple databases, as required by its
	information customers. This process measures either the state of IQ
	within a database or data collection or the IQ produced by a current
	process.
(P3) Measure Nonquality	This process defines how to establish the business case for Information
Information Costs and	Quality Management. By measuring the cost of process failure,
Risks	information scrap and rework, alienation of customers and lost business,
	and missed opportunity, you can "speak with data" [] that is presenting
	in business terms to business management.
(P4) Reengineer and	This process defines how to conduct data correction projects, transform
Correct Data	information, and control data movement processes for data warehouses
	or data conversion projects. The process of data correction is NOT a
	stand-alone process, but one to be conducted along with a P5 initiative to
	improve the processes to prevent recurrence of the defects having to be
	corrected.
(P5) Improve Information	This process is required to call a methodology a "quality management"
Process Quality	methodology. This process implements the Shewhart Cycle known as
	PDCA or Plan-Do-Check-Act. It describes the tried and true process
	improvement technique defined by Walter Shewhart, and used by W.
	Edwards Deming, Joseph Juran, Masaaki Imai (in the Kaizen quality
	system) and other proven quality management systems. It is through P5
	that you deliver value to the enterprise by improving broken processes
	that cause downstream process failure and high costs of information
	scrap and rework.
(P6) Establish the	This umbrella step is not a discrete process with a defined beginning or
Information Quality	end. It describes the 14 Points of Information Quality that must be
Environment	inculcated into the culture of the enterprise to create and sustain an
	environment for business performance excellence and the habit of
	continuous improvement.

TIQM is a methodology for utilizing quality tools and techniques. Based on the implicit requirements embedded in TIQM processes, it is possible to assess existing tools for deficiencies, and it is possible to build new tools. The key to success, according to English, is the acceptance of DQ/IQ management and improvement as a continuous process. This is emphasized since continuous process improvement sometimes is confused with continuous data cleansing. However, the latter will not provide any

opportunity for correcting the root cause of errors; hence the process will not improve. Another important aspect of this methodology is how to measure accuracy. Measuring accuracy to reality can only be done by going to the object and compare data to that physical object. Measuring accuracy to surrogate sources (sources containing reference data such as postal addresses) will only reflect accuracy to the extent the surrogate source is considered accurate.

Data Quality Assessment

The Data Quality Assessment (DQA) framework was developed to assess and measure the level of data quality in a company. Pipino et al. (2002) present the framework together with a list of data dimensions, a set of objective and subjective measures, and a suggestion for how to analyze and compare these measures. In Table 3 the assessment types and descriptions are displayed.

Table 3. Assessment types in DQA (adapted from Pipino et al., 2002 © 2013 Association for Computing Machinery, Inc.Reprinted by permission)

Assessment type	Description
Subjective data quality assessments reflect the needs and experience of stakeholders.	Stakeholders may be collectors, custodians and consumers of data products. Subjective data quality is the quality of data as perceived by the stakeholders
Objective assessment can be task- independent or task-dependent.	Task-independent metrics reflect the state of data without the contextual knowledge and may be applied to any data set. Task-dependent metrics are developed in specific application contexts.

This methodology distinguishes between *subjective* and *objective* quality metrics, and Pipino et al. (2002)'s description of the differences and similarities can be summarized as follows¹:

Subjective data quality assessments are tied to the needs and experiences of the different stakeholders, such as collectors, custodians, and consumers. How stakeholders perceive the quality of data will influence their behavior.

¹ The summary is based on Pipino et al. (2002); © 2013 Association for Computing Machinery, Inc. Reprinted by permission.

Objective data quality assessment can be either task dependent or task independent. Taskdependent metrics include business rules, regulations and policies, and constraints provided by the database administrator – all of which are specific to a given context. Task-independent metrics reflect states of data without any contextual knowledge, and can as such be applied to any data set.

Subjective assessment surveys and objective assessment metrics are executed against the data set at stake, and the results are compared. Any discrepancies are then assessed in an effort to determine the root causes of these discrepancies. When the causes are determined, actions are taken to improve the level of data quality, and new comparisons are made by executing the subjective assessment surveys and the objective assessment metrics against the data set again. This is an ongoing and iterative action that should be executed continuously. The most difficult task when trying to measure and compare outcomes is defining the different data quality dimensions that align with the context at stake (e.g. in this thesis the accuracy dimension is understood, and hence defined, as a combination of the completeness and consistency dimensions). Formulating the metrics is considered fairly easy once the quality dimensions are defined. Three functional forms for developing metrics are suggested: Simple Ratio for measuring desired outcomes to total outcomes; Min or Max Operation for dimensions requiring aggregation of multiple data quality variables; and Weighted Average as an alternative for multivariate cases. The assessments and subsequent necessary actions are ongoing and iterative and meant to be executed continuously.

In addition, Pipino et al. (2002, p. 217) mention the possibility for an industry to create a de facto standard by adopting a set of data quality metrics. Such a standard would provide an opportunity to determine the level of DQ/IQ and to compare the quality levels with other industries.

2.3 DQ/IQ Assessment Challenges in Construction Engineering

Recently the attention of DQ/IQ researchers has turned to the context of construction engineering (e.g., Lin, Gao and Koronios, 2008; Tribelsky and Sacks, 2011). Several challenges related to DQ/IQ assessment in construction engineering have been identified, one of them being that accuracy in data is difficult to achieve at certain stages in the

lifecycle of engineering asset management (EAM). Engineering assets produce vast amounts of data when they are in operative mode. The problem of achieving accuracy has been identified mainly in relation to such operational data. The accuracy of these data cannot be determined on the basis of comparison with real world data (physical objects), nor can they be compared with accurate listings of data from database sources (data objects) (Neely et al. 2006). In short, the correct data is not known, hence it is hard to determine whether operational data are accurate or not.

However, my work focuses on the part of the life cycle where the outcome of the engineering design phase is handed over to the assembly phase. The following discussion explains that at certain stages in the engineering design phase, accuracy, in terms of correctness and preciseness of data (hereafter termed correct data), can be difficult to achieve and even assess.

As mentioned previously, physical engineering assets and the related asset data need not necessarily be present at the same time, especially during engineering design. Physical assets may exist only as referenced logical/functional elements in data registers and data models, since the physical asset itself may not yet be produced. Hence the correct data is not known at the time of insertion of a value into a record. As a result, it has been recognized that sometimes data handed over to assembly (e.g. drawings) does not comply with the physical asset delivered (Lin et al. 2006a). It has also been recognized that for assembly site workers, it is common to have only 70% of the necessary data on-site (that is, at the assembly site, e.g., a yard) (Tulacz 2005), which leads to time consuming searches for the needed data.

There are a number of reasons why the correct data may not be known, and the following discussion addresses issues such as the iterative nature of concurrent engineering, unique characteristics of construction engineering data, known challenges related to DQ/IQ assessment in construction engineering, and the reasons why DQ/IQ assessment tools are rarely used. I explain how several of these issues become unavoidable challenges and how these challenges inevitably lead to not knowing the correct data. These overall circumstances lead to poor DQ/IQ in data sources and drawings.

The first challenge stems from the iterative nature of concurrent engineering where tasks previously performed in sequences are now performed in parallel due to short delivery

schedules (Dobson and Martinez 2007). Hence, since many of the involved processes are interdependent, the engineers are forced to proceed with partial information, incomplete knowledge, and subjective interpolations (Blechinger et al. 2010). Further, in engineering design, "straw" values are often used for some parameters to enable analysis. Straw values represent various preliminary data values that are inserted by the engineers while they are waiting for correct data values. However, these straw values can lead to rework if they result in design errors (Tribelsky and Sacks 2011). Hence, proceeding with partial information or straw values will, unless they are identified and corrected at some point, cause poor DQ/IQ in the IS systems used as sources for the extraction of drawings.

The second challenge relates to engineering data. Engineering data possesses unique characteristics different from data related to typical business environments, such as the following features: substantial data must be collected from many different parts of the organization (Lin et al. 2007); data are collected in a variety of formats and stored in different media (Neely et al. 2006); and such data require experts with professional knowledge for interpretation (Lin et al. 2007). The latter also indicates a need for manual correction of possible data errors.

The third challenge relates to the processing of engineering data. Engineering data are process dependent and complex to integrate (Lin et al. 2007). The lack of integration between processes leads to difficulties in connecting various information (Lee et al. 2001). This can lead to time-consuming searches for existing but not easily accessible information. Tribelsky and Sacks (2011) have identified a positive correlation between the quality of information flows and the effectiveness of design documents (e.g., less rework due to more accurate drawings) and found that unstable information flows were associated with unpredictable project outcomes.

The fourth challenge stems from a lack of integration between systems: different values and different numbers of replicated items in different Information Management Systems (Blechinger et al. 2010); specialized systems bought from multiple vendors leading to a very difficult integration job for the end-users (Lin et al. 2007); and the lack of integration between business systems and technical systems making it difficult to achieve a comprehensive overview of status resulting in significant DQ/IQ consequences.

The fifth challenge, lack of timely information, is a consequence of the four challenges discussed above:

- 1) The iterative nature of concurrent engineering forces the engineers to proceed with partial information (Blechinger et al. 2010). This implies a lack of timely information.
- 2) The uniqueness of engineering data includes collecting substantial data from many different parts of the organization (Lin et al. 2007). This can be time consuming and lead to the lack of timely information, especially in light of the two following challenges.
- 3) The lack of integration between processes can lead to time-consuming searches for existing but not easily accessible information (Lee et al. 2001), which can result in the lack of timely information.
- 4) The lack of integration between systems can lead to difficulties in exporting data (Neely et al. 2006), and the resulting delay will contribute to the lack of timely information.

Taken together, it appears that the lack of timely information can be viewed as an unavoidable consequence.

Since some of these challenges discussed above are unavoidable, they necessarily have implications for the development of DQ/IQ assessment tools. While the lack of integration between processes and the lack of integration between systems are core foci for an IS researcher, it is far too risky to change processes and systems within a single project (Blechinger et al. 2010). These changes and alignments are better done at a higher organizational level. The implication is that within a single project, these challenges are also unavoidable.

Figure 2 is an adapted representation of these challenges and their interrelations.

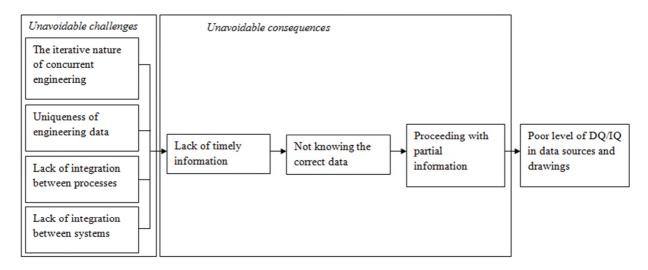


Figure 2. Connections between identified challenges and level of DQ/IQ (adapted from Paper 2)

The four unavoidable challenges on the left side of Figure 2 led to three consecutive unavoidable consequences. First, all four challenges could lead to the fifth, "lack of timely information," as explained in the numbered list under "*the fifth challenge*" above. This fifth challenge hence becomes an unavoidable consequence of the first four challenges. Second, if the information is not there when it is needed, this means the engineers do not know the correct data to insert at the time they need to insert it into a record. Third, not knowing the correct data will lead the engineers to proceed with only partial information (Blechinger et al. 2010). Inevitably, this will lead to a poor level of DQ/IQ in data sources and drawings, unless properly managed.

There is of course another possibility when the correct data is not yet known; the engineers can wait until the needed information is known, and proceed only with known, correct data values. However, due to the iterative nature of concurrent engineering, this alternative is not really an option. It will, at best, lead to delays and cost overruns because of tight project schedules. In the worst case it could lead to a complete halt of the project. Rather, the consequences of proceeding with only partial information need to be managed.

Another interesting issue is the rare use of DQ/IQ assessment tools in construction engineering. Such tools are infrequently used because engineering asset management is not considered a core business activity (Koronios and Lin 2007), hence the companies instead rely on traditional information sources such as the contents of information systems primarily designed for increasing productivity, and the knowledge of the engineers. Also,

in order for organizations to use assessment tools, available tools need to be effective, which may not be the case. A mapping of data problems to features of data quality tools was performed by Neely (1998), and Neely et al. (2006) refer to an additional review of data quality tools (reported in (CIEAM-SI-101-Team 2005), where tools were evaluated and compared. From the results of both of these studies, Neely et al. (2006) conclude that for assessing DQ/IQ in engineering asset management systems (which include engineering design), these tools are not effective.

As discussed in section 2.2.2.2, the predominant paradigm for DQ/IQ assessment is that the correct data *is* known. However, a consequence of the unavoidable challenges and their relations is that any future DQ/IQ assessment framework for construction engineering should be built on the assumption that the correct data is not known, and thus suggest guidelines for managing such a situation with the aim of achieving a sufficient level of DQ/IQ in data sources and drawings. My work aims at dealing with the consequences that are not unavoidable; that is, how to proceed with partial information in a manner that does not lead to a poor level of DQ/IQ in data sources and drawings.

2.4 Summary

In this chapter I have shown why engineering design data can be uncertain and imprecise. Further, I have emphasized that some challenges and consequences are unavoidable. I have discussed the consequences of proceeding with partial information and stated that unless managed, this will result in a poor level of DQ/IQ in data sources and drawings. Finally, I discussed why construction engineering has need for DQ/IQ assessment frameworks and tools that are based on assumptions relevant for the context.

In the next part, Part B: Solution, I discuss the research approach and development of IQS.

Part B: Solution

"Quality is not an act, it is a habit."

- Aristotle

3 Research Approach

In this chapter I first position my research approach within the research paradigms of IS. Then I present the Action Design Research approach, including its stages and the ITdominant perspective I used as the research framework.

The underlying philosophy in ADR is pragmatism (Sein et al. 2011). In pragmatism, truth lies in utility (Baskerville and Myers 2004). In accordance with the pragmatist paradigm, I established the purpose of IQS (the artifact) to be identifying quality errors in a manner that would make it possible to manage DQ/IQ in construction engineering projects. Further, the practical outcome would be continuously evaluated by the users of IQS to ensure user utility. The theoretical premises embodied in IQS could then be evaluated by evaluating IQS itself. The design process of IQS would be performed in collaboration with several employees covering a variety of roles and situated within EUMEC.

3.1 Positioning the Research Approach

The IS field focuses on the interaction between information technology (IT) and its social context for development and use (Avison and Fitzgerald 2003). Positivism has been the dominant perspective in IS research (Orlikowski and Baroudi 1991), followed by a growing interest in alternative research paradigms such as interpretivism (Walsham 1995). Benbasat et al. (1987) advocate the case research strategy as a qualitative technique that could complement other research strategies used in the IS field, and their work is frequently cited by IS researchers. Part of the reason for their call for a complementary research strategy stemmed from the fact that while quantitative research methods often require a large number of samples (nomothetic strategies), IS researchers also attempt to understand a phenomenon in its context (idiographic research strategies), which means that the researcher intensely examines a single entity or a particular event (Benbasat et al. 1987). These approaches are suitable for understanding and explanation. Other approaches, elaborated below, are change-oriented and suitable for improvement by intervention.

An approach focusing on a particular event in terms of a particular problem domain is termed action research (AR). The change-oriented approach of AR adopts the interpretivist and idiographic viewpoints of research inquiry, and in addition accepts qualitative data and analysis. The method is grounded in practical action and informs theory through solving problem situations (Baskerville 1999). Critiques of AR as a method for IS research have been raised, including the lack of scientific rigor and the argument that causal connections and explanations are difficult to establish due to an idiographic viewpoint (Baskerville and Wood-Harper 1996). As an answer to these and other critiques, McKay and Marshall (2001) developed a model of AR that includes two cycles: 1) the problem-solving cycle, and 2) the research interest cycle. In the problemsolving cycle, the aim is "improvements through making changes in a problematic situation," while in the research interest cycle, the aim is to "generate new knowledge and insights as a result of [the action researcher's] activities" (p. 50). These two cycles mean that practice and theory are both addressed. The cycles are interdependent and operate together. This model makes it possible to strengthen the rigor of the method and credibility of the knowledge or theory generated through real-life interventions by explicating reflections and learning. While AR is about improving, it does not specifically say anything about the design of artifacts. The method that explicitly does is design science research.

Design science research, or design research (DR) as the term used here, is an approach focusing on a particular problem domain related to artifact design. In DR, IT artifacts are created and evaluated with the intention to solve organizational problems. The practical implications of the designed artifact should impact the evaluation of the scientific research performed (Hevner et al. 2004). Even so, Sein et al. (2011, p. 38) stated that DR "does not fully recognize the role of organizational context in shaping the design as well as shaping the deployed artifact." They propose a new DR method; Action Design Research (ADR), which draws on AR to bring into the method a stronger emphasis on organizational intervention. In this study, one of the goals was to define design principles for a DQ/IQ assessment tool so that it would be relevant for construction engineering projects (RQ2). This could be done by developing the tool in EUMEC and then learning about possible design principles through this intervention.

Figure 3 illustrates how the research interest fits with an ADR approach.

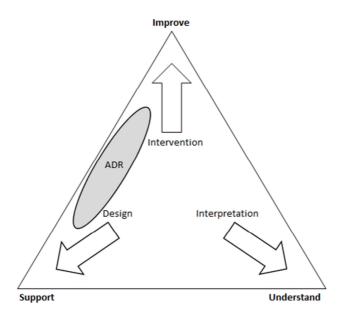


Figure 3. Positioning ADR (adapted from Mathiassen 2002, with permission from Emerald Group Publishing Limited 2013)

The figure was originally adapted from Braa and Vidgen (1999) and then presented by Mathiassen (2002). Here, the positioning of ADR is added.

Another interesting issue in collaborative research is that a variety of research approaches may be used and combined within the larger project (Mathiassen 2002). This was the case with my project. One of my approaches was a Delphi study, which are commonly used in IS research to identify and rank issues (Okoli and Pawlowski 2004). I used a Delphi study approach (Paper 1) to strengthen the argument for focusing on DQ/IQ related problems. Further, I used a case study approach to investigate the quality dimensions in EUMEC (Paper 4) for multiple reasons. First, DQ/IQ research in the construction engineering context was still very much under-investigated. In such new and little explored fields, case studies are especially appropriate (Eisenhardt 1989). Second, to be able to better understand DQ/IQ in a specific context, a case study was suitable; a phenomenon examined in a natural setting is a key characteristic of case studies (Benbasat et al. 1987). Third, since access to suitable organizations can be hard to achieve (Walsham 2006), I wanted to exploit the opportunity of access I was provided by EUMEC. Both the Delphi study approach and the case study approach I used were important to the ADR project. The result from the Delphi study was used as part of the problem formulation stage in ADR, and the opportunity to investigate the quality dimension emerged as a result of implementing the assessment tool in several projects during the ADR project.

ADR was the research approach implicitly and explicitly used through the thesis project. However, since the ADR approach was not published until March 2011 (Sein et al. 2011), it was impossible to choose ADR as the approach at the outset of the thesis project. Originally the project was framed as a prototypical DR project (Hevner et al. 2004), because an IT artifact was to be designed with the intention to solve an organizational problem. In March 2011, with ADR at hand, the team (including me) developing IQS realized ADR reflected our approach in a more accurate way. From then on, ADR was also explicitly used as the approach.

In the next section, ADR is briefly presented.

3.2 Action Design Research

ADR is a method for performing IT artifact design that emerges from interaction with an organizational context. The method considers organizational intervention a requirement for development, for it balances what the authors consider "the conflicting demands of (1) addressing a class of problems, and (2) intervening in authentic settings" (p. 39). Organizational intervention is crucial to ADR because of its underlying philosophy that the artifact is not only based on design, but also emerges from interaction with an organizational context. This view is different from existing DR methods, which focus on the technological issues related to the IT artifact and pay less attention to the impact the organizational context may have on the artifact when the two interact. Figure 4 displays the stages and principles of ADR.

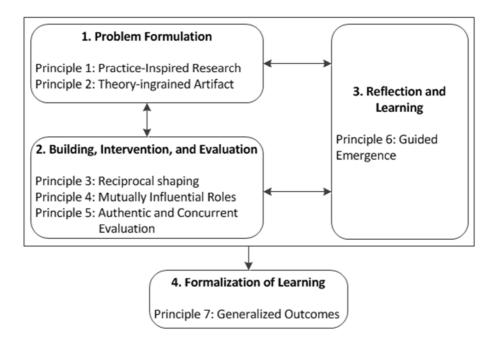


Figure 4. ADR method: stages and principles (from Sein et al. 2011, Copyright © 2011, Regents of the University of Minnesota. Used with permission.)

These four stages of ADR can be described as follows:

1. Problem Formulation

The research problem is perceived by practitioners or anticipated by researchers and provides the impetus for formulating the research effort as well as an opportunity for scholarly knowledge creation. It may include an initial empirical investigation of the problem. The specific problem is cast as an instance of a class of problems, which moves the problem-solving effort from the mere level of consulting to a level where new knowledge is generated. The problem formulation stage of this project is discussed in Section 4.3.1.

2. Building, Intervention, and Evaluation (BIE)

The IT artifact is initially designed based on the premises of stage one. It is further shaped by organizational use and subsequent design cycles. This is a continuous and iterative process that interweaves the building of the artifact, intervention in the organization, and evaluation. Design principles are articulated for the class of systems. Depending on the type of innovation, there are two end points for the design continuum: IT dominant and organization dominant. In this research the IT-dominant perspective is used as the research framework, which is illustrated in Figure 5. How the two BIE cycles were performed in this project is explained in sections 4.3.2 and 4.3.6, respectively.

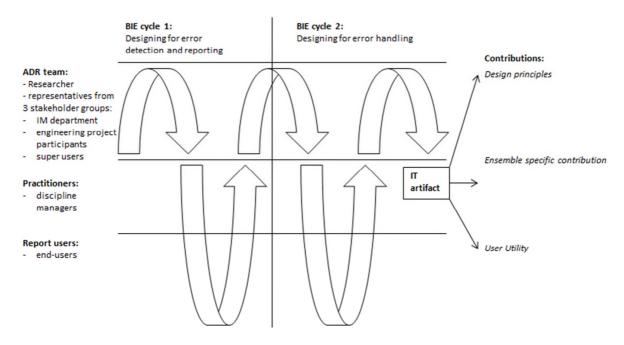


Figure 5. Generic schema for IT-dominant BIE (adapted from Sein et al., 2011, Copyright © 2011, Regents of the University of Minnesota. Used with permission.)

3. Reflection and Learning

This stage runs in parallel with the previous stages. Conceptually it moves from building a solution for a particular instance to applying that learning to a class of problems. It recognizes that the research process is about more than solving a problem. Conscious reflection is critical to identifying the contributions to knowledge. The research process is adjusted based on early evaluation results to include the increasing understanding brought about by the reflection process. The reflection and learning stages related specifically to the two BIE cycles in this project are presented in sections 4.3.3 and 4.3.7, respectively.

4. Formalization of Learning

The situated learning from an ADR project is further developed into general solution concepts for a class of field problems. By describing how the artifact design was

performed, the accomplishments realized, and the organizational outcome, the researcher formalizes the learning. The design principles that are generated contribute to theory building. This formalization of learning performed in this project is discussed in section 4.3.8.

3.2.2 Data Collection and Analysis

During the ADR project, data was collected and analyzed on a continuous basis. As mentioned, the result of the Delphi study was used in the problem formulation stage to strengthen the argument for focusing on DQ/IQ related problems. The engineers from EUMEC who represented the expert group of the Delphi study reached their level of consensus consistent with strong agreement (Schmidt 1997). This provided a high degree of certainty for the ranking and hence for the focus on DQ/IQ related problems. During the BIE iterations, data was collected in various formats: a proposal document, system charts, IQS rule definitions, requirement lists, compiled lists of decided actions related to various IQS issues, IQS reports, interface scripts, audio recorded interviews, and personal notes. The data was collected at various arenas such as in meetings where I was either a participant or a facilitator, from emails, from interviewees, or from discussions. For further details on data collection activities (e.g. who attended the meetings and who was interviewed) see Appendix A.

The analysis was performed throughout the research process in connection with data being collected. Feedback from construction engineering project participants was discussed with the participants themselves for clarification, and then further discussed and analyzed within the ADR team as part of the reflection and learning stage. Actions, such as design decisions and how to perform them, were then decided upon and performed as a response to the data collected and the following analysis. All of this was repeated several times during smaller or larger iterations within the project. The many feedbacks along with the following analysis and discussions also served as a validity measure of the ADR project. If the changes created were desirable to the practitioners, the feedback from them would confirm the validity of the project (Iversen et al. 2004).

Next, I report the emergence of IQS through the stages of ADR, including the two BIE cycles that were illustrated in Figure 5.

4 The Design and Emergence of IQS

In this chapter I describe the design and emergence of IQS. First, I present the target organization, followed by a discussion of the conceptual basis for IQS. Then, I present the emergence of IQS through the stages of ADR. Since ADR is an iterative approach, the research and practical actions were carried out back and forth through the stages. In Paper 3 the process is described in more detail. Finally, I conclude with a formal evaluation of IQS.

4.1 Target Organization

EUMEC is a European multi-discipline construction engineering company. Though the organization is European, their approximately 100 offices and 19,500 employees are located around the globe, including the US and China. The operating revenue is more than 6.5 billion USD per year. The company is a world leader in the management of global construction engineering projects, viz., engineering design, product development, procurement, completion, and complex installations for the offshore oil and gas industry. They have a significant share of the global market in their product and project domains. A majority of the employees take part in construction engineering projects. This research focuses on the intersection between the engineering design phase and the handover to the phases of assembly and completion, which are performed at the construction site. The main deliverables for handover are drawings, documents, and 3D models. The biggest projects typically cost over 150 million USD and take up to three years².

EUMEC's engineering projects are parts of larger construction projects and usually cover design-related activities. The final deliverables of an engineering project mainly consist of documents such as drawings and manuals. These are either drawings from suppliers or drawings extracted from various data sources used in engineering design, such as 2D/3D applications and engineering databases. When the design is complete, outputs from the project (e.g., drawings, documents, and 3D models) are handed over to the assembly and completion phases. It is during these latter phases that the most serious problems, especially delays, typically manifest. These phases are the same as those described in Section 2.1.

² Due to confidentiality I had to be brief on the company description

Several experts representing different engineering disciplines are needed to design large, complex, and robust-yet-delicate construction projects. Project teams are set up according to the engineering specialty and any other professional disciplines required by the individual projects. Table 4 lists the engineering disciplines typically represented in the projects at EUMEC.

Disciplines	Responsibility	
Process	Design of industrial processes; all the facts, sequences, and relations in the process and a logical placing of the different items.	
Mechanical	Design (choice of equipment and its physical layout and weight)	
Piping/Layout	Design of all piping	
Electro	Design and cabling of power distribution for electrical systems: equipment, lights, heat, etc.	
Instrument	Design of control systems, i.e., the control of various valves, machines, alarm systems, and instrumentation cables for distributing signals	
Telecom	Selection and location of radio and audio systems, alarms, etc.	
HVAC (Heating, Ventilating, and Air Conditioning)	Capacity calculations and layout for ventilation, etc.	
Safety	Various safety assessments.	
Structure (steel)	Design of steel structures, supports, outfittings like hand rails, stairs, etc.	
Architecture	Interior design	

Table 4. Engineering disciplines in EUMEC's construction projects (Paper 1)

Each engineering discipline is composed of several engineers and a manager who is responsible for the delivery as a whole. Due to the large number of interfaces and dependencies among the assets and systems to be designed and assembled, every engineering discipline depends on input from the other disciplines throughout the project. Tight project schedules mean that various activities must be performed in parallel. In such concurrent engineering (Sekine and Arai 1994), quality assurance and possible adjustments are therefore conducted both during the project and then again in the assembly and completion phases.

4.2 Conceptual Basis

A theory-ingrained artifact is a requirement for artifacts developed through ADR: they carry "traces of theory" that guide their design and development (Sein et al. 2011). TIQM and DQA were used as guiding frameworks for the development of IQS and were discussed in section 2.2.2.2. These were chosen because together they represent the origin of most assessment frameworks (i.e., TIQM) and provide good insight and explanations of issues related to practice (i.e., DQA).

IQS is an ensemble artifact, hence designing IQS involved not just the technological aspects but also the social and organizational aspects. In other words, a proper environment for DQ/IQ assessment was needed. TIQM provided a basis for that need: amongst several issues, TIQM emphasized setting the proper environment for DQ/IQ assessment. This included issues such as creating performance measures for information producers and for senior management, creating information policies, and encouraging employees to identify problems in information quality.

Further, to develop an assessment tool we needed to know what to measure. The DQA framework provided that focus, and its distinction between task-independent and task-dependent measures was especially useful for the development of IQS. Its applicability was due to the heavily context-dependent challenges identified for construction engineering, which meant that task-dependent measures would be in focus. Both frameworks emphasized the importance of solving the root of the problem. In the case of the ADR project, this meant ensuring a sufficient level of DQ/IQ in data sources before extracting and handing over drawings to assembly (how the development of IQS related to these premises is discussed in section 4.4).

Even though these two frameworks provided good guiding suggestions, they were both built on the assumption that the correct data value to insert into a record would be known by the time of insertion. None of the frameworks suggested that sometimes incorrect data values, such as straw values, would have to be inserted on purpose. Hence, to make a complete conceptual basis for the development of IQS, I brought in the characteristics of construction engineering in addition to the frameworks. Following Lempinen (2012) who used characteristics of performance measurement as part of a kernel theory, and Markus et al. (2002) who used characteristics of emergent knowledge processes as a kernel theory, I used characteristics of construction engineering projects as part of the conceptual basis. These characteristics are based on the seven unavoidable challenges and consequences previously illustrated in Figure 2, since the goal was to develop a system that would address them.

Characteristic 1: The iterative nature of concurrent engineering, which includes distinct interdependent phases conducted concurrently in the individual projects

Characteristic 2: The uniqueness of engineering data

Characteristic 3: Lack of integration between processes

Characteristic 4: Lack of integration between systems

Characteristic 5: Lack of timely information

Characteristic 6: Not knowing the correct data

Characteristic 7: Proceeding with partial information

Based on these characteristics, we derived the system requirements for IQS. Those are presented in section 4.3.2.

4.3 Designing IQS through Action Design Research

In this section, I describe the emergence of IQS. Figure 6 shows the timeline of the IQS project with the main events keyed: important actions, insights, and milestones. The description that follows is structured around the stages of ADR.

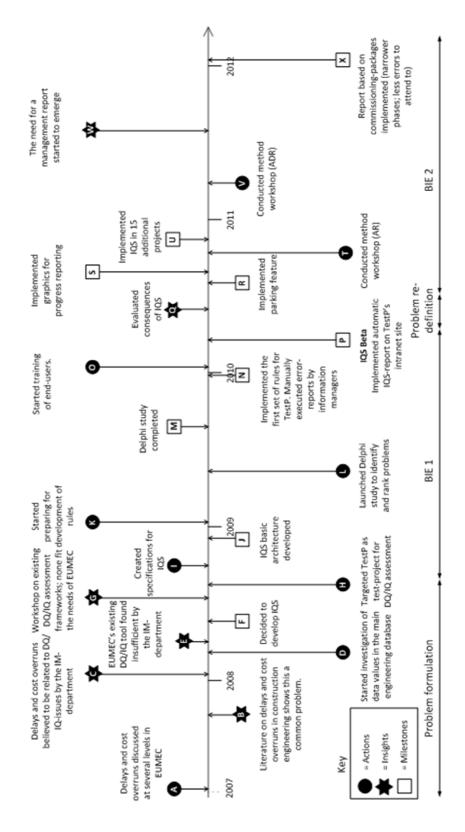


Figure 6. Project timeline (from Paper 3)

4.3.1 Problem Formulation

The problem in this study emerged from an increasing concern over delays in engineering construction projects experienced by EUMEC. The findings from the Delphi study, where DQ/IQ-related problems were deemed the most important (see Table 5), were used to strengthen the argumentation for better managing DQ/IQ and as such served as input to the problem formulation stage. In addition, collected documentation from previous projects, emails from site-workers, meetings with engineers, and interviews were all used as input to this stage. This input revealed missing information, inconsistent information, and information lacking logical coherence as the main concerns for the users of the information. It became clear that increasing the level of DQ/IQ in the construction engineering projects would be the goal for the organization.

Table 5 displays the ranking of the DQ/IQ-related problems identified the Delphi study.

Rank	Problems
1.	Delays in distribution of drawings and documents
2.	Design is based on unfinished or incorrect supplier documentation
3.	Equipment drawings change after engineering design is completed
4.	Errors and omissions in supplier drawings
5.	There are great shortcomings in the interface documentation on
	drawings, (not correct information as size, weight, tag number)
6.	GA (general arrangements/assembly) drawings are not consistent
	with the equipment
9.	Lacking interface within our organization between engineering,
	equipment, control systems, and flow diagrams (the drawings are
	not congruent)
11.	Copy projects always lead to recurring errors that we use hours to
	correct from project to project

Table 5. DQ/IQ problems identified in the target organization (from Paper 1)

We initially examined an existing DQ/IQ tool in use by the organization. The main feature of this tool was detecting missing values in the records of the main engineering database. Two problematic issues were detected: 1) often straw values were used. These values represented "guessed" values (which sometimes were incorrect), preliminary values (which had to be corrected when the engineer knew the exact value), or substitute values such as "NA" (not applicable) or "–". Straw values were not detected as missing

values, leaving a huge manual job afterwards for correction where the engineers had to assess thousands and thousands of records in order to identify such values. The other problematic issue was that 2) only the main engineering database was assessed, which meant that consistency with the other ISs used in the projects was not assessed. This was a problem, for example: 2D and 3D models were developed and stored using other systems. The models provided information input to engineering design, which was manually inserted into the main engineering database. Any discrepancies between the models and the information in the main engineering database could not be detected.

At the time, several different information systems related to engineering design had been evaluated on an organizational level. The aim was to find an information system that would embrace the needs of all business units performing engineering design, instead of having to manage different information systems for the different units. This evaluation process had just been completed, and one engineering system had been selected. The selected information system included a tool for DQ/IQ assessment. One of the business units tested this tool within the context of a live construction engineering project. They concluded that the tool ran too slowly: they had to start running the wanted report on a Friday after working hours because it took at least 24 hours for the tool to produce a complete report. This was far too much time given that they wanted daily reporting. The test team also expressed concerns related to the need for importing data from external data sources, and they suspected the running time for the report could even increase. After these concerns were raised, the testing was stopped and EUMEC decided not to use the DQ/IQ assessment feature of the selected system.

As a result of this process, EUMEC decided that a new assessment tool was needed for the following reasons: 1) the existing DQ/IQ tool used by the organization was not sufficient, and, 2) the test results reported for the new DQ/IQ tool embedded in the new engineering system showed that it was running too slow and that there were concerns related to importing needed data from other sources into the tool.

Even if the existing frameworks, TIQM and DQA, were not entirely sufficient for this context, they still provided useful information and guidance during the development of IQS. They were particularly helpful on creating an environment encouraging DQ/IQ, and on selecting suitable assessment types; in this case objective task-dependent metrics. I

describe how the developed artifact – IQS – relates to the premises of these frameworks in Section 4.4.

4.3.2 BIE Cycle 1: Designing for Error Detection and Reporting

Based on the conceptual basis discussed in Section 4.2 and the inspiration from existing research (e.g. the guiding frameworks and the importance of quality dimensions), we decided the system requirements for IQS.

System requirements and descriptions

• System requirement 1: IQS should be rule based

That IQS should be rule based was the main requirement. Rules can be defined as formal statements in a syntax code. This code can then be executed against any data set to assess for irregularities. The rationale for using a rule-based approach is as follows:

a) It would be possible to capture project requirements in rules, which also would make it possible to identify errors.

b) It would be possible to reuse rules in other projects because some rules could be applied to all projects.

c) The rules would be fairly easy to program once the project requirements were understood by everybody.

d) Developing rules would help clarify the project requirements so that it would be possible to agree on a rule set.

e) The rules would be easy to change or edit since they would be located outside the system source code. This was important for EUMEC because changes in requirements are almost inevitable during individual projects. Adjustments, changes, and additions to the construction design require immediate adjustment to the rules throughout a project. Data sources might even be added or customer systems changed, which could affect the rules.

f) The rules could be viewed as business knowledge in a way that made it possible to create a set of default rules (an archive) applicable to all projects, as well as any project-

specific rules (see b). Such an archive would also provide what Loshin (2002, p. 615) called "an archive of data quality expectations." For EUMEC, an archive of rules was important because it provided the employees (new and experienced) with a knowledge base that could be transferred between projects and also used for training. It would also be easier to estimate the cost of an information manager for an individual project when the expected level of DQ/IQ was known. The rules made this knowledge very explicit.

• System requirement 2: IQS should collect needed data from all needed information systems

Since EUMEC's old tool only assessed data from one information system, it became necessary to ensure that all needed data would be possible to collect and assess.

• System requirement 3: IQS rules should address three DQ/IQ dimensions: completeness, consistency, and logical coherence. Thus, the rules should identify which data could possibly be missing, which data could possibly be inconsistent, and which data could lack logical coherence.

The chief developer and I chose these three quality dimensions based on our experience from construction engineering projects. As previously mentioned, at EUMEC the accuracy dimension was understood as a combination of other dimensions. Hence, the need for accuracy was translated into those dimensions that were intuitive and operable for the situation: completeness and consistency. Our experience also included information from site workers who explained that missing and incorrect data/information in drawings led to time consuming searches for the correct information. We both noted that the incorrect data values often were related to inconsistencies or to the lack of logical coherence. These three dimensions of completeness, consistency, and logical coherence were verified as being needed and sufficient for rule-based assessment in paper 4, which presented the assessment of quality dimensions at EUMEC.

• System requirement 4: IQS should provide different rules for different phases of the project with details to be checked sequentially for every subsequent phase.

Due to the huge amount of data needed and the situation whereby much of the data would not be available at the start of the project, a full check of every requirement from day 1 would result in an unwieldy number of errors. The solution was to define different rules for use at different phases of the project.

The decision to develop a tool also coincided with the start-up of a new construction engineering project (TestP), which could be used as a test project for the tool. Hence, the project requirements of TestP were the requirements that needed to be captured in the first set of rules.

Defining rules

The rules were defined based on project requirements. There were two types of rules: common and project specific. Common requirements applied to every project at EUMEC and reflected internal issues such as organizational policies influencing data values and internal coding manuals. Project-specific requirements mainly reflected customer needs and contingent issues; such as the type of construction to be designed (the type of ship or oil rig) and the information systems to be used during design (customers sometimes require systems outside the organization's regular portfolio). Operating under the motto "everything that can be formalized can be assessed by rules" (chief developer), I defined sets of rules for each phase and stored them in a rule database.

Expert engineers from the various engineering disciplines were asked to explain *when* the correct data values were supposed to be known and hence could be expected to have a correct value in the database. Based on their responses, the assessment was divided into three phases aligned with existing PEM phases and with an increasing degree of detailed rule-checks at each phase. In the first phase, rules were defined to check only for the existence of data that should be known by the end of that phase. In the following phases the rules were progressively "tightened" to check more and more data fields whose values would be known by then. This reflected the increasing need for correct detail information. This process culminated in the last phase where everything needed to be correct; i.e., every data field would be checked against the determined rules.

The initial rules were implemented and executed (manually), and the reported errors were communicated to the users through email and discussed in organized meetings. These meetings brought about a need to refine the rules as the common understanding of TestP's requirements became clearer through these meetings. This BIE cycle also brought about

issues that proved to be challenging: the number of true errors was too large to handle, and the number of "false positives" reduced trust in IQS.

4.3.3 Description of IQS

Short résumé of use and purpose

The users of IQS are mainly engineers and managers participating in EUMEC's construction engineering projects. The main purpose of IQS is to identify errors in the data sources before the extraction of drawings. The errors must be corrected manually due to the need for expert engineering knowledge to correct the errors. The goal is to increase the level of DQ/IQ in the data sources.

Elaborating IQS

Figure 7 shows a schematic model of IQS.

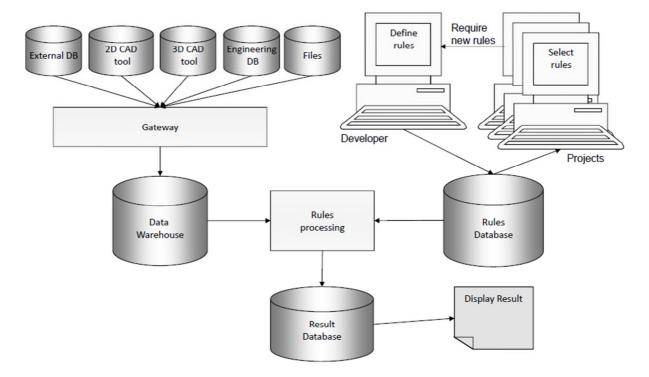


Figure 7. A schematic model of IQS

The various data sources depicted in the upper-left corner of the model are sources and systems used in engineering design. These sources consist of a main engineering database,

various CAD (computer-aided design) tools, files containing data in various formats (e.g., tif, gif, pdf, etc.), and external databases such as customer databases. Data from these sources are imported each night to an operational data warehouse. This integrated data repository is similar to an operational data store and thus contains data from disparate sources later used for reporting. The extracting, cleaning, and loading of the data is done using interfaces developed in-house. Each record in the data warehouse is marked with the ID of the project it belongs to and a contractor ID (each project may include several contractors).

In the upper-right corner, the schematic illustrates how construction engineering projects may either use existing rules (from a default set of rules applicable to all projects), or may ask for rule definitions specifically tailored to the individual project's requirements. New rules for such cases will be defined by a developer. Each requirement is separately captured in the syntax of a rule, so the rule can be executed against all required data records with the purpose of identifying any irregularities concerning the requirement.

All the rules are stored in a rule database, and each rule may be used by one or more projects. Every 24 hours, the rules are processed against data from the data warehouse. Each rule assesses data for each of the projects that have asked for this specific rule. Any deviation from the requirement specified in the rule will be stored in the Result Database. The result for each project is then displayed in detail on the project's home intranet site.

Rules and reports are elaborated in more detail in Appendix B.

4.3.4 Reflections and Learning from BIE Cycle 1

The process of building IQS, intervening in the project by implementing IQS in TestP, and evaluating its effects generated four design principles. These are shown in the left-hand column of Table 6, and their descriptions are shown in the middle column. The evaluation revealed both anticipated and unanticipated consequences, which are shown in the right-hand column, under Implications.

