POZNAŃ UNIVERSITY
OF ECONOMICS
AND BUSINESS

Michael Vaknin

Information Quality Framework for the Design and Validation of Business Processes

Doctoral dissertation

PhD Supervisor: Prof. dr. hab. Witold Abramowicz

Auxiliary PhD Supervisor: Dr. habil. Agata Filipowska

Date of submission: ……………………………………..

Supervisor's signature

Poznań 2020

Acknowledgements

This dissertation was made possible by the support and assistance of a number of people, to whom I would like to express my sincere gratitude.

First of all, I would like to express my special gratitude and appreciation to my main supervisor Prof. dr. hab. Witold Abramowicz for all support and encouragement. You have been a tremendous mentor for me. I would like to thank you for introducing me to the research world, for encouraging my research and for allowing me to grow as a research scientist.

Furthermore, I wish to express my sincere and deep appreciation to my auxiliary supervisor Dr. habil. Agata Filipowska, who is much more than an academic advisor to me. Her intelligent supervision, constructive guidance, supportive and encouraged instruction, inspiration and friendship, have been priceless and the key to my motivation and success throughout my research.

Thank you Agata, from the bottom of my heart for being there for me through all of the process of this dissertation.

Besides my supervisors, I am grateful to Prof. dr. hab. Tadeusz Kowalski, Prof. dr. hab. Sylwester Bialowas, head of DSE program, Dr. Piotr Michoń and Dr. Orit Haller-Hayon for their assistance, support and encouragement along the process.

Last but not least, a special thanks to my beloved family, who's give up on many quality hours together over the last years and for their love, support, patience, encouragement, back-up and understanding during all my research process.

Table of Contents

List of figures

List of tables

List of Abbreviations and Terms

AD	Activity Diagram
BP	Business Process
BI	Business Intelligence
BPI	Business Process Improvement
BPiDQ	Business Process including Data Quality
BPM	Business Process Management
BPMN	Business Process Management Notation
BPMS	Business Process Management System
BPR	Business Process Reengineering (Redesign)
BPSS	Business Process Support System
BPQMM	Business Process Quality Characteristics Meta-Model
CA	Competitive Advantage
CDQM	Corporate Data Quality Management
CFG	Confirmatory Focus Group
CIO	Chief Information Officer
DG	Data Governance
DGI	Data Governance Institute
DM	Dependency Matrix
DMAIC	Define, Measure, Analyze, Improve and Control
DQ	Data Quality
DQDPM	Data Quality Deficiencies Prediction Method
DQEF	Data Quality Engineering Framework
DQM	Data Quality Management
DQR	Data Quality Requirement
DSR	Design Science Research
DWH	Data Warehouse
EDQM	Enterprise Data Quality Management
EFG	Exploratory Focus Group
ERP	Enterprise Resource Planning
ETL	Extract-Transform-Load (of data)
FEDS	Framework for Evaluation in Design Science
FG	Focus Group
FR	Functional Requirement
GDPR	Global Data Protection Regulation
ICT	Information and Communication Technology
IMS	Information Manufacturing System
IP-MAP	Information Product Map
IQ	Information Quality

IS	Information System
ISO	International Organization for Standardization
IT	Information Technology
ISAD	Information Systems Analysis and Design
KM	Knowledge Management
MIS	Management Information Systems
MoRs	Matrix of Relations
NFR	Non-Functional Requirement
NISS	National Institute of Statistical Sciences
OLAP	On-Line Analytical Processing
PAIS	Process-Aware Information System
PDDQM	Process-Driven Data Quality Management
PIQ	Perceived Information Quality
PRCA	Process Root-Cause Analysis Approach
PSP/IQ	Product and Service Performance for Information Quality
QoBP	Quality of Business Process
QMS	Quality Management System
R&D	Research and Development
RE	Requirements Engineering
RW	Real World
SCA	Sustainable Competitive Advantage
SDLC	System Development Life Cycle
SRS	System Requirements Specification
TDQM	Total Data Quality Management
TQM	Total Quality Management
TTF	Task Technology Fit
UML	Unified Modeling Language
WFM	Workflow Management

Keywords

Business Process; Business Process Analysis; Business Process Improvement; Business Process Design; Information Systems; Information Technology; Information Systems Design; Data Dependency; Data Quality; Information Quality; Data Quality Dimensions; Data Quality Requirements; Focus Group.

Part I: Foundation

1 Introduction

1.1. General Background

Nowadays, in order to survive in a highly competitive and uncertain environment, organizations seek to take advantage of all their internal and external potentials, advantages, and resources while facing numerous challenges especially in efficiently and effectively executing their business processes (BPs). Therefore, they require strong information systems (IS) - business alignment (Ullah & Lai, 2011; Mondragón et al., 2013; Rahimi et al., 2014; Lahajnar & Rožanec, 2016; Alotaibi & Liu, 2017).

To sustain themselves within the market, these organizations often need to self-examine and must find ways to better manage and adapt their BPs and their structure in response to changes, trends and developments in the business environment (Daoudi & Nurcan, 2007; Sun & Zhao, 2013; Gartner, 2013a; 2016). According to Gartner researches (2013a; 2016), companies must adjust these business processes inter alia in order to reduce costs, conserve cash and out-innovate their competitors. In order to meet customer demands and stand on these challenges, organizations need to improve their agility i.e. the ability of an organization to sense environmental change and respond efficiently and effectively to change (Gartner, 2009).

Currently, one of the most interesting contemporary issues in the business processes' domain is business process (BP) design and validation. Recently, the main focus of researchers has been placed on the BPs alignment and business process support systems (BPSS) with regard to data and information quality aspects (Madnick et al., 2009; Soffer, 2010; Ullah & Lai, 2011; Mondragón et al., 2013; Zhu et al., 2014; Vaknin & Filipowska, 2016; 2017; Heinrich et al., 2017). In a global business environment and in the age of the digital market, firms must learn the best way to exploit information systems' capabilities for business performance improvement and operation flexibility (Biehl, 2007).

1.2. Research Motivation

1.2.1. The organizational challenges

Every organization has to manage a number of processes (Dumas et al., 2018). In order to be successful, organizations often need to self-examine and adapt their business processes and their structure in response to changes, trends and developments, opportunities and threats in the business environment (Daoudi & Nurcan, 2007; Kirchmer, 2010; Sun & Zhao, 2013; Gartner, 2013a; 2016). A business process (BP) refers to a sequence of activities that are performed in coordination in an organizational and technical environment (Weske, 2012; 2019). These activities jointly realize a business goal, reflected by data items in process representation (Soffer & Wand, 2007).

Business processes consume and produce data and information and are based on information flows (English, 2001; Schultz, 2006), and "many business processes leave their 'footprints' in transactional information systems" (Aalst, 2005, p.198). An information system (IS) can be defined as “a set of interrelated components that collect (or retrieve), process, store, and distribute information to support decision making and control in an organization” (Laudon & Laudon, 2018; p.48).

Most organizations use data and information in two ways: transactional/operational use (“running the business”), and analytic use (“improving the business”) (Loshin, 2011). In today’s business environment, any organization exposed to competition must define its business process information needs, identify appropriate resources, filter out the information, assimilate it, and use it to its advantage. The growing field of business process management (BPM) focuses on methods and tools for designing, enacting, and analyzing business processes (Russell et al., 2016).

Many organizations face many challenges such as the rise in frequency of goods ordered, quick decision-making, fast information transfer, the ability to adapt to changes in demand, improve their customers' satisfaction and others. IS has been harnessed to manage BPs to deal with these challenges (Ko et al., 2009; Alotaibi & Liu, 2017). Furthermore, IS are pervasive in all forms of business organizations (Ibrahim et al., 2013; Zhu et al., 2014; Abbasi et al., 2016; Bai et al., 2018) and IT infrastructure is found to have a significant effect on customers' focus and responsiveness (Bhatt & Emdad, 2010). Many organizations seek ISs development projects that will provide them with a competitive advantage (Stair & Reynolds, 2010) and it can be regarded as a key instrument for organizing and managing the firm (Ulrich, 2014). The increasing integration of information systems for the operation of business processes provides the basis for innovative data analysis approaches (Chintalapati et al., 2014).

Basically, IS are complex and implemented within an organization for the purpose of improving the effectiveness and efficiency of that organization (Hevner et al., 2004). Successful accomplishment of an information system (IS) project is a timeless goal and remains a crucial challenge for IS researchers and practitioners (Glowalla & Sunyaev, 2015; Johnson & Mulder, 2016). The interactions among business processes, information flows and information systems are critical to the success of information systems implementation (Heinrich et al., 2017; Laudon & Laudon, 2018). Furthermore, in most organizations IS supports the major business functions and business processes are increasingly becoming more supported, controlled and/or monitored by information systems through business process support systems (BPSS) (Aalst, 2005; Wetzstein et al., 2007; Soffer, 2010; Ullah & Lai, 2011; Ibrahim et al., 2013; Rahimi et al., 2014; Kerpedzhiev et al., 2017; Weske, 2012; 2019). Additionally, building a responsive Information systems (IS) infrastructure depends highly on an appropriate determination of business process information needs (Al-Mashari & Zairi, 1999; Rahimi et al., 2014). The collection, representation, and effective use of organizational information are important to business because these activities facilitate the increasingly important analysis needed for business operations and business analytics (Loshin, 2011; Storey et al., 2012).

1.2.2. Business processes and information systems alignment

The alignment issue is a central part of Information systems analysis and design (ISAD) processes, which are basic topics in the IS curriculum (Iivari et al., 2006). Over the years, many researchers (e.g. Regev et al., 2005; Madnick et al., 2009; Soffer, 2010; Ullah & Lai, 2011; Loucopoulos & Heidari, 2012; Rahimi et al., 2014; Zhu et al., 2014; Alotaibi & Liu, 2017; Heinrich et al., 2017) have been focused on the alignment aspects between business processes (BPs) and business process support systems (BPSS). A BPSS provides the possibility of running a business in a new, process-oriented way, which is more effective and efficient (Regev et al., 2005). Business processes (BPs) and information systems (ISs) mutually affect each other. However, the complex interrelations between them are not adequately understood and considered in development so far (Heinrich et al., 2017). In order to achieve and maintain a correspondence between them and to benefit fully from the potentials of modern IS, both need to be understood thoroughly and coherently.

Missing alignment between BPs and ISs may raise quality issues and problems which can impair the quality of BPs and ISs performance (Heinrich et al., 2017). While the need for the alignment of BPs and their support ISs has been emphasized and discussed, there is still a great need for systematic approaches, methods and tools to improve BPs and to deal with BPs design with regard to data quality aspects (Madnick et al., 2009; Heinrich & Paech, 2010; Soffer, 2010; Cappiello et al., 2013; Mondragón et al., 2013; Zhu et al., 2014). Moreover, a research in leading business and management journals has revealed that there is little exploration into the theory of process design (Patel, 2007) and only general work conducted in relation to process quality (Heravizadeh et al., 2009).

One prominent topic closely related to the alignment of BPs and their support ISs is data or information quality (DQ/IQ). Since data and information are critical resources and critical assets, used to conduct every organizational operation, the information flow in BP has become a critical issue and it should be examined with a focus on quality aspects (Redman, 2004; Soffer & Wand, 2007; Hamzah et al., 2014; Dumbleton & Munro, 2015: KPMG, 2018).

The correctness, effectiveness, and efficiency of the BPs supported by ISs are becoming vital to the organization (Aalst, 2008; Recker et al., 2011; Timmerman & Bronselaer, 2019). Such processes, in practice, can suffer from issues such as a poor level of data quality along their activities or in communication between processes. Data quality is a discipline focused on ensuring that the condition of data is fit for use in existing business operations and processes (Gartner, 2016). Most organizations depend on quality information for everyday business operations (Baškarada & Koronios, 2014). In addition, there is an evident need to incorporate data quality considerations into the whole data cycle, encompassing managerial and governance as well as technical aspects (Blake & Mangiameli, 2011; Sadiq, 2013).

According to Komai et al. (2017) the success of IS/IT system development largely depends on the System Requirements Definition (SRD) phase. A substantial part of IS/IT system development failures can be attributed to problems that arise during systems analysis (Wand & Weber, 2002; Iivari et al., 2006) caused by poor requirements (Info-Tech Research, 2008). Although these two components are central topics in the IS discipline and even though the challenge of improving them exists, it remains somewhat at the periphery of research and has received relatively little attention (Iivari et al., 2006). Systems development methodologies, methods or processes in the context of solving real-world business processes problems with data quality concerns at analysis and design stages, are still required (Iivari et al., 2006; Mondragón et al., 2013).

Data quality (DQ) is an important area of research and investment in information systems. It is an integral and inseparable issue from information systems domain, and remains a persistent problem in practice and a challenge for research, supported by the potentially high impact of poor data in organizations (Tayi & Ballou, 1998; Xu et al., 2002; Blake & Mangiameli, 2011; Laranjeiro et al., 2015: KPMG, 2018; Timmerman & Bronselaer, 2019).

The extent to which data can be trusted has often set apart successful IS projects from failed ones. Low level of data creates missed opportunities and customer satisfaction issues, which can lead to other problems in organization performance (Hiskey, 2018). Furthermore, DQ is a critical issue to organizational success and increases the effectiveness of business processes (Tee et al., 2007; Madnick et al., 2009; Otto, 2011; Shariat Panahy et al., 2013; Glowalla & Sunyaev, 2013a; 2014a; Goetz et al., 2015). Moreover, it is a critical factor for achieving strategic and operational business goals. Among these goals are improved decision-making (Batini & Scannapieca, 2006; Vaisman, 2006; Ofner et al., 2012; Shariat Panahy et al., 2013; Zhu et al., 2014; KPMG, 2016;2018; Belhiah et al., 2016; Gharib et al., 2018).

According to Forrester research report (2015), the data quality solutions market is growing because more enterprise architecture professionals see data quality to address their top challenges (Goetz et al., 2015). Improving DQ often requires modifying business processes, and enriching them with additional activities (Cappiello et al., 2013; KPMG, 2016). In addition, observations in organizations show that DQ considerations are not taken into account when it comes to deciding about BP redesign and it is expected that most of those initiatives will fail due to a lack of attention to quality aspects, and DQ in particular (Ofner et al., 2012).

1.2.3. The entanglement and failure syndrome of IS/IT projects

The high rate of problematic or failed information system (IS) projects is an ongoing problem with high relevance for researchers and practitioners (Cecez-Kecmanovic et al. 2014; Glowalla & Sunyaev, 2015). Many recent studies (Nasir & Sahibuddin, 2011; Ibrahim et al., 2013; Imtiaz et al., 2013; Nwakanma et al., 2013; Rajkumar & Alagarsamy, 2013; Chandrasekaran et al., 2014; The Standish Group, 2015; 2018; Komai et al., 2017; Hiskey, 2018;

Rosato, 2018) show that most of information systems or information technology (IS/IT) projects have "challenged" or "failed", in the combination of budget and/or schedule overruns and/or for not meeting customers' expectations and users' requirements. Challenged IS/IT projects are projects that were completed but late, over budget, and did not meet the target specifications and/or with less than the required features and functions. Failed projects were cancelled prior to completion or delivered and never used (The Standish Group, 2018).

For many years, the CHAOS reports have been published by Standish Group every year since 1994 and are a snapshot of the state of the IS/IT projects and software development industry and their evaluations cover countries all over the world (Gaikema et al., 2019). The results indicate that basically there is still work to be done around achieving successful outcomes from IS/IT and software development projects. For example, the 2018 CHAOS Report has been studied over 50,000 IS/IT projects around the world, ranging from small-size to large-size systems, and shows that 70% of projects in 2018 were challenged or failed and only 30% were successful projects; i.e. conducted on time, on budget and met user requirements with required features and functions (The Standish Group, 2018). Moreover, according to Consortium for IT Software Quality (CISQ) report (Krasner, 2018), the cost of poor quality software in the US in 2018 was approximately $2.84 trillion.

Table 1 summarizes the outcomes of IS/IT projects over five years (2011-2015) using the definition of success factors (on time, on budget with a satisfactory quality result).

Table 1. Project resolution results from CHAOS research for years 2011 to 2015

	2011	2012	2013	2014	2015
SUCCESSFUL	29%	27%	31%	28%	29%
CHALLENGED	49%	56%	50%	55%	52%
FAILED	22%	17%	19%	17%	19%

[Source: The Standish Group, 2015]

Furthermore, many of Standish Group chaos reports (e.g. in 2013; 2014; 2015; 2018) show that incomplete or unclear requirements and specifications errors are the critical factors and the most common cause of failure in IS/IT development projects and they are the top factors about why IS/IT projects are impaired and ultimately cancelled.

Other empirical studies (Wand & Weber, 2002; Moody et al., 2003; Kappelman et al., 2009; Nasir & Sahibuddin, 2011; Kaur & Sengupta, 2011; Nwakanma et al., 2013; Komai et al., 2017; Gaikema et al., 2019) show that more than half of errors which occur during IS development are the result of inaccurate or incomplete requirements and errors in systems specifications. Even when these projects are completed, many are no more than a mere shadow of their original specification requirements (The Standish Group, 2014) and many of them have failed to achieve the expected benefits (Sadiq, 2013). In addition, according to KPMG Survey (2016), on average about 70% of all IS/IT related projects fail to meet their objectives. According to Info-Tech Research Group (2008) 60-80% of IS/IT project failures are attributed to poor requirements gathering and analysis and flawed requirements trigger 70% of IS/IT project failures.

Verner (2008) points out that most IS/IT experts agree that failures in IS/IT developments occur far more often than they should. Most of these projects are expensive, with a difficult process affected by a series of problems including poor project management, cost and schedule overruns and poor-quality software. Verner (2008) also summarizes the main factors described in the literature that can be attributed to problems and failures in IS/IT projects include poor user requirements and poor requirements specification. Moreover, poorly set requirements and lack of requirements understanding are closely linked to lack of customer involvement (Rajkumar & Alagarsamy, 2013).

1.2.4. The main challenges and motivation

Data is the foundation of the digital economy (Otto & Österle, 2015). Data is both a critical asset and resource in every organization and it is important for decision making processes (Tayi & Ballou, 1998; Redman, 2004; Tee et al., 2007; Loshin, 2011; Dumbleton & Munro, 2015; KPMG, 2018; Zhang et al., 2019). Decision-making involves large data volumes and includes a wide variety of decision-tasks (Shankaranarayanan & Cai, 2006). Moreover, the quality of the data affects the quality of decisions (Jung, 2004; Price & Shanks, 2005; Bagchi et al., 2006; Vaisman, 2006; Frank, 2008; Xingsen et al., 2009; Ofner et al., 2012; Cappiello et al., 2013; Liu et al., 2020). In particular, the data flow in processes and along their activities is the basis for data requirements representing in information systems (IS) at analysis and design stages.

Many organizations are suffering from poor data quality (DQ) (Wand & Wang, 1996; Laranjeiro et al., 2015) and they are starting to realize that poor DQ is hurting them. In addition, many researchers (e.g. Redman, 1998; 2004; Haug et al., 2011; Ofner et al., 2012; Glowalla & Sunyaev, 2013a; Laranjeiro et al., 2015) declare that poor DQ increases operational cost and huge amounts of resource and money are spent in organizations due to poor data quality and for detecting and correcting errors to improve the quality of the data.

According to Haug et al. (2011), 75 percent of organizations have identified costs stemming from poor data. In addition, Gartner research (2011), shows that organizations cope with significant cost impact from data quality deficiencies and there is a growing awareness in all industries of data quality issues. Hence, especially in difficult economic times, organizations of all types must focus on controlling and improving DQ to minimize disruptions and losses and as a way of reducing costs and improving efficiency (Gartner, 2011). However, few organizations know how to address the issue or where to begin and for this reason, data quality methodologies, methods and tools can provide added value to organizations as a means for guidance.