Design Principle	Description	Implications
Allow for inconsistency	Inconsistencies caused by straw values or blanks should be allowed at appropriate early phases.	 Reduced number of reported errors (anticipated) The engineers stopped using time on manual identification of inconsistencies and focused on errors possible to correct at the time (anticipated)
Allow for incompleteness	Incompleteness caused by straw values or blanks, should be allowed at appropriate early phases.	 Reduced number of reported errors (anticipated) The engineers stopped using time on identification of incompleteness through the original tool and focused on errors possible to correct at the time (anticipated)
Allow for lack of logical coherence	Lack of logical coherence caused by straw values or blanks should be allowed at appropriate early phases.	 Reduced number of reported errors (anticipated) The engineers stopped using time on manual identification of possible errors related to logical coherence and focused on errors possible to correct at the time (anticipated)
Phase-based reporting (report only identified errors applicable for current phase)	Preliminary incorrect values should not affect the daily report until the appropriate project phase is reached.	 The IQS report used as a working list for correction of identified errors (anticipated) The number of reported errors perceived was still too large. Engineers found it difficult to prioritize errors to correct first, leading to demotivation (unanticipated) IQS generated "false positives" leading to reduced trust of IQS and consequently to demotivation (unanticipated)

 Table 6. Implications of BIE cycle 1 (adapted from Paper 3)

The first three principles, to allow for inconsistency, incompleteness, and lack of logical coherence, had anticipated outcomes. The engineers were generally satisfied with several aspects of the system. They appreciated that IQS assessed not only completeness, but also inconsistency and logical coherence. They also mentioned that they obtained a better total overview because the assessed data were collected from all relevant data sources.

"Assuming the rules are correctly defined, IQS is a nice tool." (electro engineer)

"If we try to do it manually [assessing data] it takes a lot of time and you also have to know exactly what you are looking for. IQS detects almost everything and fast." (discipline manager) "It is so much easier to use the IQS report to identify discrepancies between the main engineering system and the 3D modeling tool, it provides a better overview than if you try to do it manually." (mechanical engineer)

The principles affected the way the engineers worked: they stopped using time on manual identification of possible errors and focused on reported errors. Other than that, it was "business as usual," since the existing engineering design process was already based on the insertion of preliminary data that were best-guess values, random temporary values, or blanks. Thus, these three design principles only reflected existing engineering design practices.

However, the fourth principle, phase-based reporting, proved to be a challenge. Weekly focus session discussions, participant observation, and semi-structured interviews revealed that the engineers at EUMEC appreciated the intentions of the report and used it as a working list for correcting errors. The report enabled them to focus directly on the errors instead of manually trying to identify them first. However, the implementation also revealed unanticipated consequences. First, despite this principle which was intended to reduce the number of reported errors, engineers complained that the number was still unmanageably large. As a result they found it difficult to prioritize which errors to correct first. The sheer number of errors was demotivating. Second, the system identified some data values as incorrect when they were actually correct. These "false positives" reduced trust in IQS. Both these unanticipated consequences demotivated the engineers from using the system.

4.3.5 Problem Re-Definition

The unanticipated consequences of the first BIE-cycle led me to take another look at the problem. The large number of false positives in the report initiated a thorough review of the rule definitions. I was able to weed out some of them by identifying inaccuracies in the rule definitions. Even after that, I was still left with the majority of the false positives. After several discussions with the managers and engineers of each engineering discipline, where we dissected each occurrence of a false positive, I realized that most of these errors actually represented exceptions to the requirements. The discipline managers had agreed upon these exceptions with the customer.

The identification of exceptions revealed another problem: it was difficult to determine the common denominators of these exceptions that could be captured in a manageable set of rules and thus exclude them from the report. It was of course possible to define a rule for every exception, but the huge number of errors made this a staggering task. Besides, new exceptions would almost certainly arise in the future. We had to find a more manageable way to deal with existing and future exceptions.

The other unanticipated consequence was the unmanageably large number of true errors that IQS was reporting. That led us to think of better ways to group errors for reporting. The goal was to find a solution that would both decrease the amount of errors to manage at the time and also help engineers with prioritizing errors. This redefinition of the problem led to a second cycle of BIE.

4.3.6 BIE Cycle 2: Designing for Error Handling

Most of the false positive errors represented exceptions to the requirements. The solution for handling false positives was simply to remove those errors from the report and "park" them in a separate table which was called the "parking table". The rationale was this: these errors were detected by the project-specific rules whose definition was based on a formalized set of requirements. To handle exceptions required departing from this formalized set, which was a serious action. Hence, to keep track of all these exceptions, they had to be first reported as errors and then handled later – the parking table achieved this. The table contained the errors, the causes of the errors, and the signature of an authorized person who was either the concerned discipline manager or someone delegated by the manager. At the hand-over phase, this information could then be attached as a supplement to the deliverables. This report helped explain why the data deviated from the original set of specifications. This solution was formalized in a design principle called the "parking principle".

Handling the large number of true errors was solved by dividing the original phases into even smaller units of reporting. After searching for a unit of optimal size, we settled on "commissioning packages", a well-known term in engineering projects. A commissioning package is "a practical scope of work unit within a system or subsystem for commissioning, constituting a functional unit which can be tested by commissioning to confirm its suitability for operation" (NORSOK 1999, p. 3). The documentation for all systems belonging to a certain functional unit possesses the same delivery date. A project has several such commissioning packages. By grouping together all data belonging to a commissioning package, the number of reported errors is significantly smaller than the number of errors reported for several units in that project phase. In addition, the delivery dates provide a basis for prioritizing the most urgent packages first. Based on these considerations, the original report phases were divided into several narrower phases based on commissioning packages and their delivery dates. This led to a revision of the phase-based principle by explicitly adding the qualification that the phases should be optimally narrow.

4.3.7 Reflections and Learning from BIE Cycle 2

The engineers were satisfied with both solutions to the challenges that arose from BIE cycle 1: the parking principle and narrowing the reporting unit to commissioning packages.

Commenting on the parking of errors solution, an electrical engineer said: "It is nice to get rid of these false errors in a quick way. If we use some time to identify the exceptions in the first place, we will have more time to focus on correcting real errors."

A typical feedback on the solution to report errors by commissioning packages was: "Now it is possible to prioritize which errors to correct first. The small amounts of errors per package also result in higher motivation for correcting them simply because it feels possible to actually be able to correct everything." (mechanical engineer)

The evaluation and reflection led to the revision of one of the design principles derived from BIE cycle 1 and the addition of a new design principle. Table 7 lists the revised set of design principles. The changes from the previous set are shown in italics.

Design Principle	Description
Allow for inconsistency	Inconsistencies caused by straw values or blanks should be
	allowed at appropriate early phases.
Allow for incompleteness	Incompleteness caused by straw values or blank, should be
	allowed at appropriate early phases.
Allow for lack of logical	Lack of logical coherence caused by straw values or blanks
coherence	should be allowed at appropriate early phases.
Phase-based reporting	Preliminary incorrect values should not affect the daily report until the appropriate project phase is reached. <i>The</i> <i>phases should be narrow enough to produce a manageable</i>
Parking of errors	number of errors and provide means for prioritizing. "False positives" occurring as a result of (legitimate)
1 urking of errors	deviation from original (and still applicable) requirements,
	should be removed from the error report and saved with an
	explanation for deviation to be handled later.

Table 7. Revised set of design principles (adapted from Paper 3)

4.3.8 Formalization of Learning

The learning from an ADR project should be further developed into general solution concepts for a class of field problems, in this case DQ/IQ management in construction engineering projects. The theoretical concepts that emerged from the design and development of IQS were the five design principles. These are further discussed in Chapter 6.

4.3.9 Summary

The projected system design departed substantially from the existing DQ/IQ assessment tool. The BIE form is best characterized as IT-dominant. Figure 8 captures the two BIE cycles, each of which addressed specific design challenges. The right-hand side of the figure summarizes the different contributions of the IQS project.

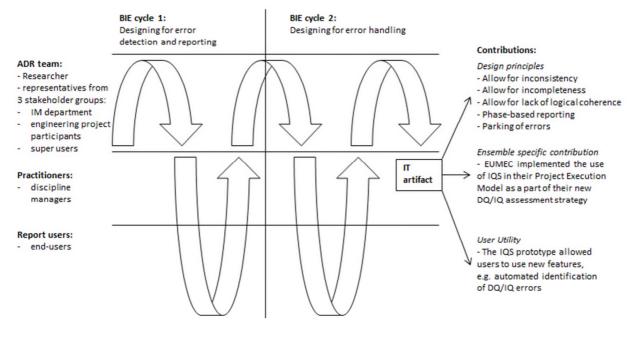


Figure 8. BIE cycles and contributions (from Paper 3)

Data collection activities can be found in Appendix A. They are presented in a table structured around the stages of ADR. IQS is discussed in more detail in Appendix B, including a reference to Appendix C which provides an example of rule syntax.

The next section discusses the use of the guiding frameworks.

4.4 Use of Guiding Frameworks

This section discusses how the selected frameworks, Total Information Quality Management (TIQM) and Data Quality Assessment (DQA), were used to guide the actions performed during the IQS design project.

4.4.1 TIQM and IQS

Table 8 shows how IQS relates to the premises of TIQM processes.

Process	Related to IQS	
(P1) Assess Data Definition and Information Architecture Quality	Defined which data sources were needed for assessment. Defined an Information Architecture that could handle needed data types and store data in a way that made them retrievable for comparison. Defined assessment outcome reports.	
(P2) Assess Information Quality	Defined project-specific assessment rules based on project requirements. Defined a set of default rules that can be used as a basis for any construction engineering project.	
(P3) Measure Nonquality Information Costs and Risks	Performed a Delphi study to identify and rank the problems having the most negative impact on the profit margins. Several DQ/IQ problems were identified, including the top six problems.	
(P4) Re-engineer and Correct Data	Corrections needed to be performed manually due to needed expert determination of correct data values. Daily error reports were presented to the engineers.	
(P5) Improve Information Process Quality	Established a methodology for evaluating and refining IQS, including the necessary alignments of various processes and weekly meetings for each project. The activities related to IQS were included in the organization's Project Execution Model (PEM).	
(P6) Establish the Information Quality Environment	Information managers were appointed per project. Support and training were provided. The development was and will continue to be performed in close collaboration with several stakeholders (especially project users and managers).	

Table 8. Mapping IQS to the premises of TIQM processes by English (1999, 2003)

As can be seen from Table 8, IQS met the requirements of TIQM to some extent: processes were defined for how to assess DQ/IQ to meet the requirements of the engineers (e.g. which data to collect and how to present error messages in an understandable way) and how to measure the level of quality. Further, processes for evaluating and refining IQS were established, and the activities related to IQS were included in EUMEC's PEM. Information managers are now appointed per project, and together with the PEM requirements this arrangement established an environment encouraging information quality.

However, some issues were not addressed. The absence of automated correction of errors by IQS is one such issue, due to the nature of engineering data which requires engineering expertise to determine the correct data values. Another important issue was the lack of possibility to measure accuracy by comparing data to a physical object, because some physical assets only existed as referenced logical/functional elements in data registers and data models since those physical assets were not yet produced. A third issue was measuring the costs of poor quality (or nonquality) information. The result of the Delphi study strengthened the argument for why this focus on DQ/IQ was needed. But even if the level of DQ/IQ was increased, this study does not provide any measurements on the relation between increased DQ/IQ and its impact on project performance.

4.4.2 DQA and IQS

DQA proposes the concept of subjective and objective assessment. In the case of IQS, TestP's project requirements were captured in formalized statements used as executable rules. The outcome of these rules formed the basis for task-dependent (objective) assessment. However, the outcome also needed subjective assessment. When the outcome included errors disputed by the engineers, their subjective assessment revealed some errors that were not really errors. Further investigation of these "errors" led to determination of their root cause: these "errors" were exceptions from stated requirements. In alignment with DQA, actions were taken for improving DQ/IQ: the parking –principle was established. Task dependent metrics was developed to measure the state of each rule and to compare outcomes between projects (Figure 9). Task independent metrics were not an issue since it was the contextual characteristics that were the challenge.

DQA also proposes defining relevant quality dimensions, in this case Completeness, Consistency, and Logical Coherence. Even though the authors of DQA suggest different measuring functions for different quality dimensions, this has not been an issue in this thesis. The reason is because no matter which dimension was involved, it was the consequence of each error that was important, i.e. delays and cost overruns. Depending on the possible impact on project performance resulting from each type of error (i.e. each rule), the rules should be weighted: the higher the impact, the higher the weight. Instead, a simple ratio per rule (using percentage) was calculated (Figure 9). That solution makes it possible to compare projects, but without knowing how severely the impact of each rule outcome might affect the project. Table 9 shows how IQS relates to the premises of DQA.

Table 9. Mapping IQS to the premises of subjective and objective data quality assessment proposed by Pipino et al. (2002)

Assessment type	Related to IQS
Subjective data quality assessments reflect the needs and experiences of stakeholders	The needs were the requirements of the individual construction engineering projects. Experienced engineers revealed some "errors" were exceptions from requirements. Such subjective assessments were necessary throughout the project.
Objective assessment can be task- independent or task-dependent. Task- dependent metrics are developed in specific application contexts	Task-dependent assessment was in focus since contextual characteristics were the challenges. Task- dependent assessment included the ISs used in construction engineering projects within EUMEC, interfaces toward external ISs (for example, the customer ISs), business rules such as use of the organization's PEM, and projects requirements.

Finally, following the suggestion of Pipino et al. (2002) to create a de facto standard, we determined a default set of assessment rules that could be used in any construction engineering project at EUMEC. Parts of that set were used for the comparison of projects (see Figure 9, next section).

4.5 Formal Evaluation of IQS

The development of IQS was continuously evaluated throughout the ADR process. This evaluation resulted in changes and additions to the design, and in the end the users expressed satisfaction with the system. Still, it is expected that new features will emerge in the future. As seen in the timeline, a management report providing progress information, and comparative assessment over time is already about to emerge.

To further evaluate whether IQS indeed helped improve DQ/IQ, the project (TestP) that used IQS was compared with two other projects (projects A and B) that did not use it. The result is displayed in Figure 9 and the details concerning the comparison were discussed in Paper 5.

The indices of comparison displayed in Figure 9 are as follows:

Pe = the total number of possible errors for each rule (i.e., number of records checked for that rule since each rule can only identify one error per record)

Te = the total number of errors indicated by that rule

Pcte= Percentage of errors (Pe/Te)

SumPe = Sum of Pe for each project

SumTe = Sum of Te for each project

PctSum = SumPe/SumTe

Thus, it was possible to compare the total percentage of errors between each project as well as granulate the result by comparing the percentage of errors per rule per project. Also, the rules were grouped into the three dimensions of quality: completeness, consistency, and logical coherence. Thus, it was possible to test whether IQS helped improve these dimensions in addition to overall quality.

			TestP		Pr	Project A		Å	Project B	
			# of			# of			# of	
		# of	records		# of	records		# of	records	
Rules	Dimension	records	with	% of	records	with	% of	records	with	% of
		checked	error	errors	checked	error	errors	checked	error	errors
		(Pe)	(Te)	(Pcte)	(Pe)	(Te)	(Pcte)	(Pe)	(Te)	(Pcte)
Rule3: Tags without drawing references	Completeness	7259	73	1,01	> 4468	42	0,94	11837	267	2,26
Rule6: Missing site code	Completeness	4502	75	1,67	3629	94	2,59	8642	1018	11,78
Rule7: Missing area code on tag	Completeness	9041	8	0,09	5146	56	1,09	13702	143	1,04
Rule8: Missing GA and Equipment Layout references										
in MEL	Completeness	1685	482	28,61	585	234		1714	1035	60,39
Rule9: Missing weight	Completeness	4630	899	19,42	3245	1889	58,21	9270	4085	44,07
Rule10: Missing mounted-on	Completeness	3942	108	2,74	2672	508	19,01	7800	1888	24,21
Rule13: Missing EX-class	Completeness	3839	325	8,47	3939	724	18,38	11484	4117	35,85
Rule14: Missing IP-grade	Completeness	3839	445	11,59	3939	723	18,35	11484	3917	34,11
Rule15: Missing or inadequate description on tag	Completeness	4629	39	0.84	3245	155	4,78	9270	513	5,53
Rule16: Discipline code missing	Completeness	8964	0	0,00	5120	5	0,10	12969	15	0,12
Rule18: Missing mounted-on on valve	Completeness	688	10	1,45	573	9	1,05	1470	0	00'0
Rule19: Missing Manufacturer	Completeness	4527	310	0,83	4578	1059	23,13	13077	3927	30,03
Rule20: Missing Model	Completeness	7649	319	4,17	4578	1074	23,46	13077	3987	30,49
Rule17: Missing or illegal PO-number	(Completeness)/consist	4506	128	2,84	3245	207	6,38	9270	1294	13,96
Rule5: Incorrect drawing type in GA- and/or Layout- field	Consistency	2363	28	1,18	2048	80	3,91	6214	1169	18,81
Rule11: Documents with incorrect DocType	Consistency	10869	36	0,33	7449	42	0,56	16385	1414	8,63
Rule1: Area code different from master tag	Logical Coherence	2741	26	0,95	2663	1734	65,11	8520	5626	66,03
Rule2: X, Y, Z outside area limits	Logical Coherence	669	79	11,30	5 432	14	3,24	1209	323	26,72
Rule4: System code different from master tag	Logical Coherence	2848	17	0,60	980	17	7,86	3613	491	13,59
Rule12: Drawing referenced by tag, but not defined										
by Document Control	Logical Coherence	85802	395	0,46	39080	8736	22,35	72971	1826	2,50
TOTAL		175022	3802	2,17	101614	17459	17,18	243978	37055	15,19
	TOTAL PER DIMENSION	NOIS								
	Completeness	65194	3093	4,74	45717	6569	14.37	125796	24912	19.80
	Consistency	17738	192	1,08	12742	329	2,58	31869	3877	12,17
	Logical Coherence	92090	517	0,56	43155	43155 10561	24,47	86313	8266	9,58

Figure 9. Comparison of projects (from Paper 5)

The findings showed that TestP's level of DQ/IQ was significantly better overall than that of the two other projects. While TestP had 2.17% errors identified in total, the numbers for projects A and B were 17.18% and 15.19% respectively (see percentage indications in the TOTAL row in Figure 9). The superiority of TestP extended to the quality dimensions, i.e., level of completeness, consistency, and logical coherence (see the TOTAL PER DIMENSION section in Figure 9).

For every rule, with three exceptions, TestP had the lowest percentage of errors. The three rules are circled in Figure 9. Since TestP was not yet completed – while the other two projects were completed – the number of errors identified by these three rules could decrease towards the delivery date. Even if this was not the case, the number would certainly not increase since all items required were registered in the engineering databases; hence, all missing or partial data values related to the item records were already identified by IQS. Therefore, it was possible to conclude that TestP was overall at a much better state in terms of DQ/IQ than the two other projects because the rules had helped identify what needed to be addressed in the various phases of the project. Fewer errors in data sources meant fewer errors in the drawings and specifications that would later be handed over to assembly, which would further implicate fewer delays and cost overruns.

The next part, Part C Contribution, includes a presentation of each research publication in Chapter 5. The contributions of this thesis are discussed in Chapter 6, and the conclusions are drawn in Chapter 7.

Part C: Contribution

"Truth is what stands the test of experience"

- Albert Einstein

5 Research Publications

This thesis comprises five papers. Table 10 displays a list of the papers. The full text of each publication is available in Appendix D.

No	Paper	Publication outlet
1	Westin, S. and Päivärinta, T. (2011). Information Quality in Large Engineering and Construction Projects: A Delphi Case Study	Proceedings of the 19 th European Conference on Information Systems (ECIS), Helsinki, Finland. (http://aisel.aisnet.org/ecis2011/273)
2	Westin, S. (forthcoming). Data and Information Quality Research in Engineering Construction Projects: A Review of Literature	Proceedings of the 21 st International Conference of Information Systems Development (ISD), Prato, Italy, 2012 (http://www.springer.com/computer/information+syst ems+and+applications/book/978-1-4614-7539-2)
3	Westin, S. and Sein, M.K. (submitted 2013). The Design and Emergence of a Data/Information Quality System	Under review
4	Westin, S., Holen, O.B. and Päivärinta, T. (2012). Data and Information Quality Dimensions in Engineering Construction Projects	Proceedings of the 18 th Americas Conference on Information Systems (AMCIS), Seattle, Washington, USA. (http://aisel.aisnet.org/amcis2012/proceedings/DataIn foQuality/1)
5	Westin, S. and Sein, M.K. (2013). Improving Data Quality in Construction Engineering Projects: An Action Design Research Approach	Journal of Management in Engineering (http://ascelibrary.org/doi/abs/10.1061/(ASCE)ME.1 943-5479.0000202)

The numbers displayed in the first column of Table 10 are used in this thesis when referencing the papers (Paper 1, Paper 2, etc.).

For each paper, I present a summary that includes the research focus and a subsequent description of the findings.

5.1 Paper 1: A Delphi Study

Westin, S. and Päivärinta, T. (2011). Information Quality in Large Engineering and Construction Projects: A Delphi Case Study

The main problems experienced at EUMEC at the outset of this research were delays and cost overruns related to their construction engineering projects. In order to deal with these problems, it was necessary to identify why they occurred. Therefore, the aim of Paper 1 was to identify and rank the problems with the most negative impact on the profit margins of EUMEC's construction engineering projects. A Delphi study was chosen as the appropriate approach for the purpose. The Delphi study process leaned on the phases and recommendations of Schmidt (1997) and Okoli and Pawlowski (2004). After phases of brainstorming, narrowing down, and ranking, the result was a ranked list of 18 problems. The list was decided with strong consensus among the participants.

5.1.1 Findings

Eight problems in the ranked list were related to DQ/IQ, including the top six (see Table 5, Section 4.3.1). All eight DQ/IQ problems referred to various problems occurring in different kinds of drawings or in data sources. The paper concluded that these results contributed to the existing literature on delays and cost overruns in construction engineering by providing more details on DQ/IQ-related issues. These were issues that had been relatively ignored so far in this context. The focus from here turned to DQ/IQ-related challenges in construction engineering.

5.2 Paper 2: A Literature Review

Westin, S. (forthcoming). Data and Information Quality Research in Engineering Construction Projects: A Review of Literature

The aim of Paper 2 was to provide a comprehensive literature review based on a systematic approach, while focusing on DQ/IQ research in construction engineering. The review was performed after it was established that DQ/IQ problems were indeed the most important ones. The purpose of the review was to learn the breadth and content of the research on the topic. The goal was to acquire knowledge that could be used as guidance when acquiring or designing a DQ/IQ assessment tool. The method chosen for the review was based on the eight-step guidelines proposed by Okoli and Schabram (2010) and the concept matrix suggestions by Webster and Watson (2002). The review covered articles published all years up until September 2011. In the end, only nine articles were included in the review, which indicated a need for more research in this area.

5.2.1 Findings

Despite the fact that only a small number of articles were found, it was fairly easy to identify challenges related to the topic of construction engineering. Six challenges were identified: the iterative nature of concurrent engineering; the uniqueness of engineering data; the lack of integration between processes; the lack of integration between systems; the lack of timely information, and the lack of relevant DQ/IQ assessment frameworks and tools. The paper established relations between these challenges and argued that the first five challenges would have to be viewed as unavoidable. Hence, relevant DQ/IQ assessment frameworks and tools could only be achieved by considering these unavoidable challenges and suggesting guidelines for how to mitigate their consequences.

5.3 Paper 3: The Design and Emergence of IQS

Westin, S. and Sein, M. K. (under review). The Design and Emergence of a Data/Information Quality System

Paper 3 aimed at describing the design, emergence, and implementation of a DQ/IQ assessment tool, IQS. The system development was performed through a collaborative effort between researchers and practitioners at EUMEC, using a live construction engineering project as the environment for the development. Presented as an ADR project, Paper 3 provided detailed insight on how the project progressed through an emergent process.

5.3.1 Findings

Five design principles for DQ/IQ assessment tools for use in construction engineering were defined: allow for inconsistency, allow for incompleteness, allow for lack of logical coherence, phase-based reporting, and parking of errors. The principles will mitigate the problem of not knowing the correct data values to insert at the time of insertion to a record. They will also help in handling the inevitable huge number of identified errors. In addition, this study also contributed to the method space by feeding formalized outcomes in ADR Stage 4 back to Stage 1, informing the problem formulation for the next iteration of BIE. Hence, the paper suggested considering whether a feedback path from Stage 4 to Stage 1 should be added to the ADR process.

5.4 Paper 4: Identifying Quality Dimensions

Westin, S., Holen, O. B. and Päivärinta, T. (2012). Data and Information Quality Dimensions in Engineering Construction Projects

"Quality dimensions" is a fundamental concept in DQ/IQ research. The aim of paper 4 was to identify needed and sufficient quality dimensions in construction engineering projects at EUMEC.

For each construction engineering project using IQS, a set of assessment rules had been defined by me and other information managers. Each set of rules had been adjusted to the individual project's requirements. These rule sets were used to analyze which quality dimensions were addressed by them. For the purpose of analysis, the 11 most important dimensions as identified by Ge et al. (2011), and in addition the "logical coherence" dimension, were used as a basis for categorization and discussion. Altogether, 246 DQ/IQ assessment rules used by 12 EUMEC projects were investigated to determine which quality dimensions were assessed by rules. In addition, it was investigated whether other dimensions of those identified by Ge et al. were somehow addressed in EUMEC by other means than assessment rules.

5.4.1 Findings

A total of seven quality dimensions were identified: accessibility, security, relevance, completeness, consistency, timeliness, and logical coherence. However, only three dimensions -- completeness, consistency, and logical coherence -- were directly handled by rules. The dimensions in use but not handled by rules were: "accessibility," which was indirectly handled through dimensions; "security," which was handled by organizational policy; "relevancy," which was indirectly handled by the rules since they were defined on the basis of project requirements; and "timeliness," which was handled by delivery dates captured by the rules but not assessed by rules. The paper concluded that the findings contributed to research on DQ/IQ assessment in construction engineering by highlighting the importance of the three dimensions and the potential of controlling these dimensions better in the projects.

5.5 Paper 5: Evaluating IQS

Westin, S. and Sein, M. K. (2013). Improving Data Quality in Construction Engineering Projects: An Action Design Research Approach

In accordance with ADR requirements, IQS was continuously evaluated by users and project participants during the stages of BIE (Building, Intervention, and Evaluation), and Reflection and Learning. Paper 5 aimed at a more formal evaluation. IQS was evaluated

by comparing the level of DQ/IQ in TestP with the level of DQ/IQ in two projects not using IQS. A common set of carefully selected rules was used to assess all three projects to make it possible to compare the outcomes. The comparison was performed on several levels: per rule, per quality dimension, and in total.

5.5.1 Findings

The result of the comparison was encouraging: the level of DQ/IQ in TestP was significantly higher overall than in the two other projects. This implied that a tool based on the design principles on which IQS was built is likely to help improve DQ/IQ in data sources, and hence improve DQ/IQ in drawings.

5.5 Overall Summary of the Papers

The five papers represent the overall story told in this thesis. Each of the papers contributed to this thesis in different ways. Paper 1 gave the argument for narrowing the focus from delays and cost overruns to DQ/IQ related problems. Paper 2 identified DQ/IQ challenges in construction engineering and strengthened the case for needing new assessment frameworks and tools. Paper 3 reported the development and emergence of IQS, and Paper 4 reported the quality dimensions used in EUMEC's projects. Finally, Paper 5 reported a separate formal evaluation of IQS. Table 11 summarizes these papers' contributions to the thesis.

Paper	
no.	Contributions to thesis
1	This paper provided an empirical basis for the focus of this thesis. Its findings were used to strengthen the argument for focusing on DQ/IQ when the problem was delays and cost overruns (Chapter 1).
2	This paper provided part of the theoretical basis for this thesis. It argued that certain unavoidable challenges in construction engineering led to a poor level of DQ/IQ. It also found from previous research that existing DQ/IQ assessment frameworks were not sufficient for this context. These insights substantiated the need for new frameworks and tools (Chapter 2).
3	This paper provided empirical insight on the design, development and implementation of the DQ/IQ assessment tool. In addition to the tool, five design principles were defined. These can guide future practitioners in their work, and they provide enhanced knowledge to researchers in this field. Further, this paper provided the basis for understanding the positioning of the research approach discussed in Chapter 3. Finally, this paper provided useful insight to the use of guiding frameworks. This insight was brought into the thesis to discuss related learning (Chapter 4).
4	This paper provided empirical insight on quality dimensions and their meanings in construction engineering. Three dimensions were identified as needed and sufficient for rule based assessment in EUMEC's construction engineering projects. In addition, four dimensions were addressed by other means such as through organizational policy or indirectly through other dimensions (Chapters 2, 3, and 6).
5	This paper provided a formal evaluation of the tool developed. By this, it brings into the thesis a <i>successful</i> story of a DQ/IQ assessment tool designed for construction engineering (Chapter 4).

Table 11. Contributions of the papers to this thesis

In Chapter 6, I discuss these contributions and show how knowledge is further enhanced when viewing the findings in conjunction with each other.

6. Contributions

In this chapter I discuss the contributions of this thesis. Theoretical and practical contributions are presented separately although they overlap to some extent.

6.1 Theoretical Contributions

The theoretical contributions include those specifically aimed at increasing knowledge in the field of DQ/IQ research in construction engineering. Table 12 summarizes these contributions, and in the rest of this section they are further discussed.

Theoretical contribution	Description
Defining a set of design principles for DQ/IQ assessment tools for use in construction engineering projects	The design principles are: allow for inconsistency (DP-1), allow for incompleteness (DP-2), allow for lack of logical coherence (DP-3), phase-based reporting (DP-4), and parking of errors (DP-5).
Identifying quality dimensions needed and sufficient for rule- based assessment in construction engineering projects	The identified quality dimensions are: completeness, consistency, and logical coherence. Even though more dimensions were covered at EUMEC, these were the only dimensions in need for rule-based assessment.
Revealing the relationship between unavoidable challenges and unavoidable consequences in construction engineering	The identified unavoidable challenges include: the iterative nature of concurrent engineering, the uniqueness of engineering data, lack of integration between processes, and lack of integration between systems. The unavoidable consequences include: lack of timely information, not knowing the correct data, and proceeding with partial information.

6.1.1 Design Principles

The main contributions to knowledge are the five design principles that emerged during the ADR project. They are presented in three different tables because they aim to mitigate three different needs: 1) the engineers need to be able to proceed with partial information even if the result is a poor level of DQ/IQ in the data sources in terms of inconsistencies, incompleteness, and lack of logical coherence (Table 13); 2) this will result in a large number of errors that need to be managed (Table 14); and 3) "false positives" - in the

sense that some errors correctly identified as errors are not really errors but legal exceptions from project requirements – also need to be managed (Table 15).

Table 13 displays design principles for proceeding with partial information.

Design Principle	Explanation
<u>DP-1</u> . Allow for inconsistency	Description: Inconsistencies caused by straw values or blanks should be allowed at appropriate early phases. Supplementary comment: Data values that are required to be identical could differ at certain early phases because the correct value may not yet be known. However, the engineers need to insert straw-values in order to proceed.
<u>DP-2</u> . Allow for incompleteness	Description: Incompleteness caused by straw values or blanks should be allowed at appropriate early phases. Supplementary comment: Required data values may be missing at certain early phases because the correct value may not yet be known. However, the engineers need to insert straw-values in order to proceed.
<u>DP-3.</u> Allow for lack of logical coherence	Description: Lack of logical coherence caused by straw values or blanks, should be allowed at appropriate early phases. Supplementary comment: Data values completely or partially depending on other data values may not reflect logical coherence between them at certain early phases because one or more of the data values may not yet be known. However, the engineers need to insert straw- values in order to proceed.

Table 13. Design principles for proceeding with partial information

As discussed previously, the correct data to insert in a data record may not yet be known at the early stages of a construction engineering project. Still, the engineers have to proceed to keep up with the project schedule. The consequence is proceeding with only partial information and/or straw values. At this point, inconsistencies, incompleteness, and lack of logical coherence in data sources would be the unavoidable result. The design principles displayed in Table 13 were hence necessary to keep the project flowing. They also contribute to knowledge since these issues have not been addressed previously in the literature. However, as a consequence, a huge number of errors will be identified. For performing manual correction on a huge number of errors, the mere number of reported errors needs somehow to be reduced to a manageable amount. In addition, the errors need to be prioritized so the engineers can correct the more important problems first.

Table 14 displays the design principle that provides a manageable number of reported errors.

Design Principle	Explanation
<u>DP-4.</u> Phase-based reporting	Description: Preliminary incorrect values should not affect the daily report until the appropriate project phase is reached. The phases should be narrow enough to produce a manageable number of errors and provide means for prioritizing. Supplementary comment: Reports are developed for displaying identified errors. However, the number of reported errors has to be manageable for manual corrections. The appropriate project phase is determined in collaboration with the engineers.

Table 14. Design principle providing a manageable number of reported errors

Keeping the number of reported errors at a manageable level is of vital interest for keeping up the motivation for performing corrections. Design principle 4 helps in managing the huge number of errors identified by displaying them in smaller sets based on some common denominator, in this case PEM phases and commissioning packages. This necessity has not been reflected in previous frameworks; hence design principle 4 is a contribution to knowledge. For prioritizing errors, delivery dates were captured by the rules and displayed in the reports, which made it easier for the engineers to choose which errors to correct first.

Table 15 displays the design principle that handles exceptions from requirements.

Table 15. Desig	gn principle	handling exceptions
-----------------	--------------	---------------------

Design Principle	Explanation
<u>DP-5.</u> Parking of errors	Description: "False positives," occurring as a result of (legitimate) deviation from original (and still applicable) requirements, should be removed from the error report and saved together with an explanation for deviation to be handled later. Supplementary comment: An issue adding to an unmanageable number of identified errors was these "false positives". These records can be moved to a "Parking" table. As long as the record exists in the parking table the record will be omitted from the report. This results in two advantages: 1) the number of errors reported will decrease, and 2) the contents of the parking table can be used as documentation later in the project.

Since exceptions from requirements are common in construction engineering, it is crucial to keep the project flowing even under such conditions. Errors caused by these exceptions need to be first detected, then considered by an engineer who will officially approve the exception, and finally moved away from the report together with the approval comments. The error records that are moved to the parking table will decrease the number of errors reported (since they are not reported again). At the same time it will be possible to view all exceptions and if necessary re-assess them. The need for such a feature has not been reported in any existing framework; hence this design principle is a contribution to knowledge.

The rationale for defining these principles was because no existing tool or framework specifically accommodated construction engineering. Nor were any of these principles mentioned in any previous literature. In that sense, these principles are similar in nature with those presented by Lempinen (2012) for designing performance dashboards. His design principles also aimed at mitigating a lack of proper tools for his purpose: a tool that would equally consider the measurement design methods and the information systems development methods. In my study, the design principles will mitigate the consequences of not knowing the correct data in construction engineering. Existing assessment frameworks have to consider not knowing the correct data if they are to be usable in construction engineering. My design principles add useful guidance in this respect.

6.1.2 Quality Dimensions

Three quality dimensions were identified as needed and sufficient for rule-based assessment in construction engineering projects at EUMEC: Completeness, Consistency, and Logical Coherence. Based on the size of TestP (the construction engineering project during which IQS was developed) and the additional projects in which IQS was implemented, it seems this finding would be applicable for the general class of DQ/IQ assessment tools in construction engineering. However, other dimensions were in use as well at EUMEC, but assessment rules were not defined specifically for those. In the further discussion, I address all dimensions used.

First, I discuss dimensions in use but not assessed by rules; Accessibility, Security, Relevance, Timeliness, and Accuracy.

Accessibility at EUMEC was understood as the availability of needed information. This understanding was in line with Ge et al. (2011). However, availability was not covered by any rules; rather, unavailable information resulted in errors identified by other rules. If the needed information was unavailable, the consequence would be that the data records contained straw values, or blanks. Hence, this would be detected by rules belonging to one of the three dimensions identified as sufficient.

Security was maintained by providing access to IQS reports only to project participants. This was in line with the understanding of this dimension as presented in Lee, Strong, Kahn and Wang (2002, p. 135), where they state that "the system must be accessible but secure." This dimension was hence not covered by any rule implemented in IQS; rather it was covered by organizational policies.

Relevance was understood as relevant information. This understanding was in line with Ge et al. (2011). Whether the information was perceived as relevant or not at EUMEC was based on project requirements. An assessment of these requirements was performed prior to selecting or defining the rules for use in each project. This meant no rules were developed to check for relevancy; rather data relevance was determined based on the prior assessment of requirements and the following choice of rules.

Timeliness was understood as whether the data were delivered on time to the customer or to internal receivers. "Delivered on time" was also a common understanding for this dimension (Ge et al. 2011). At EUMEC this dimension was covered by delivery dates set for each applicable record. The delivery dates were captured by the rules, displayed in the reports, and used for prioritizing errors. However, a timeliness dimension was not established; rather these dates were used across dimensions.

Accuracy was understood as a combination of completeness, consistency, and logical coherence. This understanding is also pointed out by Ge et al.: "users presume that inaccurate information consists of incomplete and inconsistent information."

Second, I discuss the dimensions assessed by the rules; Completeness, Consistency, and Logical Coherence.

Completeness was understood as all data values identified as mandatory from assessing project requirements were to be inserted in the records. This understanding was in line with Ge et al. (2011). Rules capturing the project requirements related to completeness were defined.

Consistency was understood as whether data values representing the same issue are identical across different tables or different data sources. This conception was in line with the common understanding of the consistency dimension (Pipino et al. 2002). Rules assessing consistency issues were defined.

Logical Coherence was understood as two or more values should not conflict with each other. This was in line with Singh et al.'s (2009) definition of coherence and similar to Blechinger et al.'s (2010) understanding of a type of inconsistency. This thesis highlights the importance of logical coherence as a separate dimension in construction engineering. Rules were defined to assess for the lack of logical coherence.

Summary

Abstracted to a general level, it is possible from the discussion to give an overall picture of how quality dimensions should be handled in construction engineering (Figure 10).

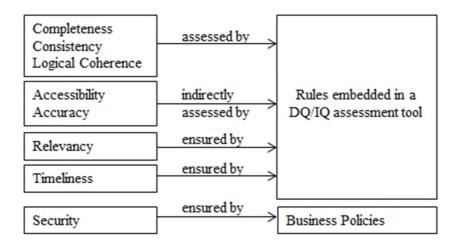


Figure 10. Handling quality dimensions in construction engineering

The left side of Figure 10 displays quality dimensions, and the right side displays how these dimensions were addressed. As shown, completeness, consistency and logical coherence were dimensions directly assessed by rules. This means that rules were defined to assess data values for divergences related to these three dimensions. These three dimensions also covered the dimensions of accessibility and accuracy as understood at EUMEC; hence they were indirectly assessed by rules addressing the first three dimensions. Further, two dimensions were ensured by the rules: relevance, because the rules reflected project requirements; and timeliness, because delivery dates were captured by the rules and could be used to prioritize errors. Finally, business policies ensured access to project data only by project participants. Hence, the security dimension was covered by these policies.

I believe the above discussion illustrates that even if Neely and Cook (2008) concluded that quality dimensions were an over-researched area in DQ/IQ literature, it is still necessary to investigate dimensions, their meaning and their use when approaching a new context.

6.1.3 DQ/IQ Assessment Challenges and Their Consequences

Challenges identified from the literature on construction engineering were discussed in section 2.3. These challenges led to a poor level of DQ/IQ in data sources and drawings (Figure 2, section 2.3). Further, it was shown that these challenges really consisted of four unavoidable challenges resulting in three unavoidable consequences. The latter included not knowing the correct data. While previous literature simply pointed out that the correct data were not known by the engineers (Neely et al. 2006), my study showed that this was an unavoidable consequence caused by four unavoidable challenges.

These conclusions also imply some interesting consequences for DQ/IQ assessment. An error detecting tool would first have to identify errors in data sources and then report them. However, due to the not-knowing-the-correct-data challenge, the data cannot be complete, or accurate, or logically coherent at the start of the project. In order to keep the projects on schedule, straw values have to be used when the correct data is unknown. This is done knowingly, and if managed properly will not harm the project's progress. IQS manages the consequences of using straw values. IQS also speeds up the process of error detection and brings consistency and uniformity to the process by using rules that automatically detect errors. Figure 11 illustrates how the results of unavoidable challenges and consequences were mitigated by the use of IQS.

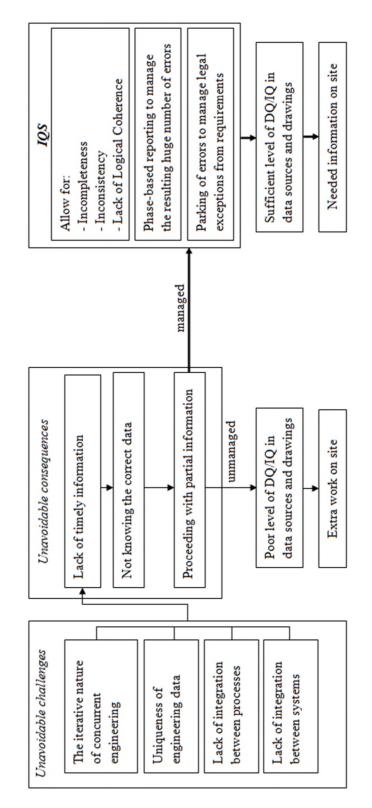


Figure 11. Challenges and how they are mitigated by IQS

The left side of Figure 11 shows the unavoidable challenges. These challenges led to the consequences displayed in the *unavoidable consequences* box in the middle of the figure. This middle alternative will lead to a poor level of DQ/IQ in data sources and drawings, unless managed. Poor DQ/IQ will lead to extra work on site, which will cause delays in the project. The right side illustrates a version managed by the use of IQS. The design principles embedded in IQS provide means for managing the consequences, and hence the level of DQ/IQ will definitely increase, hopefully to a sufficient level. This again will provide on-site workers with the information they need, which will prevent delays caused by a poor level of DQ/IQ.

The challenges faced by the engineers are unique compared to the challenges in other contexts. All that is reported as error is not error. Some are "straw" values because no one knows the correct data at that time. Others are not really errors, but legal exceptions from requirements.

IQS thus represents a shift in the paradigm of how DQ/IQ errors are handled in construction engineering. IQS does not improve DQ/IQ by correcting errors. It simply detects errors and reports them in a manageable way. The actual correction of the errors is still the responsibility of the users – engineers in this case – and has to be performed manually.