Generally, the idea of integrating DQ issues into BP models as such is not new. Different approaches and measures have been developed over the years to cope with information quality assessment needs, but the notion of information quality within BP design and design methods has received relatively little attention (Patel, 2007; Blake & Mangiameli, 2011; Ofner et al., 2012). Hence, there is a great challenge concerning provision of methods, tools and methodologies that would support the design of processes to be robust and avoid problems related to data quality dimensions (Soffer, 2010; Ofner et al., 2012; Cappiello et al., 2013; Sadiq, 2013; Alshazly et al., 2014; Batini & Scannapieca, 2016; Jaya et al., 2017). The existing literature in business processes design and validation lacks models and methods for dealing with quality of information flows on BPs (Soffer 2010; Cappiello et al., 2013; Sadiq, 2013; Sun & Zhao, 2013; Zhu et al., 2014; Jugulum, 2016; Jaya et al., 2017; Heinrich et al., 2018).

A high-level of DQ plays a significant role and is essential in a variety of relevant areas that use computing capabilities and infrastructures of IS. For example, in the healthcare sector, which is an industry that collects a large amount of data, health data quality is essential for the proper delivery of health services especially in emergency departments (Juddoo & George, 2018; Vanbrabant et al., 2019) and for centralizing patients' clinical data from distributed research networks (Juárez et al., 2019). Furthermore, data quality is critical to adequately performing management accounting (MA) tasks (Knauer et al., 2020).

DQ has an essential role in advanced information technologies (ITs) as part of an era of Industry 4.0 (Otto & Österle, 2015). Poor DQ affects analytical results from business intelligence (BI) tools, extract, transform and load (ETL) tools and data warehouses (DWH) and causes severe losses to organizations (Souibgui et al., 2019; Zhang et al., 2019). Moreover, DQ is an essential requirement to ensure the reliability and quality of Machine Learning-based Software Systems (Foidl & Felderer, 2019). In the Big Data era, DQ faces many challenges and high-quality data are the precondition for analyzing and using big data and for guaranteeing the value of the data for achieving effective decision-making and improving many business functions (Abdullah et al., 2015; Cai & Zhu, 2015; Laranjeiro et al., 2015; Abbasi et al., 2016; Khan & Vorley, 2017; Yang et al., 2017; Juddoo & George, 2018). In addition, data mining as one of the most important sources of knowledge needs high quality data to mine, but data of sufficient quality is often lacking (Xingsen et al., 2009).

Today DQ goes big-time with Global Data Protection Regulation (GDPR) which entered into effect in Europe in 2018 and likely spills over into the US businesses (Hiskey, 2018). Today, organizations need to cope with novel legal requirements regarding the processing of user data. The accumulation of numerous data about an individual and the resulting data processing can have negative consequences on individual privacy (Kurtz et al., 2018). Any organization with a single customer, employee, or other party living in the EU will be required to explain what data they have, locate it, correct it, explain where they got it, and, if requested, delete it. Otherwise, they are exposed to lawsuits or face potentially huge fines (Hiskey, 2018).

The purpose of this research is to explore the linkage between two topics: business process design and information quality, and will focus on the importance of considering information quality aspects in BP design success. Furthermore, this research aims to develop a method for identifying potential failures in BP, in particular, at the BP analysis and design stage, and to expose their effects and results. To the best of our knowledge, these two issues were not discussed enough jointly, and therefore the following analysis is considered a novelty.

In summary, the following significance conclusions were collected from the above research motivation description:

1) *The correctness, effectiveness, and efficiency of the business processes supported by information systems* ***are becoming vital to the organization****.*

2) ***Data quality is critical to organizational success****, for achieving strategic and operational business goals and for improved decision-making.*

3) ***Poor design*** *can lead the process to fail and achieve undesired and poor outputs not in the process goals set in terms of data items perspective, or lead to a deadlock situation.*

4) *The existing literature in business processes design and validation* ***lacks models and methods*** *for dealing with quality of information flows on business processes.*

5) ***Challenges concerning supporting the robustness of processes design still exist****.*

1.3. Problem Description

1.3.1. The problem statement

Many organizations have made significant investments in IS/IT projects and business process redesign and improvements initiatives (The Standish Group, 2014). However, 40% of all business initiatives failed to achieve their targeted benefits and goals because of poor DQ or due to lack of attention to DQ aspects in organization information flows (Gartner, 2011). Basically, mapping information flows is a process for analyzing and presenting how information is transferred from one point to another within an organization (Hibberd & Evatt, 2004). Hence, mapping information flows among processes is critical for processes validation and success. Moreover, as more BPs become automated, DQ becomes the rate limiting factor for overall process quality (Gartner, 2011).

Data quality is a critical element of today's business success and in managing data within an organization as more organizations become dependent on data-driven insight (Experian research, 2015; Otto & Österle, 2015; Jaya et al., 2017). The low quality of data in information systems poses enormous risks to business operations and decision making (Liu et al., 2020). Poor DQ can have a negative impact on the performance of BPs and thereby the success of companies (Otto et al., 2009; Heinrich et al., 2018). In addition, poor DQ increases operational cost because time and other resources are spent detecting and correcting errors (Redman, 1998; 2016). Most organizations depend on the quality of information for everyday business operations. However, data quality problems impede companies from obtaining the best value from data (Liu et al., 2020) and many of them are now starting to assess and improve the quality of their information (Baškarada & Koronios, 2014).

According to Gartner Research report (2011), poor data quality is a primary reason for 40% of all business initiatives failing to achieve their targeted benefits and dooms many IS/IT projects. Companies routinely make decisions based on remarkably inaccurate or incomplete data, a leading cause of the failure of high-profile and high-cost IS/IT projects. Furthermore, 83% of the respondents of a survey conducted by Experian research (2016) state that poor data quality has hurt their business objectives, and 66% report that poor DQ has had a negative impact on their organization in the last twelve months. Another report reveals that 84% of the CEOs are concerned about the quality of the data they use for the decision-making process (KPMG, 2016).

Many studies (e.g. Haug et al., 2011; Glowalla & Sunyaev, 2013a; Laranjeiro et al., 2015; Moore, 2018) show that huge amounts of resource and money are spent in organizations due to poor DQ or to improve the quality of the data. Gartner's data quality survey (Moore, 2018) indicates that poor DQ is hitting organizations where it hurts and was responsible for an average annual financial loss of $15 million in 2017. According to Forbes (2017), the average financial impact of poor DQ on businesses is $9.7 million per year. Opportunity costs, loss of reputation and low confidence in data may push these costs higher.

Hence, IS/IT leaders can make a significant contribution to their organizations by expanding the initiatives that address DQ issues in their organizations (Gartner, 2016; Moore, 2018). Other Gartner survey (2013b) states that organizations estimated that, on average, they are losing about $14.2 million annually because of DQ issues. Further, of the 140 companies surveyed, 22% estimated their annual losses resulting from bad data at $20 million. Moreover, as more BPs become automated, DQ becomes the rate limiting factor for overall process quality. Overall, it is estimated that poor DQ costs the U.S. economy about $3.1 trillion per year (IBM Big Data & Analytics Hub, 2016; Redman, 2016). The reason bad data costs so much according to Redman (2016) is that decision makers, managers, knowledge workers, data scientists, and others must accommodate it in their everyday work.

1.3.2. The problem's implications

The importance of high data quality and the need to consider data quality in the context of business processes are well acknowledged (Glowalla & Sunyaev, 2013a). To make informed and effective decisions, it is crucial to assess and assure the quality of the underlying data (Heinrich et al., 2018; KPMG, 2018).

The design of many processes is done subject to the assumption that there are no problems related to the quality of the information and assumed perfect values used through data items in the process. Nevertheless, processes in practice, can suffer from quality aspects such as poor level of DQ along their activities or in communication between processes. Furthermore, process design without taking quality considerations into account is probably expected to fail. As a consequence of poor design or DQ, a process can fail and achieve undesired and poor outputs that are not in the process goals set, or can lead the process to be in a deadlock situation. ***The general result is a low level of process quality.***

Poor data quality has negative impacts on almost all the enterprises and can be a major cause for damages or losses of organizational processes and thereby the success of organizations (Tee et al., 2007; Otto et al., 2009; Storey et al., 2012; Cappiello et al., 2013). Hence, the information flow among processes is critical for processes validation and success. Moreover, using ISs as a means for real-world and processes representation, and as an infrastructure for information flows management in and among process networks creates a real need to ensure data quality (Soffer, 2010). Obviously, DQ in enterprise information systems (IS) is key for commercial success (Röthlin, 2004). Furthermore, the BP quality also depends on the quality of its input and output information objects (Heinrich & Paech, 2010). Such failures in process design can later lead to failures, deficiencies and errors in IS design stage. Hence, we have to check the input and output data values represented by data items before recording them into IS to ensure DQ and IS to work properly and presenting desired data values in high quality.

Failures in IS in terms of DQ dimensions can also be the result of human mistakes and omissions, for example, missing or skipping activities, making human mistakes and errors or reporting incorrect information in them, etc.; however, it can also be a result of a low level of process analysis and design, for example, providing a wrong data value reflection or missing representation of the data items which may be required by the process rules or failing to reflect an external or internal event with an expected data value. Since we assume that the process is valid, our research focuses on the second type of failures i.e. DQ failures due to low quality of process design and will not deal with failures, errors or problems related to human activities or caused by humans; e.g. employees or end users.

For the convenience of the reader, we illustrate the research problem idea in the next two Figures. **Figure 1** depicts a pyramid of data usage levels in a typical BP, while **Figure 2** depicts the information flows within BPs.

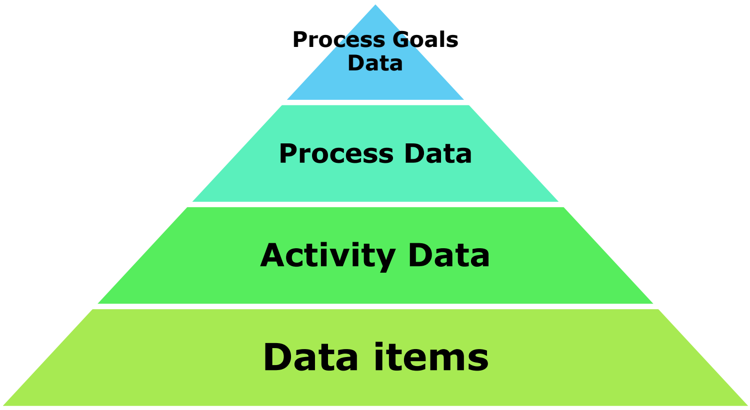


Figure 1. Levels of data usage in a typical business process

[Source: own study]

Basically, a standard BP consists of a set of activities and each activity involves sets of data objects which include inputs and output data, respectively. In the process, data dependencies are created when an output data set is used in principle as the input to the subsequent activity or activities and so on. In principle, our DQ activities will be covered and operated at all levels of the pyramid.

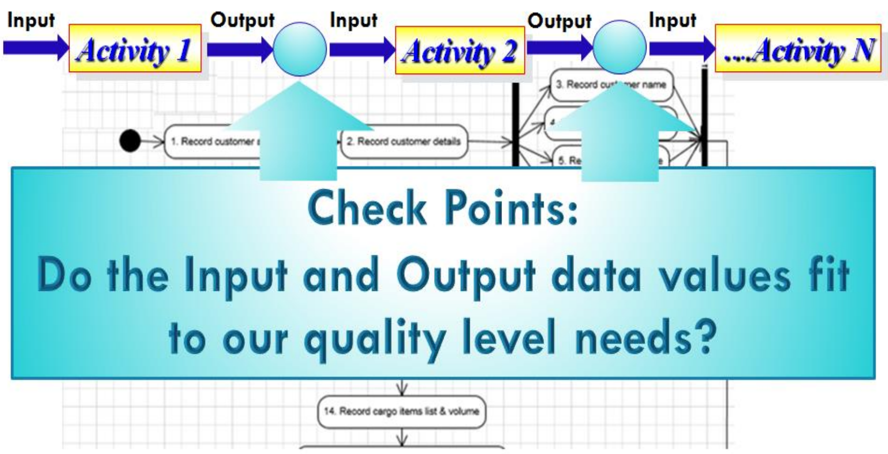


Figure 2. Problem illustration - Linkage between DQ and BP

[Source: own study]

The implications can be summarized as follows: the information flow among processes is a critical success factor for BP design and validation since BPs consume and produce information and are based on information flows. Moreover, using IS as an infrastructure for information flow management in and among BP networks, creates a real need to ensure DQ. In fact, we first need to know what kind of quality aspects we are interested in and then include those aspects into BPs design. Since the DQ depends on the design of BPs involved in generating the information, the potential problems of information quality should be taken into account into BPs and IS requirements framework and DQ level should be defined. Hence, the practical implication is expected to lead to focusing on suitable and applicable methods tied to BP during the process design stage, to provide a set of information quality requirements as part of a basis of BP requirements, to be supported by an IS.

1.4. Research Thesis

An in-depth study and analysis of the domain problems and needs of the current situation in organizations, including the existing methods in the literature in the field of DQ, led us to the conclusion that an improved and wider method is needed and can be provided. Hence, we formulate the following research thesis to be covered by realizing research goals:

The new suggested method improves the quality of business process design and helps BP analysts and designers focusing on potential failures of data dependencies and their impact on quality requirements at an earlier stage of the information systems design than currently existing methods in the data quality domain.

1.5. Research Goals and Research Questions

1.5.1. Research goals

This dissertation focuses on the importance of considering information quality aspects in BP design at the analysis and design stage. It aims to cope with various kinds of DQ problems relating to business processes data flows and design phase errors and their possible effects as described in the previous sections.

Recall, the main conclusion derived from the research problem is there is a lack of knowledge, models and methods for ensuring data quality in BPs design as a basis and surface for information systems quality. Therefore, the dissertation proposes a conceptual model and method, as artifacts, based on analyzing the impacts of information quality dimensions and information quality requirements on BP success. Thereby, these artifacts enable process and IS analysts, designers and practitioners, to predict potential failures and DQ deficiencies in business processes design and prevent them in advance.

To address these defined artifacts, the following five research goals have been set:

- ***Goal 1:*** *Identify a set of information quality aspects, problems and requirements that impact business process design quality;*
- ***Goal 2:*** *Design a conceptual model for constructs of information quality assessment;*
- ***Goal 3:*** *Design and develop a new method for information quality assessment and to predict data quality deficiencies and potential failures in business processes design and prevent them in advance;*
- ***Goal 4:*** *Develop a case study and collect information quality requirements to verify and validate the utility of the implemented method;*
- ***Goal 5:*** *Evaluate and demonstrate the application of the method in practice through focus group sessions.*

The new developed method devoted to BP and IS analysts, designers and practitioners, for improving DQ of BP design (w.r.t. DQ dimensions) and to achieve a high level of BP performance. It can help them with identifying potential failures in BPs at the process analysis and design stages by assessing the impact of information quality dimensions on the BP.

We validate deliverables of this research through the example of the sale of sea export service process as a case study to demonstrate its application in organization and evaluate the validity criteria in practice.

1.5.2. Research questions

Based on the above research goals, the focus of the dissertation is on three detailed research questions linked to these goals:

- ***RQ 1:*** *What kind of information quality aspects, problems, requirements and constructs impact the business processes design phase? (for Goal 1 and Goal 2)*
- ***RQ 2:*** *How can we predict potential failures and data quality deficiencies in business processes design and prevent them in advance? (for Goal 3)*
- ***RQ 3:*** *How can we test and demonstrate the validation and utility of the developed method in business processes design? (for Goal 4 and Goal 5)*

1.6. Research Methodology

1.6.1. Design science research (DSR)

A good starting point for design science in IS domain is provided by March & Smith (1995). They define it as an attempt to create things that serve human purposes, as opposed to natural and social sciences, which try to understand reality. The realm of IS research is about the confluence of people, organization, processes and technology (Hevner et al., 2004).

March & Smith (1995) distinguished between design sciences and natural sciences. Natural science is aimed at understanding *what reality is*. It develops a "set of concepts, or specialized language, with which to characterize phenomena". Natural science is concerned with explaining "how and why things" are. This is also called *descriptive research*. Design science (DS), also called *constructive research,* is concerned with "devising artifacts to attain goals". These artifacts can be built and evaluated, which are marked as the design science activities *build* and *evaluate*, but also theorized and justified, marked as natural science activities *theorize* and *justify*.

Based on this categorization, March & Smith (1995) and later, Hevner et al. (2004), introduced a design science framework with two axes, namely research activities and research outputs (see **Figure 3**) in which they related design science IS products to natural science research and design science research (DSR). Basically, different research activities, related to each of the different outputs, can be carried out. Furthermore, different cells have different objectives and different methods are appropriate in different cells.

		Research Outputs			
		Constructs	***Models***	***Methods***	***Instantiations***
Research Activities	***Build***				
	Evaluate				

Figure 3. A design science research framework

[Source: March & Smith, 1995; p. 255; Hevner et al., 2004]

Concerning **research activities**, March & Smith (1995) and later, Hevner et al. (2004), identify *build* and *evaluate* as the two main design processes issues in DS. Build refers to the construction of constructs, models, methods and instantiations i.e. artifacts demonstrating that they can be constructed. Evaluate refers to the development of criteria and the assessment of the output's performance against those criteria. Parallel to these two research activities in DS, March & Smith (1995) add the natural and social science couple activities, which are *theorize* and *justify*. This refers to the construction of theories that explain how or why something happens. In the case of IS/IT research this is often an explanation of how or why an artifact works within its environment. Justify refers to theory proving and requires the gathering of scientific evidence that supports or refutes the theory.

Concerning **research outputs**, according to March & Smith (1995) and later, Hevner et al. (2004), the outputs of design science research are artifacts, which are broadly categorized into four research design science products or outputs: constructs, models, methods, and instantiations. *Constructs* (or concepts) form the vocabulary of a domain. They constitute a conceptualization used to describe problems within a domain. A *model* is a set of propositions or abstraction or statements expressing relationships between constructs as a representation of the solution space. In design activities, models represent situations as problem and solution statements. A *method* is a set of steps (an algorithm or guidelines) used to perform a task. Methods are based on a set of underlying constructs (language) and a representation (model) of the solution space. An *instantiation* is the realization of an artifact in its environment. Instantiations operationalize constructs, models and methods. The relationships between these artifacts are illustrated in **Figure 4**.

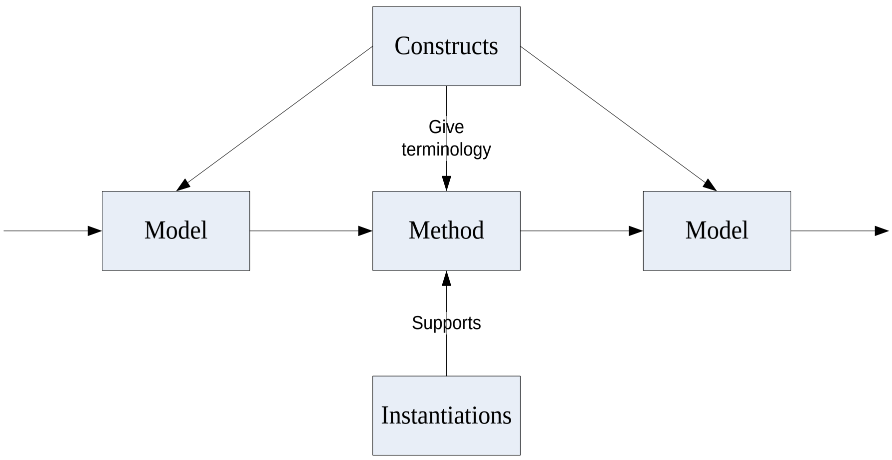


Figure 4. The relationships between IS artifacts

[Source: Hevner et al., 2004]

As a matter of fact, all mentioned types of artifacts have been developed and described in detail in this dissertation. The artifacts that result from this dissertation and their main characters are described in **Table 2**.