What underlie this error handling paradigm are the five design principles that emerged from the development of IQS. The first three principles (the three allows) keep the project on schedule by allowing shortcuts and only reporting identified errors where the correct data values should have been known at that point in time. The next principle, phase-based reporting, sets the correct point in time and reduces the number of errors reported so that engineers can prioritize the errors that need to be corrected. The last principle, parking of errors, keeps the project on schedule by handling exceptions to requirements that would be detected as errors. IQS recognizes this and does not report these errors further. Hence, the project continues to flow without impediment.

6.2 Practical Contributions

The practical contributions include those specifically aimed at informing practice. Table 16 summarizes these contributions, and in the rest of this section I discuss them further.

Practical contribution	Description
Designing an assessment tool based on the five defined design principles	The DQ/IQ assessment tool was designed specifically for use in construction engineering projects. The tool was named Information Quality System (IQS) and was implemented in 16 projects altogether at EUMEC during this research.
Identifying and ranking problems causing delays and cost overruns	The ranked list provided a common understanding of which problems to focus on when aiming at decreasing delays and cost overruns, and it offered more detailed information on DQ/IQ problems to practitioners than previous research did.

Table 16. Summary of the practical contributions

6.2.1 Assessment Tool

The main contribution to practice is the development and implementation of IQS. EUMEC's existing DQ/IQ tool was not sufficient to manage DQ/IQ to the extent needed. The result was delays and cost overruns in their construction engineering projects. IQS provided EUMEC with a fully functional DQ/IQ assessment tool that was better than the old one in several ways. A comparison of the two tools is shown in Table 17.

IQS	Old tool
The assessment rules captured all project	Assessed only for missing values
requirements possible to formalize in	
statements	
A de facto set of standard rules that can be	Did not include a standard set of rules
used in all projects was developed, in	
addition to the project specific rules	
Data from several systems were assessed	Data from only one system could be
	assessed
Errors were detected automatically	Could only detect some types of errors
	(missing values), other errors had to be
	detected manually
Error reports were automatically updated	The tool only ran on command from the
every 24 hours	engineers
Detected errors were reported in a way that	Provided no means for prioritizing
made it possible to prioritize them	

Table 17. IQS compared with EUMEC's old tool

In addition, a good environment for DQ/IQ thinking (English 1999) was created in EUMEC by implementing processes related to DQ/IQ assessment in PEM and making it possible to employ IMs in all projects where needed.

Finally, IQS was evaluated beyond the opinions of the users; by comparing the level of DQ/IQ in a project (TestP) using the tool with two projects not using the tool, it was found that TestP's level of DQ/IQ was significantly better than the other two projects. This finding adds to existing literature on this topic by providing proof-of-concept (Gregor and Jones 2007) of the design principles embedded in IQS.

6.2.2 Identified and Ranked Problems

This thesis contributes to EUMEC by identifying and ranking 18 problems causing delays and cost overruns. Eight of the problems were related to DQ/IQ in data sources and/or drawings, including the top six (see Table 5, section 4.3.1). In Table 18 the eight problems are displayed in the first column. The second column displays a comment on the impact of IQS related to that problem, and the third column indicates whether the problem was solved (Y) or not (N).

Problems	Comment	Solved
Delays in distribution of drawings and documents	This is related to the internal communication of documentation. If for some reason it is delayed, the engineers have to either wait, or proceed with only partial information. The errors occurring from this situation are detected by IQS.	Y
Design is based on unfinished or incorrect supplier documentation	Errors occurring in EUMEC's data sources because of this problem are detected by IQS.	Y
Equipment drawings change after engineering design is completed	This is not handled by IQS, but if re-design is performed IQS will detect errors when new requirements are embedded in the rules.	N
Errors and omissions in supplier drawings	This is not handled by IQS, but using IQS makes it easier for EUMEC to know exactly what information to require from the supplier (because the requirements are embedded in the rules).	N
There are great shortcomings in the interface documentation of drawings (not correct information regarding size, weight, tag number)	These shortcomings are detected by IQS.	Y
GA (general arrangements/assembly) drawings are not consistent with the equipment	This is not handled by IQS, but could be solved in succeeding projects if a) as-built drawings confirm the final solution, or b) the GAs are manually checked against the physical equipment before shipping off to assembly.	N
Lacking interface within our organization between engineering, equipment, control systems, flow diagrams (the drawings are not congruent)	IQS provides the possibility to collect data from all needed sources. It is thus possible to detect inconsistencies between data values used for decided types of drawings.	Y
Copy projects always lead to recurring errors that require hours to correct from project to project	If the source project uses IQS, the level of DQ/IQ in the copy project will match the source.	Y

Table 18. DQ/IQ problems in EUMEC and the impact of IQS

In addition to providing EUMEC with a well-founded argument for focusing on DQ/IQ, the results also provided a more detailed explanation of DQ/IQ related problems causing delays and cost overruns than previous research. For example, Dai et al. (2009) refer only to "drawing errors," and Toor and Ogunlana (2008) refer to "errors and omissions in design documents." Neither study refers to any specific types of drawings or errors, whereas Paper 1 identified different types of drawings with different types of errors.

As shown in Table 18, not all problems were solved. Those problems not solved are further discussed in relation to the research limitations (Section 7.2).

7. Conclusion

This thesis explored how to design DQ/IQ assessment tools with a focus on a specific context, namely construction engineering. It has demonstrated how DQ/IQ can be managed in construction engineering. Based on a set of design principles that emerged during the ADR project, the result was realized through the development and implementation of IQS.

In this section, I summarize the findings, highlight the limitations of this study, and propose directions for future research.

7.1 Summary

As stated in the introduction, the research question addressed by this thesis is: *How can DQ/IQ be managed in construction engineering?* To answer this question, I took an action design research approach to design a DQ/IQ assessment tool in a European construction engineering company. Although DQ/IQ assessment frameworks provided great support for the issues to consider for such assessment, they were not entirely sufficient for this context.

This study revealed connections between unavoidable challenges and their consequences in construction engineering, which had some important implications for DQ/IQ assessment. It was established that due to these challenges, engineers had to proceed with partial information in order to keep a project flowing. If this information quality proceeded unmanaged, the result was a poor level of DQ/IQ in data sources and drawings. By developing and implementing a DQ/IQ assessment tool, IQS, I have shown how it is possible to proceed with partial information and still be able to achieve a sufficient level of DQ/IQ in data sources and drawings in the end. The five design principles embedded in IQS were: allow for inconsistency, allow for incompleteness, allow for lack of logical coherence, phase-based reporting, and, parking of errors.

In addition, the quality dimensions needed and sufficient for rule-based DQ/IQ assessment in construction engineering were established: completeness, consistency, and, logical coherence.

7.2 Limitations

As with any empirical study, this thesis has its limitations. They should be carefully considered when looking at this thesis as a whole.

First, this study was motivated mainly to reduce delays and cost overruns in large construction engineering projects. The investigation first led to errors in drawings and data sources and eventually to DQ/IQ. The solution to this problem was to build IQS. The tool was evaluated and found that it indeed improved DQ/IQ. However, this study did not go "the last mile" by assessing whether this improvement in DQ/IQ also improved project performance.

Second, the errors were not weighted. Some errors might lead to more severe consequences than others, especially concerning delays and cost overruns. This issue was not examined since it required a thorough investigation at the assembly site for several projects. Given the timeframe of a construction project and the fact that several projects would be needed, this was simply not doable within the timeframe of a PhD project.

Third, the emerging call for a management report was not responded to within the time frame of this thesis. A management report providing a comparative assessment of the levels of DQ/IQ over time to upper management would have added an important contribution.

Fourth, the ADR project was still ongoing at the time this thesis was written, which means the design principles may not yet have reached their final form. Refinement of the principles may yet occur and new principles may be added. The artifact was also implemented in several other projects in the organization. It is more than likely that when the artifact has been in use for a longer period of time and in several projects, changes and additions to the design principles will emerge.

Fifth, not all DQ/IQ-related problems identified in the Delphi study were solved (Table 18). Three of the eight problems still remain:

1) "Equipment drawings change after engineering design is completed." Equipment drawings may change if the producer of the equipment has made updates to the

physical equipment after their (old) drawing was handed over to EUMEC. The engineering design may then have been based on a previous version of the physical equipment. If engineering design has been completed, the error cannot be detected before the physical equipment arrives at the assembly site.

2) "Errors and omissions in supplier drawings." This is not fully handled by IQS. Detecting *incorrect* data values in supplier drawings is not handled because they are impossible to reveal before the equipment arrives at the assembly site. However, it is possible to detect *missing* values (omissions) in drawings since they lead to missing values in data sources. This is possible because IQS has the requirements of the project embedded in its assessment rules. Using IQS hence makes it easier for EUMEC to know exactly what information to require from the supplier.

3) "GA (General Arrangements/Assembly) drawings are not consistent with the equipment." This concerns GA drawings developed for individual equipment. They commonly display dimensions of the equipment, and not necessarily the equipment location. Inconsistencies between the GA drawing and the equipment itself cannot be detected before the physical equipment arrives at the assembly site.

These three problems are related to the supplier drawings that engineering designers rely on but cannot verify until the physical equipment exists. The challenge common to all of these problems is that the physical equipment may not arrive before the engineering design is completed and assembly has started. A solution would further increase the level of DQ/IQ.

7.3 Future Research

The limitations affect both practice and knowledge. The possibilities and directions they offer for future research are discussed together in the following, since they overlap to a great extent.

First, this thesis provided EUMEC with a DQ/IQ assessment tool – IQS – which resulted in an increased level of DQ/IQ for TestP. However, the question remains as to whether this improvement in DQ/IQ also improved project performance. This would be a fruitful, and essential, extension of projects such as this. Future studies may focus on the impacts of DQ/IQ assessment systems, more specifically those related to the correlation between DQ/IQ on the one hand, and delays and cost overruns on the other. So far, I have gathered the opinions of experts who believe that this correlation would be positive, but it would be useful to prove this empirically. One possible approach would be to measure employee-hours used on the assembly site to see if they decrease when the DQ/IQ problems are resolved before the drawings are handed over to assembly. However, this approach is not a simple task since these types of construction projects are one-of-a-kind and difficult to compare. An alternative would be to measure extra man-hours used on-site specifically for locating missing information or correcting information.

Second, to provide a single aggregated measure of the level of DQ/IQ would require weighting of the various variables (the rules identifying errors in this case) (Pipino et al. 2002). Some errors might lead to more severe consequences than others. In the case of this thesis, the severity of errors would lead to delays and cost overruns to various degrees. To identify which errors result in what degree of delay would require a thorough investigation at the assembly site for several projects. Such an investigation could provide the information needed to weight the rules and could increase the understanding of the relationship between the level of DQ/IQ and cost overruns.

Third, while this thesis provides valuable insights into the processes and challenges of DQ/IQ assessment in construction engineering, there is especially one issue that was not accounted for: management reports. As in this study, management reports communicating the level of DQ/IQ and providing comparative assessment over time have been called for before (Pipino et al. 2002). However, a management report relies on the successful weighting of errors in order to represent a useful report tool.

Fourth, the design principles may not have reached their final form. To further increase knowledge of how these principles might evolve, further research is needed. In addition to perform follow-up research on this matter in EUMEC, it would be of great importance to extend the research to other organizations and by other researchers. The findings from this thesis could be useful to other researchers as a starting point for such extended research.

Fifth, from EUMEC's viewpoint it would be of great interest to find a solution to the three DQ/IQ related problems not solved. If the equipment is shipped off from the supplier to the assembly site, the only place to detect any discrepancies between the equipment and the drawings is at the assembly site. Problems similar to these have been

reported in other studies (Dai et al. 2009; Toor and Ogunlana 2008), but no solutions have been suggested. Hence, this is an area of interest for future research.

Part D: Reflection

"Logic will get you from A to Z; imagination will get you everywhere."

-Albert Einstein

8. Reflections

In this chapter I provide final reflections on the following issues: DQ/IQ assessment frameworks in construction engineering, the emergence of design principles, the ADR method, and, what the practitioners learned from this journey. Finally, I discuss some advantages and disadvantages of being an insider performing research.

8.1 DQ/IQ Assessment Frameworks in Construction Engineering

When starting the development of IQS, the aim was not to produce new general knowledge on DQ/IQ assessment. The goal was to manage DQ/IQ in construction engineering, a context that proved to be a challenge. However, TIQM and DQA both offered guidelines that were used to focus on what was needed in DQ/IQ assessment. Both emphasized the need for a continuous and iterative mode when performing quality assessment. TIQM extended this mode by including a requirement for an information quality environment. Such an environment includes information policies and using information quality as a performance measure. The response at EUMEC was to include the use of IQS in PEM. This reflected both information policies in terms of various requirements, and the possibility to measure performance related to DQ/IQ.

DQA distinguished between task dependent and task independent measures. That provided an opportunity to focus on task dependent measures. In construction engineering, the errors most difficult to identify and correct are mainly errors occurring due to the nature of the context. Context independent errors, such as errors wrongly identified because of incorrect syntax in rules, or incorrect interfacing between IQS and other ISs, were considered by the practitioners to be a matter for IT/IS-people, not for construction engineers.

Neither of the two frameworks considered not reporting particular errors during a given phase until appropriate. Neither did they consider allowing some errors to occur. As shown, IQS did not report every identified error from day one, simply because the number of errors would be too large to handle. The design principles defined on the basis of the IQS development represent a necessary addition to existing frameworks if they are to be used in construction engineering.

8.2 Design Principles: Why Did They Emerge?

Three of the identified design principles, the three "allows," were all caused by not knowing the correct data as a characteristic of construction engineering. As a result, a high number of errors were reported, making it very difficult for the engineers to address all of them. The engineers complained that their motivation for correcting errors hence was decreasing. This challenge was the reason why the two last design principles were defined: phase-based reporting, and parking of errors. Phase-based reporting was based on project phases and commissioning packages and decreased the number of reported errors. In addition, when assessing the errors, it was revealed that some errors, although correctly identified as errors, were not real errors but legal exceptions from requirements. These were then managed by the parking-principle. But why did the mere number of reported errors decrease the engineers' motivation for correcting errors, even if they all knew how important it was to achieve a sufficient level of DQ/IQ?

The situation is illustrated in Figure 12. The amount of generated information (in this case the error report) is depicted as the large circle. The amount of information considered by the engineers (parts of the error report) is depicted by the medium circle. Finally, the number of errors actually corrected by the engineers is depicted as the small circle.

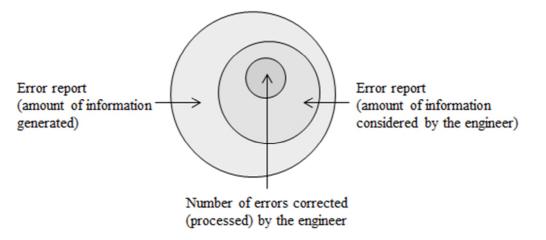


Figure 12. Amount of generated information versus amount of processed information

As shown, the number of errors considered by the engineers is smaller than the number of errors reported. Further, the number of errors corrected is smaller than the number of errors considered. If all errors reported were considered and corrected, all three circles

would be the same size. Why was the number of errors perceived as too high to handle? I turned to the literature and existing theories in an effort to find an answer.

The answer may be "information overload." Information overload is a phenomenon meaning "receiving too much information." Researchers have found that the capability to perform adequate decision making increases as the amount of information received increases. But that is true only up to a certain amount of information received (O'Reilly 1980). After that point, performance will rapidly decline even if further information is provided (Chewning Jr and Harrell 1990). In their review of literature on this topic, Eppler and Mengis (2004) presented lists of both the symptoms and causes of information overload. The symptoms included demotivation and loss of differentiation. Amongst the causes of information overload, the complexity of tasks and tasks interdependencies were found. The engineers expressed demotivation when the error report included too many errors at a time. Each of these errors represents a complex task when it has to be corrected, and each of the correction tasks may depend on information from other engineering disciplines. Based on this complexity, demotivation was a natural consequence. By narrowing the phases by using commissioning packages as the timeframe, the number of errors reported at a time decreased. This reduction to a manageable amount increased the engineers' motivation again. By introducing the possibility of sorting commissioning packages by delivery date, it became possible to differentiate between issues that needed attention as soon as possible and issues that could wait.

Finally, I want to add another issue that came up during these reflections. In their list of symptoms, Eppler and Mengis also found that the "relationship between details and overall perspective is weakened and peripherical cues get overestimated" (p. 333). This symptom of information overload adds to the importance of weighting rules in the future as mentioned in Limitations, section 7.2. As I see it, correcting data errors manually is very time consuming, yet unavoidable. It is hence important to correct the more serious errors first before using time on the less serious errors. Hence, the engineers need a way to decide which errors are more important, and that should be a part of the information they get from the error report.

8.3 Method

ADR studies are expected to generate generalized outcomes of the problem instance and the solution instance, and derive design principles (Sein et al. 2011). I elaborate how developing IQS met these expectations below.

The *class of problems* addressed by this study was DQ/IQ management in construction engineering. This framing answers the calls to IS researchers to extend DQ/IQ research to new contexts (Madnick et al. 2009) and to provide practitioners with tools that can assess the level of DQ/IQ (Pipino et al. 2002). In addition, this study identified DQ/IQ challenges unique to the context of large construction engineering projects.

The *class of solution* addressed by this study was the extant approaches used to solve DQ/IQ problems. These approaches were mainly represented by the frameworks used to guide the development of IQS, namely DQA (Pipino et al. 2002) and TIQM (English 1999, 2003). These approaches were built on the assumption that the correct data value to insert in a record was known by the time of insertion. The paradigm was error detection and correction. The approach of this study was shifting the paradigm to error handling.

The development of IQS resulted in a set of *design principles* that 1) mitigate the problem of not knowing the correct data values to insert by the time of insertion to a record, and 2) handle the reporting of an inevitable huge amount of identified errors.

This study also contributed to the method space. In ADR, Stage 4 is framed as the final stage where outcomes are formalized when the project is completed. In this study, the formalized outcome was fed back to Stage 1 and informed the problem formulation for the next iteration of BIE. For example, initially the succeeding phases of PEM were used as a basis for implementing stricter rules for error checking. This generated one of the design principles: phase-based reporting. When this was implemented in the BIE stage, we realized that the phases were too long: the number of errors was still too large, making it difficult for the engineers to manage and correct the errors. This turned the focus back to the problem formulation stage and the problem was re-formulated as: what is the optimal phase to make the reported number of errors manageable? The answer was to report per commissioning package based on the delivery dates of the packages. ADR was

therefore modified by adding a feedback path from Stage 4 to Stage 1 to capture unanticipated consequences that were vital for new iterations of BIE (Figure 13).

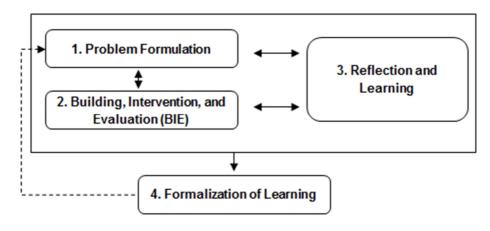


Figure 13. Revised ADR model

The dotted arrow in Figure 13 illustrates the suggested feedback path.

Finally, some reflections on the nature of ADR did occur during the project, and I will now elaborate on them. My experience with ADR has been that it combines AR and DR very convincingly and therefore provides researchers with more opportunities during a single project than before. ADR felt like it was designed to allow for gliding back and forth between the demand for relevance and the need for intervention characterizing AR, on the one hand, and the rigor and little attention to organizational intervention characterizing DR on the other. Whether it is the characteristics of AR or DR that are the most pronounced, will heavily depend on the class of problem under investigation: the more practical the problem, the more on the AR side, the more conceptual, the more on the DR side, and this balance will change during a single project. At any time, researchers can focus on what is needed at that moment and still be within the requirements of ADR. An example of this follows:

In the beginning of a practice-oriented IS design project, not just specifications from practice are needed. The researchers need to understand the context, and this can only be done through a collaborative joint effort between the researchers and the organization. In the beginning phase, the researchers might suggest various solutions, and even facilitate meetings within the organization where divergent viewpoints may be discussed. This intervention is done to increase relevance (the AR side). Then, the researchers may "slide" all the way to DR approaches, rigorously following what Sein et al. (2011) consider "the

separating and sequencing of key steps" (p. 38), which characterize DR. The researchers are not *staying* in that position however. They could bring the material back to somewhere midway between AR and DR, discussing preliminary work with developers from the organization regarding changes, additions, or tunings, before sliding all the way over to AR for testing *together* with the users, or for collecting/bringing out new or changed requirements. This moving back and forth between perspectives can be utilized throughout the design project, and is best illustrated in ADR by the principles of the BIE stage where reciprocal shaping, mutually influential roles, and authentic and concurrent evaluation are the ingredients. In my opinion, the sliding nature of ADR provides the capability for making use of the best of both AR and DR.

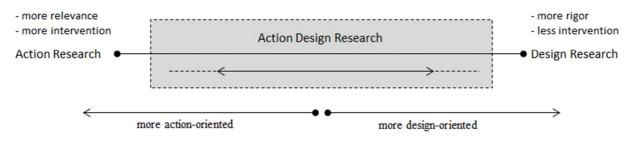


Figure 14. The sliding nature of ADR

Figure 14 illustrates the fluidity of the researcher's position that is explained above. With AR on the one side and DR on the other, it is possible to see how ADR exploits both approaches.

8.4 What Did the Practitioners Learn From the ADR Project?

In addition to the knowledge of theory and technological advances brought to any ADR project by researchers (Sein et al. 2011), the employees at EUMEC brought experiencebased knowledge about work practices and professional knowledge of construction engineering to this project. Their knowledge was used in many ways, such as to define rules, assess work processes, and help me enhance *my* general understanding of construction engineering projects. In addition, they also learned something from this project.

First, on an organizational level, EUMEC implemented procedures, milestones and work processes for DQ/IQ assessment in their PEM (Project Execution Model). This created

the means for establishing an environment for DQ/IQ assessment as suggested by English (1999) in the TIQM framework. Since the PEM represents part of the internal requirements for all projects, all projects were committed to follow PEM. IQS reports (even if by the time of writing this thesis, they still have potential for improvement) provided a possibility to measure progress in the projects, and they were used by all involved engineering disciplines as well as on a project management level.

Second, engaging in critical reflections of their own practices through many meetings over several years also implied learning related to the practitioners' own expertise. I have not collected enough data to support this opinion in a scientific manner, but I still think it is interesting to reflect upon this issue. The most conspicuous learning was the importance of achieving a common understanding of project requirements across engineering disciplines in the early phases of a project. In my mind, a common understanding of the requirements was such an obvious necessity. Hence, I was rather surprised when during TestP I repeatedly had to facilitate meetings for achieving this common understanding even on requirements that did not change. The reason why I think the practitioners learned something from this was because later, when engineers from TestP were transferred to other projects, those engineers requested meetings in early phases of their "new" projects with the aim of achieving a common understanding of the individual project's requirements.

Third, the degree of support needed from information managers was surprisingly high for follow-up with the engineers when they were about to correct errors displayed in the IQS report. I was repeatedly called for, to sit next to an engineer while he or she would go through the reported errors and correct them. Even if I, as an information manager, could not provide any suggestion whatsoever for a correct value, they still liked me to be there. While the degree of support needed declined after several months, it was still high in the end. So what did they learn? They learned that the roles and responsibilities of information managers were needed in any project. Hence, these roles were included in PEM as well, and the projects from then on had to include these roles in their budgets.

8.5 Being an Insider

In a continuum of research membership (participation) ranging from less involved participants to a complete member researcher (Adler and Adler 1987), I am definitely in

the last group. I have been employed at EUMEC for several years, and the problem investigated was also a problem included in my responsibilities as an information manager. Within the context of qualitative research, it has been discussed whether a researcher should be a member of the population he or she is investigating (Corbin Dwyer and Buckle 2009). To me, being a complete member has brought both advantages and disadvantages for my research.

One of the advantages has been that I was familiar with the terminology in the field of construction engineering to a higher degree than I would have been if I were a complete outsider. Construction engineers are highly educated and use a very specific, and complex, terminology. It would have been very time consuming to learn this terminology from scratch. Another time-saving advantage has been that I was already familiar with existing processes and information systems used in EUMEC. Both of these advantages are parts of the preunderstanding that characterizes the complete member researcher (Coghlan 2001). A third advantage was the access I had to every data source and drawing used by any project. This advantage would be hard to gain for an outsider given that much of the data and documentation are confidential. I have not experienced any restrictions from the company besides the wish for anonymity in publications.

On the other hand, I also experienced disadvantages tightly linked to my preunderstanding. First of all, at the early stages of my research, I did not really understand the need for detailed analysis of "everything." I already knew the answer, why bother? I later found the exact same feeling reported by Mathiassen and Sandberg (2012). Anna (Sandberg), a PhD-student working full time at Ericsson, reflected on the need for data collection and detailed analyses: "It was difficult to see the value of time-consuming data collection and detailed analyses to arrive at something I already knew" (p. 9). However, my preunderstanding would soon be questioned.

My understanding of EUMEC's processes had been growing over the years; however, I did not necessarily possess the correct understanding. When interviewing other employees and managers, I found differences in the understanding of processes not only between them and me, but among the employees and managers as well. This forced me to investigate existing processes further, including talking to several of the people performing work related to those processes to see what was really going on. The most disturbing consequence was that this experience led me to think about areas where my

understanding was *not* challenged by anyone's opinion. Did I not notice? Did I interview and speak with the right people? Did I observe the needed actions and did I interpret them correctly? I do not think I know the correct answer to these questions, but they forced me to reflect thoroughly on my own actions, thoughts, and opinions through the remaining parts of this journey. I also made sure my analyses and following opinions were confirmed by other employees.

I started this journey as a practitioner. During the process I became a reflective practitioner. I ended this journey having learned the craft of being a practicing researcher.

References

- Abate, M. L., Diegert, K. V. and Allen, H., (1998) A hierarchical approach to improving data quality, *Data Quality Journal*, (4: 1): 365-369.
- Ackoff, R. L., (1989) From data to wisdom, *Journal of Applied Systems Analysis*, (16): 3-9.
- Adler, P. and Adler, P. (1987). Membership roles in field research (Vol. 6). Newbury Park, CA: Sage.
- Agapiou, A., Flanagan, R., Norman, G. and Notman, D., (1998) The changing role of builders merchants in the construction supply chain, *Construction Management & Economics*, (16: 3): 351-361.
- Alavi, M. and Leidner, D. E., (2001) Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues, *MIS Quarterly*): 107-136.
- Amadi-Echendu, J. E., Willett, R., Brown, K., Hope, T., Lee, J., Mathew, J., ... Yang, B.-S. (2010). What is engineering asset management? *Definitions, Concepts and Scope of Engineering Asset Management* (pp. 3-16): Springer.
- Assaf, S. A. and Al-Hejji, S., (2006) Causes of delay in large construction projects, International Journal of Project Management, (24: 4): 349-357.
- Avison, D. and Fitzgerald, G. (2003). Information systems development: methodologies, techniques and tools: McGraw Hill.
- Ballou, D., Wang, R., Pazer, H. and Tayi, G. K., (1998) Modeling information manufacturing systems to determine information product quality, *Management Science*, (44: 4): 462-484.
- Baskerville, R. L., (1999) Investigating information systems with action research, *Communications of the AIS*, (2: 3es): 4.
- Baskerville, R. L. and Myers, M. D., (2004) Special issue on action research in information systems: making IS research relevant to practice-foreword, *MIS Quarterly*, (28: 3): 329-335.
- Baskerville, R. L. and Wood-Harper, A. T., (1996) A critical perspective on action research as a method for information systems research, [Article], *Journal of Information Technology (Routledge, Ltd.)*, (11: 3): 235-246.
- Batini, C., Cappiello, C., Francalanci, C. and Maurino, A., (2009) Methodologies for data quality assessment and improvement, *ACM Computing Surveys*, (41: 3): 1-52.
- Benbasat, I., Goldstein, D. K. and Mead, M., (1987) The Case Research Strategy in Studies of Information Systems, *MIS Quarterly*, (11: 3): 369-386.
- Bertossi, L., Rizzolo, F. and Jiang, L., (2011) Data quality is context dependent, *Enabling Real-Time Business Intelligence*): 52-67.
- Blake, R. and Mangiameli, P., (2011) The Effects and Interactions of Data Quality and Problem Complexity on Classification, *Journal of Data and Information Quality*, (2: 2): 1-28.

- Blechinger, J., Lauterwald, F. and Lenz, R. (2010). Supporting the Production of High-Quality Data in Concurrent Plant Engineering Using a MetaDataRepository. *Proceedings of the Americas Conference on Information Systems*, August 12 -15, Lima, Peru.
- Braa, K. and Vidgen, R., (1999) Interpretation, intervention, and reduction in the organizational laboratory: a framework for in-context information system research, *Accounting, Management and Information Technologies*, (9: 1): 25-47.
- Braimah, N. and Ndekugri, I., (2009) Consultants' perceptions on construction delay analysis methodologies, *Journal of Construction Engineering and Management*, (135: 12): 1279-1288.
- Chen, J. H. and Chen, W. H., (2012) Contractor Costs of Factoring Account Receivables for a Construction Project, *Journal of Civil Engineering and Management*, (18: 2): 227-234.
- Chen, M., Ebert, D., Hagen, H., Laramee, R. S., Van Liere, R., Ma, K. L., . . . Silver, D., (2009) Data, information, and knowledge in visualization, *Computer Graphics and Applications, IEEE*, (29: 1): 12-19.
- Cherns, A. B. and Bryant, D. T., (1984) Studying the client's role in construction management, *Construction Management and Economics*, (2: 2): 177-184.
- Chewning Jr, E. G. and Harrell, A. M., (1990) The effect of information load on decision makers' cue utilization levels and decision quality in a financial distress decision task, *Accounting, Organizations and Society*, (15: 6): 527-542.
- CIEAM-SI-101-Team. (2005). Data Quality in Asset Management. *Project Report No.1. School of Computer and Information Science*: University of South Australia.
- Coghlan, D., (2001) Insider action research projects: implications for practicing managers, [10.1177/1350507601321004], *Management Learning*, (32): 49-60.
- Corbin Dwyer, S. and Buckle, J. L., (2009) The space between: On being an insideroutsider in qualitative research, *International Journal of Qualitative Methods*, (8: 1): 54-63.
- Dai, J., Goodrum, P. M. and Maloney, W. F., (2009) Construction Craft Workers' Perceptions of the Factors Affecting Their Productivity, *Journal of Construction Engineering & Management*, (135: 3): 217-226.
- de Corbiere, F. (2009). Data Quality and Interorganizational Information Systems: The Role of Electronic Catalogues. *Proceedings of the Americas Conference on Information Systems*, August 6 - 9, San Francisco, California.
- Deming, W. E. (1986). Out of the crisis. Massachusetts Institute of Technology Press, Cambridge, MA.
- Dobson, D. and Martinez, R., (2007) Integrated engineering, *ABB Review*: SPEC. REP.): 48-52.
- Eisenhardt, K. M., (1989) Building Theories from Case Study Research, Academy of Management Review, (14: 4): 532-550.
- English, L. P. (1999). Improving data warehouse and business information quality: methods for reducing costs and increasing profits: John Wiley and Sons, Inc, New York, New York, USA, 1999.

- English, L. P., (2003) Total information quality management : a complete methodology for IQ management, *Dm Review (Now : Information Management)*, (9): 1-7.
- Eppler, M. J. and Mengis, J., (2004) The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines, *The information society*, (20: 5): 325-344.
- Fisher, C., Lauría, E., Chengalur-Smith, S. and Wang, R. (2012). Introduction to information quality: AuthorHouse.
- Frimpong, Y., Oluwoye, J. and Crawford, L., (2003) Causes of delay and cost overruns in construction of groundwater projects in a developing countries; Ghana as a case study, *International Journal of Project Management*, (21: 5): 321-326.
- Ge, M., Helfert, M. and Jannach, D. (2011). Information quality assessment: Validating measurement dimensions and processes. *Proceedings of the European Conference on Information Systems*, June 9 11, Helsinki, Finland.
- Gregor, S. and Jones, D., (2007) The anatomy of a design theory, *Journal of the Association for Information Systems*, (8: 5): 312-335.
- Gärdenfors, P. (2004). Conceptual spaces: The geometry of thought. Cambridge, Massachusetts: MIT press.
- Helfert, M. and Hossain, F. M. Z. (2010). An approach to monitoring data quality product oriented approach. *Proceedings of the Americas Conference on Information Systems*, August 12 -15, Lima, Peru.
- Hevner, A. R., March, S. T., Jinsoo, P. and Ram, S., (2004) Design Science in Information Systems Research, *MIS Quarterly*, (28: 1): 75-105.
- Iversen, J. H., Mathiassen, L. and Nielsen, P. A., (2004) Managing Risk in Software Process Improvement: An Action Research Approach, *MIS Quarterly*, (28: 3): 395-433.
- Koronios, A. and Lin, S. (2007). Information quality in engineering asset management. In L. Al-Hakim (Ed.), *Information quality management: theory and applications* (pp. 221-251): Hershey, PA: Idea Group Publishing.
- Kumar, R., Markeset, T. and Kumar, U., (2006) Implementation and execution of industrial service strategy: A case study from the oil and gas industry, *Journal of Quality in Maintenance Engineering*, (12: 2): 105-117.
- Lee, Y. W., (2004) Crafting rules: Context-reflective data quality problem solving, Journal of Management Information Systems, (20: 3): 93-119.
- Lee, Y. W., Allen, T. and Wang, R. (2001). Information products for remanufacturing: tracing the repair of an aircraft fuel-pump. *Proceedings of the International Conference on Information Quality.*
- Lee, Y. W., Strong, D. M., Kahn, B. K. and Wang, R. Y., (2002) AIMQ: a methodology for information quality assessment, *Information & Management*, (40: 2): 133-146.
- Lempinen, H. (2012). Constructing a Design Framework for Performance Dashboards Nordic Contributions in IS Research (pp. 109-130): Springer.
- Lim, C. S. and Mohamed, M. Z., (2000) An exploratory study into reccuring construction problems, *International Journal of Project Management*, (18: 3): 267-273.

- Lin, S., Gao, J. and Koronios, A., (2006a) Key data quality issues for enterprise asset management in engineering organisations, *International Journal of Electronic Business*, (4: 1): 96-110.
- Lin, S., Gao, J. and Koronios, A. (2006b). The need for a data quality framework in asset management. *Proceedings of the 1st Australasian Workshop on Information Quality (AUSIQ)*, Adelaide.
- Lin, S., Gao, J., Koronios, A. and Chanana, V., (2007) Developing a data quality framework for asset management in engineering organisations, *International Journal of Information Quality*, (1: 1): 100-126.
- Ling, F. Y. Y., Pham, V. M. C. and Hoang, T. P., (2009) Strengths, Weaknesses, Opportunities, and Threats for Architectural, Engineering, and Construction Firms: Case Study of Vietnam, *Journal of Construction Engineering and Management*, (135: 10): 1105-1113.
- Long, N. D., Ogunlana, S., Quang, T. and Lam, K. C., (2004) Large construction projects in developing countries: a case study from Vietnam, *International Journal of Project Management*, (22: 7): 553-561.
- Loshin, D. (2002). Rule-based data quality. *Proceedings of the eleventh international conference on Information and knowledge management*, McLean, VA, USA, November 4-9, 2002.
- Madnick, S. E., Wang, R. Y., Lee, Y. W. and Zhu, H., (2009) Overview and Framework for Data and Information Quality Research, *Journal of Data and Information Quality*, (1: 1): 1-22.
- Markus, M. L., Majchrzak, A. and Gasser, L., (2002) A Design Theory for Systems That Support Emergent Knowledge Processes, *MIS Quarterly*, (26: 3): 179-212.
- Mathiassen, L., (2002) Collaborative practice research, *Information Technology & People*, (15: 4): 321-345.
- Mathiassen, L. and Sandberg, A., (2012) How a professionally qualified doctoral student bridged the practice-research gap: a confessional account of Collaborative Practice Research, *European Journal of Information Systems*, (22: 4): 475-492.
- McKay, J. and Marshall, P., (2001) The dual imperatives of action research, *Information Technology & People*, (14: 1): 46-59.
- Memon, A. H., Abdul Rahman, I. and Abdul Azis, A. A., (2011) Preliminary Study on Causative Factors Leading to Construction Cost Overrun, *International Journal of Sustainable Construction Engineering and Technology*, (2: 1.
- Miller, R. and Lessard, D., (2001) Understanding and managing risks in large engineering projects, *International Journal of Project Management*, (19: 8): 437-443.
- Neely, M. P. (1998). Data Quality Tools for Data Warehousing-A Small Sample Survey. Proceedings of the Third International Conference on Information Quality
- Neely, M. P. and Cook, J. (2008). A Framework for Classification of the Data and Information Quality Literature and Preliminart Results (1996-2007). *Proceedings* of the Americas Conference on Information Systems (AMCIS).

- Neely, M. P., Lin, S., Gao, J. and Koronios, A. (2006). The Deficiencies of current data quality tools in the realm of engineering asset management. *Proceedings of the Americas Conference on Information Systems*, August 4 6, Acapulco, México.
- NORSOK. (1999). Z-007 Mechanical Completion and Commissioning. <u>http://www.standard.no/no/Fagomrader/Petroleum/NORSOK-Standard-Categories/Z-MC--Preservation/Z-007/</u>.
- O'Reilly, C. A., (1980) Individuals and information overload in organizations: is more necessarily better?, *Academy of Management Journal*, (23: 4): 684-696.
- Ogunlana, S., Lim, J. and Saeed, K., (1998) Desman: A dynamic model for managing civil engineering design projects, *Computers & structures*, (67: 5): 401-419.
- Okoli, C. and Pawlowski, S. D., (2004) The Delphi method as a research tool : an example, design considerations and applications, *Information and Management*, (42: 1): 15-29.
- Okoli, C. and Schabram, K., (2010) A Guide to Conducting a Systematic Literature Review of Information Systems Research, *Sprouts: Working Papers on Information Systems*, (26: 10): 1-49.
- Orlikowski, W. J. and Baroudi, J. J., (1991) Studying Information Technology in Organizations: Research Approaches and Assumptions, *Information Systems Research*, (2:1): 1-28.
- Otto, B., Hüner, K. and Österle, H. (2009). Identification of business oriented data quality metrics. *Proceedings of the 14th International Conference on InformationQuality, Germany*.
- Pipino, L. L. and Lee, Y. W. (2007). Applying IT to Healthcare: Humans, Errors, and a Data Quality Perspective. *Proceedings of the Americas Conference on Information Systems*, August 10 - 12, Keystone, Colorado.
- Pipino, L. L., Lee, Y. W. and Wang, R. Y., (2002) Data Quality Assessment, *Communications of the ACM*, (45: 4ve): 211-218.
- Ramaswamy, M., (2006) On the phenomenon of information dilution, *Issues in Information Systems*, (VII: 2): 289-292.
- Rivas, R. A., Borcherding, J. D., González, V. and Alarcón, L. F., (2010) Analysis of factors influencing productivity using craftsmen questionnaires: Case study in a Chilean construction company, *Journal of Construction Engineering and Management*, (137: 4): 312-320.
- Rowley, J., (2007) The wisdom hierarchy: representations of the DIKW hierarchy, *Journal of Information Science*, (33: 2): 163-180.
- Sambasivan, M. and Soon, Y. W., (2007) Causes and effects of delays in Malaysian construction industry, *International Journal of Project Management*, (25: 5): 517-526.
- Schmidt, R. C., (1997) Managing Delphi Surveys Using Nonparametric Statistical Techniques, *Decision Science*, (28: 3): 763-774.
- Sein, M. K., Henfridsson, O., Purao, S., Rossi, M. and Lindgren, R., (2011) Action Design Research, *MIS Quarterly*, (35: 1): 37-56.

- Sekine, K. and Arai, K. (1994). Design Team Revolution: How to Cut Lead Times in Half and Double Your Productivity: Productivity Press, Portland, OR.
- Shankaranarayanan, G., Wang, R. Y. and Ziad, M. (2000). Ip-map: Representing the manufacture of an information product. *Proceedings of the 2000 Conference on Information Quality*.
- Singh, P., Park, I., Lee, J. K. and Rao, H. R., (2009) Information sharing: a study of information attributes and their relative significance during catastrophic events, *Cyber-Security and Global Information Assurance: Threat Analysis and Response Solutions, IGI Publishers.*
- Strong, D. M., Lee, Y. W. and Wang, R. Y., (1997) Data quality in context, *Communications of the ACM*, (40: 5): 103-110.
- Sunavala, K. P., (2008) The value of RAM, *ABB Review*, (SPEC. REP: Process Automation Services & Capabilities): 74-78.
- Talburt, J. R. (2011). Entity resolution and information quality. Burlington: Morgan Kaufmann Pub.
- Tayi, G. K. and Ballou, D. P., (1998) Examining data quality, *Communications of the ACM*, (41: 2): 54-57.
- Toor, S.-U.-R. and Ogunlana, S. O., (2008) Problems causing delays in major construction projects in Thailand, *Construction Management & Economics*, (26: 4): 395-408.
- Tribelsky, E. and Sacks, R., (2011) An empirical study of information flows in multidisciplinary civil engineering design teams using lean measures, *Architectural Engineering and Design Management*, (7: 2): 85-101.
- Tulacz, G. J., (2005) Worrying about labor shortage, *Engineering News Record*, (Nov. 14, 2005, (www.enr.com) (Sep. 29, 2011).
- Vician, C. (2011). A Resource Review for Health Care Quality Execution: Business Intelligence Strategy. *Proceedings of the Americas Conference on Information Systems*, August 5 - 7, Detroit, Michigan.
- Walsham, G., (1995) The emergence of interpretivism in IS research, *Information Systems Research*, (6: 4): 376-394.
- Walsham, G., (2006) Doing interpretive research, *European Journal of Information Systems*, (15): 320-330.
- Wambeke, B. W., Hsiang, S. M. and Liu, M., (2011) Causes of variation in construction project task starting times and duration, *Journal of Construction Engineering and Management*, (137: 9): 663-677.
- Wand, Y. and Wang, R. W., (1996) Anchoring data quality dimensions in ontological foundations, *Communications of the ACM*, (39: 11): 86-95.
- Wang, R. Y., Lee, Y. W., Pipino, L. L. and Strong, D. M., (1998) Manage your information as a product, *Sloan Management Review*, (39: 4): 95-105.
- Wang, R. Y. and Strong, D. M., (1996) Beyond Accuracy: What Data Quality Means to Data Consumers, *Journal of Management Information Systems*, (12: 4): 5-33.
- Webster, J. and Watson, R. T., (2002) Analyzing the Past to Prepare for the Future: Writing a Literature Review, [Editorial], *MIS Quarterly*, (26): xiii-xxiii.