Table 2. Artifacts resulting from this dissertation

Artifact	Description
Constructs	Chapters 1 and 2 provide the general background, concept formation, vocabulary, principles, and terminology used in this dissertation. Additionally, they provide clear definitions, classifications, models, concepts and principles, DQ dimensions etc.
Models	Based on the introduced constructs, Chapter 3 presents a conceptual model - high-level description of the research problem scope, to summarize the terms and constructs and their relationships within the process by using UML class diagram. It draws the general picture, defines assumptions and emphasizes the benefits of the model. The developed model is a basis for our main artifact i.e. the DQDP method.
Methods	In sub-chapter 3.6, the Data Quality Deficiencies Prediction (DQDP) method developed in this research, is presented. The method is presented formally using the set of DQ requirements based on the existing knowledge of BP and IS design and quality aspects.
Instantiations	Sub-chapter 3.4 and Chapter 4 present the evaluation of the DQDP method, using many examples and scenarios from real-world with focus groups evaluation in order to verify whether the proposed method is able to solve the identified research problem.

[Source: own study]

A design process usually iterates over two activities: first, designing an artifact that improves something for stakeholders and subsequently empirically investigating the performance of that artifact in its context (Wieringa, 2014). Vaishnavi & Keuchler (2004; 2015) introduced a general cycle process of DSR methodology through five steps. The five steps are presented in **Figure 5** and although it suggests a sequential order, there can be some overlap in the steps and several iterations can also take place, especially in the development and evaluation phase.

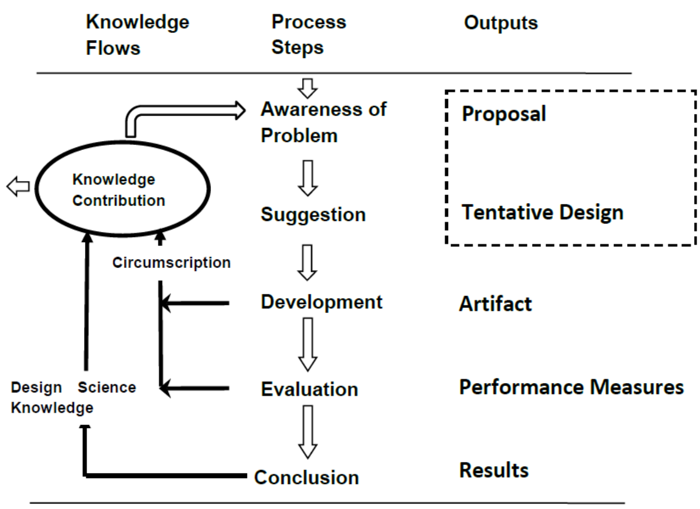


Figure 5. General cycle process of design science research (DSR)

[Source: Vaishnavi & Kuechler, 2004; 2015]

In a generic way, the steps can be explained as follows. The first step, problem awareness, is the realization that there is a problem in business, society or science. Once the problem has been defined, one can start to investigate the problem at hand a bit further and search for any available literature and then to suggest a possible design solution in the form of an artifact – step 2. In step 3, the tentative artifact, which should solve the identified problem, is developed and implemented. After building the (prototype of the) artifact, it needs to be evaluated against predefined evaluation criteria (step 4). During the process of developing and evaluating the artifact, questions might be raised that require a re-formulation of the problem resulting in further iterations (step 5). Moreover, the development and evaluation process are iterative, as the developed artifact is not expected to be right the first time.

1.6.2. Design science research framework

Hevner et al. (2004, including March as a co-writer) take the framework of March & Smith (1995) a step further: they state in a comparable way that two paradigms are important in IS research, behavioral science and design science: The behavioral science paradigm seeks to develop and verify theories that explain or predict human or organizational behavior. The design science paradigm seeks to extend the boundaries of human and organizational capabilities by creating new and innovative artifacts. Basically, it is fundamentally a problem-solving paradigm. According to Hevner et al. (2004), the design science research addresses important unsolved problems in unique or innovative ways or solved problems in a more effective and efficient way.

The objective of the design science approach in Information Systems (IS) research is the creation and evaluation of IS/IT artifacts, intended to solve these identified real-life organizational problems (Prat et al., 2014). Design problems assume a context and stakeholder goals and call for an artifact such that the interactions of artifact and context help stakeholders to achieve their goals (Wieringa, 2014). Furthermore, in the design-science paradigm, knowledge and understanding of a problem domain and its solution are achieved in the building and application of the designed artifact. Generally, IS professionals are engaged in the design and implementation of information technology artifacts aimed at improving the performance of business organizations (March & Storey, 2008).

Hevner et al. (2004) proposed a framework and set of guidelines for DSR to understand, execute and evaluate IS research and artifacts. The framework (**Figure 6**) compares the two mentioned paradigms and positions next to the problem space. DSR seeks to enhance human knowledge with the creation of innovative artifacts (Hevner & Chatterjee, 2015). The research methodology taken in this dissertation follows DSR paradigm and based on Hevner et al. (2004) framework, since the goal of the research focusing on design and developing an artifact in IS domain and thus making it a design science research type.

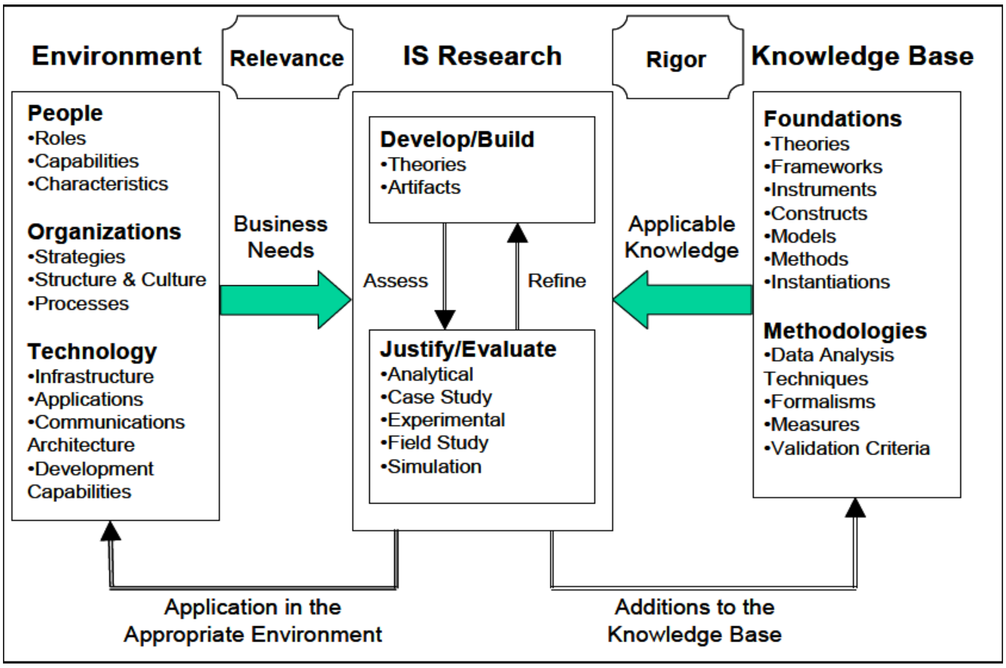


Figure 6. Information system research framework

[Source: Hevner et al., 2004]

The business environment consists of people, organizations, and technology and it defines the problem space. IS research develops and builds theories and artifacts, justifies them, and evaluates them respectively through analytical or empirical testing methods. The IS research knowledge base consists of foundations and methodologies. The environment presents business needs as subjects of IS research, which in turn applies findings to the appropriate environment; this cycle assures the relevance of research. The knowledge base can be fed by new insights as a consequence of IS research and offers applicable knowledge to IS research, which in turn adds its findings to the knowledge base; this cycle assures the rigor of the research. The goal of DSR is utility. The development of a particularly novel artifact with high utility will be a contribution to knowledge, even if the full understanding of why the artifact works is partial and incomplete (Gregor & Hevner, 2013). The instantiation as a result of the design science research efforts will influence and change the environment in which it is implemented.

1.6.3. Evaluation in design science research (DSR)

Evaluation of design artifacts is a key activity and crucial in DSR (March & Smith, 1995; Hevner et al, 2004; Vaishnavi & Kuechler, 2004; 2015; Peffers et al. 2008; Wieringa, 2014; Brandtner et al., 2016; Venable et al., 2016) and requires researchers to rigorously demonstrate the utility, quality, and efficacy of a design artifact using well-executed evaluation methods (Hevner et al, 2004; Venable et al., 2016). For this reason, we chose to adopt the Framework for Evaluation in Design Science (FEDS) approach demonstrated by Venable et al. (2016) for the evaluation phase.

FEDS is a novel evaluation framework uniquely suited to use in DSR. An important feature of the FEDS framework is its focus on the two key purposes of evaluation in DSR: the *utility aspect of the artifact* in the environment, and also the *quality of the knowledge* contributed by the construction of the artifact.

Basically, in FEDS framework we distinguish between two main dimensions of evaluation: first, the functional purpose of the evaluation i.e. *Formative* vs. *Summative* and second, the paradigm of the evaluation i.e. *Artificial* vs. *Naturalistic*. The functional purpose of formative evaluations is to help improve the outcomes of the process under evaluation. The functional purpose of *summative evaluations* is to determine the extent to which the outcomes match expectations and requirements, for example, certification, progress, or even the effectiveness of the process itself (Venable et al., 2016).

The second dimension i.e. the paradigm of the evaluation makes a distinction between artificial evaluation and naturalistic evaluation made by Venable et al. (2016). Both naturalistic and artificial evaluation methods can be used for formative and/or summative evaluations. These two dimensions form the basis of the FEDS framework and are fully orthogonal to each other and aids DSR researchers by offering a strategic view of DSR evaluation.

Artificial evaluation may be empirical or non-empirical (e.g. logical/rhetorical) being used to test design hypotheses. Artificial evaluation includes laboratory experiments, simulations, criteria-based analysis, theoretical arguments, and mathematical proofs. Artificial evaluation is often the simplest, most straightforward, and least costly form of evaluation. It often affords very precise language in its findings. Since it usually controls for the obvious confounding variables, it is less susceptible to misinterpretation and bias (Venable et al., 2016).

Naturalistic evaluation explores the performance of a solution technology in its real environment (i.e., real people, real systems, and real settings), and typically embraces all complexities of human practice within a real organization. Naturalistic evaluation methods typically include case studies, field studies, field experiments, surveys, action research, etc. (Venable et al., 2016). In addition, naturalistic evaluation offers more critical face validity and assures more rigorous assessment of the effectiveness of the artifact.

Basically, the FEDS evaluation design process is comprised of four steps[1]:

(1) Explication of the goals of the evaluation,

(2) Choosing the evaluation strategy or strategies,

(3) Determining which properties to evaluate, and

(4) Designing the individual evaluation episode(s).

The following is a description of the activities and examination in each of the above four steps that should be taken in the evaluation stage of the design artifact i.e. our new DQDP method:

(1) Explicating the goal(s) of the evaluation:

In this step, the evaluation includes examination of the existence of (at least) two key purposes of DSR: (1) the utility aspect of the artifact in the real environment, and (2) also the quality of the knowledge contributed by the construction of the artifact.

(2) Choosing the evaluation strategy or strategies:

At this step, the identified evaluation strategy is based on *Human Risk and Effectiveness* strategy since the major design risk is user oriented i.e. corporate concerns for collecting the problems constructs, requirements and feedbacks from BP and IS analysts, designers and practitioners, since the critical goal of the evaluation is to rigorously establish that the utility and benefit of the new method will prove themselves in real-world situations.

(3) Determining which properties to evaluate:

[1] Chapter 4 deals extensively with the implementation of the FEDS approach and its components.

At this step, the evaluation entails choosing the general set of features, goals, and requirements of the method that are to be subject of evaluation. In our case, the method goals and requirements have been checked in manner of its potential improvements with DQ dimensions and their impacts on overall quality of BP and IS design.

(4) Designing the individual evaluation episode(s):

In this step we need to identify and analyze some evaluation episodes, how and when these evaluation episodes will be conducted as well as the evaluation criteria, potential constraints, resources allocation and their availability in the identified environment and the impacts on the process.

In fact, this research is guided by design science (DS) approach since it guides the development of artifacts that are both practice-oriented and theory ingrained, and it provides the right balance between research rigor and relevance in research. In addition, evaluation helps to determine how well a designed artifact or ensemble of artifacts achieves its expected environmental utility and can further elaborate the knowledge outcomes by discerning why an artifact works or not.

The conclusion that follows from the above description, is that the evaluation of our research artifact should be under *summative evaluation* category since its functional purpose is to determine the extent to which the artifacts or outcomes match practitioners' expectations and requirements. Furthermore, it is more suited to the *naturalistic paradigm,* since it explores the performance of the artifact in its real-world environment.

In addition, naturalistic evaluation methods typically include case studies or field studies to justify and evaluate the suggested artifacts, so we decided to use **case study** as the research method for these two activities. The case study is an empirical method that uses a mix of quantitative and qualitative evidence to examine a phenomenon in its real-life context (Yin, 2014; Zhu et al., 2014; Karlsson, 2016). The in-depth inquiry of a single instance or event can lead to a deeper understanding of *why* and *how* that event happened. **This method is widely used in data quality researches (Zhu et al., 2014).** We examined in depth process networks of international forwarding and moving industry while using the developed method.

The evidence provided indicates that the FEDS framework and evaluation design process should help future DSR researchers (especially for novice DSR researchers) to design and improve their DSR evaluation activities.

1.6.4. Design science research guidelines

Hevner et al. (2004) presented a set of seven guidelines, which were elaborated upon later on by Österle & Otto (2010), for good design science research (DSR) that should be adhered to by researchers while building and applying artifacts within the discipline of IS. We applied this set of guidelines in our research in order to achieve an effective and suitable design science research. **Table 3** presents the set of guidelines with a short description adopted from Hevner et al. (2004) and indicates how these guidelines have been applied in this dissertation.

Table 3. Design-science research (DSR) guidelines

Guideline	Description	Apply the guideline in this dissertation
1. Design as an Artifact	DSR must produce a viable artifact in the form of a construct, a model, a method, or an instantiation.	The conceptual model in **Figure 31** (cf. sub-chapter 3.2) and proposed *DQDP method* in **Table 23** (cf. sub-chapter 3.6) meet this requirement.
2. Problem Relevance	The objective of DSR is to develop technology-based solutions to important and relevant business problems.	The important and relevant problem of IQ in BPs design is described in detail in the motivation section (cf. sub-chapter 1.2) and in problem description section (cf. sub-chapter 1.3). The developing solution (DQDP method) for BP and IS analysts, designers and practitioners for predicting potential failures in the BP design (cf. sub-chapter 3.3), aims to eliminate DQ deficiencies while examining the effect of failures on dependencies among different data values based on DQ dimensions to improve the design of a BPs as a key challenge in future IS/IT systems.
3. Design Evaluation	The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods.	The presented artifact has been validated based on evaluation methods, requirements and evaluation criteria in subject to FEDS framework proposed by Venable et al. (2016) for rigorous evaluation In DSR (cf. Chapter 4). The evaluation is rather explorative and based on scenarios and examples, from other domains which exist in a real-world environment.
4. Research Contributions	Effective design-science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies.	Basically, two types of research contributions are provided: the design artifact itself i.e., a new method (DQDP) that enables the solution of unsolved problems in the environment. The method is devoted to BP and IS analysts, designers and practitioners, for improving data quality of business process design (w.r.t. DQ dimensions) and to achieve high level of BPs performance. Foundations i.e., extend and improve the existing knowledge base of the domain, and design evaluation knowledge methodologies i.e., based on using three evaluation methods.

5. Research Rigor	DSR relies upon the application of rigorous methods in both the construction and evaluation of the design artifact.	The presented artifact has been validated based on evaluation methods i.e. case study analysis, focus groups and comparison with other methods and approaches, and FEDS Framework proposed by Venable et al. (2016) for rigorous evaluation in DSR (cf. Chapter 4).
6. Design as a Search Process	The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.	Choosing the best method based on its utility, quality and efficacy and based on comparison with other existing methods and approaches. The proposed solution has been developed iteratively. Multiple reviews were part of the research process to ensure that the method has been revised and has reached the desired quality. Nevertheless, the proposed method needs to be extended in the future and adapted to concrete usage scenarios as appropriate.
7. Communication of Research	Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences.	The new proposed method is adapted to BP and IS analysts and designers and it was presented among practitioners and managerial audiences. These artifacts enable process and BP and IS analysts, designers and practitioners, to predict potential failures and DQ deficiencies in BPs design and prevent them in advance. The artifacts that result from this dissertation have been and will be communicated in different publications. For instance, the description of the motivation and the need for the method was presented during the 18th International BIS conference in Leipzig, Germany and during the 2nd International conference for PhD students in Poznan, Poland and during the 13th International ILAIS conference in Tel Aviv, Israel last year. The concept and main constructs of the method were published in SOEP journal in Poland. The readers of these publications have been both BP and IS which are technology-oriented as well as management-oriented.

[Source: own study based on Hevner et al., 2004]

In summary, design science research requires the creation of an innovative, purposeful artifact for a special problem domain. The artifact must be evaluated in order to ensure its utility for the specified problem. In order to form a novel research contribution, the artifact must either solve a problem that has not yet been solved or provide a more effective solution. Both the construction and evaluation of the artifact must be done rigorously, and the results of the research presented effectively both to technology-oriented and management-oriented audiences.

1.7. Dissertation Structure

Figure 7 describes the dissertation structure and the main components of each part.

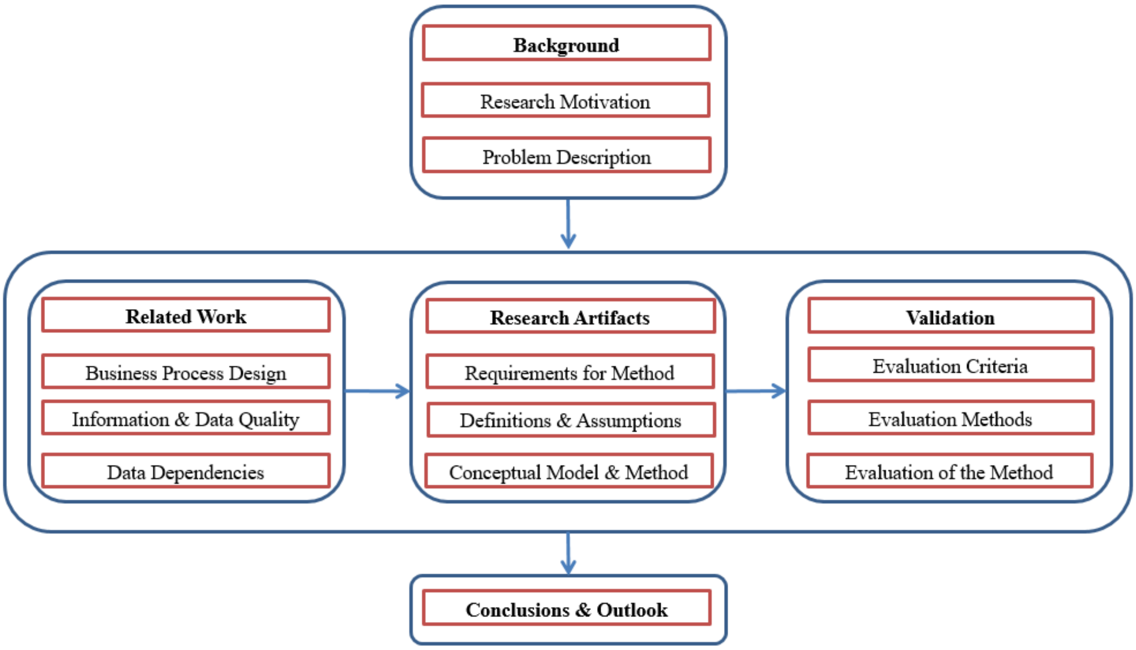


Figure 7. Dissertation structure

[Source: own study]

This dissertation is divided into four parts. Every part consists of at least one chapter with a few sub-chapters.