- Wixom, B. H. and Todd, P. A., (2005) A theoretical integration of user satisfaction and technology acceptance, *Information Systems Research*, (16: 1): 85-102.
- Yeo, K. and Ning, J., (2002) Integrating supply chain and critical chain concepts in engineer-procure-construct (EPC) projects, *International Journal of Project Management*, (20: 4): 253-262.

Appendix A – Data Collection Activities

Stages	Principles Activities		Participants	Documentation
		Collecting documentation	1 information manager	 emails from site-workers internal coding manual external coding manuals project contracts
		Assessing collected drawings vs. project requirements	End-users, site workers (assembly), 2 information managers	 lists of missing or incorrect values in drawings contractual project requirements
	÷	Participating in principal IT/IS coordination meetings	Experts in various engineering systems, developers, department managers, 2 information managers	- minutes of meetings from 7 meetings lasting 2-3 hours each - personal notes
Problem formulation stage	Practice-inspired research	Participating in bi-weekly IT/IS coordination meetings	2 information managers, IT/IS coordinators, document control managers	 compiled list of decided actions from all meetings (54 meetings, 1-2 hours each) personal notes
		Targeting one specific engineering project ("TestP") as test-project for DQ/IQ assessment	3 information managers	- personal notes from 1 meeting
		Creating system chart for TestP including organizational and customer systems	2 information managers	 minutes of meeting from 2 meetings 4 emails 3 ppt-files
		Conducting semi-structured interviews	1 information manager, 3 experienced engineers	 - 3 interviews conducted, lasting 1-2 hours each - personal notes - typed out within 24 hours of each interview
Ĕ		Identifying and ranking problems	1 information manager, a group of 25 expert engineers	 lists of problems lists of comments analysis documentation
		Assessing existing DQ/IQ tool	2 information managers	 printouts of outputs personal notes
	Theory-ingrained artifact	Conducting a review of DQ/IQ research in general	1 information manager	 collection of articles personal notes
		Conducting a review of DQ/IQ research in the context of construction engineering projects	1 information manager	 set of identified publications analysis documentation
		Selecting existing DQ/IQ assessment frameworks for guidance	1 information manager	- DQA article - TIQM article and book
		Presenting and discussing existing DQ/IQ assessment frameworks in workshops	4 information managers, 1 document control manager	 presentation (.ppt) audio recording of workshops personal notes
		Conducting method workshop	4 information managers, 1 document control manager, 2 developers	 presentation (.ppt) audio recording of workshop personal notes

Stages	Principles			Documentation		
	Reciprocal shaping	Creating a proposal for IQS	1 IT/IS coordinator, 1 developer, 2 information managers (IM)	- the proposal document		
	sha	Developing IQS architecture	2 IMs	- IQS system chart		
	Я	Developing IQS basic rules	2 IMs	- list of IQS rules		
	ncurrent evaluation	Participating in meetings concerning TestP's DQ/IQ related requirements	IMs, engineering discipline managers, document control (DC) manager	 List of requirements (4 meeting, 1.5 - 3 hours) 4 emails personal notes 		
		Participating in weekly IQS follow-up meetings with focus on training	IMs, engineering discipline managers, DC manager, end-users	- compiled list of decided actions from all meetings (19 meetings, 1 hour each) - personal notes		
		Participating in weekly IQS follow-up meetings, with focus on clarifications	1IM, engineering discipline managers or their delegates, end-users	 compiled list of decided actions from all meetings (22 meetings, 1 hour each) 81 emails 		
		Refining IQS rules according to (emergence of new understanding of) TestP's requirements	IMs, engineering discipline managers, DC manager, DC in project, end-users	- minutes of meeting (20 meetings, 1 hour each) - 89 emails - 3 discussions (recorded)		
	nd co	Implementing the first IQS report	3 IMs, 1 internal SharePointSite expert	- IQS report		
B	Mutually influential roles; Authentical and concurrent evaluation	Implementing IQS in 15 additional projects	3 IMs, 5 DCs, 5 engineers	- minutes of meetings (6 meetings, 1 hour each) - 48 emails - IQS reports		
		Investigating use of quality dimensions	2 IMs	- 246 assessment rules - descriptions of dimensions and how they were understood at EUMEC		
		Refining IQS basic architecture, interfaces towards customer applications, interfaces towards internal applications, and report features	3 IMs, 1 interface coordinator	- scripts - minutes of meetings (5 meetings, 1.5 hours each) - 14 emails		
		Conducting interviews	2 VPs, 1 site manager, 5 discipline managers, 1 QA manager	9 hours of interviews audio recorded		
		Developing progress graphics for reported errors in IQS	1 IM	 printouts of graphs minutes of meeting (1 meeting, 1 hour) 2 emails 		
		Starting discussion on the need for management report	IMs, developers, project (TestP) management	- minutes of meetings (9 meetings, 2 hours each) - 3 emails		
earning	Guided emergence	Acknowledging the need for narrower phases, starting development of commissioning-reports	2 IMs, end-users	 minutes of meetings (6 meetings, 1-2 hours each) 24 emails commissioning report 		
on and		Acknowledging the need for parking feature	3 IMs, engineering discipline managers, end- users	- minutes of meetings (2 meetings, 2 hours each) - 75 emails		
Reflection and learning		Acknowledging the need for anchoring the use of IQS in the organizational Project Execution Model (PEM)	QA-manager, DC manager, 4 IMs, IM department manager, PEM-design-responsible	- minutes of meetings, (3 meetings, 2-3 hours each) - 9 emails		

Appendix B – Rule Description

The individual construction project's requirements are captured separately in the syntax of a rule. The rule can be executed against all required records for the purpose of identifying any irregularities concerning the requirement. For example, "Rule2: X,Y,Z is outside area limits" means that the coordinates indicating exactly where the item is located are not in accordance with the area value chosen (areas are imaginary cubes indicating various areas of the whole construction and determined by coordinates named as D1, D2,...etc.). If the coordinates of the item are outside the coordinates of the chosen area, one of the chosen values (coordinates or area) must be incorrect. Only the opinion of an expert can tell which one is correct; the assessment tool will merely report this as an error (an example of the syntax code is shown for this rule in Appendix C).

Requirements that recur between projects are put together in a default set of rules, which provide the starting point of assessment for each new project. Every defined rule is saved in a rule database from which individual projects select the rules they need. They may, in addition, ask for new rules to cover their specific requirements. When a rule is chosen by a project, it is marked with the project ID for the system to know which rules to execute against that project's data. Several projects may choose the same rule.

Every night, the rules in the rule database are executed against the data in the data warehouse. After processing, the identified DQ/IQ errors are saved in a result database, which becomes the data source for Excel-based IQS reports for all projects. The reports displayed on the home intranet site of each individual project identify only errors pertinent to that project. The users can display the report in multiple formats, such as per engineering discipline, per product, per date of delivery, per commissioning package, etc. An example of a report displaying a selected work package is shown in Table 19.

deliv_date	work_pkg	error_msg		Ι	Μ	Р	Grand Total
2012_12_28	Work	Rule3: Tags without drawing references	4	12	96		112
	Package:	Rule10: Missing mounted-on		1			1
	9999-Z99	Rule15: Missing or inadequate description on tag		1			1
		Rule17: Missing or illegal PO-number		3	13		17
		Rule13: Missing EX-class	32	14	22		68
		Rule14: Missing IP-grade	32	14	22		68
		Rule2: X, Y, Z outside area limits	4	25	31	1	61

Table 19. An IQS report displaying identified errors for a work package

The first column of the report shows the delivery date to the customer for information related to the selected work package (shown in column 2). Column 3 displays the rules that have identified errors. The next four columns display the affected engineering disciplines (E= Electro, I =Instrument, M= Mechanical, and P=Process) and the number of errors related to that discipline. Column 8 shows the total number of errors identified by the rule executed.

In a typical use case, an electrical engineer may want to further investigate the four identified errors shown under E (column 4) for Rule 2: X, Y, Z outside area limits. The engineer can double-click the number '4' in the column in order to display more details (see examples of additional information in Table 20). The first column tells the user exactly which item the error is related to. The second column displays the values of the field in the item record where the error was identified, in this case the value of the coordinate field. The actual coordinate values for the item related to row 1 are as follows: X=806.000, Y=795.000, Z=219.000. In column 5 (area), the value M60 refers to an imaginary cube determined by its coordinates. If the coordinates of the item are outside the coordinates of its area, either the area value is incorrect or the coordinate value is incorrect. The rest of the columns are commonly chosen to provide the user with enough information to determine which value is incorrect and what the correct value is. Some items might have to be excluded from assessment by this rule for reasons of impracticality or because no rule syntax was possible. Such items are then set aside in a parking table together with the rule ID, and are not checked further against that specific rule. At any time, a parked item can be revoked on request. A question may also arise regarding who

is responsible for this item. Column 7 (resp._eng.) provides the name of the person to ask for the correct weight or any additional information needed.

object							
_name	field_value	description	discpl	area	work_pkg	resp_eng	error_msg
GD11	X=806.000	Starter for	Е	M60	Work	John Smith	Rule2: X, Y, Z
1222	Y=795.000	el.motor for			Package:		outside area
	Z=219.000	centrifugal			9999-Z99		limits
		fan					
GD22	X=806.000	Starter for	E	M60	Work	John Smith	Rule2: X, Y, Z
2333	Y=795.000	el.motor for			Package:		outside area
	Z=219.000	centrifugal			9999-Z99		limits
		fan					
GD33	X=806.000	Starter for	Е	M60	Work	John Smith	Rule2: X, Y, Z
3444	Y=795.000	el.motor for			Package:		outside area
	Z=219.000	centrifugal			9999-Z99		limits
		fan					
EC777	X=862.123	Drillers	Е	M60	Work	John Smith	Rule2: X, Y, Z
888	Y=808.170	control cabin,			Package:		outside area
	Z=239.758	air-			9999-Z99		limits
		conditioning					
		system					

Table 20. IQS detail report

Appendix C – Example of Syntax Code

Example of Syntax Code Rule2: X, Y, Z outside area limits SELECT tbl.project_code, tbl.contractor code, tbl.tagno, 'X=' VARCHAR(tbl.x_min)+ ',Y='+VARCHAR(tbl.y_min)+ ',Z='+VARCHAR(tbl.z_min), tbl.description, tbl.upd by, tbl.upd date, tbl.upd code, tbl.po no, tbl.discipline code, tbl.product_code, tbl.area, tbl.system no, tbl.site code, " tbl.status3, FROM table1 tbl, table2 b WHERE tbl.project code = b.project code tbl.area = b.area AND tbl.x min > 0AND AND tbl.y min > 0AND tbl.z min > 0(tbl.x min NOT BETWEEN b.x min AND b.x max AND OR tbl.y min NOT BETWEEN b.y min AND b.y max OR tbl.z min NOT BETWEEN b.z min AND b.z max) (tbl.eq pack no = 'NA' AND OR tbl.eq_pack_no = " OR tbl.eq_pack_no = tbl.tagno) tbl.upd_code <> 'D' AND tbl.project_code = 'NN' AND

AND tbl.contractor_code = 'ZZ'

Appendix D – Research Publications

No	Title	Publication outlet
1	Westin, S. and Päivärinta, T. (2011). Information Quality in Large Engineering and Construction Projects: A Delphi Case Study	Proceedings of the 19 th European Conference on Information Systems (ECIS), Helsinki, Finland. (http://aisel.aisnet.org/ecis2011/273)
2	Westin, S. (forthcoming). Data and Information Quality Research in Engineering Construction Projects: A Review of Literature	Proceedings of the 21 st International Conference of Information Systems Development (ISD), Prato, Italy, 2012 (http://www.springer.com/computer/information+syst ems+and+applications/book/978-1-4614-7539-2)
3	Westin, S. and Sein, M.K. (submitted 2013). The Design and Emergence of a Data/Information Quality System	Under review
4	Westin, S., Holen, O.B. and Päivärinta, T. (2012). Data and Information Quality Dimensions in Engineering Construction Projects	Proceedings of the 18 th Americas Conference on Information Systems (AMCIS), Seattle, Washington, USA. (http://aisel.aisnet.org/amcis2012/proceedings/DataIn foQuality/1)
5	Westin, S. and Sein, M.K. (2013). Improving Data Quality in Construction Engineering Projects: An Action Design Research Approach	Journal of Management in Engineering (http://ascelibrary.org/doi/abs/10.1061/(ASCE)ME.1 943-5479.0000202)

Association for Information Systems AIS Electronic Library (AISeL)

ECIS 2011 Proceedings

European Conference on Information Systems (ECIS)

10-6-2011

INFORMATION QUALITY IN LARGE ENGINEERING AND CONSTRUCTION PROJECTS: A DELPHI CASE STUDY

Soffi Westin

Tero Päivärinta

Recommended Citation

Westin, Soffi and Päivärinta, Tero, "INFORMATION QUALITY IN LARGE ENGINEERING AND CONSTRUCTION PROJECTS: A DELPHI CASE STUDY" (2011). *ECIS 2011 Proceedings*. Paper 273. http://aisel.aisnet.org/ecis2011/273

This material is brought to you by the European Conference on Information Systems (ECIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ECIS 2011 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

INFORMATION QUALITY IN LARGE ENGINEERING AND CONSTRUCTION PROJECTS: A DELPHI CASE STUDY

- Westin, Soffi, University of Agder, Dept. of Information Systems, Serviceboks 422, 4604 Kristiansand, Norway, <u>soffi.westin@uia.no</u>
- Päivärinta, Tero, University of Agder, Dept. of Information Systems, Serviceboks 422, 4604 Kristiansand, Norway, tero.paivarinta@uia.no

Abstract

This Delphi study identifies problems that have significant impacts on profits gained from large engineering and construction projects in a European company. Information quality gained remarkable weight among the identified problems. The problems were ranked in accordance to their estimated impact on the project profit margins. Within a consolidated list of 125 problems identified altogether, the final ranking of the top 18 problems was strongly agreed upon by an expert panel. The panel involved experienced engineering and management professionals throughout the construction project supply chain. Among the top 18, eight problems, including the top six, concerned information quality. The results address a need for increased focus on information quality challenges in the target organization and provide a detailed account of such challenges in comparison to the previous literature on information quality in engineering and construction.

Keywords: Information quality, data quality, engineering and construction project, Delphi study.

1 Introduction

During the era of increasing business globalization, large-scale engineering and construction projects have also become global (Xue, Wang, Shen, & Yu, 2007). Whereas even local construction projects are often delayed and exceed their budgets (Al-Momani, 2000; Assaf & Al-Hejji, 2006; Chan & Kumaraswamy, 1997; Faridi & El-Sayegh, 2006; Long, Ogunlana, Quang, & Lam, 2004; Toor & Ogunlana, 2008), global and inter-organizational construction supply chains require even more coordination and effective utilization of information technologies (Xue, et al., 2007).

This article discusses about results from a Delphi study, in which 38 experts identified problems which cause significant impacts on profits gained from global engineering and construction projects in a European, multi-discipline, construction and engineering company. In slight contrast to previous literature on construction and engineering projects and related challenges, information quality gained remarkable weight among the identified problems in this case. The information quality issues in our target organization also represent a more detailed view on this issue than previously reported (cf. Dehlin & Olofsson, 2008; Toor & Ogunlana, 2008; Wantanakorn, Mawdesley, & Askew, 1999; Xue, et al., 2007). While the results illustrate how information quality is a problem in this case, we believe that the study represents also interesting insight into potential information quality issues in construction and engineering, which has, as a field, been less prominently present in the contemporary literature on data and information quality in general (cf. recent reviews by Batini, Cappiello, Francalanci, & Maurino, 2009; Madnick, Wang, Lee, & Zhu, 2009). This paper thus focuses on the research question:

What are the main information quality challenges in global engineering projects?

The rest of the paper is organized as follows: Section 2 gives a brief description of previous literature on information quality among the general-level problems in large construction projects. Section 3 introduces the target organization and describes the research process, and section 4 displays the results. In section 5 we discuss our findings and section 6 concludes with suggestions for further research.

2 Background: Information Quality and Large Construction Projects

The definitions of data quality and information quality have been characterized by a lack of consensus when it comes to distinguishing between the two. Data quality often (but not always) refers to technical issues while information quality usually refers to non-technical issues (Madnick, et al., 2009). In this paper we use the term information quality to cover both aspects.

The total quality management movement (e.g., Deming, (1982); Juran & Goferey(1999) has greatly influenced on the information quality research (Madnick et al., (2009). According to this approach,, information can be regarded as a "product" which is "manufactured" in organizations (Madnick, et al., 2009) and later on consumed by its users within and across organizations. *Information quality* has been defined broadly through its "fitness for use" (Wang & Strong, 1996), in relation to the recognized purposes of information use and user groups. Whereas the early research focused on query techniques on multiple data sources and data warehouses in the end of 1980s, the research field has later on spread to a number of new application areas, such as customer resource management, knowledge management, supply chain management and enterprise resources planning (Madnick, et al., 2009).

Batini et al. (2009) identify altogether 28 quality dimensions in their review of data quality methodologies. Four core dimensions, *accuracy, completeness, consistency* and *timeliness* have been emphasized frequently throughout the methodologies. Since the definitions of the dimensions vary in the literature it is important to declare what definitions we make use of. *Completeness* means that all values required for a certain record should be recorded (Batini, et al., 2009). *Consistency* refers to the

values of data, which are expected to be identical in similar situations (Wand & Wang, 1996). In our context, it means that for the occurrences of the same data in different registers the values should be identical. *Timeliness* is often defined as whether the data is out of date but can also be defined as availability of output in time, i.e. the time the data are actually used (Wand & Wang, 1996). In this sense, it corresponds also to the dimension of speed below.

Other potential quality dimensions in previous research have been currency, volatility, uniqueness, appropriate amount of data, accessibility, credibility, interpretability, usability, derivation integrity, conciseness, maintainability, applicability, convenience, speed, comprehensiveness, clarity, traceability, security, correctness, objectivity, relevancy, reputation, ease of operation, and interactivity (Batini, et al., 2009).

In the field of large construction projects, several researchers have reported frequent delays as the main problem to be tackled (Long, et al., 2004; Sambasivan & Soon, 2007; Toor & Ogunlana, 2008). Delays are costly (Faridi & El-Sayegh, 2006), and cause negative impact on the profit margins (Ling, Pham, & Hoang, 2009). In a recent study, Toor & Ogunlana (2008) reviewed the literature on construction project delays and identified 75 problems encountered in a number of countries. The most common problems include: lack of resources, lack of adequate communication, poor contractual management, design delays, changed orders, deficiencies in public agency organizing, deficiencies in planning and scheduling, inadequate site planning and control, lack of experienced subcontractors, and poor resource estimations. During the first decade of the new millennium, some new types of problems have also emerged, including: lack of contractor experience, slow decision-making by owners, owner's lack of experience, escalation in material prices, lack of labour, complex and changing legal systems, and lack of design standardization. Toor & Ogunlana (2008) categorize the 75 problems further to eight categories, based on the entities related to projects, such as: problems related to clients, designers, project management and consultants, contractors, labour, finance, contract, communication, site and environment, and miscellaneous.

Within this exhaustive list of problems and categories, however, few information quality problems are mentioned. In fact, Toor & Ogunlana (2008, pp. 400-401) mentioned only three problems among their list of the 75, which we could recognize to represent information quality in light of the definition above:

- Confusing and ambiguous requirements (from the client), which we would categorize as a problem of information interpretability, concerning the requirements documentation.
- Errors and omissions in design documents (by the designers), which we categorize as a problem of information accuracy and completeness, concerning the design documents.
- Incomplete contract documents, which we thus categorize also as a problem of information completeness.

In addition, Toor & Ogunlana (2008) identify lack of IT use for information, coordination, and interface management and, in general, poor quality control over projects also among the 75 problems. These may involve more specific issues representing certain information quality dimensions, if studied further in context.

In the scarce literature discussing information quality in construction projects in more detail, a study in Thailand found that incomplete drawings were a major cause of delays in 75% of the projects (Ogunlana, Promkuntong, & Jearkjirm, 1996). Information completeness thus has played a very important role at least in one national context. Lim & Mohamed (2000), in an exploratory study, identified that "waiting for information", i.e., information timeliness and speed, was a major reason for delays in large construction projects in the UK.

Beyond the literature focusing on project delays, a few other articles also mention information quality as a potentially significant issue in the construction projects. For example, Wantanakorn et al. (1999), refer to a 10-item list of causes for human errors in project management, which they apply to construction projects. One of those items is information quality, which in this context refers to poor quality of managerial instructions and procedures from one person to another (ibid.). Hjelt & Björk (2007) mention information quality while evaluating adoption and use of an electronic document management system in a large construction project. To evaluate information quality, they mention the Delone & McLean (DeLone & McLean, 2003) model for information system success as their "framework of understanding". Delone & McLean's concept of information quality includes the quality dimensions of accuracy, timeliness, completeness, relevance, and consistency (ibid., p. 15). Furthermore, it proposes that information quality has a causal correlation to system use, user satisfaction, individual impacts and organizational impacts (ibid.). However, leaving out the direct impact of information quality on the organization, which is included in the Delone & McLean (2003) model, Hjelt & Björk (2007) discuss mainly about role of information quality in the user acceptance of the document management system. While low IT utilization in construction projects has been identified as a significant issue as such (Toor & Ogunlana, 2008), Hjelt & Björk (2007) thus highlight importance of perceived information quality as a *mediating* issue on project performance. Xue et al. (2007) also mention information quality, in general, as one of the issues which may impact on coordination and integration of construction supply chains.

3 Target Organization and Research Process

3.1 Target organization

Our target organization is a European, multi-discipline engineering and construction company (EUMEC; this artificial acronym is used here to anonymize the target organization) with capabilities related to global management, design, procurement, completion and generally execution of complex installations for the oil and gas industry. EUMEC delivers engineering design for construction projects and possesses a significant share of global markets in its product and project domains. The most employees are involved in engineering and construction projects. The biggest projects may typically cost more than 100 million Euros and take up to three years of calendar time.

Engineering projects usually cover the design-related phases of a construction project. The final deliveries for an engineering project mainly consist of documentation, such as drawings and manuals. These are extracted from various data sources like engineering data bases. When the design is completed, outputs from the projects (e.g. drawings, documents and three-dimensional (3D) product models) are handed over to the assembly and completion phases. The latter phases are those, during which the most serious problems, especially delays, are typically manifested.

3.2 Engineering disciplines

Several experts representing different engineering disciplines are needed to design large, complex, robust and yet delicate constructions. Project teams are set up according to engineering and other professional disciplines varyingly required by the individual projects. Table 1 describes the engineering disciplines typically represented in the projects of our target organization. Every engineering discipline has a discipline manager. A discipline usually consists of several engineers, and their manager is responsible for the delivery as a whole. Every engineering discipline depends on input from the other disciplines throughout the project, due to a great number of interfaces and dependencies among the artefacts and systems to be designed and assembled. Together with tight project schedules this means that various activities must be performed in parallel despite that quality control of the results would be easier with sequential task organization. During such concurrent engineering (Sekine & Arai, 1994), quality assurance and possible adjustments are therefore conducted both during the project and then, again, in the assembly and completion phases.

Disciplines	Responsibility
Process	Design of industrial processes; all the facts, sequences and relations in the process and a logical placing of the different items.
Mechanical	Design (choice of equipment and its physical layout and weight).
Piping/Layout	Design of all piping.
Electro	Design and cabling of power distribution for electrical systems: equipment, lights, heat, etc.
Instrument	Design of control systems, i.e. the control of various valves, machines, the alarm systems, and instrumentation cables for distributing signals.
Telecom	Selection and location of radio and audio systems, alarms etc.
HVAC (Heating, Ventilating and Air Conditioning)	Capacity calculations and layout for ventilation etc.
Safety	Various safety assessments.
Structure (steel)	Design of steel structures, supports, outfitting like hand rails, stairs etc.
Architecture	Interior design.

 Table 1.
 Engineering disciplines in EUMEC's construction projects.

3.3 Research Process

A case study approach was chosen because it provides a possibility to understand a phenomenon in its context, and allows a single entity examination (Benbasat, Goldstein, & Mead, 1987). To achieve a preliminary understanding of the problems encountered in the projects, a pilot study including four interviews of key personnel in EUMEC was conducted. As a conclusion from the pilot, EUMEC wanted to acquire more detailed information regarding problem areas with the most negative potential impact on the profit margins. Therefore, a Delphi study was chosen as the main method of collecting data, revealing the problems through an anonymous, balanced and unbiased process, achieving a consensus over the problems, and prioritizing the problems to tackle by a representative expert panel. The Delphi study process leaned on the phases and recommendations of Schmidt (1997) and Okoli & Pawlowski (2004).

45 experts were initially asked to participate in the Delphi study of whom 38 accepted the invitation and 25 followed the study until the end (Table 2). To be regarded as an expert, a candidate should have been a project manager, an engineering manager, or an engineering discipline leader of at least one major project. The list of participants so far was then shown to key personnel in the organization in order to make necessary supplements, and a few experienced engineers, expert site workers and commissioning workers were then added based on these peer opinions. To make sure they had the capacity and willingness as well as sufficient time to participate, an invitation to participate in the study was e-mailed to 45 people who matched one or more of these criteria.

Through e-mail, the panel members were asked to state at least 6 problems, which had occurred in the assembly and completion phases of an engineering and construction project, and to give some brief problem descriptions. This resulted in a list of 217 problems. To obtain the overview needed to perform a consolidation and paring down of the original 217 problems list, we had to categorize them. Initially we used the categories as presented by Toor and Ogunlana (2008), and allocated the problems to those. However, this categorization was less meaningful for EUMEC's problems. There were no problems related to financial issues, such as "Shortage of funding" or "High interest rate". The problems that might come close to this group were more related to budgeting, such as "Lack of control in the registering of hour-lists". In addition, several problems concerning drawings and documents were identified, indicating that this should be a category of its own. None of the Toor & Ogunlana's (2008) miscellaneous problems was identified in this study. Nevertheless, 13 problems had to be allocated to this category while they did not fit any other category.

Discipline	No of respondents
Assembly and completion	2
Electro	1
Engineering manager	5
HVAC	1
Instrument/Telecom	1
Mechanical and Weight	4
Piping/Layout	2
Process	1
Project manager	4
Safety	2
Structure	2
Sum	25

Table 2.Number of final respondents representing the roles and engineering disciplines in
EUMEC

Hence, we conducted a new categorization of the problems. This final categorization is displayed in Table 3 below. Thereafter, the most important problems were selected for further analysis. The consolidated list of problems (now categorized) from round one was emailed to the participants and they were asked to pick the 15 problems they regarded as the most important in relation to the project profits. They were also encouraged to comment, this time as to *why* they picked the problems they did. The categorization was not contradicted by any of the participants; some commented that the list was now more perspicuous, and the categories were comprehensible. This phase was completed by 29 participants.

After this phase the researcher can either eliminate all problems that were not selected by a simple majority of the participants, or, arbitrarily pare the list. Schmidt (1997) points out that the researcher should not be the one that decides the top issues. In our study a simple majority was not reached in this phase, which could be due to the rather large number of problems. Since a manageable number of items to rank are around 20 (ibid), and our results showed that the top 18 problems were voted for by at least 7 participants, a decision was made to select those problems. ("at least 6 participants" would have added 7 more problems and "at least 5 participants" would have added yet another 6 problems) To avoid the danger of the researcher *deciding* the top issues, this choice of the top problems was confirmed as representative by the participants.

For the ranking exercise, the problems in the consolidated list were presented in a random order in four different versions, and these versions were equally divided among the participants to avoid potential bias caused by the order of listing of the items (O'Neill, Scott, & Conboy, 2009; Schmidt, 1997). The rankings were analysed in the SPSS software (Statistical Package for the Social Sciences) in order to determine the level of consensus amongst the participants. Kendall's W, which measures the level of consensus among the rankings, was .151, representing "very weak agreement" (Schmidt, 1997). 25 participants participated in the second round of ranking, given the results from the first round, but 3 of the answers had to be rejected because of irregular ranking (e.g. leaving some problems unranked). After the second round, Kendall's W was now measured to .858 which is consistent with "strong agreement". According to Schmidt (1997), this was an appropriate exit point. Approximately 50% of the participants provided their qualitative comments and groundings on the second ranking round as well, confirming that the experts had put a good thought to the process as a whole, instead of merely selecting the overall ranking from the first round as such.

4 Results

Table 3 displays the final categorization together with the number of problems in each category respectively for the 125 consolidated problems list and the top 18 problems list.

Problem category	# Identified problems	# Top 18 problems
Problems related to documents and drawings (PDoc&Draw)	20	6
Problems related to copying of projects (PCopyProj)	5	1
Problems related to experience (PExper)	4	2
Problems related to budgeting (PBudge)	6	0
Problems related to procedures (PProced)	30	1
Problems related to changes (PChange)	7	2
Problems related to responsibilities/contracts/management (PR/C/M)	22	5
Problems related to design (PDesign)	6	0
Problems related to coding manuals and tag-numbering (PTagno)	5	0
Problems related to Data registers (engineering systems and 3D modelling application) (PDatReg)	8	0
Problems related to engineering (PEngin)	6	0
Problems related to equipment (PEquip)	6	1
Total	125	18

Table 3.Summary of problem categories for consolidated list and the top 18 problems.

The ranking result is displayed in table 4, together with descriptions of the identified problems. R is the relative rank and M represents the mean of the rankings. The problems marked with '*)' are those related to information quality, in light of the information quality dimensions by Batini et al. (2009).

5 Discussion

In comparison to the previously identified major challenges causing delays in construction and engineering projects (Toor & Ogunlana 2008), our target organization clearly experiences significant challenges with information quality. From table 4, we could identify 8 problems which relate directly to information quality (Table 5) – including the six top problems. In general, this finding highlights importance of information quality in our target organization more and in a greater detail than the general-level literature. In the following, we will discuss about the eight information quality problems in light of the previous research and outline the future managerial efforts to tackle those problems.

The problem of delays in distribution of drawings and documents (#1, table 5) is well in line with the exploratory results by Lim & Mohamed (2000), who also recognized that "waiting for information" was a major reason for delays in the British projects. The timeliness/speed issue relates also to supplier-delivered design documentation (#2) and equipment drawings (#3), which were identified as problem categories of their own. If the information is not available in time, engineering databases will either contain omissions or incorrect data (#4 also relates to this issue). The deliveries are extracted as drawings (mainly), from these databases, and if they are not complete and accurate, delays will occur on the assembly site due to extra use of work hours to correct errors or to recover missing information. Hence, our target organization would greatly benefit from improved tools and practices which would help to control and enhance the speed of delivery and timeliness of drawings and other design documents; monitoring them both inside the company and in the subcontractor relations.

п	Problem	Durklaur	Commente	M
R	category PDoc&Draw	Problem Delays in distribution of drawings and	Comments The engineers are waiting for documentation they need for	M
1	PDoc&Draw	documents *) Design is based on unfinished or incorrect supplier documentation *)	proceeding with their work. If the needed documentation does not arrive with the engineers on time, the proceeding design work might be based on incomplete documentation. In addition, if the documentation is on time, the	1,68
2	2.01		information presented in the supplier documentation could be incorrect.	2,64
3	PChange	Equipment drawings change after engineering design is completed *)	When engineering design is completed, it adds a lot of work if the equipment drawings are changed because that could lead to change and alteration of the overall design.	3,32
4	PDoc&Draw	Errors and omissions in supplier drawings *)	Even if the designers discovers the errors and omissions in time to adjust the design before it reach completion, it still demands a lot of extra work. Comment from an engineer: One of the biggest problems we have is the <i>amount</i> of errors in third party drawings.	5,36
5	PDoc&Draw	There are great shortcomings in the interface documentation on drawings, (not correct information as size, weight, to a number) *)	When there are shortcomings in the interface documentation more time is used during the assembly phase. The interfaces might not even fit and adjustments have to be done.	5 50
5	PDoc&Draw	tag number) *) GA (General Arrangements/Assembly) drawings are not consistent with the equipment *)	The General Arrangement drawing displays, among other things, where different equipment should be located. If there are disparities between the equipment on the drawing and the equipment at hand, it	5,59
6	PEquip	Errors in delivered equipment are revealed in the end phases of the projects	takes more time to complete the assembling. Errors in delivered equipment revealed in the end phases of the projects leads to re-adjusting or even rebuilding of the equipment,	6,05
7	PR/C/M	Insufficient delivery specifications, especially on internal and external interface (who does what)	which add work hours to the project. This problem concerns the interfaces between the internal design and the external equipment (e.g. if the equipment needs a bracket of some kind, does it come with the bracket or does the organization design the bracket? E.g. is the bracket/equipment the interface, or is the steel structure/bracket the interface?)	7
9	PDoc&Draw	Lacking interface within our organization between engineering, equipment, control systems, flow diagrams (the drawings are not congruent) *)	The drawings are not consistent between the different engineering disciplines which create confusion and errors.	9,5
10	PProced	Mismatch between engineering plan and equipment plan	The organisation has one execution plan for engineering design and another for product development. In large projects where both plans have to be used simultaneously there seems to be a mismatch which could lead to delays.	9,68
11	PCopyProj	Copy projects always lead to recurring errors that we use hours to correct from project to project *)	When copying data registers from one project to the next, errors are copied as well if they are not fixed.	10,41
12	PR/C/M	Underestimation of work scope resulting in deliveries being ignored or a lack of someone in charge	Some engineers states that work scope is underestimated in terms of work hours, and that this could lead to deliveries being ignored or nobody is appointed to be in charge of that delivery.	11,09
13	PR/C/M	The demanding price for detail engineering scope is too low in lump sum projects	Setting a fixed price for a detail engineering scope is difficult and the engineers commented that the prices are too low when compared to the work hours used.	12,14
14	PR/C/M	We are too "kind" to the customers and often they get much more than they actually pay for	"Personnel need greater knowledge concerning the content of the contract, so we do not give them more than they actually bought. We need to be more cynical and more like a pedlar!"	14,09
15	PChange	The scope changes and new elements are being sneaked in	"Details concerning the scope are seldom given and the contractor is "forced" to take ownership because the contract describes "system responsibility"".	15,09
16	PExper	Our organization has no procedure that is followed when it comes to transfer of experience between projects	"We have tried to make a system for transfer of experience, but we have not succeeded."	15,18
17	PExper	Important personnel leave the projects before installation and completion	"It is common that project personnel are transferred to a new project before completion. It is very important that these people give information to other people in the project to avoid using time on unnecessary search for information that already exist."	16,77
18	PR/C/M	Being seated in three different locations in one town results in more errors in the projects	The engineers argued that they would like to be seated together in one location whenever that was possible, but due to shortage of office spaces that was difficult to accomplish.	17,73

Table 4.	Ranked results;	*) = Information	Quality problem
----------	-----------------	------------------	-----------------

R	Problems	Quality dimension(s)
1.	Delays in distribution of drawings and documents *)	Timeliness/Speed
2.	Design is based on unfinished or incorrect supplier documentation *)	Timeliness, Completeness,
		Accuracy
3.	Equipment drawings change after engineering design is completed *)	Timeliness, Accuracy
4.	Errors and omissions in supplier drawings *)	Accuracy, Completeness
5.	There are great shortcomings in the interface documentation on drawings,	Accuracy
	(not correct information as size, weight, tag number) *)	
6.	GA (General Arrangements/Assembly) drawings are not consistent with the	Accuracy
	equipment *)	
9.	Lacking interface within our organization between engineering, equipment,	Consistency
	control systems, flow diagrams (the drawings are not congruent) *)	-
11.	Copy projects always lead to recurring errors that we use hours to correct	Accuracy
	from project to project *)	

Table 5.Information quality problems and the associated quality dimension(s)

Management of changes in equipment drawings (#3) in relation to on-going project documentations causes another challenge of information timeliness, realizing as an accuracy problem in the end. The problem results in inconsistency between the engineering design and the content of the drawing. The problem itself refers to an internal equipment drawing change, which is caused by changes in actual equipment, developed and produced by a separate department in EUMEC using the same data registers.

The issues of accuracy and completeness were highlighted especially in relation to supplier documents alongside the timeliness (#2, #4). However, accuracy issues were also related to general assembly drawings (#6), interface documentation (#5), and utilization of information from previous project designs (i.e., copy projects, #11), which are not supplier-originated problems. This observation is well in line with Toor & Ogunlana's (2008) category of "errors and omissions in design documents". In EUMEC, such errors and omissions take place in relation to varying categories of documentation throughout the projects. The high rankings of these problems are also well in line with the previously observed significance of drawing incompleteness in 75% of Thai construction projects (Ogunlana, et al., 1996). In EUMEC, the accuracy and completeness problems should be observed effectively both during the projects, to hinder subsequent designs and equipment assemblies based on inaccurate information, and after the projects, to ensure that the future projects would not copy the accuracy errors from the previous ones.

Inconsistency of drawings between different engineering disciplines (#9) was a problem with no direct correspondence to the literature on construction project problems viewed above. This issue highlights a need for focusing on co-operation between the engineering disciplines to ensure that their design documents and drawings would follow common conventions from the viewpoint of the whole design, in addition to being internally consistent and accurate.

The information quality problems (Table 5) altogether involved the four core quality dimensions (Batini, et al., 2009): accuracy, completeness, consistency, and timeliness. The dimensions could also be connected to particular types of design documentation in this case; e.g. flow diagrams and General Arrangement drawings. The timeliness dimension was especially connected to the speed of information delivery as an important factor to ensure timeliness, which could perhaps be expected in the context of concurrent engineering. The results give a starting point for further information management -related actions in our target organization for ensuring better information quality in their engineering projects.

The Delphi method does not reveal the potential cause-effect relationships between the identified issues, which remains as a shortcoming of this study. Hence, our next step is to delve deeper into *why*

the organization has these information quality related problems, and what can be done, more specifically, to reduce the causes to them. The results also give a basis for choosing more detailed information quality assessment and coordination techniques (cf. e.g. those identified by Batini, et al., 2009) fitting to the information types and quality dimensions involved in this case. Especially, the target organization has already started an implementation of an information quality system, built upon the existing product and project data registers, which monitors the key indicators of these quality dimensions more actively already in the early phases of the projects.

The issues of ambiguous requirements and incompleteness of contract documentation, which were among previously observed information quality problems in construction projects (Toor & Ogunlana, 2008), were less prominent in our case. Emergence and foci of particular information quality problems in construction projects may vary from time to time and context to context. However, whereas studies focusing on information quality in large construction projects remain scarce, further research efforts are needed to identify the most common or stickiest problems at the industry level.

Information quality of managerial decision-making in projects (Wantanakorn, et al., 1999) gained no visibility in our results. This could imply that such information is already satisfactory in EUMEC, or that such information is not yet recognized to play a role related to delays or profit margins. In contrast to previous research discussing perceived information quality as an intermediating factor having impact on IT adoption (Hjelt & Björk, 2007) or supply chain integration (Xue, et al., 2007) in construction projects, our expert panel results imply direct impacts of information quality on the project profits in EUMEC. Here, management of information quality has become prioritized as a revenue-bringing issue in its own right, instead of only representing a partial cause for other mediating issues. This finding is well in line with the Delone & McLean (2003) model, in which such a direct causal relationship between information quality and organizational impacts has been previously tested explicitly in the retail industry (DeLone & McLean, 2003, p. 14; Teo & Wong, 1998).

The most of the remaining top issues (#7, #8, #12-18) could be mainly found among the previously identified problems of construction and engineering projects. Whereas the focus of this paper remains on the top-ranked information quality problems, we do not discuss about the remaining ones in more detail here. Problem #10, "Mismatch between engineering plan and equipment plan", although reminding a problem of information consistency at first glance, refers here to organizing of internal project execution plans between the engineering department and the equipment department. In large projects, where both plans are used simultaneously, there seems to be a mismatch which could lead to delays. The plans and planning practices are now being worked on to be revised and aligned.

6 Summary and conclusion

Our Delphi study in EUMEC resulted in an initial list of 125 problems decreasing the profit margins of large engineering and construction projects. A strong consensus on the relative importance of the top 18 problems was reached. Among the top 18, eight problems, including the top six, relate to information quality. This result contributes by providing more detail into the previous literature, which has hitherto mentioned the information quality problems in construction and engineering projects largely at a general level, without relating them explicitly to information quality research or quality dimensions. A concretization and prioritization of information quality problems in context provide a basis for targeting concrete management and systems development initiatives for their mitigation. Previous studies on information quality in this field are scarce, only indicating that it might have importance among many factors. Hence, future research efforts will be needed to estimate whether our results would imply a more general-level contemporary problem in the whole field of large-scale engineering and construction projects or whether the great prominence of information quality issues was only contextually specific to our target organization.

We plan to continue this research in cooperation with EUMEC by following their subsequent actions to manage information quality better through a dedicated information quality system initiative for

engineering projects. The system will include concrete and maximally automated controls of the most important quality dimensions and measures, focusing on the issues found in the Delphi study. Additional field research efforts on contemporary information quality in large engineering and construction projects will be needed to bring up more generalizable knowledge on the important quality dimensions, tools, and practices for this industry in more detail.

References

- Al-Momani, A. H. (2000). Construction delay: a quantitative analysis. International Journal of Project Management, 18(1), 51-59.
- Assaf, S. A., & Al-Hejji, S. (2006). Causes of delay in large construction projects. International Journal of Project Management, 24(4), 349-357.
- Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. ACM Computing Surveys, 41(3), 1-52.
- Benbasat, I., Goldstein, D. K., & Mead, M. (1987). The Case Research Strategy in Studies of Information Systems. MIS Quarterly, 11(3), 369-386.
- Chan, D. W. M., & Kumaraswamy, M. M. (1997). A comparative study of causes of time overruns in Hong Kong construction projects. International Journal of Project Management, 15(1), 55-63.
- Dehlin, S., & Olofsson, T. (2008). An evaluation model for ICT investments in construction projects. Journal of Information Technology in Construction, 13, 343-361.
- DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. Journal of Management Information Systems, 19(4), 9-30.
- Deming, W. E. (1982). Out of the crisis: Cambridge: Massachusetts Institute of Technology Press.
- Faridi, A. S., & El-Sayegh, S. M. (2006). Significant factors causing delay in the UAE construction industry. Construction Management & Economics, 24(11), 1167-1176.
- Hjelt, M., & Björk, B.-C. (2007). End User Attitudes towards EDM Use in Construction Project Work - A Case Study. Journal of Computing in Civil Engineering, 21(4), 289-300.
- Juran, J., & Goferey, A. B. (1999). Juran's Quality Handbook (5 ed.): McGraw-Hill, New York.
- Lim, C. S., & Mohamed, M. Z. (2000). An exploratory study into reccuring construction problems. International Journal of Project Management, 18(3), 267-273.
- Ling, F. Y. Y., Pham, V. M. C., & Hoang, T. P. (2009). Strengths, Weaknesses, Opportunities, and Threats for Architectural, Engineering, and Construction Firms: Case Study of Vietnam. Journal of Construction Engineering and Management, 135(10), 1105-1113.
- Long, N. D., Ogunlana, S., Quang, T., & Lam, K. C. (2004). Large construction projects in developing countries: a case study from Vietnam. International Journal of Project Management, 22(7), 553-561.
- Madnick, S. E., Wang, R. Y., Lee, Y. W., & Zhu, H. (2009). Overview and Framework for Data and Information Quality Research. Journal of Data and Information Quality, 1(1), 1-22.
- O'Neill, S., Scott, M., & Conboy, K. (2009). What's technology got to do with it? A Delphi study on collaborative learning in distance education. Paper presented at the 17th European Conference in Information Systems, Italy, http://www.ecis2009.it/papers/ecis2009-0134.pdf.
- Ogunlana, S. O., Promkuntong, K., & Jearkjirm, V. (1996). Construction delays in a fast-growing economy: comparing Thailand with other economies. International Journal of Project Management, 14(1), 37-45.
- Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool : an example, design considerations and applications. Information and Management, 42(1), 15-29.
- Sambasivan, M., & Soon, Y. W. (2007). Causes and effects of delays in Malaysian construction industry. International Journal of Project Management, 25(5), 517-526.
- Schmidt, R. C. (1997). Managing Delphi Surveys Using Nonparametric Statistical Techniques. Decision Science, 28(3), 763-774.
- Sekine, K., & Arai, K. (1994). Design Team Revolution: How to Cut Lead Times in Half and Double Your Productivity: Productivity Press, Portland, OR.

Teo, T. S. H., & Wong, P. K. (1998). An Empirical Study of the Performance Impact of Computerization in the Retail Industry. Omega, 26(5), 611-621.

- Toor, S.-U.-R., & Ogunlana, S. O. (2008). Problems causing delays in major construction projects in Thailand. Construction Management & Economics, 26(4), 395-408.
- Wand, Y., & Wang, R. Y. (1996). Anchoring Data Quality Dimensions in Ontological Foundations. Communications of the ACM, 39(11).
- Wang, R. W., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems, 12(4), 5-33.
- Wantanakorn, D., Mawdesley, M. J., & Askew, W. H. (1999). Management errors in construction. Engineering Construction & Architectural Management, 6(2), 112-120.
- Xue, X., Wang, Y., Shen, Q., & Yu, X. (2007). Coordination mechanisms for construction supply chain management in the Internet environment. International Journal of Project Management, 25(2), 150-157.

Data and Information Quality Research in Engineering Construction Projects: A Review of Literature

Soffi Westin

University of Agder, P.O. Box 422, 4604 Kristiansand, Norway Soffi.Westin@uia.no

Abstract

This article presents a review of the research on Data and Information Quality (DQ/IQ) assessment in engineering construction. Through a review of 445 articles on the topic, only nine were found in the context of engineering construction. The analysis of these nine articles revealed six challenges in performing DQ/IQ assessment in this context; the iterative nature of concurrent engineering, the uniqueness of engineering data, lack of integration between processes, lack of integration between systems, lack of timely information, and lack of relevant DQ/IQ assessment frameworks and tools. The specific contributions of this paper are the identification of DQ/IQ challenges in engineering construction, their consequences, and implications for the development of relevant DQ/IQ assessment frameworks and tools. Additionally, it also identifies an area where Information Systems research can contribute, thereby extending the reach of the discipline.

1 Introduction and related research

Within the field of Information Systems (IS), Data and Information Quality (DQ/IQ) is a familiar concept. This paper presents a review of the research on DQ/IQ assessment in the context of engineering construction. The following presentation of related research in the areas of engineering construction and DQ/IQ, will also provide the motivation for this paper.

1.1 Engineering Construction

Delays and cost overruns are commonplace in construction projects (Long et al. 2004; Sambasivan and Soon 2007; Toor and Ogunlana 2008), and several researchers have reported delays as the most costly problem (Dyrhaug 2002; Faridi and El-Sayegh 2006; Ling et al. 2009). The problem has occurred in many countries, contributing to a high cost of construction (Memon et al. 2011). One example of the magnitude of these issues is from Norway, where a committee was appointed to investigate reasons for this problem in the offshore development projects on the Norwegian Continental Shell. The analysis showed 29.9 billion NOKs (\approx 5 billion USD) in cost overrun for 13 projects between 1994 and 1998 (Dyrhaug 2002). Project management issues (e.g., use of parallel activities) rather than technical issues were seen as the main reasons for the problem.

The effects of construction delays are not confined to the construction industry only, but could also influence the overall economy of a country. In the United Arab Emirates construction plays a major role and contributes 14% to the GDP (Gross Domestic Product) (Faridi and El-Sayegh 2006). In Norway, the Oil and Gas industry contributed 25% to the GDP in 2006 (Wolf and Pollitt 2009). Therefore, it is essential to define the most significant causes of delays in order to avoid or minimize their impact (Faridi and El-Sayegh 2006).

Several causes for delays and cost overruns have been identified, and errors in drawings remain a severe challenge. Toor and Ogunlana (2008) presented 75 problems encountered in a number of countries, including "errors and omissions in design documents". Dai, et al. (2009) identified errors in drawings as the second most significant problem, and lack of needed information related to drawings as another top ten factor in construction projects located in the United States. These findings are in line with a study by Tulacz (2005), which found construction workers commonly to have only 70% of the necessary data on-site. Westin and Päivärinta (2011) identified that DQ/IQ problems related to documents and drawings had the most negative impact on the profit margins of the projects executed by a European Engineering and Construction company. They also point out that the drawings are mainly extracted from engineering databases and conclude that the problems encountered with the drawings are related to the low level of DQ/IQ in the data sources.

1.2 Data and Information Quality

There is a lack of consensus on the definitions of data quality and information quality that distinguishes the two. Data quality often (but not always) refers to technical issues, while information quality usually refers to non-technical issues (Madnick et al. 2009). This paper does not distinguish, and uses the term DQ/IQ, adopting Wang and Strong's (1996) definition as "fitness for use" in relation to the recognized purposes of information use and user groups. In the IS literature,

2

the cost of an insufficient level of DQ/IQ has been found to be huge (Ramaswamy 2006; Strong et al. 1997), and problems have been observed across numerous corporations (Ramaswamy 2006; Wand and Wang 1996). Awareness of these issues has grown rapidly over recent years, and research is now covering several topics (Madnick et al. 2009) and contexts (Y. W. Lee 2004). Data warehousing, health care registers, and retailing- and internet-related systems (e.g., booking, online stores, etc.) are all contexts that have been studied to a certain extent. Several assessment frameworks exist and their similarities and differences have been compared by Batini, et al., (2009). Common for the mentioned contexts is the assumption of consistency and accuracy of the data records from the moment they are inserted, and most frameworks have been developed to support these assumptions.

To address the issue of persistence of DQ/IQ problems IS researchers have called for more investigation specifically in new and different contexts (Madnick et al. 2009). Researchers also called for providing practitioners with tools that can assess the level of DQ/IQ and provide comparative assessment over time to senior management (Pipino et al. 2002).

Based on the above discussion, the following literature review is focused on to two questions:

What are the challenges related to DQ/IQ assessment in the context of engineering construction?

What do these challenges imply for the development of DQ/IQ assessment frameworks and tools?

2 Research method

Literature reviews are conducted for a variety of reasons; learning the breadth of research in a topic is one of them (Okoli and Schabram 2010). The purpose of this literature review was to identify DQ/IQ research performed in the context of engineering construction projects; hence, one focus was the breadth of research on this topic. The method chosen was based on the eight-step guideline proposed by Okoli & Schabram (2010) and the concept matrix suggestions by Webster & Watson (2002).

To cover a broad range of fields three literature databases; EBSCO, Inderscience publishers, and Scopus were chosen. These databases include fields such as engineering construction, project management and IT/IS. Scopus alone includes 428 engineering journals of which several are ranked amongst the top journals (e.g., 'Journal of Construction Engineering and Management' and 'Automation in Construction').

To focus specifically on IT/IS 18 publication outlets were chosen for searches based on two criteria: 1) journals and conferences frequently referred to in DQ/IQ literature, and, 2) significant journals and conferences in the IS field. These outlets are:

- ACM Computing Surveys
- AMCIS (Americas Conference on Information Systems)
- CAIS (Communications of the Association for Information Systems)
- ECIS (European Conference on Information Systems)
- European Journal of Information Systems
- ICIQ (International Conference on Information Quality)
- ICIS (International Conference on Information Systems)
- Information Systems Journal
- Information Systems Research
- International Journal of Information Quality
- Journal of Data and Information Quality
- Journal of Information Technology
- Journal of Management Information Systems
- Journal of Strategic Information Systems
- Journal of the Association for Information Systems
- MIS Quarterly
- PACIS (Pacific Asia Conference on Information Systems)
- SJIS (Scandinavian Journal of Information Systems)

All search terms are listed in Table 2.1.

Table 2.1 Search terms

IT/IS outlets	All outlets (EBSCO, Inderscience
	Publishers, and Scopus)
"data quality"	
("data quality" AND "engineering")	("data quality" AND "engineering project")
("data quality" AND "construction")	("data quality" AND "construction project")
"quality of data"	
("quality of data" AND "engineering")	("quality of data" AND "engineering pro- ject")
("quality of data" AND "construction")	("quality of data" AND "construction pro- ject")
"information quality"	
("information quality" AND "engineering")	("information quality" AND "engineering project")
("information quality" AND "construction")	("information quality" AND "construction project")
"quality of information"	
("quality of information" AND "engineer-	("quality of information" AND "engineering
ing")	project")
("quality of information" AND "construc-	("quality of information" AND "construction
tion")	project")

An important criterion for the types of studies to include was that the search terms would be found in either abstract, title or keywords. All outlets were searched for all years. Some studies were rejected due to various reasons: misfit of context; a promising abstract in English, then the rest of the article available only in a foreign language (e.g., Chinese); and some "studies" were proposals. The remaining studies were submitted to full text analysis, after which several more studies were eliminated. Backward and forward (Bw/Fw) searches were then conducted based on the identified studies. The searches were performed from late July 2011 to September 2011.

3 Results

Table 3.1 displays the search result for the IT/IS outlets, the non IT/IS outlets, and a final Bw/Fw search.

Table 3.1 S	earch results
-------------	---------------

Search objects	# of unique	# of relevant
	hits	studies
IT/IS outlets	408	6
Non IT/IS outlets	37	2
Bw/Fw search	-	1
Total	445	9

In total for all search terms, 445 studies were found; however, after the practical screening only eight studies qualified as DQ/IQ research in the context of engineering construction projects. Bw/Fw searches resulted in the addition of one study. A total of nine studies fulfilled the requirements. All studies were peer reviewed, except from Dobson and Martinez (2007), which is an ABB review and not a scientific publication. However, practice could provide necessary information on the topic of this literature review, and this paper was included for that reason, as well as fulfilling the search criteria.

With the first research question in mind, these nine studies were analyzed with the aim to identify any reported challenges related to DQ/IQ assessment. Based on this analysis six challenges were identified. These challenges were: the iterative nature of concurrent engineering, the uniqueness of engineering data, lack of integration between processes, lack of integration between systems, lack of timely information, and, lack of relevant DQ/IQ assessment frameworks and tools in the context of engineering construction projects. Table 3.2 below lists the nine studies and the challenges they elaborated.

Article	Context details			Challe			
		The iterative nature of concurrent engineering	The uniqueness of engineering data	Lack of integration between pro- cesses	ack of integration between systems	Lack of timely information	Lack of relevant DQ/IQ assessment frameworks and tools
Y. Lee, et al., 2001	Assessed the process of remanufacture; an air-craft fuel pump			X	x		
Neely, et al., 2006 Dobson and Mar-	Assessed current DQ/IQ tools in the realm of engi- neering asset management Considerations on globali-	Х	X	X	х		Х
tinez, 2007	sation in plant engineering (large & global engineer- ing projects)	Х			Х		
Lin, et al., 2007 Lin, et al., 2008	Assessed engineering asset management in Australian engineering organisations Explored DQ/IQ issues in	Х	X	Х	X		X
	engineering asset man- agement (AM) and devel- oped and tested a frame- work for AM	Х	X	Х	X		X
Blechinger, et al., 2010	Proposed how to support production of high-quality data in concurrent plant engineering	Х	X	Х	х	Х	Х
Wingkvist, et al., 2010 Tribelsky and	Assessed quality in tech- nical documentation of: -a mobile phone -parts of a warship Assessed information						X
Sacks, 2011 Westin and	flows in 14 civil engineer- ing projects, all part of a major airport construction project Ranked problems experi-	Х		Х		X	
Päivärinta, 2011	enced in engineering construction projects with- in one organisation (off- shore installations)	Х		Х		X	

Challenge 1: The iterative nature of concurrent engineering

The globalization of engineering leads to shorter delivery schedules and force a break in the traditional linear engineering model (Dobson and Martinez 2007). The resulting iterative nature of concurrent engineering creates challenges for DQ/IQ assessment. Many of the processes in concurrent engineering are interdependent. This fact forces the engineers to proceed with partial information, incomplete knowledge and subjective interpolations (Blechinger et al. 2010). Late changes in requirements or changes caused by design errors – erroneous or missing information – lead to rework (Tribelsky and Sacks 2011). Rework initiates more iteration, which, in turn, leads to delays. Delays – in terms of waiting for information – reduce the profit margins of the projects (Westin and Päivärinta 2011). To further complicate the situation, the engineering asset management process, which consists of several interdependent stages, is not only iterative, but also requires collaboration with other processes in the organization, such as Supply Chain Management and Supplier Relationship Management (Shien Lin et al. 2007; Neely et al. 2006).

Challenge 2: The uniqueness of engineering data

One of the unique characteristics of engineering data is that the "correct answer" for the values to be inserted into the engineering asset data record is not known at the time of insertion (Neely et al. 2006). Physical engineering assets and the related asset data need not necessarily be present at the same time. For example, in engineering design, engineering assets may exist only as referenced logical/functional elements in data registers and data models. In addition, data are collected in a variety of formats and have to be shared among various technical and business systems. Some of the data is textual and, as such, unverifiable by the current automated processes (Shien Lin et al. 2007; Neely et al. 2006). This is different from typical business environments, where data often is provided in fixed formats and shared only among relevant business systems (Shien Lin et al. 2007).

Challenge 3: Lack of integration between processes

Lack of integration between processes leads to difficulties in connecting various information (Y. Lee et al. 2001). This could lead to a time-consuming search for existing but not easily accessible information. Tribelsky and Sacks (2011) identified a positive correlation between the quality of information flows and the effectiveness of design documents (e.g., less rework due to more accurate drawings) and found that unstable information flows were associated with unpredictable project outcomes. Westin and Päivärinta (2011) identified a similar problem; one of the top-ranked problems reflected missing information flows between teams contributing to the projects, which led to drawings that were not congruent. This problem was also identified by Blechinger et al. (2010), who indicated the importance of notifying users of work-copies when changes in the master Information Management Systems (IMS) occurred. Data captured in a variety of formats, processed in isolation, and stored in legacy systems is usually process-dependent, which makes it difficult to reuse data in other processes (Neely et al. 2006). Management

of all engineering asset data processes has to be performed across all significant lifecycle stages, and there is a need for synchronization of asset lifecycle management with other enterprise processes (Shien Lin et al. 2007).

Challenge 4: Lack of integration between systems

Several DQ/IQ problems stem from a lack of integration between systems: different values and different numbers of replicated items in different Information Management Systems (Blechinger et al. 2010); data that cannot be exported because they are captured with sensors only readable by specially designed monitoring systems (Neely et al. 2006); specialized systems bought from multiple vendors leading to a very difficult integration job for the end-users (Shien Lin et al. 2007); and lack of integration between business systems and technical systems making it difficult to achieve a comprehensive overview of status, resulting in significant DQ/IQ consequences.

Challenge 5: Lack of timely information

Lack of timely information is identified as a major challenge in three of the studies (Blechinger et al. 2010; Tribelsky and Sacks 2011; Westin and Päivärinta 2011). The resultant waiting causes delays in progress of the project (Blechinger et al. 2010) and leads designers to shift attention to other projects, which in turn leads to extra time spent on recapitulation when work on the project is resumed (Tribelsky and Sacks 2011). A delay in receiving necessary information causes the design to be based on incomplete information (Westin and Päivärinta 2011), which in turn could lead to rework. This is similar to what happens when engineers are forced to proceed with partial information due to process interdependencies and short delivery schedules, as mentioned earlier.

Challenge 6: Lack of relevant DQ/IQ assessment frameworks and tools

Existing DQ/IQ solutions and tools focus on typical business systems, such as financial systems, customer databases, and CRM systems (Neely et al. 2006). These tools can be categorized as 1) auditing tools, which aim to create a variance report where data that do not conform to the business rules can be manually examined; 2) cleansing tools, which aim to automatically verify and correct data; and 3) migration tools, which physically move data from one location to another (Neely et al. 2006). These tools are primarily directed at the quality dimension of accuracy; hence, the "correct answer" has to be known in order to perform the comparison of inserted data with real world data. However, this is not sufficient for DQ/IQ assessment in engineering construction (Blechinger et al. 2010; Shien Lin et al. 2007; Neely et al. 2006). A more relevant framework has been developed and tested by S. Lin et al. (2008). Their framework contributes to the lifecycle of engineering asset management (AM), and addresses the challenges identified from an AM lifecycle point of view.

4 Discussion

In order to perform DQ/IQ assessment involving large amounts of data, a relevant tool is needed. This review reveals that most existing DQ/IQ assessment frameworks and tools are developed on the assumptions that the "correct answer" is known by the time of data insertion, which is not the case in engineering construction (Neely et al. 2006). The AM framework (S. Lin et al. 2008), represents an exception. The framework considers the identified challenges and contributes extensively to the AM part of engineering construction. However, the poor level of DQ/IQ in drawings mentioned in the introduction is related to engineering design as a whole (not solely AM). Hence, engineering construction projects are still in need of relevant DQ/IQ assessment frameworks and tools. The question that arises is: which concepts need to be included in such a framework to make it suited to this specific context?

When analyzing the findings of this review, an interesting issue emerged: some of the identified challenges seem to be unavoidable in engineering construction. Therefore, an assessment tool should be developed on the assumption that these challenges are unavoidable and their consequences need to be handled appropriately. Below is a discussion on which identified challenges are unavoidable, and associated implications for the development of relevant DQ/IQ assessment frameworks and tools.

The iterative nature of concurrent engineering is unavoidable. This is also the case for the uniqueness of engineering data. Lack of integration between processes and lack of integration between systems are core foci for an IS researcher. However, it is far too risky to change processes and systems within a single project (Blechinger et al. 2010). These changes and alignments are better done at a higher, organizational level. The implication is that within a single project these challenges are also unavoidable. Apart from the identified lack of relevant DQ/IQ frameworks and tools, only one challenge remains: lack of timely information. Assuming several reasons exist for *why* information is lacking, *elements* of the answer could include the relationship between the identified challenges. This is elaborated below:

1) The iterative nature of concurrent engineering forces the engineers to proceed with partial information (Blechinger et al. 2010). This implies a lack of necessary information. 2) The uniqueness of engineering data is characterized by the fact that the "correct answer" is not always known at the time of data insertion (Neely et al. 2006). This means the information is not there when first needed, (i.e. there is a lack of timely information). 3) Lack of integration between processes could lead to time-consuming searches for existing but not easily accessible information (Y. Lee et al. 2001), which could also lead to lack of timely information. 4) Lack of integration between systems could lead to difficulties with exporting data (Neely et al. 2006), and the resulting delay would again contribute to lack of time-

ly information. Taken together, it appears that lack of timely information could also be viewed as unavoidable, at least in some situations.

What are the consequences then? Since these four challenges could all lead to lack of timely information, engineers may have to proceed with only partial information (Blechinger et al. 2010). The consequences will be incompleteness and/or inconsistencies in information used in the projects. Research findings indicate that engineering drawings are mainly extracted from such systems (Westin and Päivärinta 2011). It was also found that errors in drawings comprise one of the most significant problems related to productivity factors (Dai et al. 2009). Hence, proceeding with partial information will have negative impact on the level of DQ/IQ in the data sources and the extracted drawings. The connections between the different challenges and the consequences are summed up in Figure 4.1, where dotted arrow means "could lead to" and solid arrows mean "leads to".

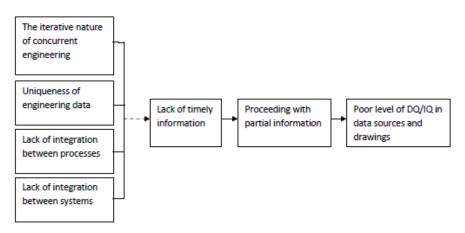


Fig. 4.1 Connections between identified challenges and level of DQ/IQ

A DQ/IQ framework has to include concepts and guidelines that consider these issues and help overcome the challenges.

5 Implications and conclusion

This review has identified several challenges for DQ/IQ in the engineering construction context. The key challenge now is to develop a relevant assessment framework and a tool to implement that framework. In doing so, it is important that such tools have to diverge from the common assumption in existing DQ/IQ research (i.e. that the "correct answer" is known at the time of data insertion). In the present context, the assumption is the opposite: the "correct answer" is not always known. This has some implications: 1) DQ/IQ assessment frameworks and tools have to allow for partial information and, consequently, incompleteness and inconsistencies in the data sources during early phases of engineering construction projects, and 2) thereafter, further guidelines would have to suggest a way to mitigate the unavoidable poor level of DQ/IQ in later project phases.

Developing a DQ/IQ assessment tool requires expertise in the domain of engineering construction. That requires close collaboration between the researcher and the engineers; hence, access to an engineering construction organization is needed. Since the goal is to develop a tool, suitable methods could be Design Research (DR), or Action Design Research (ADR). Both methods emphasize close collaboration between researchers and organizations. DR posits that it is through the building and application of such a tool (the artifact) that the problem domain is understood and its solution achieved (Hevner et al. 2004). ADR emphasizes that such a tool (the artifact) emerges from interaction with the organizational context (Sein et al. 2011). No matter which method is chosen, engineering construction is in need of the development of a DQ/IQ assessment tool. Finally, the surprising finding that only 9 papers are found on this topic is in itself intriguing. This calls for more research in this area. IS research can contribute by building up a research agenda in DQ/IQ in engineering construction. Findings from general DQ/IQ research in IS will be very useful for this agenda.

References

- Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. ACM Computing Surveys, 41(3), 1-52.
- Blechinger, J., Lauterwald, F., & Lenz, R. Supporting the Production of High-Quality Data in Concurrent Plant Engineering Using a MetaDataRepository. In Americas Conference on Information Systems, August 12 -15, Lima, Peru, 2010
- Dai, J., Goodrum, P. M., & Maloney, W. F. (2009). Construction Craft Workers' Perceptions of the Factors Affecting Their Productivity. *Journal of Construction Engineering & Management*, 135(3), 217-226, doi:10.1061/(asce)0733-9364(2009)135:3(217).
- Dobson, D., & Martinez, R. (2007). Integrated engineering. ABB Review(SPEC. REP.), 48-52.
- Dyrhaug, Q. (2002). A generalized critical success factor process model for managing offshore development projects in Norway. Doctoral thesis, Norwegian University of Science and Technology (NTNU), Trondheim, Norway,
- Faridi, A. S., & El-Sayegh, S. M. (2006). Significant factors causing delay in the UAE construction industry. *Construction Management & Economics*, 24(11), 1167-1176.
- Hevner, A. R., March, S. T., Jinsoo, P., & Ram, S. (2004). Design Science in Information Systems Research. MIS Quarterly, 28(1), 75-105.
- Lee, Y., Allen, T., & Wang, R. Information products for remanufacturing: tracing the repair of an aircraft fuel-pump. In *International Conference on Information Quality*, 2001 (pp. 77-82): Citeseer

- Lee, Y. W. (2004). Crafting rules: Context-reflective data quality problem solving. Journal of Management Information Systems, 20(3), 93-119.
- Lin, S., Gao, J., & Koronios, A. (2008). A data quality framework for engineering asset management. Australian Journal of Mechanical Engineering, 5(2), 209-219.
- Lin, S., Gao, J., Koronios, A., & Chanana, V. (2007). Developing a data quality framework for asset management in engineering organisations. *International Journal of Information Quality*, 1(1), 100-126.
- Ling, F. Y. Y., Pham, V. M. C., & Hoang, T. P. (2009). Strengths, Weaknesses, Opportunities, and Threats for Architectural, Engineering, and Construction Firms: Case Study of Vietnam. Journal of Construction Engineering and Management, 135(10), 1105-1113.
- Long, N. D., Ogunlana, S., Quang, T., & Lam, K. C. (2004). Large construction projects in developing countries: a case study from Vietnam. *International Journal of Project Management*, 22(7), 553-561.
- Madnick, S. E., Wang, R. Y., Lee, Y. W., & Zhu, H. (2009). Overview and Framework for Data and Information Quality Research. *Journal of Data and Information Quality*, 1(1), 1-22.
- Memon, A. H., Abdul Rahman, I., & Abdul Azis, A. A. (2011). Preliminary Study on Causative Factors Leading to Construction Cost Overrun. *International Journal of Sustainable Construction Engineering and Technology*, 2(1).
- Neely, M. P., Lin, S., Gao, J., & Koronios, A. The Deficiencies of current data quality tools in the realm of engineering asset management. In *Americas Conference on Information Systems, August 4 - 6, Acapulco, México, 2006*: Proceedings of the Twelfth Americas Conference in Information Systems
- Okoli, C., & Schabram, K. (2010). A Guide to Conducting a Systematic Literature Review of Information Systems Research. Sprouts: Working Papers on Information Systems, 26(10), 1-49.
- Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data Quality Assessment. Communications of the ACM, 45(4ve), 211-218.
- Ramaswamy, M. (2006). On the phenomenon of information dilution. *Issues in Information Systems*, VII(2), 289-292.
- Sambasivan, M., & Soon, Y. W. (2007). Causes and effects of delays in Malaysian construction industry. *International Journal of Project Management*, 25(5), 517-526.
- Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action Design Research. MIS Quarterly, 35(1), 37-56.
- Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. Communications of the ACM, 40(5), 103-110, doi:10.1145/253769.253804.
- Toor, S.-U.-R., & Ogunlana, S. O. (2008). Problems causing delays in major construction projects in Thailand. *Construction Management & Economics*, 26(4), 395-408.
- Tribelsky, E., & Sacks, R. (2011). An empirical study of information flows in multidisciplinary civil engineering design teams using lean measures. Architectural Engineering and Design Management, 7(2), 85-101.
- Tulacz, G. J. (2005). Worrying about labor shortage. Engineering News Record, Nov. 14, 2005, (www.enr.com) (Sep. 29, 2011).

- Wand, Y., & Wang, R. Y. (1996). Anchoring Data Quality Dimensions in Ontological Foundations. Communications of the ACM, 39(11), 86-95.
- Wang, R. W., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 12(4), 5-33.
- Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. [Editorial]. *MIS Quarterly*, 26, xiii-xxiii.
- Westin, S., & Päivärinta, T. Information quality in large engineering and construction projects: A Delphi Case Study. In European Conference on Information Systems, June 9 - 11, Helsinki, Finland, 2011
- Wingkvist, A., Ericsson, M., Lincke, R., & Löwe, W. A metrics-based approach to technical documentation quality. In *International Conference on Information Quality*, 2010 (pp. 476-481): IEEE
- Wolf, C., & Pollitt, M. (2009). The Welfare Implications of Oil Privatization: A Cost-Benefit Analysis of Norway's Statoil. *Cambridge Working Paper in Economics No. 0907*, *EPRG Working Paper No. 0905*.

The Design and Emergence of a Data/Information Quality System

Soffi Westin University of Agder, Norway soffi.westin@uia.no

Maung K. Sein*

University of Agder, Norway and Luleå Teckniska Universitet, Sweden *maung.k.sein@uia.no*

Abstract. Construction engineering is characterized in the literature as a field where the consequences of poor data quality and information quality (DQ/IQ) is particularly severe. Yet, the field lacks sufficient DQ/IQ assessment frameworks and tools. To address this shortcoming, we applied an action design research (ADR) approach to develop and implement a DQ/IQ assessment tool. The multi-year research project took place in a European construction engineering company, and lasted from 2007 to 2012. We drew upon insights from the literature on DQ/IQ assessment and related challenges in construction engineering, as well as practical lessons learned from managing DQ/IQ in the target organization. Our research contributes to knowledge by developing a set of design principles. We also contribute to practice by solving an actual problem. Finally, we also contribute to the method space by discussing possible modification to the ADR method.

Key words: data quality, information quality, construction engineering, action design research.

1 Introduction

The problem came to light when site managers in a global construction engineering company started complaining to project managers that drawings that the assembly sites were receiving from the detailed engineering design section contained errors. Since the actual physical construction starts in these sites with assembly of the numerous physical parts (also known as assets) and since the assembly was done based strictly on these drawings, any inaccuracy in the drawings significantly delayed the rest of the project. The company (anonymized as European multi-discipline construction engineering company – EUMEC) operates in the oil and gas sector and has assembly sites all over the world.

Project managers reacted to these complaints in the time honoured manner: they sent more people to the assembly sites to try to work out the problems. These work hours represented cost overruns, and the problems eventually delayed the handover of the installations to the customers. A typical project undertaken by EUMEC takes up to 3 years and costs in the region of 100 million Euros. The impact of delays in the project can be huge.

Alarmed at this trend, EUMEC started a number of discussions at several levels in the period 2007-2008. Many of these meetings were cross-departmental. Information managers (IM) from the IT/IS department attended engineering meetings where these problems were referred to as "problems on site", "problems with equipment interfaces", and "problems with finding the right parts and where they belonged". The first author of this paper was one of the IMs. To reflect this first hand engagement in the project, we will write the rest of the paper in first person plural.

We studied the existing literature on delays and cost overruns in construction engineering projects and found that these problems were quite prevalent. Amongst several reasons indicated were errors and omissions in drawings (Rivas et al. 2010; Toor and Ogunlana 2008). Slowly, it dawned on us that these errors could have something to do with the low data quality in the engineering data bases. After all, the drawings in the target organization were generated precisely on these data. However, we could not yet devote our attention to examining this hunch. Our regular work load was high, and much of it was spent in handling day-to-day problems on ongoing projects. Projects are committed to milestones with deadlines for deliveries. Penalties and bonuses were tied to these milestones. It was thus very important that the projects reach these milestones on the agreed date, fulfilling the agreed delivery requirements. It was quite understandable that management prioritized fire-fighting.

This management priority left little time for us to investigate whether our suspicion about the relationship between data quality and errors in engineering drawings was correct. However, the management mandate to reduce the need for fire-fighting also serendipitously handed us the key to turn the focus back on the data quality issue. We started an investigation on the matter, studied documentation of requirements and assessed the level of several completed projects. Our findings appeared to support our insight that the main reason for the delays in the projects was in fact related to the level of DQ/IQ in the engineering data bases. Some of these issues could actually be avoided had we the time to perform some serious assessment of the data quality at an earlier stage of the projects. There was an urgent need to have in place a mechanism or tool to assess the data quality at EUMEC. The company was using a tool, but the IM department found this tool to be inadequate.

At this juncture the IM who is the first author of this paper entered a doctoral program and the search for an appropriate tool became her dissertation work. Originally framed as a prototypical Design Research project (Hevner et al. 2004), its later stages were conducted as an Action Design Project (Sein et al. 2011). The resultant artifact was a new tool we called Information Quality System or IQS in short. In keeping with the principles of ADR, IQS represents a class of problems, namely, Data/Information Quality in construction engineering projects.

In the rest of this paper, we tell the story of the design, emergence, evaluation and implementation of IQS. First, we present a brief review of DQ literature to introduce the knowledge gap that we seek to bridge in this paper. Next, we describe the construction engineering process, specifically situating it in EUMEC to help the reader understand the context of our research. We then interpret and analyze our project through the lens of design research and action design research approaches. As we stated above, IQS is cast as a class of data quality problem: we show how our project generates design rules for this class of problem. Following this analysis, we present the theoretical contributions of the project and end the paper with a reflection on the practical and theoretical implications.

2 The Research Problem: Data Quality

DQ/IQ has been defined in various ways. Data Quality often (but not always) refers to technical issues, while Information Quality usually refers to non-technical issues (Madnick et al. 2009). For our purpose we do not distinguish, but adopt the "fitness of use" perspective (Wang and Strong 1996)which is based on the intended purposes of the users of information.

In the IS literature, the cost of an insufficient level of DQ/IQ has been found to be very high (Ramaswamy 2006; Strong et al. 1997). This problem has been observed across numerous organizations (Ramaswamy 2006; Wand and Wang 1996). Awareness of these issues has grown rapidly over recent years. DQ/IQ research has moved on from technical issues as query techniques on multiple data sources and data warehouses in the 1980s, to a number of new application areas, such as knowledge management (Madnick et al. 2009)and health care registers (e.g. Pipino and Lee 2007; Vician 2011).

To address the issue of persistence of DQ/IQ problems IS researchers have called for more investigation specifically in new and different contexts (Madnick et al. 2009). Researchers have also called for providing practitioners with tools that can assess the level of DQ/IQ and provide comparative assessment over time to senior management (Pipino et al. 2002).

Recently the attention of DQ/IQ researchers has turned to the context of construction engineering (e.g. Lin et al. 2008; Tribelsky and Sacks 2011). Construction engineering is a field where the consequences of poor DQ/IQ are particularly severe. Even if the body of DQ/IQ literature includes several assessment frameworks (e.g. English 1999, 2003; Lee et al. 2002; Pipino et al. 2002), these frameworks and tools cannot be used for sufficient assessment in this field. Their focus is on typical business systems and primarily on accuracy of the data (Neely et al. 2006). This means the "correct answer" has to be known in order to perform the comparison of inserted data with real world data. This is an assumption not valid for construction engineering (ibid). Hence, construction engineering is in need for assessment frameworks and tools adapted to their specific situation.

Despite the mentioned shortcomings, existing frameworks can nevertheless provide useful insight on processes and principles related to DQ/IQ assessment in construction engineering. Our search through the literature identified two frameworks that are particularly germane to our research, namely, (a) Total Information Quality Management (TIQM) proposed by English (1999, 2003), and, (b) Data Quality Assessment (DQA) proposed by Pipino et al. (2002). Next, we describe these frameworks.

2.1 TIQM

TIQM, as proposed by English (1999, 2003) is a methodology for exploiting quality tools and techniques. As such, it is a means to assess existing tools for deficiencies, and, to build new tools based on the implicit requirements embedded in TIQM processes. The framework is based on principles, techniques and processes used for quality management in general, which English considers equally useful for information quality management.

The framework encompasses six processes. The first three processes are on assessment, the fourth is on data correction, and the fifth is an improvement process to prevent recurrence of errors. Finally, the sixth process is an overall process of culture transformation which aims at establishing a good environment for DQ/IQ improvement over time.

The key to success is the acceptance of DQ/IQ management and improvement as a continuous process. This is emphasized since continuous process improvement sometimes is confused with continuous data cleansing. The latter will not provide any opportunity for correcting the root cause of errors, hence, the process will not improve. Another important aspect is measuring accuracy which in reality can only be done by comparing the data to the physical object. Use of surrogate sources will only reflect accuracy to the extent the surrogate source is considered accurate. For construction engineering design, accuracy is almost impossible to measure since the physical object is often not yet produced. It is the consequences of this problem we aimed at mitigating.

2.2 DQA

The Data Quality Assessment (DQA) framework was developed by Pipino et al. (2002) to assess and measure the level of data quality in a company. The framework consists of a list of data dimensions, a set of objective and subjective measures, and a suggestion on analyzing and comparing these measures. The methodology distinguishes between *subjective* and *objective* quality metrics.

Subjective data quality assessments are tied to the needs and experiences of the different stakeholders, such as collectors, custodians, and consumers. How stakeholders perceive the quality of data influence their behavior. Objective data quality assessment can be either task dependent or task independent. Task-dependent metrics include business rules, regulations and policies, and constraints provided by the database administrator – all of which are specific to the given context. Task-independent metrics reflect states of data without any contextual knowledge, and can as such be applied to any data set.

DQA proceeds as follows. Subjective assessment surveys are conducted and objective assessment metrics are executed against the data set at stake, and the results are compared. Any discrepancies are then analyzed in an effort to determine the root causes of these discrepancies. When the causes are determined, actions are taken to improve the level of data quality, and new comparisons are made. This process is repeated continuously. The most difficult task when trying to measure and compare outcomes is defining the different data quality dimensions that aligns with the context at stake. Formulating the metrics is considered fairly easy once the quality dimensions are defined. Three functional forms of developing metrics are suggested: Simple Ratio for measuring desired outcomes to total outcomes, Min or Max Operation for dimensions requiring aggregation of multiple data quality variables, and, Weighted Average as an alternative for multivariate cases.

In addition, Pipino et al. mention the possibility for an industry to create a de facto standard by adopting a set of data quality metrics. This will provide an opportunity to determine the level of DQ/IQ and to compare the quality levels with other industries.

3 The Research Context: Construction Engineering and the Target Organization

In this section we first provide a brief overview of the nature and challenges of construction engineering projects in general. Then we focus more specifically on EUMEC and how the organization carries out its projects.

3.1 The Nature of Construction Engineering Projects

Large construction engineering projects are amongst other things characterized by the large number of people involved, the need for different kinds of expertise, and huge costs. The projects are complex and challenging to manage (Franco et al. 2004). In addition, short delivery schedules have resulted in a break in the traditional linear engineering model which lead to the current iterative nature of concurrent engineering (Dobson and Martinez 2007). This causes challenges for DQ/IQ management because while the processes involved are interdependent, the tasks that previously were executed in sequences are now executed in parallel. The result is that the engineers are forced to proceed with incomplete information (Blechinger et al. 2010). Another serious challenge is that existing DQ/IQ assessment frameworks and tools are based on the assumption that "the correct answer" is known by the time of data insertion to a record. This assumption makes these frameworks insufficient for use in construction engineering literature; five of the most relevant challenges have been identified in a review of the construction engineering literature; five of the most relevant challenges listed on the left most side are invariant contexts of construction engineering. These lead to lack of timely information. Faced with delays, and the pressure of short delivery schedules (Dobson and Martinez 2007),

engineers are forced to proceed with incomplete information (Blechinger et al. 2010) which leads to poor DQ/IQ.

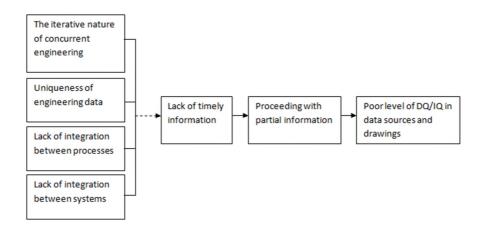


Figure 1: Connections between challenges and level of DQ/IQ (Westin, 2012)

In carrying out its projects, EUMEC follows a generic process that is the norm in the construction engineering industry. In the following we describe this process.

3.2 Engineering projects at EUMEC

EUMEC is a European, multi-discipline construction engineering company. It operates in the oil and gas industry and its core business is construction engineering design related to complex offshore installations such as oil rigs and drill ships, and, product development. The company delivers engineering design for construction projects and possesses a significant share of global markets in its product and project domains. Most of its employees are involved in construction engineering projects. The biggest projects may typically cost more than 100 million Euros and take up to three years of calendar time.

Engineering projects are parts of a construction project and usually cover design-related activities. Engineering projects have phases of their own that are usually described in some sort of project execution model (Dyrhaug 2002). Figure 2 depicts the particular model used at EUMEC.

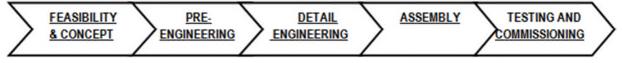


Figure 2: Project phases

In the Feasibility & Concept phase, alternative solutions are identified, and one or two design concepts are selected for further work. At the end of this phase one concept is chosen and some requirements related to this concept are frozen. In the System Definition phase, drilling processes and equipment concepts are frozen, together with layout and main structure. Towards the end of this phase, global design and equipment design are completed based on the chosen concept. In the Detail Engineering phase the design details are completed. The final deliverables of this phase mainly consist of documents, such as drawings and manuals. The drawings are either from suppliers or extracted from various data sources used in engineering design such as 2D/3D applications and engineering data bases. When the design is complete, these deliverables are handed over to the Assembly phase.

It is at this phase, Assembly, that the rubber hits the road, so to speak. Actual physical work, mainly mechanical in nature, is done here based on the drawings received from Detail Engineering. A test called Factory Acceptance Test (FAT) is performed and is used for checking if the equipment meets the requirements and is fully functional. The System Completion phase includes system assembly, commissioning

and close-out. "System" here means the entire construction which makes up a system of equipment, pipelines, valves, etc.

The most serious problems, especially delays, typically occur in these latter phases. The crux of the challenge for EUMEC, therefore, is to improve the data quality in the drawings handed over from detail engineering phase to assembly phase.

3.2.1 Project teams

Project teams at EUMEC are set up with members from organizational units called "engineering disciplines". Several experts from different engineering disciplines are needed to design large, complex, robust and yet delicate constructions. The exact composition of the teams varies depending on the requirements of the individual projects. Table 1 lists the engineering disciplines typically represented in the projects at EUMEC. Each engineering discipline is composed of several engineers and has a manager who is responsible for the delivery as a whole. Every engineering discipline depends on input from the other disciplines throughout the project, due to a large number of interfaces and dependencies among the artifacts and systems to be designed and assembled. Tight project schedules means that various activities must be performed in parallel. In such concurrent engineering (Sekine and Arai 1994), quality assurance and possible adjustments are therefore conducted both during the project and then, again, in the assembly and completion phases.

Disciplines	Responsibility
Process	Design of industrial processes; all the facts, sequences and relations in the process and a logical placing of the different items.
Mechanical	Design (choice of equipment and its physical layout and weight).
Piping/Layout	Design of all piping.
Electro	Design and cabling of power distribution for electrical systems: equipment, lights, heat, etc.
Instrument	Design of control systems, i.e. the control of various valves, machines, the alarm systems, and instrumentation cables for distributing signals.
Telecom	Selection and location of radio and audio systems, alarms etc.
HVAC (Heating, Ventilating and Air Conditioning)	Capacity calculations and layout for ventilation etc.
Safety Various safety assessments.	
Structure (steel)	Design of steel structures, supports, outfitting like hand rails, stairs etc.
Architecture	Interior design.