Part I provides the foundation, general background, and concept formation for this dissertation. It consists of two chapters. *Chapter 1*, gives an introduction and broad background about data and information quality and their application in business process design and summarizes the research motivation and problem description. Additionally, our research thesis, research goals and questions, as well as the research methodology are presented.

Chapter 2 presents the related work and defines terms that are relevant for the context of this dissertation. The concept of data quality dimension and its role in business process design are discussed. In addition, the chapter introduces the concept of data dependency, its elements and mechanisms. It also provides the necessary terminology for the remaining chapters of this dissertation. Finally, *Chapter 2* discusses current efforts, related work and models in the area of business process management and design and data and information quality.

Part II with *Chapter 3* presents the proposed artifacts. Firstly, it introduces the conceptual model and describes the general idea and lists definitions and assumptions, as well as presents a generic information model that integrates all necessary information for the needs of the method process. Secondly, in sub-chapter 3.6, the main artifact, named Data Quality Deficiencies Prediction (DQDP) method, is presented. It describes the requirements analysis on data quality (with the focus on DQ dimensions), and the general idea and lists of definitions and assumptions of the method. In addition, it describes and evaluates current efforts against the identified requirements to support business practitioners and designers' expectations during process design.

Part III with *Chapters 4 and 5* provides the validation and evaluation steps of research results and the validation of the developed DQDP method regarding the information quality. It describes the proposed method evaluation by implementing it in case study, focus groups sessions and based comparison with other methods and approaches in subject to FEDS approach for evaluation. Then in *Chapter 5* the discussion is provided.

Finally, **Part IV** with *Chapter 6* summarizes the dissertation, draws major conclusions, some major contributions and limitations, and provides an outlook on further research topics.

2 Related Work

This chapter describes the existing knowledge base in the thesis domains and the basis and concept formation for this thesis. It focuses on two domain topics: business process management and design and information quality (sub-chapters **2.2** and **2.3** respectively).

Each of the sub-chapters introduces the basic terminology and definitions related to business processes and information quality and discusses relevant aspects of business process management system (BPMS), while the subset of the involved topics that are particularly relevant to the context of this thesis is discussed with special regard to the quality of BP design. Current efforts, models and methods, problems and challenges in both topics are presented in detail.

Our research goals include, inter alia, identifying the set of possible information quality aspects, problems and requirements that impact BP design quality and develop a new method for IQ assessment to identify and predict potential failures in BPMS design and prevent them in advance. Hence, the focus in the literature review is to reveal the existing gap in knowledge base of aspects of BP design quality and the existing models and methods to cope with such problems and challenges.

2.1. Introduction

The economy today is driven by market forces in rivalry, dynamic and volatile environment (Ulrich & Smallwood, 2004; Ketchen & Short, 2012). Over the last couple of decades, organizations have faced rapid changes in the business environment that have negatively affected business performance (Ullah & Lai, 2011; Sun & Zhao, 2013).

Awareness of importance of the BPs has steadily increased in developed economies from the mid-eighties (Lahajnar & Rožanec, 2016) and all contemporary organizational structures, more or less, emphasize the important role of BPs (Harmon, 2014). BPs are frequently called organization's strategic assets (Smart et al., 2009; Falge et al., 2012).

Michael Porter (1985), who is considered the pioneer of competitive strategy, argues that a firm should achieve a sustainable competitive advantage (SCA). Competitive advantage (CA) is based on the ability to respond to evolving opportunities, which depends on business processes or capabilities. The information revolution is changing the nature of business and in recent years many enterprises increasingly rely on data and information to support business processes, decision-making processes and to create a competitive advantage (Porter & Millar, 1985; Shankaranarayanan & Cai, 2006; Karel & Richardson, 2010; Heinrich et al., 2018). Information is a critical asset and essential resource in the information age, used to conduct every organizational operation (Earl, 1997; Tayi & Ballou, 1998; Redman, 2004; Dumbleton & Munro, 2015; KPMG, 2018).

Business success involves choosing the right resources and capabilities to build them, manage them carefully, and exploit them fully. One of the five main steps recommended by Porter & Millar (1985), which can be taken in order to exploit opportunities created by the information revolution, is assessing the information intensity of products and processes i.e. identifying and assessing information needs and information uses, as inputs and outputs, in transforming products and processes through activities in value chain. These information resources serve as a basis for information requirements definition and enable organizations to gain information advantage (Porter & Millar, 1985; Auster & Choo, 1996).

In general, high-quality data is an important competitive factor for enterprises, to make informed and effective decisions and to help them to achieve high market position (Vaisman, 2006; Falge et al., 2012; Arachchi et al., 2015; Heinrich et al., 2018). English (2000) notes the emerging discipline of 'Enterprise Data Quality Management' (EDQM) whereby the organization develops and adopts a set of consistent technology processes, which institutionalize DQ as a strategic asset, and BPs to achieve a competitive advantage. To confirm their products and processes quality, organizations should include the assessment of activities' aspects and requirements of information quality to ensure information quality, since information is a critical resource in products and processes. Here, we focus on processes only; in particular, in analysis and design processes with implementation of information quality aspects and requirements.

2.2. Business Process Management (BPM)

2.2.1. Definition of business process

Over recent years, managing the business processes (BPs) of organizations has become increasingly important (Alotaibi & Liu, 2017). Processes are the arterial system within organizations and in inter-organizational supply networks (Dumas et al., 2018). Organizations are becoming increasingly aware of the meaning of a comprehensive treatment and management of their BPs due to an intense competition in the global market (Lahajnar & Rožanec, 2016).

The concept of a business process has been well defined by Davenport & Short (1990) and others (Harmon, 2007; Jeston & Nelis, 2014; Kirchmer, 2017; Weske, 2012; 2019) as a collection of logically related and structured activities or tasks that in a specific sequence produces a product or service for a particular customer or customers. The activities of business process represent the work of a person, an internal system, or the process of a partner company (Havey, 2005).

This general definition has been widely adopted in the literature on the design and management of BPs. Other researchers (e.g. Hammer & Champy, 1993; Falge et al., 2012; Dumas et al., 2018) extend this definition and emphasize the customer-orientated aspect of a BP as "a collection of activities that takes one or more kinds of input and creates an output that is of value to the customer". The idea of process as a series of added value tasks has also been raised by Patel (2007) for the customer's benefit and for revenue stream for the business. Moreover, an organization's current performance depends upon its BPs' collective ability to achieve its fundamental objectives and realize particular business goals (Shaw et al., 2007; Weske, 2012; 2019).

In general, organizations define functional goals, objectives and achievable outcomes. Based on their value-adding activities, BPs jointly performed to achieve organization goals, objectives and outcomes (Hammer & Champy, 1996; Zoet et al., 2011; Lohrmann & Reichert, 2013; González-Pérez et al., 2014; Weske, 2012; 2019). Furthermore, almost every single activity in an organization is very tightly bound with information seeking and information processing (Auster & Choo, 1996; Tee et al., 2007). BPs are becoming more and more complex and IS support or even automate the execution of business transactions in modern companies (Chintalapati et al., 2014). In particular, IS has been used to manage BPs and to support this effort by providing supply chain information and facilitating communication among supply chain partners as well as managers' decision when coordinating the network (Biehl, 2007; Alotaibi & Liu, 2017).

Each BP is enacted by a single organization, but it may interact with BPs performed by other organizations (Weske, 2012; 2019). Globalization and competitiveness force many organizations to change their structure and improve the way of executing business processes (Daoudi & Nurcan, 2007). Furthermore, companies have to design their BPs in an appropriate way in order to meet customer demands (Heravizadeh et al., 2009).

The BP domain is characterized by several types of stakeholders with different knowledge, expertise, and experience like Process owner, Process designer, Process participant, etc. (Weske, 2012; 2019). This research aims to focus on and support the analysts and designers of BPs as well as processes practitioners in organizations, with an emphasis on the expectations, needs and requirements of all process' participants in terms of IQ aspects in order to achieve a higher quality at the process level.

2.2.2. Levels of business processes

Weske (2012; 2019) describes in depth different levels that can be identified in business process management, ranging from high-level business strategies to implemented BPs. These levels are depicted in **Figure 8**.

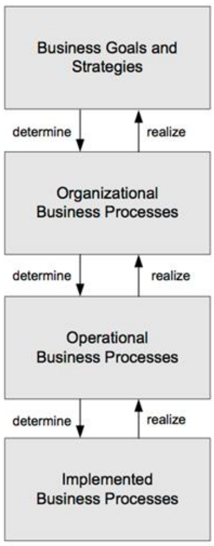


Figure 8. Levels of business processes

[Source: Weske, 2012; 2019]

The following four bullets provide an explanation of each of these levels.

- **<u>Business Goals and Strategies</u>**

At the highest level, the business goals and strategies are defined. Business goals refer to the long-term objectives of the company, while business strategies refer to plans for achieving these goals. Business goals describe long-term concepts to develop a sustainable competitive advantage (SCA) in the market (e.g. cost leadership for products in a certain domain). In fact, the business strategy is broken down to operational goals. These goals can be divided into a set of sub-goals that contribute to the realization of the defined business strategy (e.g. reducing the cost for supplied materials).

- **<u>Organizational Business Processes</u>**

At the second level, organizational business processes can be found. Organizational BPs are high-level processes that are typically specified in textual form by their inputs, their outputs, their expected results, and their dependencies on other organizational BPs. These BPs act as supplier or consumer processes. An organizational BP to manage incoming raw materials provided by a set of suppliers is an example of an organizational BP. Informal and semiformal techniques are used at these high levels. The strategy of a company, its goals, and its organizational BPs can be described in plain text, enriched with diagrams expressed in an ad-hoc or semiformal notation.

- **<u>Operational Business Processes</u>**

At the third level, operational business processes can be found. While organizational BPs characterize coarse-grained business functionality, typically there are multiple operational business processes required that contribute to one organizational BP. Basically, operational BPs are described by BP models.

- **<u>Implemented Business Processes</u>**

In operational BPs, the activities and their relationships are specified, but implementation aspects of the BP are ignored. Hence, at the bottom level, the implemented BPs can be found while the operational BPs are the basis for developing implemented BPs. There are multiple ways to implement BPs, ranging from written procedures and policies of the organization to the use of process enactment platforms. In any case, implemented BP refers to a specification that allows the enactment of the process on a given platform. Implemented BPs contain information on the execution of the process activities and the technical and organizational environment in which they will be executed (Weske, 2012; 2019).

Since our research aims to support BPs analysts and designers as well as processes practitioners in organizations, it focuses at the third and fourth levels of BPs i.e. on the activities and their relationships and the implementation aspects with an emphasis on the expectations, needs and requirements of all process' participants in terms of IQ aspects, in order to achieve a higher quality at the process level to support business goals and its long-term objectives.

2.2.3. Business process lifecycle

Business process management activities can be arbitrarily grouped into some categories or phases. Generally, there is no uniform view in the literature, on the number of phases in the BP lifecycle (Wetzstein et al., 2007; Alotaibi & Liu, 2017). The typical BP lifecycle is presented by Weske (2012; 2019) as shown in **Figure 9** and it is an extension of what was presented by Aalst et al. (2003).

Basically, the BP lifecycle consists of four phases that are related to each other. The phases are organized in a cyclical structure, showing their logical dependencies. These dependencies do not imply a strict temporal ordering in which the phases need to be executed. Many design and development activities are conducted during each of these phases, and incremental and evolutionary approaches involving concurrent activities in multiple phases are often implemented (Weske, 2012; 2019).

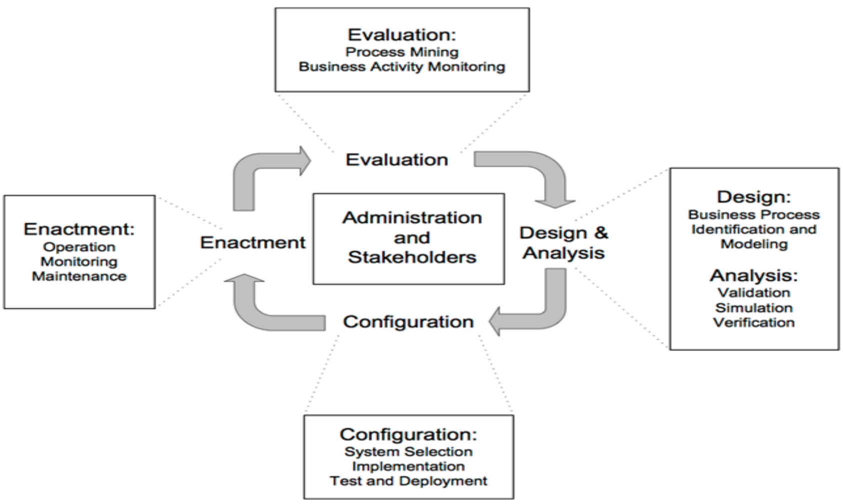


Figure 9. Business process life cycle

[Source: Weske, 2012; 2019]

The following bullets provide an explanation of each phase of BP life cycle according to Weske (2012; 2019).

- **Design and Analysis**

The BP lifecycle starts from the design and analysis phase, in which surveys on the BPs and their organizational and technical environment are conducted. Based on these surveys, BPs are identified, reviewed, validated, and represented by BP models (Weske, 2012; 2019). BP modeling techniques as well as validation, simulation, and verification techniques are used during this phase. BP modeling is the core technical sub phase during process design. Based on the survey and the findings of the BP improvement activities, the informal BP description is formalized using a BP modeling notation.

Once an initial design of a BP is developed, it needs to be validated. A workshop is a useful instrument to validate a BP. The participants of the workshop will check whether all valid BP instances are reflected by the BP model. Simulation techniques can also be used to support process validation. Simulation of BPs allows stakeholders to walk through the process in a step-by-step manner and to check whether the process exposes the desired behavior. Most management ISs provide a simulation environment that can be used in this phase (Weske, 2012; 2019).

- **Configuration**

Once the BP model is designed and verified, the BP needs to be implemented. There are different ways to do so. It can be implemented through a set of policies and procedures that the employees of the enterprise need to comply with. In this case, a BP can be realized without any support by a dedicated BPM system. In case a dedicated IS is used to realize the BP, an implementation platform is chosen during the configuration phase. The BP model is enhanced with technical information that facilitates the enactment of the process by the business process management system (BPMS).

The system needs to be configured according to the organizational environment of the enterprise and the BPs whose enactment it should control. This configuration includes the interactions of the employees i.e. end users with the system as well as the integration of the existing ISs with the BPM system. The last point is very important, since in today's business organizations, most BPs are supported by existing ISs. The configuration of a BPM system might also involve transactional aspects. Once the system is configured, the implementation of the BP needs to be tested (Weske, 2012; 2019).

At the process level, integration and performance tests are important for detecting potential run-time problems during the configuration phase. Once the test sub phase is complete, the system is deployed in its target environment (Weske, 2012; 2019).

- **<u>Enactment</u>**

Once the system configuration phase is completed, BP instances can be enacted. The process enactment phase encompasses the actual run-time of the business process. BP instances are initiated to fulfill the business goals of a company. Initiation of a process instance typically follows a defined event, for instance, the receipt of an order sent by a customer.

The BPM system actively controls the execution of BP instances as defined in the BP model. Process enactment needs to lead to a correct process orchestration, guaranteeing that the process activities are performed according to the execution constraints specified in the process model.

A monitoring component of a BPM system visualizes the status of BP instances. Process monitoring is an important mechanism for providing accurate information on the status of business process instances. This information is valuable, for instance, to respond to a customer request that inquiries about the current status of his case.

Detailed information on the current state of process instances is available in a BPM system. State information can be used to visualize and monitor process instances. Most BPM systems provide monitoring information that is based on states of active BPs. During BP enactment, valuable execution data is collected, typically in some form of log file. These log files consist of ordered sets of log entries, indicating events that have occurred during BPs. Start of activity and end of activity is typical information stored in execution logs (Weske, 2012; 2019).

- **<u>Evaluation</u>**

The evaluation phase uses information available to evaluate and improve BP models and their implementations. Execution logs are evaluated using business activity monitoring and process mining techniques. These techniques aim at identifying the quality of BP models and the adequacy of the execution environment.

For instance, business activity monitoring might identify that a certain activity takes too long due to shortage of resources required to conduct it. Since this information is useful also for business process simulation, these phases are strongly related (Weske, 2012; 2019).

- **Administration and Stakeholders**

The administration and stakeholders element of the chart does not refer to a specific phase of the BPM lifecycle. There are numerous artifacts at different levels of abstraction in BP management scenarios that need to be organized and managed well. For example, structured storage and efficient retrieval of artifacts regarding BP models and information on BP instances as well as the organizational and technical execution environment need to be taken into account. Especially in large organizations with hundreds or thousands of BP models, a well-structured repository with powerful query mechanisms is essential. In addition to BPs, knowledge workers with their organizational roles and skills, as well as the information technology landscape of the enterprise, need to be represented properly.

In this context, we adopt a view on BPM worded by Leymann & Altenhuber (1994). They distinguish two fundamental aspects, namely the *build time* aspect and the *run-time* aspect of managing BPs. The build time aspect focuses on the design and creation stages of the BP and the run-time aspect focuses on its execution stage. Using this distinction, we regard BPM as the field of designing and controlling BPs. This research aims to focus on the building stage, namely, the design and creation of the BP based on defined requirements.

2.2.4. Business process requirements

The most challenging task of the BP analysts and designers is to define the specific information requirements that must be met by the chosen IS solution (Laudon & Laudon, 2018). Requirements include anything that an activity requires to initiate, continue or to be completed. Sadiq et al. (2003) claim that in the area of BP analysis and design, the requirements and specification of data and data flows between process activities is required, since they have a significant impact on the process activities. This is an essential step that must be considered beforehand (Sadiq et al., 2003).

Furthermore, the requirements phase is the most critical phase of the information system development life cycle (SDLC). Basically, IS quality refers to the extent that the system is capable of performing its requirements and it is suitable for use both from a technical and a design perspective (Knauer et al., 2020). Wrong or missing requirements lead to wrong or incomplete systems (Alshazly et al., 2014). Both researchers and practitioners have widely recognized the need for an IS to deliver real quality to its stakeholders (Blaine & Cleland-Huang, 2008). The quality of the requirements phase affects the overall quality of the subsequent phases and hence, the IS product. Writing good system requirements specification (SRS) is an important determinant of system quality (Alshazly et al., 2014).

At the most basic level, the information requirements of a new system involve identifying who needs what information, where, when, and how. Requirements analysis carefully defines the objectives of the new or modified system and develops a detailed description of the functions that the new system must perform. Faulty requirements analysis is a leading cause of systems failure and high systems development costs (Laudon & Laudon, 2018). Furthermore, requirements analysis imposes constraints on the design or implementation such as performance requirements, quality standards, or design constraints (Glinz, 2007).

The interactions among BPs, information flows and ISs are critical to the success of ISs implementation (Laudon & Laudon, 2018). However, as a consequence of poor requirements, a process can fail and achieve undesired and poor-quality results as outcomes. In addition, faulty requirements analysis remains a major problem during ISs development (Wand & Weber, 2002; Iivari et al., 2006; The Standish Group, 2014; 2015; Komai et al., 2017).

BP requirements analysis is concerned with the definition, analysis, and formalization of the requirements that a potential information system must have to accomplish specific organizational needs. Traditionally, requirements analysis has focused mainly on the system and its interactions with the users (Donzelli & Bresciani, 2004). The overall emphasis of analysis is gathering data on the existing system, determining the requirements for the new system, considering alternatives within these constraints, and investigating the feasibility of the solutions. The primary outcome of systems analysis is a prioritized list of systems requirements (Stair & Reynolds, 2010).

Despite the fact that there is no common definition of the requirements analysis process, four tasks to be performed have been identified (Pohl, 1996):

- Requirements Elicitation,
- Requirements Analysis and Negotiation,
- Requirements Specification/Documentation,
- Requirements Validation.

The successful implementation of a new information system relies, in fact, on a firm understanding of its application context, and, above all, on the ability to transform the needs of the application context into the requirements for the new system. In addition, there is a need to redesign the application context around the new information system in order to better exploit its capabilities and avoid negative reactions from users. Both are very difficult tasks (Donzelli & Bresciani, 2004). Moreover, it has been recognized that quality requirements are essential factors influencing the success of ISs and their development (Chung et al., 1999; Nwakanma et al., 2013; The Standish Group, 2014; 2015).

2.2.4.1. Functional and non-functional requirements

BPs constitute a source of requirements and surface for IS analysis and design and for the development of IS requirements as mentioned above. In every current requirements' classification, we find a distinction between requirements concerning the functionality of an IS and other requirements derived from BP analysis (Glinz, 2007; Laudon & Laudon, 2018).

A **functional requirement** (FR) defines a function of software or IS or its component. Functional requirements specify what a system does. A function is described as a set of inputs, behavior, and outputs. Functional requirements may be calculations, technical details, data manipulation and processing and other specific functionality that show how a use case is to be fulfilled. They are supported by non-functional requirements, which impose constraints on the design or implementation (such as performance requirements, security, or reliability).

Non-functional requirement (NFR) is a requirement which specifies criteria that can be used to judge the operation of an IS, rather than specific behaviors. This should be contrasted with functional requirements that specify specific behavior or functions (Glinz, 2007). Typical non-functional requirements are reliability, scalability, maintainability, robustness, and quality of an IS and processes (Mylopoulos et al., 1992). Quality requirements describe how well those functional requirements are accomplished (Ewusi-Mensah, 1997; Blaine & Cleland-Huang, 2008). Although process validation applies to all aspects of process specification, one of the most critical aspects of BP design and analysis is specification of data requirements and data flow between process activities (Sadiq et al., 2003).

Basically, our research is supposed to focus on and support the second group i.e. the non-functional requirements since it aims to cope with quality aspects and to define set of data and information quality requirements of all process' participants, in order to achieve a higher quality at the process level to meet business goals and its long-term objectives.

2.2.5. Business process management

Business process management (BPM) is a top-down methodology and systematic approach designed to organize, manage, analyze, reengineer, improve and automate the processes running in an organization (Wetzstein et al., 2007; Jeston & Nelis, 2014; Alotaibi & Liu, 2017). Furthermore, BPM has been identified as a top business priority for organizations and building BP capability is seen as a major challenge for senior executives (Recker et al., 2006; Alotaibi & Liu, 2017). In addition, Gartner report (2010) confirmed the significance of BPM with the top issue for CIOs identified for the sixth year in a row being the improvement of BPs.

BPM enables organizations to align internal business functions with customer needs, and helps executives determine how to direct, monitor and measure company resources (Chintalapati et al., 2014). BPM activities aim to create efficient and effective BPs which can be adapted in a rapidly changing business environment (Alotaibi & Liu, 2017). When properly executed by qualified professionals, BPM has the capacity to reduce costs, enhance efficiency and productivity, and minimize errors and risk – thereby protecting and optimizing corporate resources (Chintalapati et al., 2014).

The definitions of BPM range from IS/IT focused views (e.g. technologies for BP automation) to BPM as a holistic management practice (Falge et al., 2012). According to Aalst et al. (2003), BPM is defined as "supporting business processes using methods, techniques and software to design, enact, control and analyze operational processes involving humans, organizations, applications, documents and other sources of information". Furthermore, BPM is a structured approach to analyze and continually improve fundamental activities such as manufacturing, marketing, communications and other major elements of a company's operations (Zairi & Sinclair, 1995). BPM includes concepts, methods, and techniques to support the design, administration, configuration, enactment, and analysis of BPs (Weske, 2012; 2019).

BPM is a management discipline and has evolved from past management theories and practices, such as total quality management (TQM), just-in-time (JIT) principles, business process re-engineering (BPR) and Lean management for the maintenance and improvement of BPs. It requires and enables organizations to manage the complete revision cycles of their processes, from process design to monitoring and optimization, and to change them more frequently to ensure consistent outcomes and to take advantage of improvement opportunities (Gartner, 2006; Zoet, et al., 2011; Alotaibi & Liu, 2017; Dumas et al., 2018). Importantly, BPM is not about improving the way individual activities are performed. Rather, it is about managing entire chains of events, activities and decisions that ultimately add value to the organization and its customers (Dumas et al., 2018).

According to Harmon (2004), BPM refers to "aligning processes with the organization's strategic goals, designing and implementing process architectures, establishing process measurement systems that align with organizational goals, and educating and organizing managers so that they will manage processes effectively". Furthermore, according to Gartner research experts (2006) BPM requires organizations to shift to process-centric thinking and to reduce their reliance on traditional territorial and functional structures" and it represents a fundamental change in how businesses manage and operate their processes". BPM treats processes as core assets that directly contribute to enterprise's performance by driving operational excellence and business agility (Gartner, 2009; Dumas et al., 2018).

Rosemann & Brocke (2015) identified six core elements (or factors) of BPM, which are heavily grounded in the literature. The six core elements of BPM are shown in **Figure 10**. Each of the six core elements represents a critical success factor for BPM. Therefore, each element, sooner or later, needs to be considered by organizations striving for success with BPM. In addition, for each of these six factors, the researchers provided a further level of detail, called 'Capability Areas'.

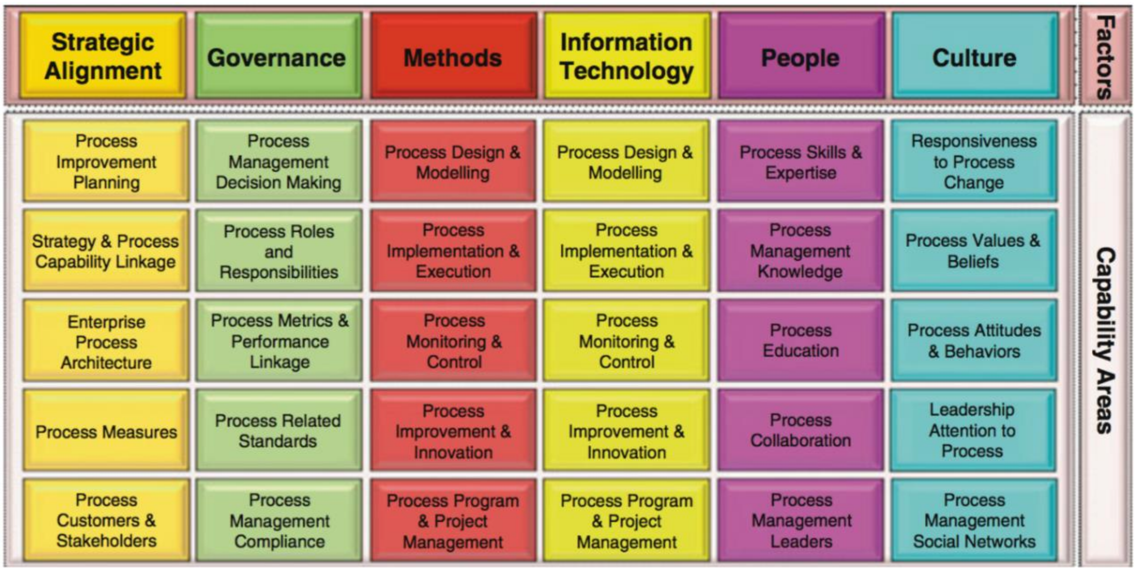


Figure 10. The six core elements of BPM

[Source: Rosemann & Brocke, 2015]

According to Rosemann & Brocke (2015):

1. *Strategic alignment* is defined as the tight linkage of organizational priorities and enterprise processes enabling continual and effective action to improve business performance.
2. *BPM governance* is dedicated to appropriate and transparent accountability in terms of roles and responsibilities for different levels of BPM (portfolio, program, project, and operations). Furthermore, it is tasked with the design of decision-making and reward processes to guide process-related actions.
3. *Methods*, in the context of BPM, have been defined as tools and techniques that support and enable consistent activities on all levels of BPM (portfolio, program, project, and operations).
4. *Information technology (IT)* refers to the software, hardware, and information systems that enable and support process activities. IT-based solutions are of significance for BPM initiatives.
5. *People:* While the information technology factor covered IT-related resources, the factor "people" comprises human resources. This factor is defined as the individuals and groups who continually enhance and apply their process and process management skills and knowledge to improve business performance.
6. *Culture*, the sixth and final BPM core element, refers to the collective values and beliefs that shape process-related attitudes and behavior to improve business performance.

Organizations can achieve additional benefits if they use ISs or software tools for supporting management of such operational processes and for coordinating the activities involved in BPs. These ISs or software became known as business process management systems (BPMS) (Ko et al., 2009). A BPMS is "a generic software system that is driven by explicit process representations to coordinate the enactment of business processes" (Weske, 2012, p.6). The entire system, and the BP with which it is associated, should be evaluated. Often, a firm can make great gains if it restructures both business activities and the related IS simultaneously (Stair & Reynolds, 2010).

BPM in systems engineering and software engineering is the activity of representing processes of an enterprise, so that the current process may be analyzed and improved in the future. The successful development and implementation of business information systems (BIS) requires an integrated approach that includes the seamless design of both the business processes and the ISs supporting the BPs (Heinrich et al., 2017). However, organizations are concerned about the successful planning and design of information systems. The concern begins with the correct determination of the critical information needs of top management and other members, and extends to the methods and techniques that will transform these information needs into ISs (Auster & Choo, 1996).

BPM can be seen as an extension of workflow management (WFM). WFM primarily focuses on the automation of BPs, whereas BPM has a broader scope: from process automation and process analysis to operations management and the organization of work. On the one hand, BPM aims to improve operational BPs, for example, by modeling a business process and analyzing it using simulation, management may get ideas on how to reduce costs while improving service levels. On the other hand, BPM is often associated with software and new technologies to manage, control, and support operational processes (Aalst, 2013).

BPM is becoming an important part of organizations' operational business, as well as of many new projects for performance improvement (Lahajnar & Rožanec, 2016) and it has received considerable attention recently by both business administration and software communities (Aalst, 2013; Weske, 2012; 2019). The software community is interested in providing robust and scalable software and ISs. Since BPs are realized in complex IT landscapes, the integration of existing ISs is an important basis for the technical realization of BPs (Weske, 2012; 2019).

2.2.6. Business process modeling

BP modeling is a significant component of BPM and plays an important role in management and BPM disciplines (Reijers et al., 2010; Alotaibi & Liu, 2017). BP modeling provides support to organizational processes using a set of technologies and techniques, different methods, standards and software tools for the design, execution, administration, and monitoring of organizational BPs and activities, which include people, organizations, applications, documents and other related information (Havey, 2005; Alotaibi & Liu, 2017).

BP modeling is a technique used to analyze and model BPs, often supported by some graphical notation, which is used by process practitioners to capture, organize and communicate information about BPs (Patel & Hlupic, 2001; Harmon & Wolf 2011; Cappiello et al., 2013). Some of the key benefits of BP modeling tools are that they provide simulation capabilities that predict where bottlenecks may occur in a process before it is deployed.

Furthermore, it focuses on finding and solving organizational problems and media incompatibilities in BPs (Nüttgens et al., 1998).

BP models are used for documentation, reorganization and standardization purposes (Dzepina & Lehner, 2018) in designing business operations, in reengineering BPs, in analyzing inter-organizational process links and in designing integrated ISs (Hammer & Champy, 1994; Soffer & Wand, 2007). BP models aim at providing a joint understanding of business processes. Therefore, they typically cover information about structure and behavior of a BP like description of activities or decisions within the process, but they do not aim at providing quality information relating to BPs (Heinrich & Paech, 2010; Heinrich et al., 2011).

The notion of a process model is foundational for BPM. A BP model is a representation that tries to capture the BPs which are essential in understanding the function and performance of an organization in reality (Ying et al., 2004). Explicit BP models expressed in a graphical notation facilitate communication about these processes, so that different stakeholders can communicate efficiently, and refine and improve them (Weske, 2012; 2019). A model helps to visualize what the important steps are in a process, how they are related to each other, which actors and systems are involved in carrying out the various steps, and at what points communication with customers and external parties takes place (Reijers et al., 2010). Moreover, it aims to capture the different ways in which a case (i.e., process instance) can be handled. In addition, BP model consists of a set of activity models and execution constraints between them. A BP instance represents a concrete case in the operational business of a company, consisting of activity instances. In fact, each BP model "acts as a blueprint for a set of BP instances, and each activity model acts as a blueprint for a set of activity instances" (Weske, 2012, p.7).

Organizations create BP models with the purpose of obtaining a simplified view of reality, which allows practitioners to understand and eventually modify a BP with the aim of incorporating improvements into it (Cappiello et al., 2013). Notations for BP description, e.g. BPMN, are becoming increasingly more important owing to the fact that the success of modeling is based both on the ability to express the different needs of the business and on the availability of a notation in which these needs can be described (Weske, 2012; 2019). Basically, a plethora of notations exists to model operational BPs (e.g., Petri nets, BPMN, UML, and EPCs). These notations have in common that processes are described in terms of activities (and possibly sub-processes). The ordering of these activities is modeled by describing causal dependencies (Aalst, 2013).

Basically, BP modeling can be divided into several aspects or perspectives including functional, structural, informational, temporal, transactional, and behavioral (Sadiq & Orlowska, 1999; Sadiq et al., 2003; Sun et al., 2006). According to Sun et al., (2006) the informational perspective defines what data are consumed and produced with respect to each activity in a BP.

BPs are complex and require people with various skills and abilities to work in a cooperative fashion. Processes will not be efficient and effective unless: they are clearly defined, individuals are adequately trained in their roles, and individuals understand how their roles fit into the overall process. Process modeling tools provide a way to describe BPs so that everyone involved in the process can understand the process. Furthermore, BP models are main artifacts for implementing BPs. This implementation can be done by organizational rules and policies, but it can also be done by a software system, using a business process management system (BPMS).

BP modeling is becoming increasingly popular. Both experts in the field of information and communication technology (ICT) and in the field of business engineering have come to the conclusion that successful (re)engineering of the involved systems starts with a thorough understanding of the BPs of an organization (Hommes & Dietz, 2001). In principle, process-aware information systems (PAIS) are used in BP modeling. PAIS is a software system that manages, supports and executes operational processes involving people, applications, and/or information sources based on process models. It combines advances in IS/IT with recent insights from management science. The typical example of PAISs are workflow management systems (WFM), case-handling systems, enterprise information systems (ERP), etc. (Recker et al., 2006; Aalst, 2009; Trčka et al., 2009). The shift from data orientation to process orientation has increased the importance of PAISs. Moreover, advanced analysis techniques ranging from simulation and verification to process mining and activity monitoring allow for systems that support process improvement in various ways (Aalst, 2009).

According to Hommes & Dietz (2001) BP modeling is deployed on a large scale to facilitate various purposes (see **Figure 11**). One of these purposes is to capture the BP when new technologies or information systems are introduced in an organization.

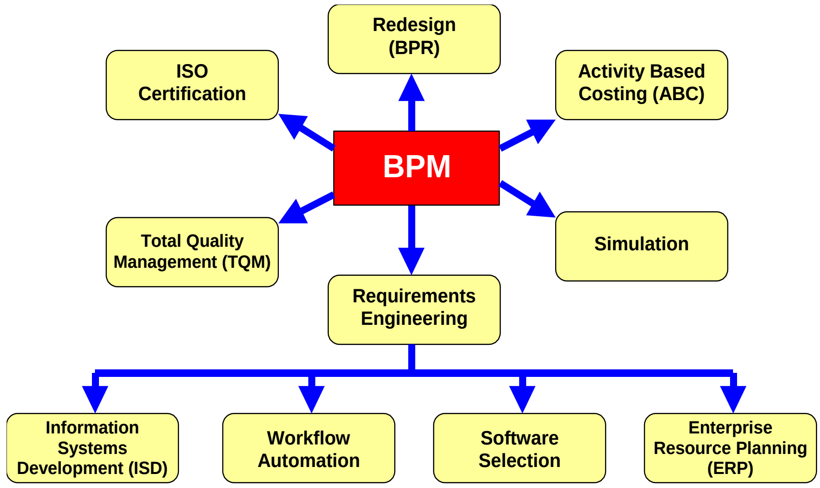


Figure 11. Business process modeling purposes

[Source: Hommes & Dietz, 2001]

In general, there are several techniques for the purpose of BP modeling. However, these can generally be divided into two categories: Static modeling and Dynamic modeling (Patel & Hlupic, 2001). A *static* BP model can be thought of as a diagrammatic representation of the process under consideration (see an example in **Figure 12**). Furthermore, there is a variety of methods and notations for the purpose of BP modeling (Giaglis, 2001; Patel & Hlupic, 2001). Static modeling enables the structure of the process to be displayed along with the flow of information between processes. In addition, static models have a deterministic nature and are independent of process sequence (Patel & Hlupic, 2001).

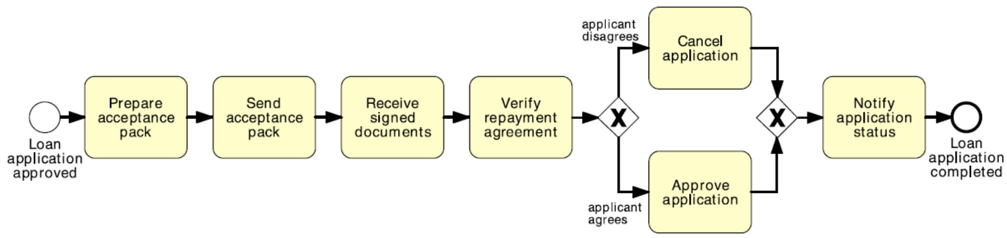


Figure 12. A process model for assessing loan applications using BPMN standard

[Source: Dumas et al., 2018]

The main advantage of using this modeling technique is that it enables an in-depth understanding of the process being modeled. The disadvantage of using a static modeling technique is that it does not facilitate the outcome of a changed process to be predicted. Furthermore, the physical aspects of the process cannot be modeled including resources and technology. However, despite the rigidity of static modeling over 80% of business process change deployments adopt this technique (Patel & Hlupic, 2001).

In comparison to static modeling, *dynamic modeling* enables a closer representation of the physical BP environment including people and equipment. Furthermore, a dynamic model facilitates the display of activities and flow of events within a process. The advantage of using dynamic modeling is that it enables the outcome of a changed process to be evaluated prior to it being implemented into the physical environment. Furthermore, resources and their movements are also considered within the dynamic model (Patel & Hlupic, 2001). In addition, the restrictions enforced by this technique mean that it is impossible to predict the outcome of a changed process - one of the reasons contributing to the failure of BP change. Dynamic BP modeling does enable the analysis and evaluation of changed processes; therefore, it is possible that this technique can help to improve the success rate of BP change deployments. In order to investigate this process requiring change is used to demonstrate both static and dynamic BP modeling (Patel & Hlupic, 2001).

Basically, our research is based on both static and dynamic modeling elements by using some techniques and notations for process modeling and representation of our case study on one hand and to analyze, evaluate and reshape changed processes prior to them being implemented into the real environment on the other hand. Furthermore, it is supposed to focus on and support the BP modeling from an informational perspective, which is the data and information that are consumed and produced respectively per each activity in a BP. In addition, it aims to cope with data and information quality aspects and DQ dimensions to create a set of DQ requirements for all process' participants, in order to achieve a higher quality at the process level and finally to meet business goals and its long-term objectives.