 Table 1: Engineering disciplines in EUMEC's construction projects (identifying reference)

4 Method

As we mentioned in the Introduction section, our research method was ADR which we followed implicitly in the earlier stages of the project and explicitly in the later stages. Figure 3 displays the stages and principles of ADR, while Figure 5 in the Discussion section displays the building, intervention, and evaluation cycles of the IQS project, and Table 8 also in the Discussion section lists the principles of ADR as well as how our project met those principles. In the rest of this section, we briefly describe the stages.

Problem Formulation: The problem is encountered by a target organization and/or acknowledged by researchers as a possibility to enhance scholarly knowledge. A research question can be defined during an initial empirical investigation of the problem. Casting the specific problem as an instance of a class of

problems moves the problem solving effort from a level of mere consulting to a level where new knowledge is generated.

Building, Intervention, and Evaluation (BIE): The initial design of the artifact is based on the premises of stage one. Through an iterative process which requires intervention in the organization, the artifact is further developed (build) and evaluated. Design principles are defined for a class of systems.

Reflection and Learning: This stage runs in parallel with the previous stages and provides the conceptual move from building a solution for a specific instance to applying that learning to a class of problems. This is where it becomes evident that the research process is more than solving a problem. To identify contributions to knowledge conscious reflection is critical. The learning from these reflections is used to adjust the research process accordingly. The adjustments are performed in a continuous manner as the understanding increases.

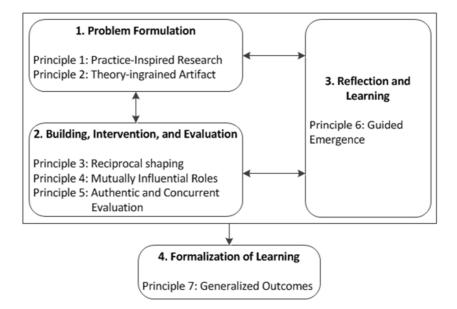


Figure 3: ADR Stages and Principles (from Sein et al., 2011)

Formalization of learning: The knowledge generated from designing the artifact is used for generalization by developing general solution concepts for a class of field problems. By describing how the artifact design was performed, the accomplishments realized and the organizational outcome, the researcher formalizes the learning. The design principles embedded in the artifact instance contribute to theory building for the class of systems.

5 The IQS Project

In this section, we will first describe the chronology of the development of IQS framing it as an ADR project. The initial phase was not explicitly conducted as ADR but can be retrospectively interpreted as such. The main reason for this is that ADR did not exist at that time and so we could not possibly start a project with that method. On reflection though, we realized that we had been implicitly following ADR although not in its complete form. After the publication of Sein et al. (2011), we were convinced that it was the right approach for IQS and decided to follow it in its entirety for the remaining phases.

Figure 4 shows the timeline of the IQS project with the main events keyed. In narrating the story of the project, we will refer to this figure and specifically the keys. We organize this section around the stages of ADR.

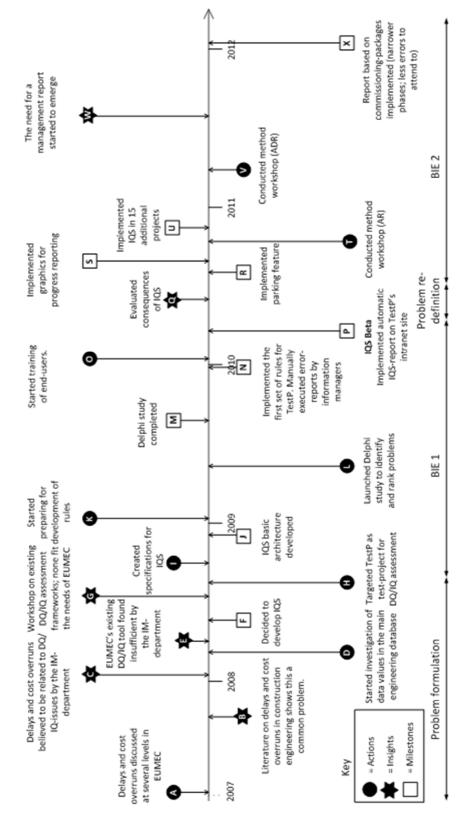


Figure 4: Timeline of the project

5.1 Problem formulation

Investigating the poor data quality in engineering databases: In the Introduction section, we described how the problem arose (pt. A), what existing literature indicated (pt. B), and how the cause was tracked down to poor DQ/IQ (pt. C). We began our investigation of how poor level of DQ/IQ affected the engineering drawings by exploring the data values in the main engineering data base (pt. D) This database featured a query tool for detecting missing values. Users could automatically check whether fields in a record contained a value. The intention was to identify blanks in fields that were subjected to internal or external (customer) requirements. An example would be to check if the site code field contained a value. Since site codes indicate where an item should be shipped for assembly, a value is required. We decided to take a closer look into the results of these checks because there was more information missing on the drawings than reported as missing by the query tool. We found that in addition to blanks, there was a widespread use of the so-called "straw values". These were characters and values totally unrelated to the definition of the expected field value. The most frequently used characters were "NA" and "-". When we asked the design engineers why they inserted these straw values, the three most common answers were (these comments were made in Norwegian which we have translated):

"The field is not subjected to any requirements of filling so I inserted one of those values ["NA" or "-"] to avoid the field from becoming a part of a "missing values report".

"I know the requirements state that a value is needed for this field, but I don't know the correct value yet. I am not sure whether to leave it blank in the meantime so I sometimes insert "NA" or "-"".

"I do not need to know this value to perform my job, so I simply insert "NA" or "-" to indicate the value is of no interest".

Management's perception of the situation is typified by the following quote:

"In the projects there is no clear philosophy for who are responsible for inserting values in the various data fields in the main engineering data base. At the management level we agree that somebody has to be the "father" of the information to be inserted, but the users responsible for the engineering assets, who we think should be those "fathers", say they have never been told to insert this information " (Engineering manager)

The outcome of our investigation can be summed up as follows. In situations where a data value had to be inserted in order to proceed, straw values were inserted. These were either a best-guess value or a random temporary value. Data values were also often omitted. In all cases, the intention was to insert the correct values as soon as they were known. On top of incomplete or meaningless straw values, our exploration revealed more problems with data quality. There were inconsistencies between different databases and lack of logical coherence between different fields in the same records of the main engineering data base.

The large number of records to assess for incorrect values was way too large to be handled manually, especially since several data sources were simultaneously in use (e.g. 3D-models and engineering systems). This, combined with the fact that the existing DQ/IQ tool could only check and flag missing values, made it extremely difficult to detect errors related to consistency or logical coherence.

The result was missing or incorrect values in the database and that led to incorrect drawings. These incorrect drawings were sent to the assembly site which created all manners of problems for the site workers. The drawings had to be corrected which entailed finding the missing or incorrect data first. The ultimate result was delays and cost overruns:

"The errors in the drawings are revealed on site and that leads to a lot of extra work" (Engineering manager)

Having found the nature of errors in the engineering database and their causes, we now turned our attention to the existing DQ/IQ tool in use at EUMEC. Our investigation revealed a number of its inadequacies and shortcomings which are summarized in Table 2, (pt. E).

Problem	DQ/IQ tool in use
Errors in drawings. The errors are related to completeness, consistency and logical coherence.	 Supported only searches for missing values Supported only assessment of the main engineering database
Rigid reporting style by allowing only analysis of predefined parameters	• Did not support project specific or changing requirements
"Correct answer" not always known by the time of data insertion resulting in insertion of straw values or blanks	• Did not detect or flag straw values as errors.

Table 2: Problems with existing DQ/IQ assessment tool

Decision to develop a DQ/IQ assessment tool: Based on the findings of our investigation described above, a decision was made to develop a tool from scratch (pt. F). The rationale to develop a new system is captured succinctly in the following quote:

"Now the best way to perform data quality checks is manual assessment. But then you have to know exactly what you are looking for, and the problem is that you don't really know. Also, manual checking is very time consuming. We need automatic assessment to cope with the huge amount of data and to always know the current status of the data quality" (Engineering discipline manager)

We now needed to find an existing DQ/IQ assessment framework that we could use to build the tool. It is here, that we ran into a roadblock: none of the existing frameworks met our needs (pt. G). New ground had to be broken. This coincided with the start-up of a new project (TestP) which we could use as a test project, (pt. H). We identified 3 main stakeholder groups: the information management (IM) department, engineering projects, and end-users. Together with the researchers, representatives from these three groups formed what we now realize as the ADR team. During the development of the system, the IM who is the first author of this paper started her doctoral work. The challenges of the innovative nature of the project transformed it into a research project framed as designing an innovative artifact to address a class of problems. As we stated earlier, the project started as a Design Research project and later evolved into an Action Design Research project. In the following, we describe in detail, this ADR project which developed the artifact called Information Quality System (IQS). This section described the first stage of ADR, namely, Problem Formulation. The next section describes the second stage, Build-Intervene-Evaluate (BIE). In this stage, the initial design of the artifact is generated and then further shaped by use and in subsequent design cycles (Sein et al. 2011). As we shall see, there were in fact two BIE cycles in our project.

5.2 BIE - Cycle 1 - Designing for error detection and reporting

An artifact developed through ADR is theory-ingrained: it carries "traces of theory" that guides its design and development (Sein et al. 2011). The precise nature of the theory remains a matter of debate. We followed Lempinen (2012) and Markus et al. (2002) and used the characteristics of construction engineering (Westin 2012) as the kernel theory. In addition, we used TIQM and DQA as guiding frameworks.

Characteristics 1: Construction engineering projects in the global contexts have short delivery times.

Characteristics 2: Construction projects have distinct interdependent phases which are conducted concurrently.

Characteristics 3: The correct data values for every field in engineering databases are most often not known to engineers at the time of insertion due to the iterative nature of concurrent engineering.

Characteristics 4: Engineers in a construction project have to proceed with incomplete and partial information.

Characteristics 5: Engineering databases are large and the data format is varied and stored in different media.

Requirement Specifications for the System: Based on the kernel theory, we were now ready to define the requirement specifications for the system (pt. I). These are described next.

Requirement Specification 1: IQS should be rule based

This was the main requirement. The rationale for using a rule-based approach is as follows:

- It is possible to capture requirements in rules, which also makes it possible to identify errors.
- It is possible to reuse rules in other projects because some rules can be applied to all projects.
- The rules are fairly easy to program once the requirements are understood by everybody. Development of rules also helps clarifying the requirements.
- The rules are easy to change or edit since they are located outside the system source code. This is important for EUMEC because changes in requirements are almost inevitable during projects. Adjustments, changes, and additions to the construction-design require immediate on-site amendments to the rules throughout a project. Even data sources may be added, or customer systems changed, which could affect the rules.
- The rules can be viewed as business knowledge in a way that makes it possible to create a set of default rules applicable to all project as well as project-specific rules. For EUMEC this is important because it provides the employees with a knowledge base that can be transferred between projects and also used for training.
- As an added "bonus" using rules makes it easier to estimate the cost of an Information Manager in a specific project. This is because rules capture the project requirements and thus the number of required rules indicates the amount of time an IM is estimated to spend on a project.

Requirement Specification 2: *IQS rules should address three DQ/IQ dimensions: completeness, consistency and logical coherence. Thus, rules should identify which data could possibly be missing, which data could possibly be inconsistent, and which data could lack logical coherence.*

"Quality dimensions" is an important aspect of DQ/IQ assessment and they represent groups of data values with the same qualities (e.g. accurate values or consistent values). Identifying quality dimensions made it easier to define and discuss issues related to DQ/IQ without referring to specific data values.

Requirement Specification 3: *IQS should provide different rules for different phases of the project with increasingly more details to be checked at every subsequent phase.*

Due to the huge amount of data needed and the fact that much of the data would not be available at the start of the project, a full check on every requirement from day 1 would result in an unwieldy number of errors. The solution was to define different rules for use at different phase of the project. By December 2008 the requirement specifications were ready. During Christmas holidays the basics of the IQS architecture were developed (pt. J). The programmer stated that "the specifications were so detailed that the programming itself did not take very much time".

Developing rules: In January 2009 we started to prepare for development of rules according to TestP's requirements (pt. K). We were inspired by statements from top management that encouraged cost saving initiatives. A quotation from the CEO illustrates this:

"Cost control on site will be a major focus in the months to come as significant amounts of money can be saved if we use shorter time on site" (CEO of EUMEC)

Spurred by this statement, we launched a Delphi study in June 2009 with the goal of identifying and ranking problems that had the most negative impact on the projects' profit margins. Of 18 ranked problems, 8, including the top 6, were DQ/IQ related. The findings of the Delphi study documented and strengthened our argument that DQ/IQ related problems were worth our close attention (pts. L and M).

We developed the rules based on project requirements, operating under the motto "*everything that can be formalized can be assessed by rules*" (chief developer). These requirements were of two types: common and project specific. Common requirements applied to every project at EUMEC and reflected internal issues such as organizational policies influencing data values and internal coding manuals. Project specific requirements mainly reflected customer needs and such contingent issues as the type of construction to be designed (types of ship or type of oil rig), and the information systems to be used during design (customers sometimes require systems outside the organization's regular portfolio).

To meet Requirement Specification 3 we had to divide the project into assessment phases. An existing procedure at EUCMEC provided us with a basis for doing this. Engineering projects at the company are required to follow an internal Project Execution Model (PEM). PEM provides an overview of all phases in a project, a detailed explanation of requirements for each phase and shows the milestones for deliveries (e.g. drawings). We asked expert engineers from the various engineering disciplines to tell us at which phase of a project were the correct data values supposed to be known and hence could be expected to be correct in the database. Based on their responses, we divided the project into three phases aligned with existing PEM phases. In the first phase, we defined rules to check only for existence of data that should be known by the end of that phase. In the following phases the rules were progressively "tightened" to check more and more data fields whose values would be known by then. This reflects the increasing need for correct detail information. This process culminated in the last phase where everything needed to be correct, i.e. every data field would be checked against the determined rules.

To illustrate this process of gradual "tightening" of rules checking, we take the example of pipelines. Larger pipelines are designed and modelled before smaller pipelines. Since the properties of larger pipelines are commonly known before the properties of smaller pipelines, data records belonging to larger pipelines can be expected to contain known and correct values before the records belonging to the smaller pipelines do. Therefore the larger pipelines' data records can be assessed before the smaller pipelines' data records. Data for smaller pipelines would be inserted into the data records with only an ID field which is needed for referencing. The remaining values could be inserted later when the engineers would have the needed information, such as the type of liquid or gas it would carry. In this phase the smaller pipelines and use the rules to identify DQ/IQ errors related to those only. The smaller pipelines could be included in the rule definition at later stages.

 Project requirement
 Rule

 All
 in the last of the last of

By the end of 2009 an initial set of rules were ready for use in TestP (pt. N). Table 3 lists two examples of

Project requirement	Rule
All main equipment have to be installed at	All equipment-records are checked to determine if the
the Yard	equipment is main equipment. If so, the rule also checks
	if the site-code-field has a value that indicates Yard (all
	site-code-field contain a value indicating where to ship
	the equipment). If the value is not Yard, an error is
	reported.
All equipment ID's should reflect	All equipment ID's are checked against the NORSOK
NORSOK standard	standard. If the value does not comply with NORSOK
	(there are many other standards) an error is reported.

 Table 3: Examples of rules

Implementing and refining the set of rules for TestP: We manually executed the rules (pt. N) and discussed the reported errors with the engineers through emails and in organized meetings. We also started the development of a daily report that would run automatically and would be visible on TestP's home intranet site. From January 2010 we participated in weekly follow-up meetings with focus on training end-users in TestP and providing them with clarifications when they asked (pt. O). These meetings brought about a need for refinement of the rules as the common understanding of TestP's requirements became clearer through these meetings. Below, we provide two examples of exchanges from these meeting:

Example 1- Clarification of requirements: "Does anybody know which site code to use for HVAC items?" (design engineer)

Explanation: The site code is used to indicate where an item (HVAC in this case) is to be assembled. Each project usually has several assembly sites identified by a unique site code. These sites can be located at different places all over the globe. If the site code for an item is incorrect, the item will be shipped to a wrong location. Shipping items are costly and time consuming to begin with. To return and re-ship items are even more costly. Hence, it is important to insert the correct site code but first one need to know the code.

Example 2 – Unclear system response: "I have referred to a P&ID for one of my items, but an error is reported: This P&ID is not referred to in the Line list. What the h.. does that mean??" (design engineer)

Explanation: P&ID is a Piping and Instrumentation Diagram which shows the piping of the process flow together with equipment and instrumentation. The Line list contains all the P&IDs used in a project. To register an item in a database, engineers need to refer to the P&ID to which the item is connected. To find the relevant P&ID, they consult the Line list. In the Line list all P&IDs used in the project are listed. The error message here was meant to indicate that the P&ID that the engineer was referring to did not exist in the Line list. In this case the error message was not very well formulated. A better formulation that would have made more sense would have been: The P&ID referred to by this item does not exist in the Line list.

In April 2010 the automatic IQS report was implemented. The aim was to provide the engineers with upto-date error reports which they could assess whenever they wanted to. This would enable them to focus on errors that they considered most important for immediate handling. This meant the beta version of IQS now was completed and rolled out (pt. P). Several types of activities followed: The training sessions from January continued; another set of weekly meetings were established with focus on further refinements of rules; a third set of meetings focused on clarifications of rules and the concept of DQ/IQ.

During this period the engineers had several questions about the project requirements related to DQ/IQ. A recurring question from them related to a rule: "*how did you know that this was a requirement*?" We had to repeatedly explain that we had assessed the requirements for the project in collaboration with discipline managers, document control manager and engineering manager. As time went by and more and more rules were developed and refined, the engineers began to perceive the error report as a working list for correcting errors to meet the requirements. They assumed all DQ/IQ related requirements were covered by the rules.

This assumption had an unfortunate implication every time a new rule was added. A new rule meant more errors reported; hence the total number of errors actually increased gradually even though the engineers had corrected the errors earlier. This surprised them greatly. During this period we also refined a number of aspects of IQS, namely, the basic architecture of IQS and the interfaces with internal and external (customer) applications.

The regular meetings we held throughout the period of implementation and refinement, also served as forum for evaluating IQS (pt. Q). The ADR team and the engineers evaluated IQS from several perspectives such as: do the principles established result in the anticipated outcome? is the number of errors manageable? do the engineers find IQS useful?. In the next section, we present the reflections from these evaluative sessions.

5.3 Reflection and Learning from BIE-Cycle 1

Building of IQS, intervening by implementing it in TestP and evaluating its effect generated four design principles. These are shown in the left-hand column of Table 4. Our evaluation revealed both anticipated and unanticipated consequences which are shown in the right-hand column of Table 4. The first three principles, allow for inconsistency, incompleteness and lack of logical coherence, had anticipated outcomes. The engineers were generally satisfied with several aspects of the system. They appreciated that IQS assessed not only completeness, but also inconsistency and logical coherence. They also mentioned that they got a better total overview because the assessed data were collected from all relevant data sources.

"Assuming the rules are correctly defined, IQS is a nice tool" (electro engineer)

"If we try to do it manually [assessing data] it takes a lot of time and you also have to know exactly what you are looking for. IQS detects almost everything and fast" (discipline manager)

"It is so much easier to use the IQS report to identify discrepancies between the main engineering system and the 3D modelling tool, it provides a better overview than if you try to do it manually" (mechanical engineer)

The principles affected the way the engineers worked: they stopped using time on manual identification of possible errors and focused on reported errors. Other than that it was "business as usual", since the existing engineering design process was anyway based on the insertion of straw values: data that were best-guess values, random temporary values, or blanks. Thus, while these three design principles epitomized the error management paradigm, they only reflected existing engineering design practices.

However, the fourth principle, phase-based reporting, proved to be a challenge. Weekly focus session discussions, participant observation and semi-structured interviews revealed that the engineers at EUMEC appreciated the intentions of the report and used it as a working list for correcting errors. The report enabled them to focus directly on the errors instead of manually trying to identify them first. However, the implementation also revealed unanticipated consequences. First, although this principle was intended to reduce the number of reported errors, engineers complained that the number was still unmanageably large. They felt that as a result, they found it difficult to prioritize which errors to correct first. The mere number of errors was in some sense demotivating. Second, the system identified some data values as incorrect when they were actually correct. These "false positives" reduced trust in IQS. Both these unanticipated consequences demotivated the engineers from using the system.

Design Principle	Consequences
Allow for inconsistency	• Reduced errors. Stopped spending time to manually identifying inconsistencies and focused on errors that could be corrected (anticipated)
Allow for incompleteness	• Reduced errors. Stopped spending time to manually identifying incompleteness and focused on errors that could be corrected (anticipated)
Allow for lack of logical coherence	• Reduced errors. Stopped spending time to manually identifying logical incoherence and focused on errors that could be corrected (anticipated)
Phase-based reporting (report only identified errors applicable for current	• The IQS report used as a working list for correction of identified errors (anticipated)
phase)	• The number of reported errors perceived as still too large. Engineers found it difficult to prioritize errors to correct first leading to demotivation (unanticipated)
	• IQS generated "false positives" leading to reduced trust in IQS and consequently to demotivation (unanticipated)

Table 4: Consequences of Building, Intervention, and Evaluation of IQS

5.4 Problem (re)Definition

The unanticipated consequences of the first BIE-cycle led us to take another look at the problem. The large number of false positives in the report initiated a thorough review of the rule definitions. We were able to weed out some of them by identifying inaccuracies in the rule definitions. Even after that, we were still left with the majority of the false positives. After several discussions with managers and engineers of each engineering discipline, where we dissected each occurrence of false positive, we realized that most of these errors actually represented exceptions from requirements. The discipline managers had agreed upon these exceptions with the customer.

Identification of exceptions revealed another problem: it was difficult to determine the common denominators of these exceptions that could be captured in a manageable set of rules and thus exclude them from the report. It was of course possible to define a rule for every exception, but the huge number of errors made it a staggering task. Besides, new exceptions would almost certainly arise in the future. We had to find a more manageable way to deal with existing and future exceptions.

The other unanticipated consequence was the unmanageably large number of true errors that IQS was reporting. That led us to think of better ways to group errors for reporting. The goal was to find a solution that would both decrease the amount of errors to manage at the time, and, help engineers in prioritizing errors. This redefinition of the problem led to a second cycle of BIE.

5.5 BIE - Cycle 2 - Designing for error handling and error management

Handling false positives: Our solution for handling false positives was simply to remove those errors from the report and "park" them in a separate table which we called the "parking table" (pt. R). The rationale was this: these errors were detected by the project specific rules whose definition was based on a formalized set of requirements. Handling exceptions require departing from this formalized set which is a serious action. Hence to keep track of all these exceptions, they had to be first reported as errors and handled later. The parking table did that. The table contained the errors, the cause of the error and the signature of an authorized person who was either the concerned discipline manager or someone delegated by the manager. At hand-over this information could then be attached as a supplement to the deliverables. This report helped explain why the data deviated from the original set of specifications. We formalized this in a design principle which we call "Parking principle".

Handling the large number of true errors reported: The phase-based principle was aimed at reducing the total number of errors reported by grouping the errors by phases. This was not achieved: the number of reported errors remained large. Our discussions with the engineers and discipline managers revealed that the phases were still a large unit to group errors. We needed a smaller unit. After searching for a unit of optimal size, we settled on "commissioning packages". A commissioning package is "a practical scope of work unit within a system or subsystem for commissioning, constituting a functional unit which can be tested by commissioning to confirm its suitability for operation" (NORSOK 1999). The deliverables for all systems belonging to a certain functional unit possesses the same delivery date. A project has several such commissioning packages. By grouping all data belonging to a commissioning package, the number of reported errors would be significantly smaller than the number of errors reported for several units in that project phase. In addition, the delivery dates would provide a basis for prioritizing the most urgent packages first. Based on these considerations, we divided the original three report phases into several narrower phases based on commissioning packages and their delivery dates (pt. X). This led us to revise the phase-based principle by adding explicitly that the phases should be optimally narrow.

5.6 Reflection and Learning from BIE-Cycle 2

The engineers were satisfied with both solutions. Commenting on the parking of errors solution, an electrical engineer said: *"It is nice to get rid of these false errors in a quick way. If we use some time to identify the exceptions in the first place, we will have more time to focus on correcting real errors"*

A typical feedback on the solution to report errors by commissioning packages was: "Now it is possible to prioritize which errors to correct first. The small amounts of errors per package also result in higher motivation for correcting them simply because it feels possible to actually be able to correct everything" (mechanical engineer)

The evaluation and reflection led us to revise one design principles derived from BIE-1 and add a new one. Table 5 lists the revised set of design principles. The changes from the previous set are shown in italics.

Design Principle	Description
Allow for inconsistency	Inconsistencies caused by best-guess values, random temporary values, or blanks, should be allowed at appropriate early phases
Allow for incompleteness	Incompleteness caused by best-guess values, random temporary values, or blanks, should be allowed at appropriate early phases
Allow for lack of logical coherence	Lack of logical coherence caused by best-guess values, random temporary values, or blanks, should be allowed at appropriate early phases
Phase-based reporting	Preliminary incorrect values should not affect the daily report until the appropriate project phase is reached. <i>The</i> <i>phases should be narrow enough to produce a manageable</i> <i>number of errors and provide means for prioritizing.</i>
Parking of errors	"False positives" occurring as a result of (legitimate) deviation from original (and still applicable) requirements, should be removed from the error report and saved with an explanation for deviation together to be handled later.

 Table 5: Revised set of Design Principles

Other events in the BIE-Cycle-2: Several enhancements to IQS were also made during BIE Cycle 2. We implemented graphics for progress reporting (pt. S) and held workshops on research methods for the ADR team and the engineers. The first was on Action Research (pt. T) and the second one was on ADR (pt. V). Participating in these workshops helped the engineers to appreciate our efforts and to have an insight into their roles in a project that can simultaneously develop a practical solution and generate academic knowledge. However, perhaps the most effective milestone was the spread of the idea of IQS. By now, news about IQS was spreading throughout EUMEC and other projects wanted the tool. From February to November 2010, 15 other projects implemented IQS on their home intranet sites (Pt. U). To make it easier for the project participants in these 15 projects to start using IQS, a default set of rules was developed. The default set contained rules applicable to all projects. Every project would then have to add project specific rules based on project specific requirements. Due to lack of Information Management resources and lack of willingness by some projects to budget such resources where available, different projects received IM support to varying degrees. One of the experiences from TestP was exactly the need for extensive support. We noted that those projects with an internal "champion" used time to learn and use IQS while those projects that lacked such champions simply used a default set of rules not fully aligned with the project's requirements for deliverables.

6 Discussion

In this section we first discuss how we used TIQM and DQA to guide our actions while designing IQS. We then present a summary of the case in the form of a figure and a table. Figure 5 depicts the BIE cycles and the outcome of the project, while Table 8 lists how the main actions of the project map to the ADR processes and adheres to the ADR principles.

6.1 Guiding frameworks and IQS

TIQM and IQS: In Table 6 we present how IQS related to the premises of TIQM processes. The first column lists the six TIQM processes, and the second column describes how IQS relates to each process.

Process	Related to IQS
P1 Assess Data Definition	Defined which data sources were needed for assessment. Defined an Information
and Information	Architecture that could handle needed data types and store data in a way that
Architecture Quality	made them retrievable for comparison. Defined assessment outcome reports.
P2 Assess Information	Defined project-specific assessment rules based on project requirements. Defined
Quality	a set of default rules that can be used as a basis for any construction engineering
	project.
P3 Measure Nonquality	Performed a Delphi study to identify and rank the problems having the most
Information Costs and	negative impact on the profit margins. Several DQ/IQ problems were identified,
Risks	including the top six problems
P4 Re-engineer and	Corrections will have to be performed manually due to needed expert
Correct Data	determination of correct data values. Daily error reports are presented to the
	engineers
P5 Improve Information	Established a methodology for evaluating and refining IQS, including the
Process Quality	necessary alignments of various processes and weekly meetings for each project.
	The activities related to IQS are included in the organization's Project Execution
	Model (PEM)
P6 Establish the	Information managers are appointed per project. Support and training are
Information Quality	provided. The development has and will continue to be performed in close
Environment	collaboration with several stakeholders (especially project users and managers)

Table 6: Mapping IQS to the premises of TIQM processes by English (1999, 2003)

As can be seen from Table 6, IQS met the requirements of TIQM to some extent: processes were defined on how to assess DQ/IQ to meet the requirements of the engineers (e.g. which data to collect and how to present error messages in an understandable way) and how to measure the level of quality. Further, processes for evaluating and refining IQS were established, and the activities related to IQS were included in EUMEC's PEM. Information managers are now appointed per project and together with the PEM requirements this will establish an environment for information quality thinking.

However, some issues were not addressed. The absence of automated correction of errors by IQS is one such issue. This was due to the nature of engineering data, which requires engineering expertise to determine the correct data values. Another important issue was the lack of possibility to measure accuracy by comparing data to a physical object. This was because some of the physical assets only existed as referenced logical/functional elements in data registers and data models since those physical assets were not yet produced. A third issue was that of measuring costs of poor quality information. Even if the level of DQ/IQ improved, this study does not provide any measurements on the relation between increased DQ/IQ and impact on project performance.

DQA and IQS:DQA proposed the concept of subjective and objective assessment. In the case of IQS, TestP's project requirements were captured in formalized statements used as executable rules. The outcome of these rules formed the basis for task-dependent (objective) assessment. However, the outcome also needed subjective assessment. When the outcome included errors disputed by the engineers, their subjective assessment revealed that they were not really errors. Rather, these were exceptions from stated requirements. In alignment with DQA, the action taken was establishing the "parking principle". Task dependent metrics was developed to measure the state of each rule and to compare outcome between projects. Task independent metrics was not an issue since it was the contextual characteristics that were the challenge.

DQA also propose defining relevant quality dimensions. In this case they were Completeness, Consistency, and, Logical Coherence. Even though DQA proposes different measuring functions for different quality dimensions, this has not been an issue in our study. The reason was that no matter which dimension was involved, it was the consequence of each error that was important, i.e. delays and cost overruns.

DQA also propose weighting of variables if the importance of each variable is well understood by the company. In our case this meant the importance of each rule could be weighted based on how severe a detected error would impact the project in terms of delays and cost overruns; the higher the impact, the heavier the weight. This could have helped prioritizing errors. Instead, a simple ratio per rule (using percentage) was calculated. That solution makes it possible to compare projects, but without knowing how severe the impact of the outcome of each rule would affect the project. Table 7 shows how IQS related to the premises of DQA.

Assessment type	Related to IQS
Subjective data quality assessments	The needs were the requirements of the individual
reflect the needs and experiences of	construction engineering projects. Experienced
stakeholders	engineers revealed some "errors" were exceptions from
	requirements. Such subjective assessments were
	necessary throughout the project.
Objective assessment can be task-	Task-dependent assessment was in focus since
independent or task-dependent. Task-	contextual characteristics were the challenges. Task-
dependent metrics are developed in	dependent assessment included the ISs used in
specific application contexts	construction engineering projects within EUMEC,
	interfaces toward external ISs (for example, the
	customer ISs), business rules such as use of the
	organization's PEM, and projects requirements.

Table 7: Mapping IQS to the premises of DQA proposed by Pipino et al. (2002)

Finally, following the suggestion of Pipino et al. (2002) to create a de facto standard, a default set of assessment rules that could be used in any construction engineering project was developed.

6.2 Summary of the Case

Figure 5 captures the two BIE cycles each of which addressed specific design challenges. The right-hand side of the figure summarizes contributions of the IQS project.

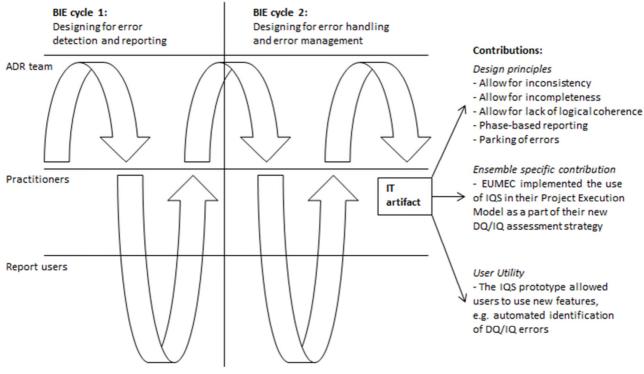


Figure 5: BIE cycles and contributions of the IQS project

Table 8 summarizes how the project meets the ADR principles.

ADR principles	The ADR process in the IQS project		Main actions
Principle 1: Practice- Inspired Research	Research was driven by the need for better DQ/IQ assessment tools for construction engineering projects.		Studied documentation Targeted TestP as the test- project for DQ/IQ assessment Conducted Delphi study for identification and ranking of problems Assessed existing DQ/IQ tool
Principle 2: Theory- Ingrained Artifact	The theory used were based on existing literature describing unique characteristics of engineering data such as not knowing the "correct answer", several disparate data sources, and the large amount of data produced. Two quality assessment frameworks were used to guide the development of IQS: DQA (Pipino, et al., 2002) and TIQM (English 1999, 2003).	-	Assessed existing DQ/IQ literature in general Conducted a review of DQ/IQ research in construction engineering, identified challenges Selected frameworks for guidance
Principle 3: Reciprocal Shaping	Errors in data sources and drawings were expected to be an ongoing problem. In collaboration with the practitioners problems were iteratively addressed and design principles were formulated.	- - -	Created requirements for IQS Developed IQS basic architecture Developed IQS basic rules
Principle 4: Mutually Influential Roles	The ADR team included researchers and representatives from three groups of stakeholders: IM department, engineering project, and end-users. One of the designers was an employee from EUMEC who was also a PhD student.	-	Participated in actions concerning training, support, requirement clarifications, etc. Refined of IQS rules
Principle 5: Authentic and Concurrent Evaluation	IQS was first evaluated within the ADR team, then in the wider setting of end-users at EUMEC, and finally through a comparison of level of DQ/IQ in TestP and two other projects not using IQS	-	Conducted interviews Assessed IQS reports
Principle 6: Guided Emergence	The preliminary design of IQS was continuously re-shaped through use and feedback from project participants. During BIE iterations refinements of IQS were performed based on anticipated and unanticipated consequences.		Developed parking feature Developed report based on commissioning packages
Principle 7: Generalized Outcomes	A set of design principles for DQ/IQ assessment systems was articulated (see Table 5). IQS was positioned as an instance of such systems.	-	Formulated design principles Prepared for dissemination of design principles (represented by this paper)

Table 8: Mapping IQS project to ADR principles

6.3 Contributions

ADR studies are expected to generate generalized outcomes of the problem instance, of the solution instance and to derive design principles (Sein et al. 2011). Below, we elaborate how in developing IQS, our study met these expectations.

The *class of problems* addressed by our study was level of DQ/IQ. The specific context was construction engineering which can be also understood as the area of concern (Mathiassen et al. 2012). By doing so, we answered calls to IS researchers to extend DQ/IQ research to new contexts (Madnick et al. 2009), and to provide practitioners with tools that can assess the level of DQ/IQ (Pipino et al. 2002). In addition, we identified DQ/IQ challenges unique to the context of large construction engineering projects.

The *class of solution* addressed by our study was the extant approaches used to solve DQ/IQ problems. These approaches were mainly represented by the frameworks that we used to guide the development of IQS, namely DQA (Pipino et al. 2002) and TIQM (English 1999, 2003). These approaches were built on the assumption that the correct data value to insert in a record was known by the time of insertion. Thus, the extant paradigm was *error detection and correction*. Our approach was shifting the paradigm to *error management*. We elaborate further on this later in this section.

The development of IQS resulted in a set of *design principles* (see Table 5) that (a) mitigate the problem of not knowing the correct data values to insert by the time of insertion to a record, and, (b) handle the reporting of an inevitably huge number of identified errors by introducing phase-based reporting and parking of identified "errors" that for some reason are not to be considered errors (i.e. false positives).

In addition, we identified quality dimensions that are relevant for construction engineering projects: completeness, consistency, and logical coherence. In the rest of this section, we elaborate on these contributions to knowledge and reflect on the IQS project in a holistic manner.

Essentially, IQS helps to balance two equally important requirements that act in opposite directions. The first is that the projects need to proceed swiftly and with minimum delay. The second is that data should be accurate so that drawings can be accurate. This requires a thorough check of the data to detect errors. That would delay the project and thus go against the first requirement. An error detecting tool would first have to correctly identify errors and then report them. The detection of errors is not straightforward. Due to the concurrent nature of construction engineering, data cannot be complete, or accurate or even coherent at the start of the project. So in order to meet the first requirement – keep the project flowing – shortcuts have to be taken in inserting data values. This is done knowingly and if managed properly, will not harm the project's progress. That is what IQS does. IQS also speeds up the process of error detection and brings consistency and uniformity to the process by using rules which automatically detects errors.

IQS thus represents a shift in paradigm on how DQ/IQ errors are defined and handled in construction engineering by bringing in the context – namely the challenges faced by engineers. All that is reported as error is not error. Some are "required errors" because no one knows the correct data at that time. Others are not really errors but are exceptions. Once the users – the engineers – understand this, errors become manageable. IQS does not improve DQ/IQ by correcting errors. It simply detects errors and reports them in a manageable way. The actual correction of the errors is still the responsibility of the users – engineers in this case.

Taken together, the 5 design principles that emerged from the development of IQS underscores the error management paradigm. The first three principles (the three "allows") keep the project flowing by allowing shortcuts and insisting only on the known correct data. The next one, phased based reporting, reduces the number of errors reported so that engineers can prioritize the errors that need to be corrected. The last one, parking of errors, keeps the project flowing by handling exceptions to requirements that would be detected as errors. IQS recognizes this. However, it also recognizes that as long as there is a record somewhere that marks these as exceptions, and identifies the person who has indicated that these are exceptions, the project continues to flow without harm.

In addition, our study also *contributed to practice* by solving an actual problem. EUMEC's projects were plagued by delays and cost overruns. We tracked the cause of the problem to low level of DQ/IQ. Our

solution was developing IQS. By comparing a project that used IQS with two projects that did not, we demonstrated that the system alleviated DQ/IQ problems.

Finally, our study also contributed to the method space. We had implicitly followed ADR as our research method at the outset and explicitly at later stages of the project after the publication of the Sein et al. (2011) paper. Nevertheless, we can suggest a modification to the method. In ADR, Stage 4 is framed as the final stage where outcomes are formalized when the project is completed. In our study, the formalized outcome was fed back to Stage 1 and informed the problem formulation for the next iteration of BIE. For example, we initially used the succeeding phases of PEM as a basis for implementing stricter rules for error checking. This generated one of the design principles, namely, phase-based reporting. When this was implemented in the BIE stage, we realized the phases were too long: the number of errors was still too large, making it difficult for the engineers to manage and correct the errors. This led to us back to the problem formulation stage and we re-formulated the problem as: what is the optimal phase to make the reported number of errors manageable? The answer was to report per commissioning package based on delivery dates of the packages. It is worth considering whether a feedback path from Stage 4 to Stage 1 should be added to the ADR process to capture unanticipated consequences that were vital for new iterations of BIE.

7 Limitations and future research directions

As with any other empirical study, our study has its limitations but these limitations also offer opportunities for further research. First, at the time of writing, the ADR project was still on-going. Thus, the design principles that we are reporting may not have reached their final form yet. Refinement of the principles may occur and new principles may be added. It is more than likely that when IQS has been in use for an extended period and in several projects, it would mutate. Consequently, changes and additions to the design principles will emerge. Following this development at EUCMEC over this extended period would be interesting and insightful. In addition, it would be of interest to extend the research to other organizations. The findings from this study could be useful to other researchers as a starting point for such extended research.

As shown in the timeline (Figure 4) the need for a management report was about to emerge at the time of writing. The need for such management reports that communicate the level of DQ/IQ and provide comparative assessment over time, has been articulated in the literature (Pipino et al. 2002). Not including such a report can be seen as a limitation of this study. This issue is related to the limitation described next.

A limitation of IQS is that the errors it reports are not weighted. Some errors may have more severe consequences than others on delays and cost overruns To be able to provide a single aggregated measure of the level of DQ/IQ, weighting of the various variables (the rules identifying errors in this case) is needed (Pipino et al. 2002). If the organization has a good understanding of the importance of each rule defined then a subjective weighting can be used (ibid). Another alternative would be to empirically determine the degree of delay caused by each error type. This can be done by gathering data at assembly sites of several projects. Such an investigation could provide the information needed to weight the rules. Either way, we can get more accurate information on which of the DQ/IQ errors are more important to weed out in construction engineering. This could also enhance the value of a management report.

Another limitation of this study is that we did not demonstrate that an increased level of DQ/IQ also improved overall project performance. This would be an essential extension of projects such as this. Future studies may focus on the impacts of DQ/IQ assessment systems in construction engineering, more specifically those related to the correlation between DQ/IQ on the one side, and delays and cost overruns on the other.

Finally, we are left with an intriguing question. An artifact such as IQS includes features that make it possible to detect errors that would not have been possible through manual searches. The question is whether the time consuming work of capturing project requirements in assessment rules outweighs the time used at the assembly site for correction of errors. More research is needed to investigate these issues.