In the next sub-chapter we will describe the idea of BP improvement and the need to design or redesign a process as a result of improvement and what the change in the organization entails following the improvement since a significant part of outputs of BPR projects concerns IS/IT and the information resources that a BP needs.

2.2.7. Business process improvement and reengineering

Business process improvement (BPI) is a systematic approach to help organizations to archive significant changes in the way they do business (Forster, 2006). The term "improvement" may take different meanings depending on the objectives of the organization (Dumas et al., 2018). Improving a BP means changing the state of its elements in order to be faster, cheaper, more flexible, or to achieve a better quality (Griesberger et al., 2011). Process improvements are critical to an organization's long-term survival and well-being and should be linked to strategic goals and objectives of the organization. Typical examples of improvement objectives include reducing costs, reducing execution times and reducing error rates (Dumas et al., 2018). Other goals of improving BPs are change or gaining competitive advantages through better processes (Griesberger et al., 2011; Falk et al., 2013). BPI is seen as essential for improving customer services as well as product and service innovation. Thus, organizations have to make the continual improvement of their BPs part of their strategies (Bhatt & Troutt, 2005; Falk et al., 2013).

Generally, there are many different terms in literature relating to the management and improvement of BPs, including Business Process Redesign (Davenport & Short, 1990), Business Process Reengineering (Hammer & Champy, 1993), or Business Process Change (Harmon, 2007). Basically, they all address the same notion of enhancing and improving the work in organizations by means of BPs but the level of change, the starting point, the frequency of change, the time and scope differ in these methodologies (Forster, 2006). The demand for improving BPs increased after the business process reengineering (BPR) [2] wave in the early 1990s and methodologies, techniques, and tools were developed (Griesberger et al., 2011).

The idea of BPR was originally conceptualized by Hammer & Champy (1993) to radically improve organizational effectiveness and productivity. BPR focuses on analysis and design of core BPs within an organization to achieve revolutionary changes and substantial improvements in their performance, productivity, and quality (Griesberger et al., 2011). In addition, it assumes starting from a blank slate and completely recreating major BPs as well as the use of IT for significant performance improvement.

BPR and IS/IT infrastructure strategies which are both derived from organization strategy need effective alignment to ensure the success of the BPR initiative (Al-Mashari & Zairi, 1999). The interrelationships between BPR and IS/IT have been widely discussed in the academic studies on management information systems (MIS) and BPM (Law & Ngai, 2007). Technology in general, and especially ISs, is a key instrument to improve BPs (Dumas et al., 2018). Furthermore, ISs play a central and critical role in BPR initiatives (Hammer & Champy, 1993; Bhatt, 2000). Hence, IS requirements are based on and derived from BP reengineering

[2] Also known as Business Process Redesign

requirements including quality requirements and can suffer from quality aspects. IS development process is based on defined functional requirements. Functional requirements specify what a system does whereas quality requirements describe how well those functions are accomplished (Blaine & Cleland-Huang, 2008). The functionality and quality of the providing system (i.e., IS quality) therefore based on the quality of data (Knauer et al., 2020).

According to Bhatt & Troutt (2005) BP improvement initiatives and data integration directly affect customer responsiveness and product/service innovation. Adequate measurement of IT infrastructure effectiveness on BPR Information and IS/IT are the information resources that a BP needs (Sabherwal & King, 1991) to create a competitive value in an organization and, therefore, they are essential assets that need to be acquired, used, managed, and measured to judge the value obtained by investment in information resources (Earl, 1997). A lack of, or poor, IS/IT infrastructure will limit or jeopardize the success of BP changes (Law & Ngai, 2007). Thus, the infrastructure effectiveness determines IS/IT deficiencies that exist when BP information resource requirements cannot be met by the current IT infrastructure capabilities.

Organizations are committed to high quality with fast and flexible responses to customer needs. This puts pressure on organizations to redesign the way in which they conducted their business and build IS to support the new processes (Venkatraman, 1994). The potential of IS/IT to provide new sources of advantage for business operations in general, and for BP re-design projects in particular, is presented in Venkatraman's model (Venkatraman, 1994), where IS/IT design enabled business transformation (**Figure 13**).

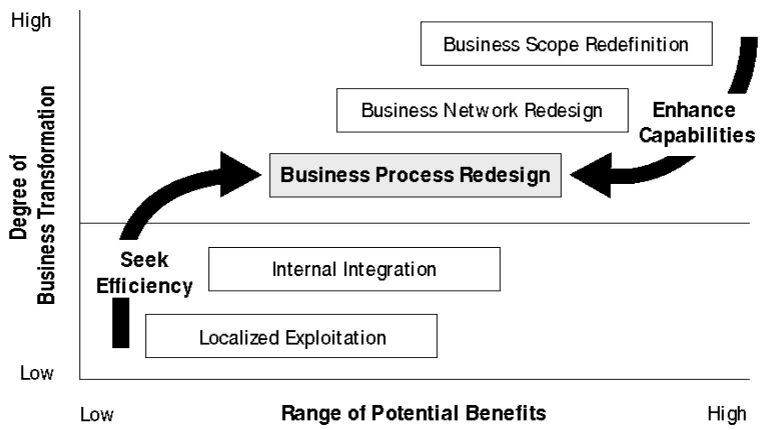


Figure 13. Five levels of IT-enabled business transformations

[Source: Venkatraman, 1994]

The first two levels, according to Venkatraman, are *evolutionary*, as they require only minimal changes to the BPs. The top three levels are *revolutionary*, as these levels require radical change to existing BPs. An organization could redesign its processes and then go on to redesign its network, stretching beyond the organization and ultimately redefining the scope of the organization. However, each higher stage requires a greater degree of organizational change. Eventually, for the organization to achieve better results, it will need to move up to the first revolutionary level and engage in BP reengineering. Venkatraman (1994) suggests that each organization first determines the level at which benefits are in line with the costs or efforts of the needed changes and then proceeds to higher levels as the demands of competition and the need to deliver greater value to the customer increases.

Basically, BP designs, and enterprise IS designs are often not well aligned (Heinrich et al., 2017). IS and BPR initiatives are complex and often tend to end in failure as mentioned above. As a complex process, managing IS development and implementation projects deserves a very high degree of attention, and is virtually impossible without taking change impacts into consideration (Ćirić & Raković, 2010). Change management is a vital component of IS development and implementation projects (Ćirić & Raković, 2010). One common mistake is to treat these initiatives in isolation from the rest of the organization. For this reason, Leavitt (1965) offered an approach to looking at organizations, called Leavitt's Diamond (**Figure 14)**. Leavitt's diamond proposes that every organizational system is made up of four main components which an enterprise depends on: Structure, People, Technology and Processes (which focus on business activities and information flows).

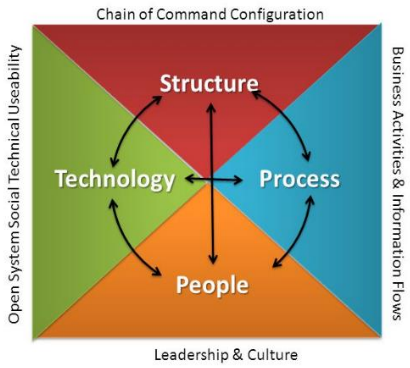


Figure 14. Leavitt's diamond - factors involved in change

[Source: Leavitt, 1965]

It is the interaction between these four components that determines the fate of an organization. Any changes in one will affect the other. This approach is now widely used for making organizational change more effective.

2.2.8. Business process quality

Quality has been the topic of research in neighboring disciplines such as manufacturing, software engineering, information management, and services management. As a result, a variety of standards and frameworks have been introduced to define, manage, assure, control and improve the quality (Heravizadeh et al., 2009). Integrating quality management with BPM concepts and evaluating their quality is very promising from a business perspective (Indulska et al., 2009; Lohrmann & Reichert, 2010; 2013; Alotaibi & Liu, 2017; Dzepina & Lehner, 2018).

Quality issues play a central role in BPM and the quality of BPs has to be taken into account at the requirement analysis stage, even though, a systematic consideration of quality requirements is still missing to ensure the alignment of quality requirements between BPs and IS/IT (Herrmann & Paech, 2006; Kedad & Loucopoulos, 2011). Yet, in BP modeling and design the quality dimension of a process is often neglected (Heravizadeh et al., 2009).

BP quality refers to components of a BP, to the process as a whole as well as to the context of the process (Heinrich & Paech, 2010). Components of a BP are the activities, the actors performing these activities, the information objects handled and created by the process as well as the resources necessary for execution. Furthermore, the quality of a BP highly affects the success of an organization (Heinrich & Paech, 2010).

In order to meet customer demands, organizations have to design BPs in an appropriate way. In particular, four essential process competencies have been discussed in operations management: process cost, process flow time, process flexibility, and process quality (Anupindi et al., 1999; Heravizadeh et al., 2009). While each of them has been subject to dedicated research, there is only little general work on process quality (Heravizadeh et al., 2009). In this context, process quality refers to "the ability of a process to produce and deliver quality products" (Heravizadeh et al., 2009, p.80). It covers aspects such as accuracy, conformance to specification, and reliability. Furthermore, quality aspects are part of an approach introduced by Filipowska et al. (2009) to evaluate and assess BPM methodologies. The researchers combine evaluation criteria in three dimensions: the purpose of methodology, structure, and quality. The “purpose” evaluates reasons for the application of the BPM methodology, the dimension “structure” gives an assessment of whether the methodology is adopted to the studied problem area, and the dimension “quality” gives a general assessment of its quality.

Heravizadeh et al. (2009) introduce the process root cause analysis approach (PRCA). It combines goal-oriented and activity-oriented process modeling for an explicit description of quality aspects of a process. They also built the quality of business process (QoBP) framework, a holistic framework for capturing the quality dimensions of a process. In particular, the framework helps modelers in identifying quality attributes of a specific process by using quality dimensions of BPs. Furthermore, Heravizadeh et al. (2009) also identify four generic quality categories of BP quality and populate them with quality requirements: quality of functions, quality of input and output objects, quality of non-human resources and quality of human resources.

A function is a basic block in a BP that corresponds to an activity or task which needs to be executed. The input and output of functions within a BP capture both, the physical and informational objects that are consumed and produced by it. Inputs and outputs differ in their significance to the overall process quality. Functions may be executed by non-human resources such as machines, devices, systems or software applications or by human resources (e.g. employees) (Heravizadeh et al., 2009). In fact, QoBP framework puts the emphasis mainly on the quality of the physical aspects in the process analysis and does not deal with the aspects or dimensions of the DQ in it. This research planned to focus on the second category i.e. the quality of data and information objects that are consumed and produced by activities on BP.

Heinrich & Paech (2010) have identified set of BP quality characteristics from software product quality standards, and classified them in a BP quality characteristics meta-model (BPQMM) (see **Figure 15**). The Meta-Model provided attributes and measures of these characteristics grouped by components and visualizes their dependencies. A BP quality attribute is an inherent property of a BP that can be distinguished quantitatively or qualitatively by human or automated means.

BP quality according to Heinrich & Paech (2010), refers to the components of a BP, to the process as a whole as well as to the context of the process. The context of a BP covers the conditions of use as well as the organizational environment. Components of a BP are the activities, the actors (which performing these activities), the information objects and data objects handled and created by the process as well as the resources necessary for execution. To each component of a BP they associated a set of quality characteristics. For information objects they took from the ISO/IEC 25012 (2008) DQ characteristics.

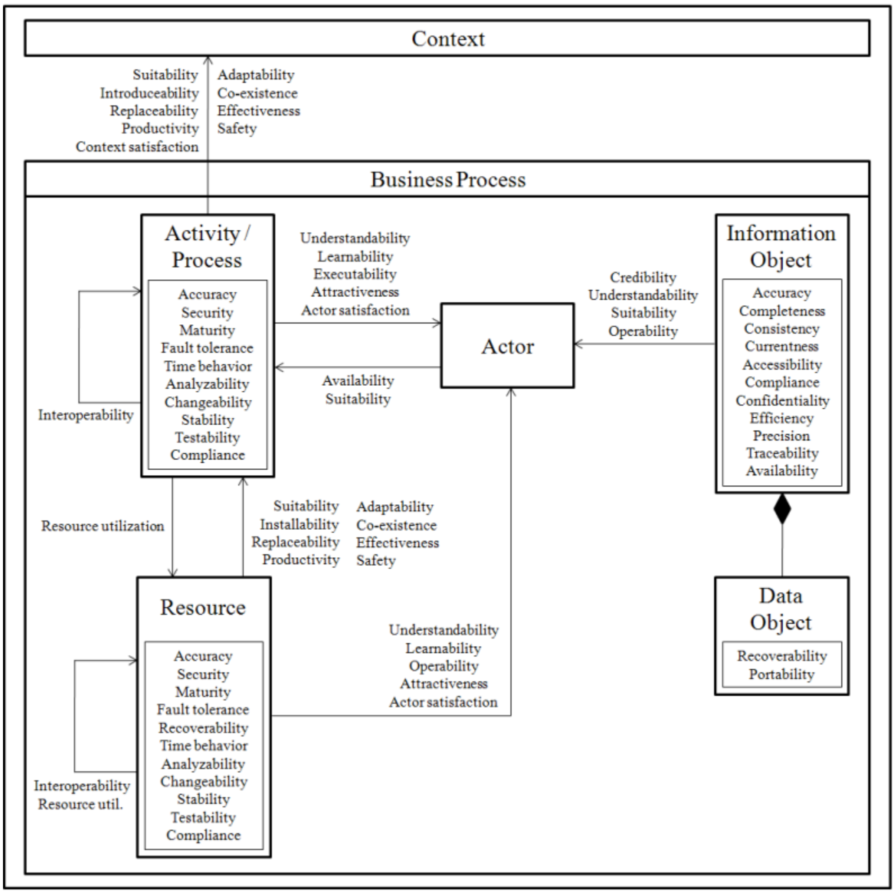


Figure 15. Business process quality meta-model (BPQMM)

[Source: Heinrich & Paech, 2010]

Lohrmann & Reichert (2013) proposed a framework for BP quality as a foundation to guide the development of specific quality attributes, criteria and predicates. The proposed framework is based on four central elements:

1. *Business process efficacy* means the effectiveness of a BP with respect to achieving its business goal(s) and objective(s). A BP is efficacious if its business objective is achieved for a reasonable set of states of its affecting environment.
2. *Business process efficiency* means the effectiveness of a business process with respect to limiting its impact on resources. A BP is efficient if it limits impact on resources reasonably considering the state of its affecting environment.
3. *Business process design & implementation quality* is the degree to which an actual business process model enables BP efficacy, achieves BP efficiency during design and implementation, and enables BP efficiency during its enactment.
4. *Business process enactment quality* is the degree to which a set of BP instances achieves BP efficacy and BP efficiency.

Naturally, this research corresponds to and focuses on the third element*, business process design quality,* since it strives to achieve effectiveness and efficiency during BP design and adapted it based on quality aspects and DQ requirements in particular. Basically, the quality of a BP also depends on its input and output quality. Input and output can be information objects as well as physical objects (Heinrich & Paech, 2010). Since our focus is on the quality of the information, the physical object quality is not within the scope of this research.

2.2.9. BPM summary

Quality issue plays a central and significant role in BPM and many academic efforts and studies have been done in this field so far, as can be seen in the above literature review. Basically, successful accomplishment of IS projects is dependent upon on the quality of BP design and remains a timeless goal and a crucial challenge for BP and IS researchers and practitioners (Glowalla & Sunyaev, 2015; Heinrich et al., 2017). Nevertheless, the characteristics of our research problem have not been adequately researched. Moreover, a systematic consideration of quality requirements is still missing to ensure the alignment of quality requirements between BPs and IS/IT (Herrmann & Paech, 2006; Komai et al., 2017).

Our first research goal is to identify and elicit a set of possible IQ aspects, problems and requirements that impact on BP design quality and how we can improve the BPs design phase. Although there are numerous models and methods to assess and improve a quality of data within BP and IS in the literature, these methods often do not address the original source of these problems or in principle provide a partial solution.

Our research expands an area which is relatively lacking in research and is considered to be of great importance in the information age we are in. Furthermore, the literature of BP design domains lacks new studies, methods and models, especially regarding the theoretical foundations and methods for collecting IQ requirements to improve IQ aspects and problems within BP and IS design. Hence, this research focused on the linkage between BP design and IQ, by exploring the importance of considering IQ requirements into BPs at analysis and design stage and by analyze the impacts of IQ dimensions and data dependency on IQ requirements. Based on our research goals the new suggested method can improve the quality of BP design and helps BP analysts and designers to identify potential failures of data dependencies and their impact on quality requirements at an earlier stage of the ISs design, than currently existing methods in the DQ domain.

2.3. Information Quality

2.3.1. Data, information and knowledge

To understand exactly what data or information quality means, we should first define three key terms: data, information and knowledge and their hierarchy relationship also referred to as the 'Knowledge hierarchy' or the 'Information hierarchy' (see **Figure 16**) that was first proposed by Ackoff (1989).

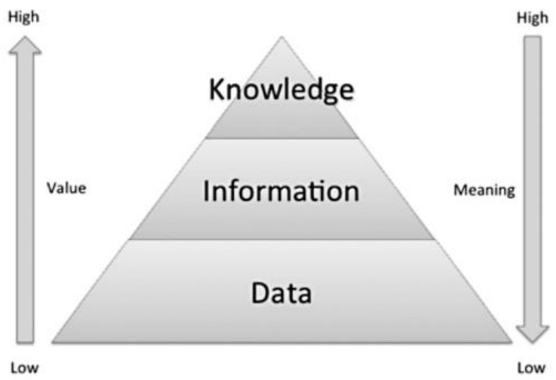


Figure 16. Data, information and knowledge hierarchy

[Source: Ackoff, 1989; Laudon & Laudon, 2018]

In general, the terms 'data' and 'information' are often used synonymously and interchangeably (Wang, 1998; Strong, 2009). However, the literature distinguishes between them and their roles. 'Data' are "streams of raw facts representing events occurring in organizations or the physical environment such as business transactions, before they have been organized and arranged into a form that people can understand and use" (Laudon & Laudon, 2018, p.48). 'Information', in contrast, concerns clusters of facts collecting in non-random way or data that have been shaped into a form that is meaningful and useful to human beings in the processes such as making decisions and fulfils end-user requirements (Bocij et al., 2015; Laudon & Laudon, 2018). Moreover, in practice, managers differentiate information from data intuitively and describe information as data that has been processed in some manner (Wang, 1998), or data whose form is for a particular use (Alter, 2002).

'Knowledge' refers to the semantic aspects of information that create knowledge and the meaning that the individual infers from the information by being interpreted and linked for a given purpose (Falge et al., 2012). Knowledge is the concise and appropriate collection of information in a way that makes it useful. Knowledge refers to a deterministic process where patterns within a given set of information are ascertained (Heilbronner &

Renzulli, 2016). In the literature, knowledge is multidimensional and it is related to experiences, values, contextual information, and expert insight that provide a framework for evaluating and incorporating new experiences and information (Nonaka & Takeuchi, 1995; Davenport & Prusak, 1998; Keskin, 2005). “It originates and is applied in the minds of knower. In organizations, it often becomes embedded not only in documents or repositories but also in organizational routines, processes, practices and norms.” (Davenport & Prusak, 1998, p.5). Nonaka & Takeuchi (1995) distinguish between 'Tacit Knowledge' and 'Explicit Knowledge': Tacit Knowledge resides within the individual, known but extremely difficult or in some cases impossible to articulate or communicate adequately. Explicit Knowledge, on the other hand, can be readily codified and communicated to others (Newell et al., 2002).

2.3.2. Definition of data quality

Data quality (DQ) is an interdisciplinary field. Existing research results show that researchers are primarily operating in two major disciplines: information systems (IS) and computer science (CS) (Sadiq, 2013; Zhu et al., 2014). Data quality research has developed rapidly in the past ten years and has become a hot research topic (Liu et al., 2020). Particularly, data or information quality (DQ/IQ) is an IS research area that seeks to apply modern quality management theories and practices to organizational information and systems. This involves building and applying conceptual frameworks and operational measures for understanding the causes and effects of DQ problems – problems ranging from DQ definition, measurement, analysis and improvement, to tools, methods, and processes (Wang, 1998).

The terms 'Data Quality' and 'Information Quality' are often used interchangeably in the information and data quality literature (Madnick et al. 2009; Strong, 2009; Glowalla & Sunyaev, 2014a) and since there has been no consensus about the distinction between data quality and information quality, there is a tendency to use 'Data Quality' (DQ) to refer to technical issues and 'Information Quality' (IQ) to refer to nontechnical issues (Zhu et al., 2014). In this research, we do not make such distinction and use the term data quality to refer to the full range of issues.

DQ is not an esoteric notion; it is particularly important for the successful execution of BPs and directly affects their effectiveness and efficiency, and also plays a major role in customer satisfaction” (Haug et al., 2011; Gharib & Giorgini, 2015). In this era quality is the most important factor in any kind of business. Furthermore, quality is an important factor in software industry (Javed et al., 2012) and enterprises need DQ to achieve high market position (Arachchi et al., 2015).

DQ is a term which describes the quality of the content of ISs and is defined as a context dependent and multi-dimensional concept. Furthermore, it describes the degree to which a set of data characteristics fulfills the requirements and is viewed as "fitness for use" by information consumers i.e., the ability of a data collection to meet end user requirements (Wand & Wang, 1996; Wang & Strong, 1996; Strong et al., 1997; Tayi & Ballou, 1998; Wang, 1998; Juran & Godfrey, 1999; Huang et al., 1999; Cappiello et al., 2004; Francalanci & Pernici, 2004; Redman, 2004; Haug et al., 2011; Caro et al., 2012; Ofner et al., 2012; Zhu et al., 2014; Laranjeiro et al., 2015; DAMA dictionary, 2017; Jaya et al., 2017; Juddoo & George, 2018). Heinrich et al. (2018) define DQ as “the measure of the agreement between the data views presented by an information system and that same data in the real-world”. Linking data and information quality activities to business needs is essential (Pierce et al., 2012).

Data and information are critical assets and strategic resources in the information age and used to conduct everyday business operations (Redman, 2004; Loshin, 2011; Baškarada & Koronios, 2014; Dumbleton & Munro, 2015; KPMG, 2018). Moreover, they are important for decisions making and planning on operational and strategic levels. However, the quality of decisions is affected by and dependent on the quality of the data (Price & Shanks, 2005; Bagchi et al., 2006; Vaisman, 2006; Xingsen et al., 2009; Haug et al., 2011; Ofner et al., 2012; Baškarada & Koronios, 2014; Gharib et al., 2018; KPMG, 2018). In this context, Karr et al. (2005) argue that DQ should always be embedded in a decision theoretic context. They recommended the following definition: "Data quality is the capability of data to be used effectively, economically and rapidly to inform and evaluate decisions".

The International Organization for Standardization (ISO) supplies an acceptable definition of 'data quality' using accepted terminology from the quality field. According to ISO 9001 standards for quality management systems, the formal definition of quality is "the degree to which a set of inherent characteristics of an object and its ability to satisfy stated and implied needs".In addition, the adjective quality applies to objects and refers to "the degree to which a set of inherent characteristics fulfils a set of requirements" (ISO 9000:2015). An object is any entity that is either conceivable or perceivable and an inherent characteristic is a feature that exists in an object. Furthermore, the definition of quality includes the word requirement, which is a need, expectation or obligation that is stated or implied by an organization, its customers, or other interested parties. Basically, there are many types of requirements in organization, e.g. quality requirements, customer requirements, management requirements, product requirements, contractual requirements, etc. (Hoyle, 2009).

The quality of an object can be determined by comparing a set of inherent characteristics against a set of requirements. If those characteristics meet all requirements, high or excellent quality is achieved but if those characteristics do not meet all requirements, a low or poor level of quality is achieved. Thus, the quality of an object depends on a set of characteristics and a set of requirements and how well the former complies with the latter (Hoyle, 2009).

A characteristic is a distinctive feature or property of something. Characteristics can be inherent or assigned and can be qualitative or quantitative. An inherent characteristic exists in something or is a permanent feature of something, while an assigned characteristic is a feature that is attributed or attached to something. Thus, we can define data to be of the required quality if it conforms to specifications or standards or if it satisfies the requirements stated in a particular specification and the specification reflects the implied needs of the users (Olson, 2003; Francalanci & Pernici, 2004; Kenett & Shmueli, 2016). Therefore, an acceptable level of quality has been achieved if the data conforms to a defined specification and the specification correctly reflects the intended use.

DQ is a critical element of today's organization success due to the increasing amounts and diversity of data processed by organizations (Glowalla & Sunyaev, 2013a; 2014a; Experian research, 2015). Awareness of DQ issues is growing globally and across all industries and many organizations are now starting to assess and improve the quality of their data and information resources (Baškarada & Koronios, 2014). In other words, DQ moves from a "nice to have" marketing component to a "must" business requirement (Hiskey, 2018) and trouble with maintaining DQ is an ongoing problem (Zatyko, 2017). According to that definition Wang & Strong (1996) defined a set of dimensions of DQ, which we will discuss later, from the consumer's point of view by means of a systematic multistage survey study. Prior to this research, DQ had been characterized by attributes identified by intuition, and selected unsystematically by individual researchers (Zhu et al., 2014).

For the convenience of the reader, we have summarized in **Table 4** the different definitions of DQ which were presented above.

Table 4. Summary of data quality definitions

#	Definition	References
1	Data that are "fit for use" by data consumers i.e., the ability of a data collection to meet end user requirements.	Wang & Strong, 1996; Strong et al., 1997
2	"Data to be of high quality if they are fit for their intended uses in operations, decision making and planning".	Redman, 2004
3	Data quality is "the capability of data to be used effectively, economically and rapidly to inform and evaluate decisions".	Karr et al. (2005)
4	"The degree to which a set of inherent characteristics of an object and its ability to satisfy stated and implied needs".	ISO 9001:2015 standard
5	“The degree to which data is accurate, complete, timely, and consistent with all requirements and business rules".	DAMA dictionary (2017)
6	“The measure of the agreement between the data views presented by an information system and that same data in the real world”.	Heinrich et al. (2018)

[Source: own study]

2.3.2.1. Data quality vs. Data governance

In the data management domain, there is a lot of terminology that is used interchangeably. However, many terms like data quality (DQ) and data governance (DG) for example, are often used as synonyms but are different terms. As mentioned above, according to DAMA dictionary (2017), the International dictionary for terms of data management domain, 'data quality' is “the degree to which data is accurate, complete, timely, and consistent with all requirements and business rules” while 'data governance' is defined as “the exercise of authority, control, and shared decision making (e.g. planning, monitoring, and enforcement) over the management of data assets”.

According to Data Governance Institute (DGI) 'data governance' is a "system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods" (Thomas, 2014).

Data governance is the discipline of cataloging and defining important data, assigning ownership of data and incorporating governance of data into the everyday BP (Zatyko, 2017). Seiner (2014) defines 'data governance' as "the execution and enforcement of authority over the management of data and data-related assets". At its core, data governance is about establishing methods, and an organization with clear responsibilities and processes to standardize, integrate, protect and store corporate data.

According to Otto (2011), data governance refers to the allocation of decision-making rights and related duties in the management of data in enterprises and it aims to maximize the value of data assets in enterprises. Furthermore, it provides all data management practices with the necessary foundation, strategy, and structure needed to ensure that data is managed as an asset and transformed into meaningful information (Knight, 2017; Firican, 2018). Data governance defines roles, and it assigns responsibilities for decision areas to these roles. It establishes organization-wide guidelines and standards for Data Quality Management (DQM), and it assures compliance with corporate strategy and laws governing data (Weber et al., 2009).

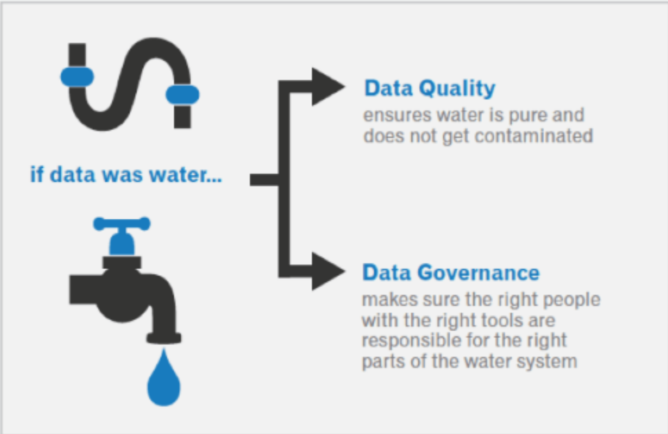


Figure 17. Data quality vs. Data governance

[Source: Zatyko, 2017]

As shown in **Figure 17**, DQ is all about keeping your data from getting contaminated by maliciously invalid information or by deterioration. Furthermore, data quality improves processes to prevent defective data from being created.

DQ is one of the most important elements of the data governance puzzle (Wassén, 2017). Data governance often sets direction for data quality and then monitors the success of DQ efforts (Thomas, 2014). In fact, these two terms are symbiotic and closely related, meaning they are interdependent and basically when one is applied, the other is asked. Quality needs to be a mandatory piece of a larger governance strategy. Without it, organization is not going to successfully manage and govern its most strategic asset: its data (Zatyko, 2017).

DAMA International lists ten major functions of Data Management in their DAMA-DMBOK (Data Management Body of Knowledge, 2014). Data governance is identified as the core component of data management, tying together the other disciplines as depicted in **Figure 18**.

Figure 18. Core functions of data management

[Source: DAMA DMBOK, 2014]

In general, effective data governance encompasses and requires a synergy between three key areas: *people*, the *processes* they use and the *technology* that supports those processes required to create a consistent and proper handling of an organization's data across the business enterprise. In other words, a data governance framework assigns ownership and responsibility for data, defines the processes for managing data, and leverages technologies that will help enable the aforementioned people and processes (Zatyko, 2017).

1. *People*: Putting the right team together is critical for data governance success. These people will be responsible for managing all aspects of the organization's data.
2. *Process*: Defining a process for how data will be controlled, audited, and monitored. The DQ processes mentioned above are an important facet of data governance and help ensure that data is accurate, consistent, and fit for purpose. Other key governance processes include security and risk management, reference and master data management, regulation and standards.
3. *Technology*: Technology helps in enabling people and streamline processes, and making sure the organization can make data-driven business decisions. Data management technologies can include things like verification, standardization, monitoring, collaboration, reporting, and identity resolution tools.

Basically, this research does not focus on data governance aspects, but will focus on its main DQ component. Moreover, it will focus on the quality of the content of the information rather than on the role of the people involved in the information processes.

2.3.2.2. *The importance and impacts of data quality for organizations*

Data quality is a critical issue in organizations and has an impact on organizational success (Umar et al., 1999; Lee et al., 2002; Madnick et al., 2009; Otto, 2011; Glowalla & Sunyaev, 2013a; 2014a). The importance of data quality and its improvement has increased with the widespread application of ISs in various sectors and it has been examined widely across different business sectors (Bai et al., 2018) and across organizations in the public sector (Tepandi et al., 2017).

While awareness of data quality has increased in recent years, there is still a lot of work to be done on this subject in organizations (Nagle et al., 2020). The development of ISs and technology during the last decades has enabled organizations to collect and store enormous amounts of data especially in the big data era (Abdullah et al., 2015; Cai & Zhu, 2015; Juddoo & George, 2018). However, as the data volumes grows, the greater the complexity and risks of poor DQ of managing them increases (Watts & Shankaranarayanan, 2009; Haug et al., 2011; Cai & Zhu, 2015). No industry, organization within any industry or any department within any organization is immune to the effects and impacts of poor DQ (Redman, 2004).

Poor DQ can have significantly negative impacts on the efficiency of an organization and may lead to substantial direct and indirect costs, while high quality data are often crucial to a company's success and for achieving business goals (Madnick et al., 2004; Batini et al., 2009; Even & Shankaranarayanan, 2007; Haug et al., 2011; Pierce et al., 2012; Glowalla et al., 2014). Organizational and managerial issues in DQ control involve the measurement or assessment of information quality, analysis of impacts on the organization, and improvement of DQ through process and systems redesign and organizational commitment to DQ (Ishikawa, 1985; Wang et al., 1993).

Defective data cause a litany of problems. According to TDWI's Data Quality Survey, almost half of companies (40%) have suffered losses, problems, or costs due to poor quality data (Abdullah et al., 2015). According to Gartner research report (2008) data quality is a business issue, not an IS/IT matter, and it requires the business to take responsibility and drive improvements. Moreover, European companies ranked poor DQ as the second-biggest business intelligence problem. The ability to create, collect, store, maintain, transfer, process and present data and information to support BPs in a timely and cost effective manner requires both an understanding of the characteristics of the data and information that determine its quality, and an ability to measure, manage and report on data and information quality (ISO 8000-62:2018). However, the cost of poor DQ analyses are difficult to conduct. A few costs, such as the cost of error detection and correction, can be measured. Other costs, such as the cost of customer dissatisfaction, are difficult to assess (Redman, 2004).

High quality information is information fit for use by information consumers (Wand & Wang, 1996; Wang & Strong, 1996; Strong et al., 1997; Wang, 1998; Huang et al., 1999; Redman, 2004) and naturally depends on application context and usage needs (Serra, 2020). This definition emerges directly from the standard fitness-for-use definition for products and services (Juran & Godfrey, 1999) and it is suitable with the declaration that "the purpose of designing an interactive system is to meet the needs of users: to provide quality in use" (Bevan, 1999; Bevan & Bogomolni, 2000). The top management shall ensure that customer requirements are determined and are met with the aim of enhancing customer satisfaction. In general, different users have different DQ requirements, and different data is of different quality (Wang et al., 1993).

Data are deemed of high quality if they correctly represent the real-world construct to which they refer (Huang et al., 1999; Juran & Godfrey, 1999; Redman, 2004). In general, data may be of poor quality because it does not reflect real-world conditions and quality must be measured in terms of user requirements (Ishikawa, 1985; Wang et al., 1993). Furthermore, poor DQ is costly (Haug et al., 2011; Glowalla & Sunyaev, 2014a). It lowers customer satisfaction, adds expense, and makes it more difficult to run a business and pursue tactical improvements such as data warehouses and re-engineering (Redman, 1997). Recent studies of data professionals indicate that a resounding 68% of DQ problems are still detected due to complaints and/or by chance (Sadiq et al., 2011; Sadiq, 2013).

2.3.3. Data quality management (DQM)

Data quality management (DQM) is an important area of research and investment in information systems, supported by the potentially high impact of poor data in organizations (Tee et al., 2007; Laranjeiro et al., 2015). DQM concerns a wide range of tasks and techniques, largely used by companies and organizations for assessing and improving the quality of their data (Serra, 2020).

Enterprises need DQM to respond to strategic and operational challenges demanding high-quality corporate data (Weber et al., 2009). According to Redman (1998), many of reengineering initiatives aim to put the right data in the right place at the right time to better serve a customer. Additionally, many of them involve DQ in one way or another. However, it is expected that most of those BP initiatives will fail due to lack of attention to DQ aspects (Wand & Wang, 1996; Ofner et al., 2012; Woodall et al., 2013). Marketing professionals know that the no. 1 risk of poor data quality is that it will negatively affect the customer experience and satisfaction (Goetz et al., 2015). Furthermore, organizations often find themselves failed in their efforts to translate data into meaningful insights that they can use to improve BPs, make smart decisions, and create strategic advantages (Zhu et al., 2014).

According to Sadiq (2013), DQM is based on three main pillars: Organizational, Architectural and Computational as shown in **Figure 19**.

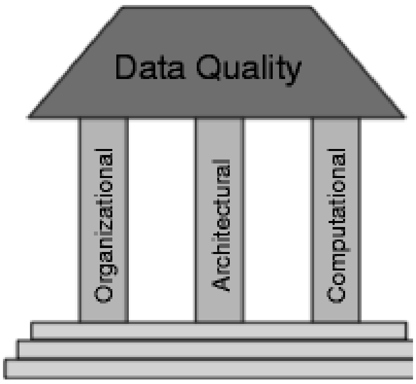


Figure 19. The three pillars of DQM
[Source: Sadiq, 2013]

Organizational refer to the development of data quality objectives for the organization, as well as the development of strategies to establish roles, processes, policies and standards required to manage and ensure that data quality objectives are met. *Architectural* describes the technology landscape required to deploy developed data quality management processes, standards and policies. *Computational* refers to effective and efficient IT tools, and computational techniques, required to meet data quality objectives. Techniques in this regard can include record linkage, lineage and provenance, data uncertainty, semantic integrity constraints, as well as information trust and credibility.

2.3.4. Data quality levels

In the DQ literature, there is almost an agreement on the distribution of DQ characteristics. The most accepted and popular distribution of DQ characteristics is based on three levels: Attributes, Dimensions and Categories.

2.3.4.1. *Data quality attributes*

DQ attributes refers to "a precise definition of the data quality parameters peculiar to the applications which will appeal to both primary and secondary data users" (Willshire & Meyen, 1997). A set of DQ attributes or characteristics are required for the objective and measurable assessment of conformance and utility, and hence for the assessment of DQ as well (Abate et al., 1998).

IQ attributes can be collected from end user requirements and expectations, functional and process context, recommendations of the literature and suggestions from experts, and the best engineering judgment will determine the selection and definition of DQ parameters. Based on DQ attributes, an organization can evaluate how well the current system satisfies the requirement made for DQ attributes (Wang & Strong, 1996; Willshire & Meyen, 1997).

2.3.4.2. *Data quality dimensions*

Data quality dimensions signify a crucial management element in the domain of DQ (Jayawardene et al., 2013). DQ dimension is a set of DQ attributes that represent a single aspect or construct of DQ and each DQ dimension is specifically related to a particular aspect of data (Wang & Strong, 1996; Cappiello et al., 2004). Moreover, DQ is presented in the literature as a multidimensional concept, even though there is no agreement on DQ dimensions (Wang & Strong, 1996; Wand & Wang, 1996; Ballou et al., 1998; NISS, 2001; Pipino et al., 2002; Röthlin, 2004; Knight & Burn, 2005; Otto & Österle, 2015).

Over the last two decades many researchers and practitioners have suggested several classifications of DQ dimensions, many of which have overlapping, and sometimes conflicting interpretations (Jayawardene et al., 2013). The most significant and useful classification in the literature was done by Wang & Strong (1996) research. They discussed how to construct specific DQ dimensions. Their group first collected 179 DQ attributes from DQ literature, researchers and consumers. The attributes are both subjective (qualitative) and objective (quantitative). They used factor analysis to consolidate these attributes into fifteen DQ dimensions. Whereas attributes represent the lowest level at which DQ problems can be identified and understood, the dimensions represent a higher level of understanding. Grouping attributes into dimensions according to Abate et al. (1998) offers the following advantages:

- Dimensions are more comprehensive than attributes;

- By grouping interdependent attributes into dimensions, DQ researchers can both minimize and organize information necessary for comprehension and interpretation; and
- Dimensions can help DQ researchers identify systemic DQ problems and are fundamental to understanding how to improve data.

Table 5 shows the dimensions of DQ suggested by Wang & Strong (1996) and later by Pipino et al. (2002) including a brief description of each.