References

- Blechinger, J., Lauterwald, F. and Lenz, R. (2010). Supporting the Production of High-Quality Data in Concurrent Plant Engineering Using a MetaDataRepository. *Proceedings of the Americas Conference* on Information Systems, August 12 -15, Lima, Peru.
- Dobson, D. and Martinez, R., (2007) Integrated engineering, ABB Review: SPEC. REP.): 48-52.
- Dyrhaug, Q. (2002). A generalized critical success factor process model for managing offshore development projects in Norway. Doctoral thesis, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.
- English, L. P. (1999). Improving data warehouse and business information quality: methods for reducing costs and increasing profits: John Wiley and Sons, Inc, New York, New York, USA, 1999.
- English, L. P., (2003) Total information quality management : a complete methodology for IQ management, Dm Review (Now : Information Management), (9): 1-7.
- Franco, L. A., Cushman, M. and Rosenhead, J., (2004) Project review and learning in the construction industry: embedding a problem structuring method within a partnership context, *European Journal of Operational Research*, (152: 3): 586-601.
- Hevner, A. R., March, S. T., Jinsoo, P. and Ram, S., (2004) Design Science in Information Systems Research, *MIS Quarterly*, (28: 1): 75-105.
- Lee, Y. W., Strong, D. M., Kahn, B. K. and Wang, R. Y., (2002) AIMQ: a methodology for information quality assessment, *Information & Management*, (40: 2): 133-146.
- Lempinen, H. (2012). Constructing a Design Framework for Performance Dashboards *Nordic Contributions in IS Research* (pp. 109-130): Springer.
- Lin, S., Gao, J. and Koronios, A., (2008) A data quality framework for engineering asset management, *Australian Journal of Mechanical Engineering*, (5: 2): 209-219.
- Madnick, S. E., Wang, R. Y., Lee, Y. W. and Zhu, H., (2009) Overview and Framework for Data and Information Quality Research, *Journal of Data and Information Quality*, (1: 1): 1-22.
- Markus, M. L., Majchrzak, A. and Gasser, L., (2002) A Design Theory for Systems That Support Emergent Knowledge Processes, *MIS Quarterly*, (26: 3): 179-212.
- Mathiassen, L., Chiasson, M. and Germonprez, M., (2012) Style composition in action research publication, *MIS Quarterly*, (36: 2): 347-363.
- Neely, M. P., Lin, S., Gao, J. and Koronios, A. (2006). The Deficiencies of current data quality tools in the realm of engineering asset management. *Proceedings of the Americas Conference on Information Systems*, August 4 - 6, Acapulco, México.
- NORSOK. (1999). Z-007 Mechanical Completion and Commissioning. <u>http://www.standard.no/no/Fagomrader/Petroleum/NORSOK-Standard-Categories/Z-MC--</u> <u>Preservation/Z-007/</u>.
- Pipino, L. L. and Lee, Y. W. (2007). Applying IT to Healthcare: Humans, Errors, and a Data Quality Perspective. *Proceedings of the Americas Conference on Information Systems*, August 10 - 12, Keystone, Colorado.
- Pipino, L. L., Lee, Y. W. and Wang, R. Y., (2002) Data Quality Assessment, *Communications of the ACM*, (45: 4ve): 211-218.
- Ramaswamy, M., (2006) On the phenomenon of information dilution, *Issues in Information Systems*, (VII: 2): 289-292.
- Rivas, R. A., Borcherding, J. D., González, V. and Alarcón, L. F., (2010) Analysis of factors influencing productivity using craftsmen questionnaires: Case study in a Chilean construction company, *Journal* of Construction Engineering and Management, (137: 4): 312-320.
- Sein, M. K., Henfridsson, O., Purao, S., Rossi, M. and Lindgren, R., (2011) Action Design Research, *MIS Quarterly*, (35: 1): 37-56.
- Sekine, K. and Arai, K. (1994). Design Team Revolution: How to Cut Lead Times in Half and Double Your Productivity: Productivity Press, Portland, OR.
- Strong, D. M., Lee, Y. W. and Wang, R. Y., (1997) Data quality in context, *Communications of the ACM*, (40: 5): 103-110.
- Toor, S.-U.-R. and Ogunlana, S. O., (2008) Problems causing delays in major construction projects in Thailand, *Construction Management & Economics*, (26: 4): 395-408.

- Tribelsky, E. and Sacks, R., (2011) An empirical study of information flows in multidisciplinary civil engineering design teams using lean measures, *Architectural Engineering and Design Management*, (7: 2): 85-101.
- Vician, C. (2011). A Resource Review for Health Care Quality Execution: Business Intelligence Strategy. Proceedings of the Americas Conference on Information Systems, August 5 - 7, Detroit, Michigan.
- Wand, Y. and Wang, R. Y., (1996) Anchoring Data Quality Dimensions in Ontological Foundations, Communications of the ACM, (39: 11): 86-95.
- Wang, R. Y. and Strong, D. M., (1996) Beyond Accuracy: What Data Quality Means to Data Consumers, Journal of Management Information Systems, (12: 4): 5-33.
- Westin, S. (2012). Data and Information Quality Research in Engineering Construction Projects: A Review of Literature. *Proceedings of the 21st International Conference on Information Systems Development (ISD2012)*, Prato, Italy, August 29-30, 2012.

Association for Information Systems AIS Electronic Library (AISeL)

AMCIS 2012 Proceedings

Proceedings

Data and Information Quality Dimensions in Engineering Construction Projects

Soffi Westin Information Systems, University of Agder, Kristiansand, Norway., soffi.westin@uia.no

Olav Birger Holen Private enterprise, Kristiansand, Norway., ol-birg@online.no

Tero Päivärinta Computer and Systems Science, Luleå University of Technology, Luleå, Sweden., Tero.Paivarinta@ltu.se

Follow this and additional works at: http://aisel.aisnet.org/amcis2012

Recommneded Citation

Soffi Westin, Olav Birger Holen, and Tero Päivärinta, "Data and Information Quality Dimensions in Engineering Construction Projects" (July 29, 2012). *AMCIS 2012 Proceedings*. Paper 1. http://aisel.aisnet.org/amcis2012/proceedings/DataInfoQuality/1

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2012 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Data and Information Quality Dimensions in Engineering Construction Projects

Soffi Westin University of Agder, Norway soffi.westin@uia.no Olav Birger Holen Private enterprise, Norway ol-birg@online.no

Tero Päivärinta Luleå University of Technology, Sweden tero.paivarinta@ltu.se

ABSTRACT

Poor data and information quality (DQ/IQ) causes delays and cost overruns in engineering construction projects. However, only little DQ/IQ research has been performed in this context. This paper explores quality dimensions in the context of engineering construction projects. The most important dimensions identified by Ge, et al., (2011) is used as a basis and compared with dimensions used in 12 large engineering construction projects in one organization. The findings show that six of these dimensions are in use in those projects: accessibility, security, relevancy, completeness, consistency, and timeliness. In addition, the findings indicate another dimension also very important in this context; logical coherence. The logical coherence dimension compares different data values and determines if there is any illogicality between them. Three dimensions are monitored by rules provided by a DQ/IQ tool, and we discuss about the contributions which such a tool can provide for an engineering construction firm.

Keywords

Data quality, information quality, quality dimensions, engineering, construction, logical coherence

INTRODUCTION

Large engineering construction projects include a variety of project types such as groundwater (Frimpong et al., 2003), highrise (Faridi and El-Sayegh, 2006; Long et al., 2004), buildings (Assaf and Al-Hejji, 2006) and offshore development projects in the oil and gas industry (Dyrhaug, 2002). The characteristics for large engineering construction projects are their need for several different kinds of expertise, they cost millions of USD and the construction sites are usually far away from where the design is carried out. The teams carrying out these projects are of a transient nature and deliver their products through temporary project-based organizations which exist only for the single project (ibid). The main output from the engineering design phase (for example the design of an oil rig) to the assembly phase (for example the assembly of the oil rig) is drawings. The information on these drawings is mainly extracted from various data sources, such as engineering databases.

Delays and cost overruns are common phenomena in construction projects (Long, Ogunlana, Quang and Lam, 2004; Sambasivan and Soon, 2007; Toor and Ogunlana, 2008) and several researchers have reported delays as the most costly problem (Faridi and El-Sayegh, 2006; Ling, Pham and Hoang, 2009). Recent research identifies lack of information related to drawings as one of the most significant factor for delays and cost overruns (Dai, Goodrum and Maloney, 2009; Westin and Päivärinta, 2011). This implies poor data and information quality (DQ/IQ) in the data sources and systems in the construction engineering field.

DQ/IQ has become important for businesses also in general. The amount of business related data and the impact of poor DQ/IQ information are growing continuously (Ramaswamy, 2006; Strong, Lee and Wang, 1997). Assessment of DQ/IQ thus becomes equally important. DQ/IQ can be assessed from the viewpoint of many quality dimensions. A quality dimension includes a definition and a set of measures for DQ/IQ evaluation. Among the most frequently mentioned are accuracy, completeness, consistency and timeliness (Carlo Batini, Cappiello, Francalanci and Maurino, 2009). In addition, a recent review identified that the most frequently mentioned dimensions are accessibility, security, relevancy, ease of understanding, reliability, ease of manipulation, and objectivity (Ge, et al., 2011).

Even though research on DQ/IQ assessment is performed in various contexts, little has been performed in the context of engineering construction. There are a few exceptions; among them are studies of DQ/IQ assessment in large engineering projects (ref. e.g. Blechinger, Lauterwald and Lenz, 2010; Westin and Päivärinta, 2011) and assessment of current DQ/IQ

tools and their relevance for this context (ref. e.g. Neely, Lin, Gao and Koronios, 2006). In the light of the serious impact of poor DQ/IQ and the early stages of DQ/IQ research in this context more research is needed. Since quality dimensions are a fundamental concept in DQ/IQ research it could be useful to increase knowledge on the need and use of such dimensions in engineering construction projects. The research question of our study hence is:

Which quality dimensions are needed in engineering construction projects and why?

To answer this question we present a case study of 12 large engineering projects in one organization. Based on the most important DQ/IQ dimensions as identified by Ge, et al., (2011) we compare existing quality dimensions with dimensions used in these projects. The findings confirm the importance and use of several existing quality dimensions. In addition the findings demonstrate the use of a dimension not mentioned in previous research.

The remaining sections of this paper present related research, the research approach and case description, the findings, the discussion of the findings, and finally the conclusion.

RELATED RESEARCH

DQ/IQ has been defined broadly through its "fitness for use" (Wang and Strong, 1996), and the definition is adopted in this research. Whereas early research focused on query techniques on multiple data sources and data warehouses in the end of 1980s, the research field has later on spread to a number of new application areas, such as customer resource management, knowledge management, supply chain management and enterprise resources planning (Madnick, Wang, Lee and Zhu, 2009), including contexts such as data warehousing (e.g. Blake and Mangiameli, 2011; Wixom and Todd, 2005), health care registers (e.g. Pipino and Lee, 2007; Vician, 2011), retailing and internet related systems (e.g. de Corbiere, 2009; Helfert and Hossain, 2010). Recently also engineering construction has become a field of interest for DQ/IQ researchers (e.g. Tribelsky and Sacks, 2011; Westin and Päivärinta, 2011).

An important aspect of DQ/IQ assessment is the use of "quality dimensions". Quality dimensions make it easier to define and discuss issues related to DQ/IQ without referring to specific data values. Rather it is groups of data values with the same qualities (e.g. accurate values; consistent values, etc). Batini et al. (2009) identify altogether 28 quality dimensions in their review of data quality methodologies. Many of the dimensions have had varying definitions and measures for data quality evaluation. However, four core dimensions; accuracy, completeness, consistency and timeliness have been emphasized frequently throughout the methodologies. Other potential quality dimensions in previous research have been currency, volatility, uniqueness, appropriate amount of data, accessibility, credibility, interpretability, usability, derivation integrity, conciseness, maintainability, applicability, convenience, speed, comprehensiveness, clarity, traceability, security, correctness, objectivity, relevancy, reputation, ease of operation, and interactivity (ibid.). In their study Ge, et al., (2011) have assessed 17 dimensions commonly used and tailored them to a set of 9 dimensions they perceive as the most important dimensions (see Table 1).

A dimension of special interest in our case is coherence or logical coherence. Previous research have discussed on a few definitions on this dimensions for example in terms of consistency in dates: "like you cannot die in the future, you cannot be born in the future" (Piprani and Ernst, 2008, p. 5); coherence in statistical data: ".....includes coherence between different data items pertaining to the same point in time, coherence between the same data items for different points in time, and international coherence" (Brackstone, 1999, p. 16); or more general: "coherence implies that two or more values do not conflict with each other" (Singh, Park, Lee and Rao, 2009, p. 6). Similar examples are presented in other studies (e.g. Bahorich and Farmer, 1995; C. Batini and Scannapieco, 2006; Berti-Équille, 2007; Ochoa and Duval, 2009; Peralta, 2008; Stadler and Kolbe, 2007; Vassiliadis, Bouzeghoub and Quix, 2000). In the field of engineering construction one study addresses this issue as a type of inconsistency using examples of various value mismatches: between adjacent items (e.g. mismatch between a pipe diameter and its belonging valve diameter); between attributes of one item having conditional functional dependencies (e.g. the type of valve constrains the allowed nominal diameter range); between items and corresponding catalogs (e.g. the valve from a specific supplier has to satisfy modeling ranges of the according catalog), and, between items in the corresponding environment (e.g. a cooling system should have cooling item types) (Blechinger, et al., 2010).

In this study we use logical coherence dimension in addition to the most important dimensions as identified by Ge, et al., (2011) as a basis for our discussion.

Quality Dimensions	Attributes of Items	
Accessibility	accessible, obtainable, retrievable, available	
Security	secure, protected, authorized access	
Relevancy	useful, relevant, applicable, helpful	
Accuracy	correct, accurate, free of error, precise	
Completeness	sufficient, complete, comprehensive, include all necessary values, detailed	
Consistency	consistent meaning, consistent structure, presented in the same format	
Timeliness	current, up to date, delivered on time, timely	
Ease of understanding	easy to understand, easy to comprehend, easy to identify the key point	
Reliability	reliable, dependable	
Ease of Manipulation	easy to manipulate, easy to aggregate, easy to combine	
Objectivity	impartial, unbiased, objective, based on facts	

Table 1. Quality Dimensions and Attributes of Items (Ge, et al., 2011)

In Table 1 we have listed these dimensions together with "Attributes of Items". "Attributes of Items" are measuring items that are suitable for measuring the related dimension (ibid).

RESEARCH APPROACH AND CASE DESCRIPTION

We have chosen the case study approach for our research. The reasons for that are multiple. First, DQ/IQ research in this context is rather new and still very much under-investigated. In such new and little explored fields case studies are especially appropriate (Eisenhardt, 1989). Second, to be able to better understand DQ/IQ in a specific context, a case study is suitable (phenomenon examined in a natural setting is a key characteristic of case studies (Benbasat, Goldstein and Mead, 1987). Third, since access to suitable organizations can be hard to achieve (Walsham, 2006), we want to exploit the opportunity of access we have been provided by our target organization.

Our target organization is a European, multi-discipline engineering construction company with capabilities related to global management, design, procurement, completion and generally execution of complex installations for the oil and gas industry. The organization delivers engineering design for construction projects and possesses a significant share of global markets in its product and project domains. The most employees are involved in engineering and construction projects. The biggest projects may typically cost more than 100 million Euros and take up to three years of calendar time.

Each project involves several engineering disciplines. Due to tight schedules engineering tasks are performed in parallel (concurrent engineering) even if the tasks are interdependent. In an environment like this, it is possible that some data values are omitted, or preliminary values used, when engineering item data is inserted for the first time in a record. The intention from the engineers is to insert the missing or correct values as soon as they are known. Meanwhile, thousands of records are inserted and it gets difficult or even impossible to manually identify incorrect values or missing values. An in-house developed DQ/IQ assessment tool is hence used by the projects to help identify any DQ/IQ insufficiencies. The tool is rule-based which means that rules are developed for assessment of data and information in various data sources.

The result is presented to the engineers in a report. Amongst other information the report contains a descriptive name of each rule and the number of related errors. Project participants can click the number of errors reported on a specific rule, and the report will display all information needed to correct the error (e.g. ItemID, which values are missing or incorrect etc). By identifying and reacting to observed DQ/IQ errors earlier the projects hope to increase DQ/IQ in extracted drawings which previously have been identified as insufficient (Westin and Päivärinta, 2011). Increasing the level of DQ/IQ in drawings should then lead to a reduction in delays and cost overruns.

The rules are defined and implemented based on internal organizational requirements, customer requirements and suggestions from the engineers. We have investigated the rules used by 12 projects within the organization. Based on Ge, et al.'s (2011) quality dimensions and measuring items, every rule have been investigated to determine its type of measuring

item and hence the related dimension. A total of 246 rules have been examined. The number of rules per project depends on the size of the project and the customer requirements. Amongst the 12 investigated projects the number of rules varied from 49 to 138.

FINDINGS

When comparing the dimensions used by the projects with dimensions identified by Ge, et al. (2011) we found that not all of Ge, et al.'s dimensions or measuring items were in use. In addition we found a dimension in use not mentioned by Ge, et al. In Table 2 column one we first list the dimensions common for the projects and Ge, et al., and then we have added the dimension only identified in the projects. Column two contains the attributes of items (measuring items) the projects used to assess the belonging dimension. The third column contains the number of rules used to assess each dimension, or an 'X' to indicate that the dimension is handled by other means (e.g. organizational policies).

Quality Dimensions	Attributes of Items	Number of rules applied
Accessibility	available	Х
Security	secure, protected, authorized access	Х
Relevancy	relevant	Х
Completeness	include all required values	131
Consistency	consistent meaning	43
Timeliness	current, up to date, delivered on time, timely	Х
Logical Coherence	two or more values do not conflict with each other	72

Table 2. Quality Dimensions and Attributes of Items in use, measured by rules (indicated by a number) or handled by other means (indicated by an X)

In the next section we explain how the different dimensions are understood and used by the projects. Further we discuss whether this is in line with Ge, et al. (2011) and how it possibly differs.

DISCUSSION

Accessibility is in the projects understood as the availability of needed information. This understanding is in line with Ge, et al., (2011), since "available" is one of their identified attributes of items. However, availability is not covered by any used rules, rather unavailable information result in errors identified by other rules. For example: imagine an engineer waiting for information related to an instrument or a cable or any type of part s/he is responsible for. The engineer is about to insert the record for this part into the database, but will not be able to fill in values for all the required fields since s/he is still waiting for that information. The record will hence be inserted with some missing values in some fields. If this record is checked by a rule concerning completeness, an error will be reported due to the missing values in some fields. From previous research we know that lack of information in drawings extracted from the databases is a common phenomenon in this context (Dai, et al., 2009; Westin and Päivärinta, 2011). Available information in time could be one of the problems.

Security is only regarded through the fact that the used DQ/IQ assessment tool provides reports only to the individual project and only project participants have access to the reports. This is in line with the understanding of this dimension as presented in Lee, Strong, Kahn and Wang, (2002, p. 135), where they state that "the system must be accessible but secure". (Ge, et al. also refers to this study for the definition of the security dimension). This dimension is hence not covered by any rule implemented in the DQ/IQ control system; rather it is covered by organizational policies.

Relevancy is understood as relevant information. "Relevant" is also one of Ge, et al.'s identified measuring item for this dimension. Whether the information is perceived as relevant or not is based on customer and internal requirements. An assessment of these requirements is performed prior to selecting rules for use in each project. The total set of chosen rules will hence be considered assessing all relevant data in the project. This means that there are no rules *checking* for relevancy; rather data relevance is determined based on the prior assessment of requirements and the following choice of rules.

Accuracy is in the projects understood as a combination of completeness and consistency. This understanding is also pointed out by Ge, et al. (section 3.1); "users presume that inaccurate information consists of incomplete and inconsistent

information". The rules identified for the dimensions completeness and consistency are hence covering the accuracy dimension.

Completeness is understood as all required values should be inserted in the records. Again, this is in line with one of the identified measuring items for this dimension. The number of rules assigned to this dimension could indicate that this dimension is perceived as very important which is no surprise; without values important information is missing. Completeness can also be viewed in conjunction with the accessibility dimension. If the engineers have to skip some values from the record while the values are not yet available, the completeness dimension and its measuring items will provide an overview of missing values. If available information in time is a common problem in our context, the rules for assessing completeness are important to make sure that every value is there by the time the drawings are extracted.

Consistency is in the engineering projects understood as whether data values representing the same issue are identical across different tables or different data sources. This is also a common understanding of the consistency dimension (Pipino, Lee and Wang, 2002). This means that the measuring item "consistent meaning" can be used because identical values would imply consistent meaning. The number of rules assigned to this dimension (43) illustrates the need for this type of quality assessment also in this context.

Timeliness is understood as whether the data is delivered on time to the customer or to internal receivers. "Delivered on time" is also one of the measuring items for this dimension (Ge, et al., 2011). In the projects this dimension is covered by delivery dates set for each record. The delivery dates are used as a grouping indicator when reporting DQ/IQ status. This means that for each delivery date the users are able to see if there are any errors reported by the rules on items that belong to a certain date. The delivery date grouping makes it easier to prioritize which errors to attend to at the moment.

Despite of their popularity in other domains (Ge, et al., 2011), *Ease of understanding, Reliability, Ease of manipulation* and *Objectivity* are dimensions that seem to be not very relevant for these projects. Maybe the reason lays in the nature of engineering construction; data values are considered understandable, objective and easy to manipulate (in the sense of inserting and editing records). However, there are some exceptions where the data is not reliable. For example when the information is not available and therefore missing or presented with a preliminary value. Then the rules for other dimensions such as completeness, consistency or logical coherence will intercept these instances and report them as DQ/IQ errors.

Logical coherence is understood as if two or more values do not conflict with each other. This is in line with Singh et al.'s (2009) definition of coherence and similar to Blechinger et al.'s (2010) understanding of a type of inconsistency. Even if the terms logical coherence and inconsistency in this sense have been discussed to a certain degree, our case seems to highlight the importance of this dimension more than previous research.

The rules used for measuring this dimension compare different values of different fields. The assessment is important to avoid delays and extra costs. Let us discuss about an example:

The main assembly site (e.g. a yard) of an oil rig is usually far away from the location of design. Smaller parts could be put together at another site before shipment to the Yard. A rule of thumb in the projects is that all *main* assets should be assembled on the Yard. In this case the rules will assess the values in the site code field and the value of main/not-main asset field. If the asset is main asset, the site code should be Yard meaning that the data values would be logically coherent between themselves. If not, if the site code is incorrect, it could in practice result in assets being shipped to wrong locations. Shipping assets is costly and takes time. To return them and re-ship them will be even more costly. By attending the possible problem by using DQ/IQ assessment rules while the assets are still at their location of origin, these extra costs are avoided.

We compared the logical coherence dimension to Ge et al.'s (2011) dimensions and measuring items to see if it was possible to use any of them. The most similar dimensions were accuracy and consistency. If the measuring items for the accuracy dimension were to be used, they would probably not indicate any error. For example, since there is no real world equivalent to compare the values (yet), the site code value would return as correct as long as it is one of the legal site codes in the project. The main/not main asset field would always return as correct since the only two possible values to insert is Y/N, and they are both legal. Hence, the two inserted values would *each* be accurate, free of error and precise. Only by assessing the combined result; i.e. logical coherence, an error could be revealed.

In some sources, the consistency dimension has included a dimension named as "(logical) coherence". However, the consistency dimension can be viewed from a number of perspectives (Pipino, et al., 2002). Common perspectives are representational consistency (data is consistently presented in the same format (Wang and Strong, 1996)), semantic consistency (consistency of the same data values across tables (Pipino, et al., 2002)), and structural consistency (consistency in the representation of similar attribute values (Levitin and Redman, 1995)). From other perspectives it is possible to imagine that logical relations could be perceived as consistency between different values. For example: if a buyer was to

choose a car model from an online system, and the car model only was produced in the color red, it should be impossible to choose the color green from the color field. The system would probably prevent any possibility of choosing any color but red as soon as the buyer had chosen a model. This is a common situation for several types of systems (you choose a flight date and then the available flight times, or you choose an author and then the available titles). But these situations are different because they are commonly dealt with by the time of insertion of data. This is not possible in engineering construction context since the data is inserted in several different ways; manually, imported from other data sources, and copied and inserted from other data sources. Also, due to a probability of not yet having received the predefined values of for example site codes, the engineers still have to be able to insert partial information without validation towards such predefinitions. Hence, the quality assessment has often to be performed *after* insertion, and some of the constraints materialize after initial data has been already put into the system. This causes the need for periodical control of DQ/IQ in engineering projects using the approach of concurrent engineering.

From previous literature we already know that lack of information related to drawings is one of the most significant factors for delays and cost overruns in engineering construction projects (Dai, et al., 2009). Knowing that this information is extracted from various data sources implies poor DQ/IQ in these data sources. So how can our findings contribute to this field? We see that the target organization, in their effort to increase the level of DQ/IQ in the projects, mainly lean on three DQ/IQ dimensions for monitoring. These dimensions are completeness, consistency and logical coherence. The result of their effort could be measured by assessment with regard to the drawings in those projects actively using the DQ/IQ tool versus drawings extracted by those projects not using the tool. We have not been able to identify research in this area pointing at these three DQ/IQ dimensions and providing suggestions of how to determine whether DQ/IQ efforts impact the extracted drawings. Whereas this paper is proposing such issues and actions, further research of ours will investigate the impact of using the implemented DQ/IQ tool, embodying the data quality rules, over time. We also hope to be able to show significant decrease in delays and cost overruns in the case organization as an ultimate result of our development.

CONCLUSION

This paper aimed at identifying quality dimensions in DQ/IQ assessment performed in the context of engineering construction projects. For the assessment and discussion of each dimension we used the most important dimensions as identified by Ge et al., (2011) in addition to a dimension we termed "logical coherence". 246 DQ/IQ assessment rules, implemented in a DQ/IQ assessment tool and used by 12 projects in one organization were investigated to determine which quality dimensions were in use. A total of seven quality dimensions were identified; accessibility, security, relevancy, completeness, consistency, timeliness, and, logical coherence. However, only three dimensions; completeness, consistency, and, logical coherence were directly handled by these rules. Accessibility was indirectly handled by rules from other dimensions (completeness and consistency) since these rules will reveal errors such as missing or incorrect values due to non-available information. Security was handled by organizational policy. Relevancy was handled by assessment of internal requirements and customer requirements prior to defining/selecting rules for the individual project. Timeliness was handled by delivery dates and these dates are used as a grouping indicator when reporting DQ/IQ status.

Logical coherence implies that two or more values do not conflict with each other. This dimension seems to appear as a very important dimension in this context, despite the fact that it is rarely discussed in previous DQ/IQ research. The number of rules applied to this dimension (72 of 246) indicates the importance of this dimension and that several sets of data values are in need for such assessment.

So, three dimensions, completeness, consistency and logical coherence, were identified as the most relevant dimensions for assessment in our target organization. If errors reported are corrected in the data sources, it is believed that extracted drawings will increase in level of quality. To determine if this is correct these drawings could be compared with drawings produced by comparable projects that do not use the DQ/IQ tool. If the level of DQ/IQ increases it should be possible to identify impact on delays and cost overruns. These three dimensions is also the most important dimensions to assess by a DQ/IQ tool, since the other dimensions in use are either used for determination of which rules to use, are handled by organizational policy, or is indirectly handled by dimensions already assessed by rules.

We believe our findings contribute to research on DQ/IQ in the field of engineering construction by highlighting the importance of the three dimensions, the potential of controlling these dimensions better in the projects, and the suggestion of how to determine whether the use of the DQ/IQ tool resulted in the wanted impact on the drawings. Even if this research has been performed within one organization, we believe our findings can be used as guidance for similar organizations in their effort to increase their level of DQ/IQ. We will continue our research in the target organization aiming at presenting future suggestions for DQ/IQ assessment concepts needed in the context of engineering construction, and hopefully be able to identify impact on delays and cost overruns.

REFERENCES

- 1. Bahorich, M. and Farmer, S. (1995) The coherence cube, *The leading edge*, 14, 10, 1053-1058.
- 2. Batini, C., Cappiello, C., Francalanci, C. and Maurino, A. (2009) Methodologies for data quality assessment and improvement, *ACM Computing Surveys*, 41, 3, 1-52.
- 3. Batini, C. and Scannapieco, M. (2006). Data-Centric Systems and Applications: Springer
- 4. Benbasat, I., Goldstein, D. K. and Mead, M. (1987) The Case Research Strategy in Studies of Information Systems, *MIS Quarterly*, 11, 3, 369-386.
- 5. Berti-Équille, L. (2007) Data quality awareness: a case study for cost optimal association rule mining, *Knowledge and information Systems*, 11, 2, 191-215.
- 6. Blake, R. and Mangiameli, P. (2011) The Effects and Interactions of Data Quality and Problem Complexity on Classification, *Journal of Data and Information Quality*, 2, 2, 1-28.
- Blechinger, J., Lauterwald, F. and Lenz, R. (2010). Supporting the Production of High-Quality Data in Concurrent Plant Engineering Using a MetaDataRepository. *Proceedings of the Americas Conference on Information Systems*, August 12-15, Lima, Peru.
- 8. Brackstone, G. (1999) Managing data quality in a statistical agency, Survey methodology, 25, 2, 139-150.
- 9. Dai, J., Goodrum, P. M. and Maloney, W. F. (2009) Construction Craft Workers' Perceptions of the Factors Affecting Their Productivity, *Journal of Construction Engineering & Management*, 135, 3, 217-226.
- 10. de Corbiere, F. (2009). Data Quality and Interorganizational Information Systems: The Role of Electronic Catalogues. *Proceedings of the Americas Conference on Information Systems*, August 6 - 9, San Francisco, California.
- 11. Eisenhardt, K. M. (1989) Building Theories from Case Study Research, Academy of Management Review, 14, 4, 532-550.
- 12. Faridi, A. S. and El-Sayegh, S. M. (2006) Significant factors causing delay in the UAE construction industry, *Construction Management & Economics*, 24, 11, 1167-1176.
- 13. Ge, M., Helfert, M. and Jannach, D. (2011). Information quality assessment: Validating measurement dimensions and processes. *Proceedings of the European Conference on Information Systems*, June 9 11, Helsinki, Finland.
- 14. Helfert, M. and Hossain, F. M. Z. (2010). An approach to monitoring data quality product oriented approach. *Proceedings of the Americas Conference on Information Systems*, August 12 -15, Lima, Peru.
- 15. Lee, Y. W., Strong, D. M., Kahn, B. K. and Wang, R. Y. (2002) AIMQ: a methodology for information quality assessment, *Information & Management*, 40, 2, 133-146.
- 16. Levitin, A. and Redman, T. (1995) Quality dimensions of a conceptual view, *Information Processing & Management*, 31, 1, 81-88.
- 17. Ling, F. Y. Y., Pham, V. M. C. and Hoang, T. P. (2009) Strengths, Weaknesses, Opportunities, and Threats for Architectural, Engineering, and Construction Firms: Case Study of Vietnam, *Journal of Construction Engineering and Management*, 135, 10, 1105-1113.
- 18. Long, N. D., Ogunlana, S., Quang, T. and Lam, K. C. (2004) Large construction projects in developing countries: a case study from Vietnam, *International Journal of Project Management*, 22, 7, 553-561.
- 19. Madnick, S. E., Wang, R. Y., Lee, Y. W. and Zhu, H. (2009) Overview and Framework for Data and Information Quality Research, *Journal of Data and Information Quality*, 1, 1, 1-22.
- 20. Neely, M. P., Lin, S., Gao, J. and Koronios, A. (2006). The Deficiencies of current data quality tools in the realm of engineering asset management. *Proceedings of the Americas Conference on Information Systems*, August 4 6, Acapulco, México.
- 21. Ochoa, X. and Duval, E. (2009) Automatic evaluation of metadata quality in digital repositories, *International Journal on Digital Libraries*, 10, 2, 67-91.
- 22. Peralta, V. (2008). *Data quality evaluation in data integration systems*. Université de Versailles (chair) Raúl RUGGIA Professor, Universidad de la República, Uruguay.
- 23. Pipino, L. L. and Lee, Y. (2007). Applying IT to Healthcare: Humans, Errors, and a Data Quality Perspective. *Proceedings of the Americas Conference on Information Systems*, August 10 - 12, Keystone, Colorado.
- 24. Pipino, L. L., Lee, Y. W. and Wang, R. Y. (2002) Data Quality Assessment, *Communications of the ACM*, 45, 4ve, 211-218.
- 25. Piprani, B. and Ernst, D. (2008) A Model for Data Quality Assessment, *Lecture Notes in Computer Science*, 5333, 750-759.
- 26. Ramaswamy, M. (2006) On the phenomenon of information dilution, Issues in Information Systems, VII, 2, 289-292.
- 27. Sambasivan, M. and Soon, Y. W. (2007) Causes and effects of delays in Malaysian construction industry, *International Journal of Project Management*, 25, 5, 517-526.

- 28. Singh, P., Park, I., Lee, J. K. and Rao, H. R. (2009) Information sharing: a study of information attributes and their relative significance during catastrophic events, *Cyber-Security and Global Information Assurance: Threat Analysis and Response Solutions, IGI Publishers.*
- 29. Stadler, A. and Kolbe, T. H. (2007). Spatio-semantic coherence in the integration of 3D city models. *Proceedings of the International ISPRS Symposium on Spatial Data Quality*, June 13 15, Enschede, The Netherlands.
- 30. Strong, D. M., Lee, Y. W. and Wang, R. Y. (1997) Data quality in context, Communications of the ACM, 40, 5, 103-110.
- 31. Toor, S.-U.-R. and Ogunlana, S. O. (2008) Problems causing delays in major construction projects in Thailand, *Construction Management & Economics*, 26, 4, 395-408.
- 32. Tribelsky, E. and Sacks, R. (2011) An empirical study of information flows in multidisciplinary civil engineering design teams using lean measures, *Architectural Engineering and Design Management*, 7, 2, 85-101.
- 33. Vassiliadis, P., Bouzeghoub, M. and Quix, C. (2000) Towards quality-oriented data warehouse usage and evolution, *Information Systems*, 25, 2, 89-115.
- 34. Vician, C. (2011). A Resource Review for Health Care Quality Execution: Business Intelligence Strategy. *Proceedings* of the Americas Conference on Information Systems, August 5 7, Detroit, Michigan.
- 35. Walsham, G. (2006) Doing interpretive research, European Journal of Information Systems, 15, 320-330.
- 36. Wang, R. W. and Strong, D. M. (1996) Beyond Accuracy: What Data Quality Means to Data Consumers, *Journal of Management Information Systems*, 12, 4, 5-33.
- 37. Westin, S. and Päivärinta, T. (2011). Information quality in large engineering and construction projects: A Delphi Case Study. *Proceedings of the European Conference on Information Systems*, June 9 11, Helsinki, Finland.
- 38. Wixom, B. H. and Todd, P. A. (2005) A theoretical integration of user satisfaction and technology acceptance, *Information Systems Research*, 16, 1, 85-102.

Improving Data Quality in Construction Engineering Projects: An Action Design Research Approach

Soffi Westin¹ and Maung K. Sein²

ABSTRACT

The topic of data and information quality (DQ/IQ) is a longstanding issue of interest in both academia and practice in the construction engineering field. Poor DQ/IQ has led to poor engineering drawings that, in turn, have led to delays and eventuality to cost overruns. In this paper, we report a study that took an Action Design Research (ADR) approach to develop and evaluate a DQ/IQ assessment tool, which we call Information Quality System (IQS), in a large global engineering and construction company. The evaluation was performed by comparing the level of DQ/IQ in a project that used IQS with two projects that did not use the tool. The result is encouraging: the DQ/IQ in the project using IQS was significantly higher overall than in the two other projects. The implication is that a tool based on the design principles on which IQS was built is likely to help improve DQ/IQ in engineering systems and, hence, in engineering drawings. Consequently, it will decrease project delays and cost overruns. More generally, our paper adds to the discourse in the literature on the use of information and communication technologies (ICT) in the construction context. Our paper illustrates another successful application of action-oriented research that can solve practical problems while generating academic knowledge. In taking a design approach, we augment the literature on the use of action research in construction engineering and management.

Keywords

Information management, Data quality, Information quality, construction engineering, action design research, frameworks

INTRODUCTION

Delays and cost overruns are common phenomena in construction projects (Toor and Ogunlana 2008). Several causes have been identified such as: waiting for information (Frimpong et al. 2003), design complexity (Lim and Mohamed 2000), delays in design information (Assaf and Al-Hejji 2006) and changes in scope (Sambasivan and Soon 2007). Recent research has specifically pointed to issues related to drawings as one of the most significant problem areas (Dai et al. 2009; Westin and Päivärinta 2011). For example, poor drawings and specifications have been identified as a significant factor that causes delays and cost overruns (Rivas et al. 2010). The variation between planned and actual task start time and start duration has been traced to errors in design and/or drawings (Wambeke et al. 2011).

Improving drawings by reducing errors has, thus, become vital to reduce delays and overruns. Engineering projects are carried out in phases. After the feasibility of alternative solutions has been evaluated in the "concept engineering" phase, the project moves to the "detail engineering" phase where the design is completed and drawings and specifications are detailed to be handed over to the "assembly and construction" crews at remote construction sites (Ogunlana et al. 1998). This design information is commonly communicated in documents, often in electronic form (Tribelsky and Sacks 2011). Since the drawings are based on data (e.g. specifications), errors in such design documentation can be traced to poor data and information quality (DQ/IQ) in the data sources and systems in the construction engineering. Clearly, there is a need to improve DQ/IQ. In order to do so, we first need quality assessment frameworks.

The substantial body of literature on DQ/IQ proposes several quality assessment frameworks (English 1999; Lee et al. 2002; Pipino et al. 2002). These frameworks aim to increase the level of DQ/IQ in information systems by identifying, and sometimes automatically correcting, data errors. The frameworks are based on the premise that the correct data values are known at the time of insertion of data values in a database record (Neely et al. 2006). This assumption of correct insertion is not valid in the context of engineering construction. Moreover, construction engineering has characteristics that raise unique challenges. For example, the iterative nature of concurrent engineering forces engineers to proceed with only partial information (Blechinger et al. 2010). The upshot is that the existing DQ/IQ framework and tools are not useful in the

¹ Ph.D. candidate, Dept. of Information Systems, Univ. of Agder (UiA), Kristiansand, Norway. E-mail: Soffi.Westin@uia.no

² Professor, Dept. of Information Systems, Univ. of Agder (UiA), Kristiansand, Norway, and Luleå University of Technology, Luleå, Sweden. E-mail: Maung.K.Sein@uia.no

engineering construction context. There is a need to develop tools that meet its specifics needs. In this paper, we address this need by taking an Action Design Research approach (Sein et al. 2011) to develop and evaluate a DQ/IQ tool.

The rest of the paper is organized as follows. First, we define the problem as the lack of appropriate DQ/IQ assessment frameworks in construction engineering and briefly examine the engineering context to justify our problem definition. In the following section, we present ADR and describe how we applied it to our research. We then describe our proposed solution to the problem, namely, development of a tool named Information Quality System (IQS) in a large global engineering and construction company. Next, we describe how we evaluated IQS, and in the following section, we discuss our findings. We conclude the paper by raising some questions for which we provide some preliminary answers.

THE PROBLEM DEFINITION

We define the problem as the lack of appropriate DQ/IQ assessment tools in the construction engineering context. We first offer a brief description of the nature of large construction engineering projects and engineering data. Then, we look at the existing DQ/IQ assessment tools as possible solutions and argue that they are inadequate for our purpose.

Construction Engineering

Large construction engineering projects are generally complex and challenging to manage (Miller and Lessard 2001). As mentioned in the introduction, delays and cost overruns are common phenomena in construction projects; in fact, several researchers have reported delays as the *costliest* problem (Ling et al. 2009; Memon et al. 2011).

Apart from the huge costs involved, large-scale construction projects are characterized by the large number of project participants and the broad variety of stakeholders (Yeo and Ning 2002). Moreover, construction sites usually are far away from where the design is carried out. This presents the continuous challenge of providing adequate project information to construction sites (Braimah and Ndekugri 2009). The teams carrying out these projects are of a transient nature and deliver their products through project-based temporary multiple organizations that exist only for the single project (Cherns and Bryant 1984). The project is usually carried out in phases; pre-engineering (engineering design needed for planning the project) and detail engineering (engineering design needed for the specific project concept) are carried out first. The outcome of the engineering phase is mainly drawings and documents, which form the basis for the construction (or assembly) phase.

Although phase-based, engineering today is iterative and concurrent. Concurrent engineering means that tasks previously performed in sequence are now performed in parallel. This break from the traditional linear engineering model is mainly attributed to the globalization of engineering, which requires shorter delivery schedules (Dobson and Martinez 2007).

Engineering data possesses some unique characteristics that are different from typical business environments. For example, data are collected in a variety of formats while in business systems data often are provided in fixed formats (Lin et al. 2007). Engineering data are also typically shared among various technical and business systems. The unique characteristic that is most relevant for our problem definition is that the "correct answer" for the values to be inserted into an engineering asset data record is not always known at the time of insertion. Therefore, in keeping with concurrent engineering, engineers have to proceed with partial data. Clearly, this can lead to poor data quality. In our search for an appropriate solution to the problem, we turn to the existing DQ/IQ assessment literature.

DQ/IQ Assessment

DQ/IQ has been defined in various ways. For our purpose, we adopt the "fitness of use" perspective (Wang and Strong 1996), which is based on the intended purposes of the users of information. DQ/IQ research has moved on from such technical issues as query techniques on multiple data sources and data warehouses in the 1980s, to a number of new application areas, such as knowledge management (Madnick et al. 2009) and health care registers (e.g. Pipino and Lee 2007; Vician 2011). Recently, attention has turned to the context of construction engineering (e.g. Lin et al. 2008; Tribelsky and Sacks 2011).

An important aspect of DQ/IQ assessment is "quality dimensions," which are groups of data values with the same qualities (e.g. accurate values or consistent values). Quality dimensions make it easier to define and discuss issues related to DQ/IQ without referring to specific data values. A number of quality dimensions have been identified; e.g. a review of data quality methodologies identified 28 (Batini et al. 2009). The four dimensions that have been emphasized most frequently are accuracy, completeness, consistency and timeliness. An additional dimension, logical coherence, was identified by Westin et al. (2012). Table 1 lists a sample of dimensions relevant to DQ/IQ in engineering organizations. The three dimensions specifically relevant to our problem definition are italicized.

Dimension	Description	Reference
Accessibility	available	Ge et al. (2011)
Security	secure, protected, authorized access	Ge et al. (2011)
Relevancy	relevant	Ge et al. (2011)
Completeness	include all necessary (required) values	Ge et al. (2011)
Consistency	consistent meaning	Ge et al. (2011)
Timeliness	current, delivered on time, timely	Ge et al. (2011)
Logical coherence	Two or more values do not conflict with each other	Singh et al. (2009)

Table 1. A sample of relevant dimensions

Dimensions such as these are the elements on which DQ/IQ assessment frameworks are built, several of which have been proposed in the literature (e.g. Lee et al. 2002). However, as we pointed out in the introduction, such frameworks require that the "correct answer" has to be known in order to perform the comparison of inserted data with real-world data. As we stressed earlier, this requirement is not feasible in an engineering project. In order to underscore our contention, we next take a closer look at the construction engineering context.

DQ/IQ Assessment in Construction Engineering

DQ/IQ problems in construction engineering lead primarily to errors and omissions in engineering drawings (Dai et al. 2009; Westin and Päivärinta 2011). Studies have been conducted on different aspects of an engineering project. However, while quite a few DQ/IQ studies have examined the assembly and operation phases (e.g. Carter and Thorpe 2006; Elazouni and Salem 2011; Murphy 2009), the earlier phases in a project have received less attention. Thus, extant literature gives us little empirical evidence on how the specific and unique characteristics of construction engineering affect DQ/IQ.

Some insight, however, is provided from a literature review conducted by Westin (2012). She lists five challenges that lead engineers to proceed with partial information, which in turn leads to poor DQ/IQ. The challenges and their connections are shown in Figure 1.

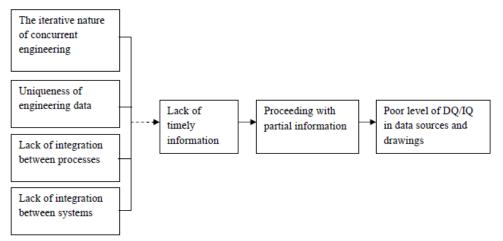


Fig. 1. Connections between challenges and level of DQ/IQ (from Westin, 2012)

In a specific construction engineering project, the first four challenges (shown on the left-hand side of Figure 1) are invariant. Since they could all contribute to "lack of timely information," this challenge too is invariant. Lack of timely information forces engineers to proceed with partial information (Blechinger et al. 2010) due to short delivery schedules (Dobson and Martinez 2007). The result is poor DQ/IQ in data sources and, consequently, incorrect drawings.

Taken together, viz., the inadequacy of the existing DQ/IQ frameworks and the lack of attention to earlier phases of engineering projects, we find few possible solutions to our problem definition. The alternative, then, is to construct our own solution. Action design research is an appropriate approach for such a solution. We, therefore, used this approach to develop a tool that we call IQS. Here, we briefly describe IQS and how we evaluated it.

OUR SOLUTION: IQS

IQS was developed in close collaboration between the researcher and the members of a specific project termed as the test project (TestP). IQS was built using design principles that emerged during the project. First, we briefly describe the organization where IQS was developed, and then present IQS before proceeding on to the evaluation.

The Organization

The organization is a European multi-discipline construction engineering company. Though the organization is European, their approximately100 offices and 19,500 employees are located around the globe, including the US and China. The operating revenue is more than 6.5 billion USD per year. The company is a world leader in management of global construction engineering projects, viz., engineering design, product development, procurement, completion and complex installations for the offshore oil and gas industry. They have a significant share of global markets in their product and project domains. A majority of the employees takes part in construction engineering projects. Our focus of research lies in the intersection between the engineering design phase and handover to the phases of assembly and completion, which are performed at the construction site. The main deliveries for handover are drawings, documents and 3D models. The biggest projects typically cost over 150 million USD and take up to three years. Several experts representing different engineering disciplines are needed to design large, complex, robust and yet delicate constructions. Project teams are set up according to engineering and other professional disciplines required by the individual projects. **Error! Reference source not found.** describes the engineering disciplines typically represented in the projects of the organization.

Discipline	Responsibility
Process	Design of industrial processes; all facts, sequences and relations
	in the process, logical placing of items
Mechanical	Design (choice of equipment and its physical layout and weight)
Piping/Layout	Design of all piping
Electro	Design and cabling of power distribution for electrical systems: equipment, lights, heat, etc.
Instrument	Design of control systems, i.e. control of various valves, machines, alarm systems, cables, etc.
Telecom	Selection and location of radio and audio systems, alarms, etc.
HVAC (Heating, Ventilating and Air Conditioning)	Capacity calculations and layout for ventilation, etc.
Safety	Various safety assessments
Structure (steel)	Design of steel structures, supports, outfitting like handrails, stairs, etc.
Architecture	Interior design

Table 2 Engineering distributes in		(from Worth and Ditrictor 2011)
Table 2. Engineering disciplines if	n the target organization's projects	(from Westin and Päivärinta, 2011)

Methodology

We used the ADR approach to develop IQS. Introduced by Sein et al. (2011), ADR creates scholarly knowledge through designing an IT artifact in a real organization. Organization intervention is crucial to ADR because of its underlying philosophy that the artifact is not only based on the design, but also emerges from interaction with an organizational context. The four stages of ADR are shown and briefly described in Figure 2.

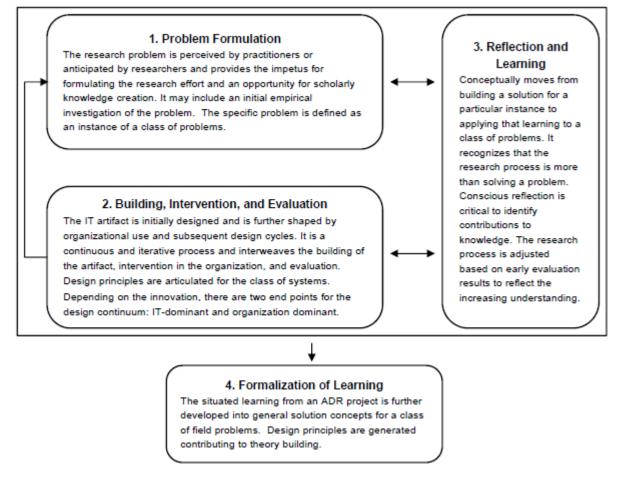


Fig. 2. Overview of ADR stages (adapted from Sein et al., 2011)

Each stage of ADR has a number of distinct tasks. These tasks are shown in Table 3. The rightmost column of the table shows how these tasks were carried out in our project. We emphasize that although these tasks are presented as stage-based, the actual process was iterative in keeping with Figure 2. To illustrate, the continuous evaluation in stage 2 almost always leads to a change in problem formulation. This is exactly what happened in our project. During the first evaluation of the problem, the result of an initial study and review of existing literature led to a reformulation of the problem from delays and cost overruns to inadequate DQ/IQ, which led to the decision to design IQS. The project generated design principles that could be used in any organization that carries out construction engineering projects. Examples are "phase-based assessment" (or reporting) and "allowing for inconsistencies" (in early phases).

ADR stages	ADR tasks	Project actions
	Identify and conceptualize the research opportunity	Original problem: delays and cost overruns experienced in the target organization. A study conducted in the target organization revealed DQ/IQ problems as the most important factor causing delay problems
Problem formulation	Formulate initial research questions	How to design a DQ/IQ tool appropriate for construction engineering context
	Cast the problem as an instance of a class of problems	DQ/IQ challenges in construction engineering
	Identify contributing theoretical bases and prior technology advances	Reviews of literature on DQ/IQ frameworks and tools, challenges, and DQ/IQ challenges in construction engineering
	Secure long-term organizational commitment	Contractual agreements with target organization
	Set up roles and responsibilities	The ADR team included several stakeholders e.g. developers, engineers and managers
	Discover initial knowledge- creation target	Develop IQS
	Select or customize BIE form	IT-dominant BIE selected since the locus of innovation comes from artifact design
Building, Intervention, and Evaluation (BIE)	Execute BIE cycle(s)	A set of requirements for IQS prepared, these included: • identification of data sources needed to collect all required data • decision of making the tool rule-based (rules can capture external and internal design requirements such as pipe-size, placing of equipment, where each item will be assembled etc.),
		• how the resulting report should be presented and the frequency of reporting
		 An engineering project (TestP) selected as a test project for development, implementation and evaluation of the tool, including these actions: Based on these requirements the architectural infrastructure of the tool (see Figure 3) the rules capturing internal and external project requirements for TestP developed IQS implemented and used by members of TestP
	Assess need for additional cycles, repeat	IQS continuously evaluated in collaboration with developers, users and managers
	Reflect on the design and redesign during the project	An ongoing activity throughout the project with the goal of identifying all DQ/IQ errors with emphasis on the detail engineering phase
Reflection and Learning	Evaluate adherence to principles	Reflected on the premises of the chosen guiding frameworks Reflected on anticipated and unanticipated consequences caused by the design (and/or changes in the design) and implementation of IQS
	Analyze intervention results according to stated goals	The level of DQ/IQ measured every day in reports available for all project participants
	Abstract the learning into concepts for a class of field problems Share outcomes and assessment	Conceptualizing IQS into a class of solutions done by identifying DQ/IQ dimensions and design principles Outcomes and assessment discussed with the target organization's practitioners
Formalization of Learning	with practitioners Articulate outcomes as design principles	 on a regular basis Examples of design principles: Allow for incomplete information Allow for illogical information (see the example in tables 5 and 6) Phase-based reporting (for example per "delivery-date", see table 5) to enable corrections of previously incomplete or incorrect information
	Articulate learning in light of theories selected	Done through research publications
	Formalize results for dissemination	Done through research publications

The System

IQS was developed mainly on two premises: the data quality assessments in practice proposed by Pipino et al. (2002) and the processes of TIQM (Total Information Quality Management) proposed by English (2003). Both emphasize the importance of solving the root of the problem. In our case, this meant that we had to ensure a sufficient level of DQ/IQ in data sources before extraction and use of drawings. Table 4 and Table 5 show how the development of IQS relates to these premises.

Table 4. Mapping IQS to the premises of subjective and objective data quality assessment proposed by Pipino et al. (2002)
Table 4. Mapping 10.5 to the premises of subjective and objective data quanty assessment proposed by Tipino et al. (2002)

Assessment type	Related to IQS
Subjective data quality assessments reflect the needs and experience of stakeholders	The needs are in our case the requirements of the individual projects. These are mainly collected from contracts and coding manuals. Together with several stakeholders such as engineers and discipline managers, we decided which assessment rules and quality dimensions were needed
Objective assessment can be task-independent or task- dependent. Task-dependent metrics are developed in specific application contexts	We focused on task-dependent assessment, which, among other factors, include the IS's used in construction engineering projects within the organization, interfaces towards external IS's (for example, the customer IS's) and business rules such as use of organization's Project Execution Model (PEM)

In addition, Pipino et al. (2002, p. 217) mention the possibility for an industry to ".....adopt a set of data quality metrics as a de facto standard....." For IQS, we translated this by developing a default set of assessment rules that could be used in any engineering project as well as an additional set of rules meant for various individual project requirements.

Process	Related to IQS
P1 Assess Data Definition and Information Architecture Quality	Defined which data sources were needed for assessment and import the data into a data warehouse
P2 Assess Information Quality	Defined project-specific assessment rules based on project requirements
P3 Measure Non-quality Information Costs and Risks	Performed a Delphi study to identify and rank the problems having the most negative impact on the profit margins. Several DQ/IQ problems were identified, including the top six problems
P4 Re-engineer and Correct Data	Corrections will have to be performed manually due to needed expert determination of correct data values. Daily error reports are presented to the engineers
P5 Improve Information Process Quality	Established a methodology for evaluating and refine IQS, including the necessary alignments of various processes and weekly meetings for each project. The activities related to IQS are included in the organization's Project Execution Model (PEM)
P6 Establish the Information Quality Environment	Information managers are appointed per project. Support and training are provided. The development has and will continue to be performed in close collaboration with several stakeholders (especially project users and managers)

The most conspicuous issue in Table 5 is the absence of automated correction of errors. This is due to the nature of engineering data, which requires engineering expertise to determine the correct data values (see P4).

IQS is rule-based. A rule is defined as a select statement that is based on a project requirement. Generally, all rules contain a default set of selected fields to display. These fields are chosen in collaboration with the users to provide sufficient information on the item reported with an error. In addition, most rules contain more selected fields depending on the type of item checked. The WHERE-clause of every rule contains syntax determining exactly which records are to be assessed. This could be, for example, only records belonging to a specific engineering discipline, only items of a specific type, only items in a specific mode of state, etc. The content of the WHERE-clause is determined by the delivery requirements related to the different items.

The requirement is captured in the syntax of the rule, so the rule can be executed against all required records with the purpose of identifying any irregularities concerning the requirement. For example, "Rule2: X,Y,Z is outside area limits" means that the coordinates indicating exactly where the item is located is not in accordance with the area value chosen (areas are imaginary cubes indicating various areas of the whole construction and determined by coordinates and named as D1, D2,... etc.). If the coordinates of the item is outside the coordinates of the chosen area, one of the chosen values (coordinates or area) must be incorrect. Only the opinion of an expert can tell which one is correct; the assessment tool will merely report this as an error. (An example of the syntax is shown for this rule in Appendix A.)

Figure 3 shows a schematic model of IQS.

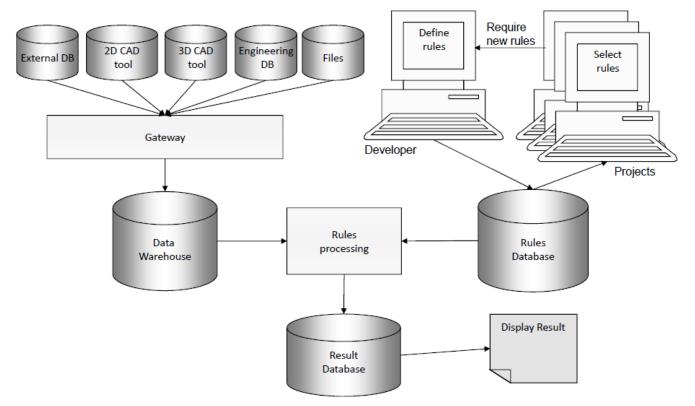


Fig. 3. A schematic model of IQS

The various data sources depicted in the upper-left corner of the model are sources and systems used in engineering design. These sources consist of a main engineering database, various CAD (Computer-Aided Design) tools, files containing data in various formats (e.g. tif, gif, pdf, etc.) and external databases such as customer databases. Data from these sources are imported each night to an Operational Data Warehouse. This integrated data repository is similar to an operational data store (ODS) and, thus, contains data from disparate sources later used for reporting. The extraction, cleaning and loading of the data is done using interfaces developed in-house. Each record in the data warehouse is marked with the ID of the project it belongs to and Contractor ID (each project may include several contractors).

The upper-right corner of the model depicts IQS rules that cover the requirements of each project. Requirements that recur between projects are put together in a default set of rules, which provide the starting point of assessment for each new project. Every defined rule is saved in a Rule DB from which individual projects select the rules they need. They may, in addition, ask for new rules to cover their specific requirements. When a rule is chosen by a project, it is marked with the Project ID for the system to know which rules to execute against that project's data. Several projects may choose the same rule.

Every night, the rules in the Rule DB are executed against the data in the data warehouse. After processing, the identified DQ/IQ errors are saved in a Result DB, which becomes the data source for Excel-based IQS reports for all projects. The reports displayed on the home intranet site of each individual project identify errors only pertinent to that project. The users can display the report in multiple formats, such as per engineering discipline, per product, per date of delivery, per work package, etc. An example of a report displaying a selected work package is shown in Table 6.

deliv_date	work_pkg	error_msg	E	Ι	М	Р	Grand Total
2012_12_28	Work Package:	Rule3: Tags without drawing references	4	12	96		112
	9999-Z99	Rule10: Missing mounted-on		1			1
		Rule15: Missing or inadequate description on tag		1			1
		Rule17: Missing or illegal PO-number	1	3	13		17
		Rule13: Missing EX-class	32	14	22		68
		Rule14: Missing IP-grade	32	14	22		68
		Rule2: X, Y, Z outside area limits	4	25	31	1	61

Table 6. An IQS report displaying identified errors for a work package

The first column of the report shows the delivery date to the customer for information related to the selected work package (shown in column 2). Column 3 displays the rules that have identified errors. The next four columns display the affected engineering disciplines (E= Electro, I =Instrument, M= Mechanical and P=Process) and the number of errors related to that discipline. Column 8 shows the total number of errors identified by the rule executed.

In a typical use case, an electrical engineer may want to further investigate the four identified errors shown under E (column 4) for Rule 2: X, Y, Z outside area limits. The engineer can double-click the number '4' in the column in order to display more details (see examples of additional information in Table 7).

object_name	field_value	description	discpl	area	work_pkg	resp_eng	error_msg
GD111222	X=806.000	Starter for el.motor	Е	M60	Work Package: 9999-	John	Rule2: X, Y, Z
	Y=795.000	for centrifugal fan			Z99	Smith	outside area
	Z=219.000						limits
GD222333	X=806.000	Starter for el.motor	Е	M60	Work Package: 9999-	John	Rule2: X, Y, Z
	Y=795.000	for centrifugal fan			Z99	Smith	outside area
	Z=219.000						limits
GD333444	X=806.000	Starter for el.motor	Е	M60	Work Package: 9999-	John	Rule2: X, Y, Z
	Y=795.000	for centrifugal fan			Z99	Smith	outside area
	Z=219.000						limits
EC777888	X=862.123	Drillers control	E	M60	Work Package: 9999-	John	Rule2: X, Y, Z
	Y=808.170	cabin, air-			Z99	Smith	outside area
	Z=239.758	conditioning system					limits

Table 7. IQS detail report

The first column tells the user exactly which item the error is related to. The second column displays the values of the field in the item record where the error was identified, in this case the value of the coordinate field. The actual coordinate values for the item related to row 1 are as follows: X=806.000, Y=795.000, Z=219.000. In column 5 (area), the value M60 refer to an imaginary cube determined by its coordinates. If the coordinates of the item are outside the coordinates of its area, either the area value is incorrect or the coordinate value is incorrect. The rest of the columns are commonly chosen to provide the user with enough information to determine which value is incorrect and what the correct value is. Some items might have to be excluded from assessment by this rule for reasons of impracticality or because no rule syntax was possible. Such items are then set aside in a parking table together with the RuleID, and are not checked further against that specific rule. At any time,

a parked item can be revoked on request. A question may also arise about who is responsible for this item. Column 7 (resp._eng.) provides the name of the person to ask for the correct weight or any needed additional information.

EVALUATION AND FINDINGS

To evaluate whether IQS indeed helped to improve DQ/IQ, we compared the project where it was used: TestP with two other projects (Projects A and B) that did not use it. Project A was an upgrade of a specific system called Mud System in use in an existing oil rig. Mud is a mixture of water, clay and chemicals used in oil drilling to lubricate and cool the oil drill, flush out mass that has been drilled out and to cover the walls of the drilling hole. Without an efficient mud system, the chance of equipment failure and hole and drilling problems will increase.

Project B was a "new-build" where the target organization was responsible for the drilling facilities of the oil rig. Drilling facilities are "structures containing systems and equipment required for drilling operations" (NORSOK 1998, p. 4). Offshore oil and gas production is challenging inter alia because of the remote and harsh environment. The drilling facilities are complex constructions and need to be able to operate at depths far below the ocean surface. For example, the well-known Troll A platform stands on the sea floor approximately 300 meters below the surface, and has an overall height of 472 meters.

TestP was an upgrade of existing drilling facilities to prolong the lifetime of the construction. This included prefabrication and offshore installation of equipment and prefabricated parts, in addition to offshore completion and commissioning.

To make a fair comparison, we identified a set of assessment rules that included customer requirements common for all three projects. Other rules covering nice-to-have information or ones used by developers only to mitigate any problems concerning interfaces or malfunctions directly related to the IS's in use were omitted from the comparison since they were not relevant to the final deliverables of the compared projects. Where the rules were not directly usable in Projects A and B, they were modified to suit their individual requirements. In short, we wanted to ensure that our comparison of the level of DQ/IQ in the projects was based on their individual requirements. With that said, more rules are required to capture *all* assessable requirements of a project. At the time of evaluation, TestP was using 137 rules, capturing more and more details in the requirements as the project progressed. These projects each represent a one-of-a-kind construction, and the total requirements can never be identical. Some of the projects using the tool today require only a total of 49 rules, which illustrates the deviation among requirements.

For this comparison, a total of 20 rules were selected as the common base. These rules reflected some of the most important requirements as identified in the target organization, for example weight, missing drawings and incorrect drawing types, and represent part of the requirements for each project.

The indices of comparison were as follows:

Pe = the total number of possible errors for each rule (i.e. number of records checked for that rule since each rule can only identify one error per record)

Te = the total number of errors indicated by that rule

Pcte= Percentage of errors (Pe/Te)

SumPe = Sum of Pe for each project

SumTe = Sum of Te for each project

PctSum = SumPe/SumTe

We could, thus, compare the total percentage of errors between each project, as well as granulate it by comparing the percentage of errors per rule per project. We also grouped the rules into three dimensions of quality identified by Westin et al. (2012), namely, Completeness, Consistency and Logical Coherence (see Table 1). Thus, we tested whether IQS helped improve these dimensions in addition to overall quality.

Figure 4 shows the result of the comparison. Note that both Project A and Project B are completed projects, while TestP was nine months away from completion at the time of comparison.

		TestP			Pı	roject A		Project B		
Rules	Dimension	# of records checked	# of errors	% of errors	# of records checked	# of errors	% of errors	# of records checked	# of errors	% of errors
Rule3: Tags without drawing references	Completeness	7259	73	1,01	4468	42	0,94	11837	267	2,26
Rule6: Missing site code	Completeness	4502	75	1,67	3629	94	2,59	8642	1018	11,78
Rule7: Missing area code on tag	Completeness	9041	8	0,09	5146	56	1,09	13702	143	1,04
Rule8: Missing GA and Equipment Layout references										
in MEL	Completeness	1685	482	28,61	585	234	40,00	1714	1035	60,39
Rule9: Missing weight	Completeness	4630	899	19,42	3245	1889	58,21	9270	4085	44,07
Rule10: Missing mounted-on	Completeness	3942	108	2,74	2672	508	19,01	7800	1888	24,21
Rule13: Missing EX-class	Completeness	3839	325	8,47	3939	724	18,38	11484	4117	35,85
Rule14: Missing IP-grade	Completeness	3839	445	11,59	3939	723	18,35	11484	3917	34,11
Rule15: Missing or inadequate description on tag	Completeness	4629	39	0,84	3245	155	4,78	9270	513	5,53
Rule16: Discipline code missing	Completeness	8964	0	0,00	5120	5	0,10	12969	15	0,12
Rule18: Missing mounted-on on valve	Completeness	688	10	1,45	573	6	1,05	1470	0	0,00
Rule19: Missing Manufacturer	Completeness	4527	310	0,80	4578	1059	23,13	13077	3927	30,03
Rule20: Missing Model	Completeness	7649	319	4,17	4578	1074	23,46	13077	3987	30,49
Rule17: Missing or illegal PO-number	(Completeness)/consist	4506	128	2,84	3245	207	6,38	9270	1294	13,96
Rule5: Incorrect drawing type in GA- and/or Layout-										
field	Consistency	2363	28	1,18	2048	80	3,91	6214	1169	18,81
Rule11: Documents with incorrect DocType	Consistency	10869	36	0,33	7449	42	0,56	16385	1414	8,63
Rule1: Area code different from master tag	Logical Coherence	2741	26	0,95	2663	1734	65,11	8520	5626	66,03
Rule2: X, Y, Z outside area limits	Logical Coherence	699	79	11,30	4 32	14	3,24	1209	323	26,72
Rule4: System code different from master tag	Logical Coherence	2848	17	0,60	980	77	7,86	3613	491	13,59
Rule12: Drawing referenced by tag, but not defined										
by Document Control	Logical Coherence	85802	395	0,46	39080	8736	22,35	72971	1826	2,50
TOTAL		175022	3802	2,17	101614	17459	17,18	243978	37055	15,19
	TOTAL PER DIMENSION									
	Completeness 65194 3093 4,74				45717	6569	14,37	125796	24912	19,80
	Consistency	17738	192	1,08	12742	329	2,58	31869	3877	12,17
	Logical Coherence	92090	517	0,56	43155	10561	24,47	86313	8266	9,58

DISCUSSION

The findings show that TestP's level of DQ/IQ was significantly better overall than for those in the two other projects. While TestP had 2.17% errors identified in total, the numbers for projects A and B were 17.18% and 15.19%, respectively (see percentage indications in TOTAL row in Figure 4). This superiority extends to the quality dimensions, i.e. level of Completeness, Consistency and Logical Coherence (see TOTAL PER DIMENSION section in Figure 4).

For every rule, with three exceptions, TestP had the lowest percentage of errors. The three rules are circled in Figure 4. Since TestP is not yet completed (while the other two are completed), the number of errors identified by these three rules could decrease towards delivery date. Even if this is not the case, the number will certainly not increase since all items required were registered in the engineering databases; hence, all missing or partial data values related to the item records were already identified and displayed in the reports. We, therefore, conclude that TestP is overall at a much better state in terms of the level of DQ/IQ than the two other projects. This speaks well for the effectiveness of IQS.

Our study has some limitations. First, the three projects were not contemporaries: TestP is yet to be completed, while projects A and B were completed three and six years ago, respectively. In that timeframe, it is reasonable to expect an increase in the quality of project deliveries simply because the organization gained more experience. However, projects A and B have quite similar data quality status (approx. 17% and 18% errors despite the three-year difference between them), while it is a significant leap to TestP (approx. 2% errors). In all other aspects, the three projects were quite similar. This further lends support for the effectiveness of IQS.

Second, the errors are not weighted. Some errors might lead to more severe consequences than others, especially concerning delays and cost overruns. We have not examined this issue yet since it requires a thorough investigation at the assembly site

for several projects. Such an investigation will also increase our understanding of the relationship between a low level of DQ/IQ in engineering data sources and delays and cost overruns.

CONTRIBUTIONS AND IMPLICATIONS

The contribution of our paper can be seen at two levels. At a specific level, it addresses an area that is relatively ignored in the construction engineering literature, namely, DQ/IQ. Existing DQ/IQ assessment frameworks and tools are developed based on assumptions that are not valid for construction engineering projects. By situating the research in construction engineering, it brings to light the importance of characteristics and features that are unique to a specific context in understanding and subsequently solving DQ/IQ problems. Moreover, by applying a design research methodology, it demonstrates the usefulness of taking a proactive approach in meeting DQ/IQ challenges. For practice, the tool itself is a useful artifact. More importantly, the design principles can guide the development and enhancement of such tools in the future. The relevance of the tool is self-evident: it was developed, implemented and tested in a real organization. That the organization was a large global construction engineering firm argues further for its relevance. Beyond demonstrating the efficacy of the tool, our findings validate the design principles on which the tool was built.

At a more general level, our paper adds to the small but rapidly growing body of literature in construction engineering management on the role of information and communication technologies (ICT) in research and practice. Scholars in the field have long bemoaned the lack of serious attention paid to ICT in the literature beyond descriptions of tools that are used (Ahmad et al. 1995). Specifically noted is the need to examine the transformational capability of ICT and the related socio-organizational context of its use (Ahmad et al. 2010; Gal et al. 2008). (This was the topic of a spirited panel discussion at the conference organized in 2007 at the University of Reading, UK to mark the 25th anniversary of one of the premium journals of the field *Construction Management & Economics.*)

The recent trend, however, is encouraging (for a thorough review specifically on BIM research, see Merschbrock and Munkvold (2012)). Several papers have gone beyond simply describing cases of implementation of an ICT tool in a construction project and have discussed organizational implications (e.g. Azhar and Ahmad 2007; Gal et al. 2008). Our paper contributes to this healthy trend. We specifically used an action-oriented approach in our project. It can be seen at first glance as an illustration of action research in construction management as described by Azhar et al. (2010). By intervening in a real organization and designing an ICT system to solve a practical problem and generating scientific knowledge, our project is similar to previously reported work in construction management (Azhar 2005; Cushman 2001; Rezgui 2007).

Our approach has a nuanced difference: we frame it as being more in line with action design research (ADR) as elaborated by Sein et al. (2011). ADR blends essential elements of action research with design research to generate design knowledge in a real practical setting. While AR requires intervention in an actual organization, it does not necessarily require design of a technological system. For example, such intervention could be a managerial action (e.g. Cushman 2001) or studying the effectiveness of implementation of technology (e.g. Rezgui 2007). Even where designing a system is an integral part of the intervention, it can be a tried and tested technology such as data warehousing (Azhar and Ahmad 2007). These types of AR can create generalizable and transferable knowledge that can be applied across a wider context than the project itself. An excellent example is the five-phase framework developed by Azhar (2005) for implementing a data-warehousing solution in a construction organization, it also requires that the research must generate new design rules and principles specifically for the design of the system itself. In this regard, it is different from straightforward design projects (e.g. Ahuja et al. 2010; Zhou et al. 2011) where actual intervention in an organization is not required.

These characteristics make ADR an exceptionally appropriate research approach for the construction engineering and management context since innovative and creative design is a vital part of new knowledge created in the field. The design, of course, is not limited to ICT, but also applies to the entire spectrum that is part of the eclectic field of construction management.

CONCLUSION

We conclude by raising some specific questions and offering some answers:

• How can the learning from the development and evaluation of IQS be transferred to other contexts?

This addresses the question of abstraction of knowledge one gains from an ADR project. In such projects, abstraction of knowledge requires that the specific artifact built must be viewed as an instantiation of a class of system and a class of

problem. IQS belongs to the general class of DQ/IQ assessment tools. The way IQS was developed and the eventual principles embedded in it could serve future researchers and developers with guidelines and principles when developing DQ/IQ assessment tools for contexts similar to ours. The context was large-scale and complex projects that involved highly sophisticated tasks that required a broad variety of expertise and involved huge amounts of data, which constituted an important part of the deliverables. It remains to be seen whether this approach would be equally effective in a smaller organization or for simpler tasks.

• To what extent have we contributed to theory building in DQ/IQ in general and to construction engineering in particular?

By locating our research in the specific context of construction engineering, and addressing its unique characteristics, we have contributed to contextualization of the concept of DQ/IQ. Using ADR as the research method was instrumental in this. However, one may argue that we have simply demonstrated the "proof of concept". Moreover, contextualization is just one aspect of DQ/IQ theory, and we only used a subset of the DQ/IQ dimensions. Much research needs to be done to substantially enhance DQ/IQ theory. For the field of construction engineering, this study particularly brings forward the importance of managing data and information in a controlled manner throughout the project. To achieve a sufficient level of DQ/IQ, adequate tools are needed for assessment. IQS provide one example of such tools, and the evaluation performed in this study indicates that there is a good foundation for using such a tool.

• What concrete guidelines emerge for practice from projects such as ours?

ADR is rooted in pragmatism, and the utility of the research outcome for practice is essential in ADR projects. Ours was no exception. In briefly commenting on the knowledge outcomes of the project above, we have also indicated the learning opportunities for developers. In addition to the design principles and lessons from the development process, the artifact itself, IQS, has relevance for practice. Yet, as ADR stresses, an artifact is in continuous emergence, and its final form will only emerge after it has been in use for a reasonable period of time. This emergence will result from the context in which it is used and the exigencies of its actual use. Design principles, therefore, will continue to emerge.

• How can a possible enhancement in project performance be measured if improvement in DQ/IQ is achieved?

Our study was motivated mainly to reduce delays and cost overruns in large construction engineering projects. Our investigation first led us to errors in drawings and specifications and eventually to DQ/IQ. Our solution to this problem was to build IQS. We evaluated the tool and found that it indeed improved DQ/IQ. However, we did not go "the last mile" by assessing whether this reduction in DQ/IQ improved project performance. This is a fruitful, and essential, extension of projects such as ours. One possible way to measure improvement in project performance could be to follow Wambeke et al. (2011) by focusing solely on quality in drawings and specifications used at construction sites. Another approach is to measure employee-hours used on site (assembly) to see if they decrease if the DQ problems are resolved before the drawings are handed over to assembly. However, this is not a simple task since all these projects are one-of-a-kind and difficult to compare. An alternative is to measure extra man-hours used on-site for locating missing information or correcting information.

REFERENCES

- Ahmad, I., Russell, J. S. and Abou-Zeid, A. (1995). Information technology (IT) and integration in the construction industry, *Construction Management and Economics*, 13, 2, 163-171.
- Ahmad, I., Sein, M. K. and Panthi, K. (2010). Challenges of integration and ICT's potentials in the globalized construction industry. *Proceedings of the Technology Management for Global Economic Growth (PICMET), 2010 Proceedings of PICMET'10:.*
- Ahuja, V., Yang, J. and Shankar, R. (2010). Benchmarking framework to measure extent of ICT adoption for building project management, *Journal of Construction Engineering and Management*, 136, 5, 538-545.
- Assaf, S. A. and Al-Hejji, S. (2006). Causes of delay in large construction projects, *International Journal of Project Management*, 24, 4, 349-357.
- Azhar, S. (2005). Information systems to support decision-making in construction owner organizations/ A data warehousing approach. Ph. D. dissertation, Florida International University, Miami.
- Azhar, S. and Ahmad, I. (2007). A case study of data warehousing implementation in a public construction owner organization. In S.-H. Song (Ed.), *Construction data warehouse* (2 ed., pp. 15-31). Seoul National Univ., Seoul, Korea.
- Azhar, S., Ahmad, I. and Sein, M. K. (2010). Action Research as a Proactive Research Method for Construction Engineering and Management, *Journal of Construction Engineering and Management*, 136, 1, 87-98.
- Batini, C., Cappiello, C., Francalanci, C. and Maurino, A. (2009). Methodologies for data quality assessment and improvement, ACM Computing Surveys, 41, 3, 1-52.
- Blechinger, J., Lauterwald, F. and Lenz, R. (2010). Supporting the Production of High-Quality Data in Concurrent Plant Engineering Using a MetaDataRepository. *Proceedings of the Americas Conference on Information Systems*, August 12-15, Lima, Peru.
- Braimah, N. and Ndekugri, I. (2009). Consultants' perceptions on construction delay analysis methodologies, *Journal of Construction Engineering and Management*, 135, 12, 1279-1288.
- Carter, N. and Thorpe, T. (2006). Internet provision on small construction sites: The i-MO solution. *Proceedings of*, Proceedings of the 6th European Conference on Product and Process Modelling - eWork and eBusiness in Architecture, Engineering and Construction, ECPPM 2006, pp. 391-396.
- Cherns, A. B. and Bryant, D. T. (1984). Studying the client's role in construction management, *Construction Management* and Economics, 2, 2, 177-184.
- Cushman, M. (2001). Action research in the UK construction industry-the B-hive project. *Proceedings of the IFIP WG 8.2* Working Conference on Realigning Research and Practice in Information System Development: The Social and Organizational Perspective, Boise State University, Boise, Id.
- Dai, J., Goodrum, P. M. and Maloney, W. F. (2009). Construction Craft Workers' Perceptions of the Factors Affecting Their Productivity, *Journal of Construction Engineering & Management*, 135, 3, 217-226.
- Dobson, D. and Martinez, R. (2007). Integrated engineering, ABB Review, SPEC. REP., 48-52.
- Elazouni, A. and Salem, O. A. (2011). Progress monitoring of construction projects using pattern recognition techniques, *Construction Management and Economics*, 29, 4, 355-370.
- English, L. P. (1999). Improving data warehouse and business information quality: methods for reducing costs and increasing profits: John Wiley and Sons, Inc., New York, New York, USA, 1999.
- English, L. P. (2003). Total information quality management: a complete methodology for IQ management, *Dm Review (Now : Information Management)*, 9, 1-7.
- Frimpong, Y., Oluwoye, J. and Crawford, L. (2003). Causes of delay and cost overruns in construction of groundwater projects in a developing countries; Ghana as a case study, *International Journal of Project Management*, 21, 5, 321-326.
- Gal, U., Lyytinen, K. and Yoo, Y. (2008). The dynamics of IT boundary objects, information infrastructures, and organisational identities: the introduction of 3D modelling technologies into the architecture, engineering, and construction industry, *European Journal of Information Systems*, 17, 3, 290-304.
- Ge, M., Helfert, M. and Jannach, D. (2011). Information quality assessment: Validating measurement dimensions and processes. *Proceedings of the European Conference on Information Systems*, June 9 11, Helsinki, Finland.
- Lee, Y. W., Strong, D. M., Kahn, B. K. and Wang, R. Y. (2002). AIMQ: a methodology for information quality assessment, *Information & Management*, 40, 2, 133-146.
- Lim, C. S. and Mohamed, M. Z. (2000). An exploratory study into recurring construction problems, *International Journal of Project Management*, 18, 3, 267-273.
- Lin, S., Gao, J. and Koronios, A. (2008). A data quality framework for engineering asset management, Australian Journal of Mechanical Engineering, 5, 2, 209-219.

- Lin, S., Gao, J., Koronios, A. and Chanana, V. (2007). Developing a data quality framework for asset management in engineering organisations, *International Journal of Information Quality*, 1, 1, 100-126.
- Ling, F. Y. Y., Pham, V. M. C. and Hoang, T. P. (2009). Strengths, Weaknesses, Opportunities, and Threats for Architectural, Engineering, and Construction Firms: Case Study of Vietnam, *Journal of Construction Engineering* and Management, 135, 10, 1105-1113.
- Madnick, S. E., Wang, R. Y., Lee, Y. W. and Zhu, H. (2009). Overview and Framework for Data and Information Quality Research, *Journal of Data and Information Quality*, 1, 1, 1-22.
- Memon, A. H., Abdul Rahman, I. and Abdul Azis, A. A. (2011). Preliminary Study on Causative Factors Leading to Construction Cost Overrun, *International Journal of Sustainable Construction Engineering and Technology*, 2, 1.
- Merschbrock, C. and Munkvold, B. E. (2012). A Research Review on Building Information Modeling in Construction—An Area Ripe for IS Research, *Communications of the Association for Information Systems*, 31, 1, 10.
- Miller, R. and Lessard, D. (2001). Understanding and managing risks in large engineering projects, *International Journal of Project Management*, 19, 8, 437-443.
- Murphy, G. D. (2009). Improving the quality of manually acquired data: Applying the theory of planned behaviour to data quality, *Reliability Engineering and System Safety*, 94, 12, 1881-1886.
- Neely, M. P., Lin, S., Gao, J. and Koronios, A. (2006). The Deficiencies of current data quality tools in the realm of engineering asset management. *Proceedings of the Americas Conference on Information Systems*, August 4 6, Acapulco, México.
- NORSOK. (1998). D-001 NORSOK STANDARD, Drilling facilities. http://www.standard.no/pagefiles/1318/d-001.pdf.
- Ogunlana, S., Lim, J. and Saeed, K. (1998). Desman: A dynamic model for managing civil engineering design projects, *Computers & structures*, 67, 5, 401-419.
- Pipino, L. L. and Lee, Y. W. (2007). Applying IT to Healthcare: Humans, Errors, and a Data Quality Perspective. *Proceedings of the Americas Conference on Information Systems*, August 10 - 12, Keystone, Colorado.
- Pipino, L. L., Lee, Y. W. and Wang, R. Y. (2002). Data Quality Assessment, Communications of the ACM, 45, 4ve, 211-218.
- Rezgui, Y. (2007). Exploring virtual team-working effectiveness in the construction sector, *Interacting with Computers*, 19, 1, 96-112.
- Rivas, R. A., Borcherding, J. D., González, V. and Alarcón, L. F. (2010). Analysis of factors influencing productivity using craftsmen questionnaires: Case study in a Chilean construction company, *Journal of Construction Engineering and Management*, 137, 4, 312-320.
- Sambasivan, M. and Soon, Y. W. (2007). Causes and effects of delays in Malaysian construction industry, *International Journal of Project Management*, 25, 5, 517-526.
- Sein, M. K., Henfridsson, O., Purao, S., Rossi, M. and Lindgren, R. (2011). Action Design Research, *MIS Quarterly*, 35, 1, 37-56.
- Singh, P., Park, I., Lee, J. K. and Rao, H. R. (2009). Information sharing: a study of information attributes and their relative significance during catastrophic events, *Cyber-Security and Global Information Assurance: Threat Analysis and Response Solutions, IGI Publishers*.
- Toor, S.-U.-R. and Ogunlana, S. O. (2008). Problems causing delays in major construction projects in Thailand, *Construction Management & Economics*, 26, 4, 395-408.
- Tribelsky, E. and Sacks, R. (2011). An empirical study of information flows in multidisciplinary civil engineering design teams using lean measures, *Architectural Engineering and Design Management*, 7, 2, 85-101.
- Vician, C. (2011). A Resource Review for Health Care Quality Execution: Business Intelligence Strategy. *Proceedings of the Americas Conference on Information Systems*, August 5 7, Detroit, Michigan.
- Wambeke, B. W., Hsiang, S. M. and Liu, M. (2011). Causes of variation in construction project task starting times and duration, *Journal of Construction Engineering and Management*, 137, 9, 663-677.
- Wang, R. W. and Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers, Journal of Management Information Systems, 12, 4, 5-33.
- Westin, S. (2012). Data and Information Quality Research in Engineering Construction Projects: A Review of Literature. *Proceedings of the 21st International Conference on Information Systems Development (ISD2012)*, Prato, Italy, August 29-30, 2012.
- Westin, S., Holen, O. B. and Päivärinta, T. (2012). Data and Information Quality Dimensions in Engineering Construction Projects. *Proceedings of the Americas Conference on Information Systems*, Seattle, Washington, USA.
- Westin, S. and Päivärinta, T. (2011). Information quality in large engineering and construction projects: A Delphi Case Study. *Proceedings of the European Conference on Information Systems*, June 9 11, Helsinki, Finland.
- Yeo, K. and Ning, J. (2002). Integrating supply chain and critical chain concepts in engineer-procure-construct (EPC) projects, *International Journal of Project Management*, 20, 4, 253-262.

Zhou, Z., Li, Q. and Wu, W. (2011). Developing a Versatile Subway Construction Incident Database (SCID) for the Safety Management, *Journal of Construction Engineering and Management*, 138, 10, 1169-1180.

APPENDIX A

Example of Syntax Code Rule2: X, Y, Z outside area limits SELECT tbl.project_code, tbl.contractor code, tbl.tagno, 'X=' VARCHAR(tbl.x_min)+ ',Y='+VARCHAR(tbl.y_min)+ ',Z='+VARCHAR(tbl.z_min), tbl.description, tbl.upd_by, tbl.upd_date, tbl.upd_code, tbl.po_no, tbl.discipline_code, tbl.product_code, tbl.area, tbl.system_no, tbl.site_code, ", ", tbl.status3, FROM table1 tbl, table2 b WHERE tbl.project_code = b.project_code tbl.area = b.areaAND tbl.x $\min \ll 0$ AND AND $tbl.y_min \ll 0$ AND $tbl.z_min <> 0$ (tbl.x_min NOT BETWEEN b.x_min AND b.x_max AND OR tbl.y_min NOT BETWEEN b.y_min AND b.y_max OR tbl.z_min NOT BETWEEN b.z_min AND b.z_max) (tbl.eq_pack_no = 'NA' AND OR tbl.eq_pack_no = " OR tbl.eq_pack_no = tbl.tagno) AND tbl.upd code <> 'D' AND tbl.project code = 'NN' tbl.contractor_code = 'ZZ' AND