Table 5. Dimensions of data quality

Dimension	Description
Accessibility	The extent to which data is available or easily and quickly retrievable.
Access Security	The extent to which access to data is restricted, and hence, kept secure.
Accuracy	The extent to which data is correct, reliable, and certified free of error.
Amount of Data	The extent to which the volume or quantity of data is appropriate for the task at hand.
Believability	The extent to which data is accepted or regarded as true, real, and credible.
Completeness	The extent to which data is not missing and is of sufficient breadth, depth, and scope for the task at hand.
Concise Representation	The extent to which data is compactly represented without being overwhelming.
Ease of Manipulation	The extent to which data is easy to manipulate and apply to different tasks.
Interpretability	The extent to which data is appropriate language, symbols and units and the data definitions are clear.
Objectivity	The extent to which data is unbiased, unprejudiced and impartial.
Relevancy	The extent to which data is applicable and helpful for the task at hand.
Representational Consistency	The extent to which data is presented in the same format and compatible with previous data.
Reputation	The extent to which data is trusted or highly regarded in terms of their source or content.
Timeliness	The extent to which data is sufficiently up-to-date and the age of the data must be appropriate for the task at hand.
Understandability	The extent to which data is clear, without ambiguity, and easily comprehended.
Value-Added	The extent to which data is beneficial and provide advantages from their use.

[Source: Wang & Strong, 1996; Pipino et al., 2002]

2.3.4.3. Data quality categories

Wang & Strong (1996) conducted a follow-up empirical research to group the above dimensions into four families of factors or categories as shown in **Table 6**. This form can help researchers to cope with so many dimensions and recognize patterns of DQ problems or deficiencies at process level (Abate et al., 1998).

Table 6. Categories of data quality dimensions

DQ category	DQ dimensions	Potential deficiencies aspects at business process level
Intrinsic DQ	Accuracy	Deficiencies or problems in BP in creating an actual and free of errors data values or trusted values that fit to process needs and reflect the real world.
	Believability	
	Objectivity	
	Reputation	
Accessibility DQ	Accessibility	Deficiencies or problems in BP in providing access to data items, available and readily obtainable data for end user.
	Access Security	
Contextual DQ	Amount of information	Deficiencies or problems in BP in producing data pertinent to the tasks of the end user at the right time and size.
	Completeness	
	Relevancy	
	Timeliness	
	Value-Added	
Representational DQ	Concise Representation	Deficiencies or problems in BP in providing structured and clear data that are easy to understand and presented correctly to the end user.
	Ease of Understanding	
	Interpretability	
	Representational Consistency	

[Source: own study based on Wang & Strong, 1996]

Wand & Wang (1996) suggest rigorous definitions of DQ dimensions by anchoring them in ontological foundations, and show how such dimensions can, in principle, provide guidance to systems designers on DQ issues. They also claim that the notion of DQ must be well understood in order to design ISs which deliver high-quality data, and an ontologically based approach to defining data may be the ticket for success in real-world systems. System construction includes design and implementation; system operation includes activities involved in producing the data, such as data capture, data entry, data maintenance, and data delivery.

Furthermore, according to Wand & Wang (1996), the dimensions which are mentioned above, can be categorized into two groups based on the definitions of internal and external views (**Table 7**). The external view of an IS is concerned with the use and effect of an information system at run-time stage. The internal views of an IS address the construction and operation necessary to attain the required functionality, given a set of requirements which reflect the external view at design time stage. In our research we are focusing on the dimensions which are relevant at design time stage.

Table 7. DQ dimensions grouped by internal or external views

View mode	DQ dimension
Internal View (design, operation)	***Data-related*** accuracy, completeness, currency, consistency, precision, reliability, timeliness
	System-related Reliability
External View (use, value)	***Data-related*** timeliness, relevance, content, importance, sufficiency, useableness, usefulness, clarity, conciseness, freedom from bias, informativeness, level of detail, quantitativeness, scope, interpretability, understandability
	System-related timeliness, flexibility, format, efficiency

[Source: Wand & Wang, 1996]

Table 7 also indicates whether a dimension is related to the data or to the system. Note, the timeliness dimension appears as related to both the internal and external views.

Wand & Wang (1996) base their approach on the notion that the role of an IS is to provide a representation of an application domain, also termed the real-world (RW) system, as perceived by the user. They define representation deficiencies in terms of the difference between the view of the real-world system as inferred from the IS and the view obtained by directly observing the real-world system by suggesting ontologically-based framework. The framework (**Table 8**) consists of four intrinsic dimensions: complete, unambiguous, meaningful and correct, with the nature and source of associated deficiency.

Table 8. Intrinsic data quality dimensions

DQ dimension	Nature of associated deficiency	Source of deficiency
Complete	Improper representation: missing IS states	Design failure
Unambiguous	Improper representation: multiple RW states mapped to the same IS state	Design failure
Meaningful	Meaningless IS state and Garbling (map to a meaningless state)	Design failure and Operation failure
Correct	Garbling (map to a wrong state)	Operation failure

[Source: Wand & Wang, 1996]

Referring to the relationships among attributes, dimensions, and categories as shown by Wang & Strong (1996), Abate et al. (1998) illustrate a hierarchy approach for analysing DQ problems as shown in **Figure 20**.

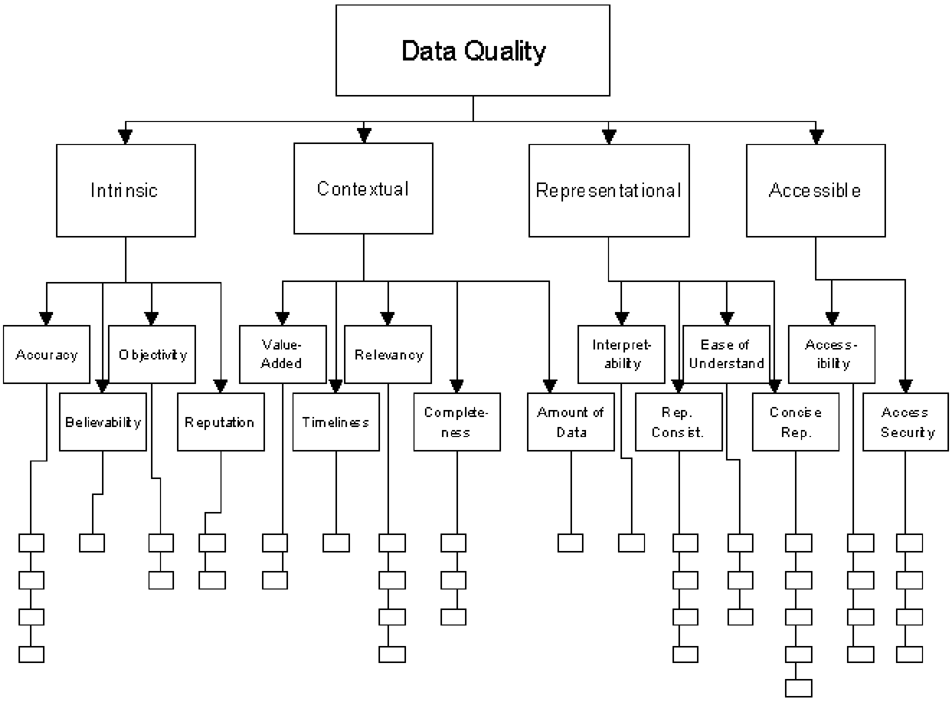


Figure 20. Hierarchy for data quality problems

[Source: Wang & Strong, 1996; Abate et al., 1998]

The lowest boxes in **Figure 20** denote the attributes or characteristics as mentioned above, representing a taxonomy of DQ metrics and the lowest level shows the most detailed view of DQ issues and problems can be identified and understood. The dimensions represent a higher (second) level of understanding – the conditions which would not occur if DQ problems become apparent at the attribute level. At the third level of understanding - the categories shown in **Table 6** - we can find the potential deficiencies aspects or weaknesses of BPs which allow the conditions at the second level to occur or exist, and, therefore, the processes responsible for the presence or absence of the second level conditions. Abate et al. (1998) recommend using this hierarchy approach as a valid basis for implementing a DQ assessment.

Since our focus in this research is on the analysis and design stage of BPs therefore the DQ deficiencies at process design stage refers to a set of dimensions derived from intrinsic or contextual DQ categories. Although there is no general agreement on which are the most important DQ dimensions, it is easy to note that the four DQ dimensions that have been considered in most of the IQ models are accuracy, completeness, consistency and timeliness (Scannapieco et al., 2005; Batini & Scannapieco, 2006; Cappiello et al., 2013; Shariat Panahy et al., 2013; Jaya et al., 2017; Yang et al., 2017; KPMG, 2018; Gharib et al., 2018).

2.3.5. Data quality problems review

2.3.5.1. Data quality problems – General aspects

Data quality is a complex problem facing all organizations (Bai et al., 2018). Discovering the quality of a data set is a fundamental task in most, if not all, DQ improvement projects (Batini et al., 2009; Zhang et al., 2014) and many organizations are suffering from poor DQ or are affected by DQ problems (Wand & Wang, 1996; Redman, 2004; E.D. Quality, 2015; Laranjeiro et al., 2015). A DQ problem can be defined as a difficulty in one or more quality dimensions that makes data unfit for use (Arachchi et al., 2015).

An important reason for addressing DQ problems is the growing need to integrate information across diverse data sources, because poor quality disturbs integration efforts (Scannapieco et al., 2005). The technological developments have implied that organizations store increasingly more data (Haug et al., 2011). Hence, DQ is crucial to organizational success due to the increasing amounts and diversity of data processed by organizations (Tee et al., 2007; Madnick et al., 2009; Otto, 2011; Glowalla & Sunyaev, 2013a) and many enterprises today are in the process of establishing corporate data quality management (CDQM) in order to meet their strategic business requirements (Falge et al., 2013).

Enterprises striving to improve DQ must establish data stewardship roles to improve data quality and move toward a culture that views data as a competitive asset rather than a necessary evil and define clear goals for DQ improvement (Heinrich et al., 2018). In the business context, poor DQ impacts the typical enterprise in many ways and levels. At the operational level, non-quality data lower customer satisfaction, lead to increased cost and lowered employee job satisfaction. At a tactical level, DQ flaws compromise decision making and reengineering projects. Finally, at a strategic level, insufficient DQ makes it more difficult to define and execute business strategies (Röthlin, 2004).

DQ directly affects the effectiveness and efficiency of BPs and plays a major role in customer and user satisfaction (English, 2001; Cappiello et al., 2014). Moreover, DQ improvement often requires changes in processes and organizational behaviors (Zhu et al., 2014) and problems with the quality of the data can have catastrophic consequences (Ofner et al., 2012; Woodall et al., 2013). However, DQ maintenance work is often neglected, and poor-quality business data constitute a significant cost factor for many organizations (Haug et al., 2011).

Organizations have increasingly invested in technology and human resources to collect, store, and process vast quantities of data. Even so, they often find themselves failed in their efforts to translate this data into meaningful insights that they can use to improve BPs, make smart decisions, and create strategic advantages (Zhu et al., 2014). The well-known 'Chaos Reports' by Standish Group (2013; 2014) declared that even though there is an increase in project success rates, IS/IT projects are still in chaos and more than half of projects are challenged and about 20% are failed. According to KPMG Survey (2016), on average, about 70% of all IS/IT related projects fail to meet their objectives.

Basically, data errors can be caused by both design and operational problems (Arachchi et al., 2015). Batini et al. (2009), provide a comprehensive analysis of existing approaches for DQ assessment and requirements identification, indicating that such approaches typically include three core aspects: data and process analysis, DQ requirements analysis, and DQ analysis. Data and process analysis includes examination of data schemas, performing interviews, and meetings with data users to reach a complete understanding of data, related constraints and rules, and processes creating or consuming the data. DQ requirements analysis often includes surveys of data users and administrators to identify quality issues, with the aim of identifying critical data sets, define DQ metrics, and set quality goals. DQ analysis then belongs to activities related to data sets exploration, assessment and profiling against the defined DQ metrics.

The new data-oriented shape of organizations inevitably imposes the need for the improvement of their DQ (Belhiah et al. 2016). In fact, growing DQ initiatives are offering increased monetary and non-monetary benefits for organizations. These benefits include increased customer satisfaction, reduced operating costs and increased revenues. However, regardless of the numerous initiatives, there is still no globally accepted approach for evaluating DQ projects in order to build the optimal business cases considering the benefits and the costs (Belhiah et al. 2016).

Eppler & Helfert (2004) proposed a "data quality cost taxonomy" (**Table 9**) that categorizes the potential costs of poor DQ to support the development of quantifiable measures of DQ costs for researchers. They outlined the list of direct and indirect costs of the quality of information divided into two main categories: costs caused by low level of DQ and costs for improving or assuring DQ.

Table 9. A data quality cost taxonomy

Data quality costs	Costs caused by low data quality	Direct costs	Verification costs
			Re-entry costs
			Compensation costs
		Indirect costs	Costs based on lower reputation
			Costs based on wrong decisions or actions
			Sunk investment costs
	Costs of improving or assuring data quality	Prevention costs	Training costs
			Monitoring costs
			Standard development and deployment costs
		Detection costs	Analysis costs
			Reporting costs
		Repair costs	Repair planning costs
			Repair implementation costs

[Source: Eppler & Helfert, 2004]

According to Gartner research (2011), in any business we can find some direct cost that is attributed to poor DQ. Redman (2004) argues that poor DQ costs the typical company at least ten percent (10%) of revenue; twenty percent (20%) is probably a better estimate and it is affecting decision-makers and planners as well.

In summary, the following significance conclusions were gathered from the above description:

1. *Many organizations are suffering from **poor data quality** and many BPM and IS/IT initiatives fail due to lack of attention to data quality (DQ) aspects.*
2. ***Unclear requirements and specifications or requirements errors** are critical factors and the most common cause of failure in information systems development projects.*
3. *We have to **check any input and output data values** represented by data items in process by DQ requirements before designing them into IS to ensure IS to work properly and presenting desired data values in high quality.*

2.3.5.2. Data quality problems – Approaches, models and frameworks

The literature on DQ has developed over the years and suggested many different approaches, models, techniques, measures and viable solutions to contend with data quality assessment needs, managing, and improving its quality (Shankaranarayanan & Wang, 2007). Most of the approaches are developed on an ad hoc basis to solve specific problems, and still the field lacks fundamental principles and comprehensive methodologies and methods for developing usable metrics in practice for organizations assessment and improvement (Lee et al., 2002; Pipino et al., 2002).

The existing models and methodologies for identifying DQ issues are inevitably user-centric because quality of data is assessed against a certain stated requirement and DQ requirements are determined in a top-down manner following organizational structures and data governance frameworks. In the current data landscape, however, users are often confronted with new, unexplored data sets that may have relevance and potential to create value (Zhang et al., 2014). For example, information quality and user satisfaction are two major dimensions of evaluating the success of information systems (DeLone & McLean, 1992).

DeLone & McLean (1992) presented a multidimensional measuring model (**Figure 21**) with interdependencies between the different success categories for evaluating the success of implementation of information systems, in order to provide a general and comprehensive definition of IS success. This model covers different perspectives of evaluating IS success. The researchers reviewed the existing definitions of IS success and their corresponding measures, and classified them into six major dimensions where information quality is one of the central dimensions as can be seen in **Figure 21**.

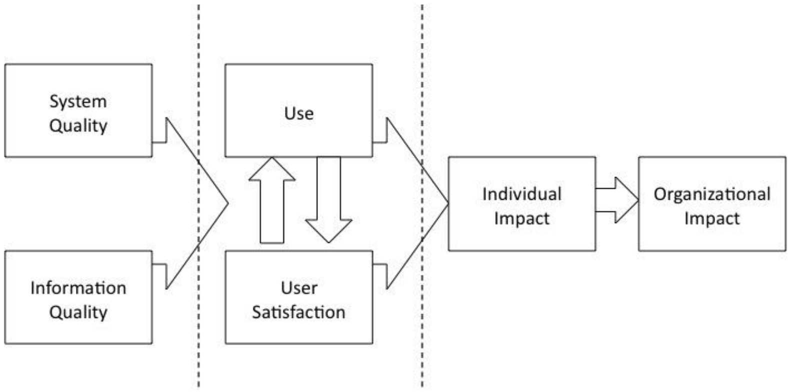


Figure 21. DeLone & McLean IS success model

[Source: DeLone & McLean, 1992]

Motivated by DeLone & McLean's call for further development and validation of their model, many researchers have attempted to extend or redefine the original model. Ten years after the publication of their first model and based on the evaluation of the many contributions to it, DeLone & McLean (2003) proposed an updated IS success model as shown in **Figure 22**. The updated model consists of six interrelated dimensions of IS success: information, system and service quality, (intention to) use, user satisfaction, and net benefits. The arrows demonstrate proposed associations between the success dimensions. The model can be interpreted as follows: a system can be evaluated in terms of information, system, and service quality; these characteristics affect the subsequent use or intention to use and user satisfaction. As a result of using the system, certain benefits will be achieved. The net benefits will (positively or negatively) influence user satisfaction and the further use of the IS.

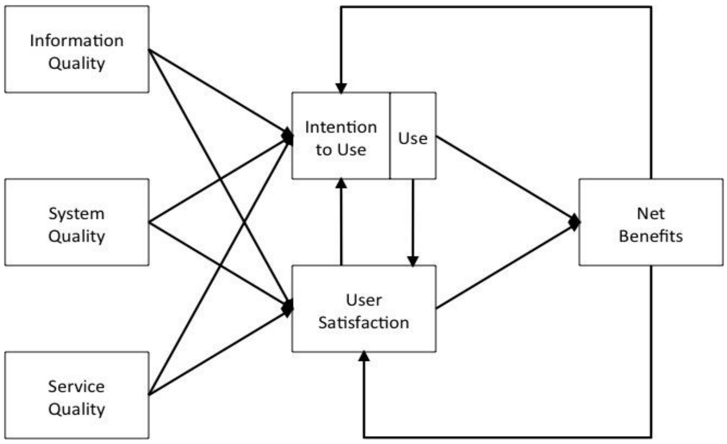


Figure 22. Updated IS success model

[Source: DeLone & McLean, 2003]

Willshire & Meyen (1997) present a data quality engineering framework (DQEF) – a Step-by-Step process an organization may tailor and customize according to its unique needs and implement it. An organization can also use it to define a model of its data environment, identify relevant DQ attributes, analyse data quality attributes in their current (or future) context, and provide guidance for data quality improvement. The researcher presents lists of some relevant objective and subjective quality indicators that have been collected in the literature as indicative of database quality.

Other researchers (e.g. Wang & Strong, 1996; Dvir & Evans, 1997; Wang, 1998; Kahn et al., 2002; Olson, 2003; Shankaranarayanan & Wang, 2007) claim that an analogy exists between quality issues in product or service manufacturing and those in information manufacturing. They suggest assessing the information quality from the product or service quality perspectives. Several approaches, methods and techniques for improving IQ based on total data quality management (TDQM) approach proposed by Wang (1998). The TDQM is a dominant approach of managing data as a product, designed for systematically managing DQ in organizations based on total quality management (TQM) approach. This approach addresses not just data but also the processes that create that data. It is based on the concept of Information Product (IP) to emphasize the fact that the information output from an information manufacturing system has a value that is transferable to the consumer (Shankaranarayanan & Wang, 2007; Gharib et al., 2018). These perspectives are limited, since they do not focus on process requirements and the design stage, and none of them links it to defined process goals and their achievement in high quality by the process.

This kind of this analogy, albeit in a different form, is suggested by Wand & Wang (1996) in terms of process design. They argue that similarly to the way the quality of a product depends on the process by which the product is designed and produced, the quality of data depends on the design and production processes involved in generating the data.

Another perspective, combining specifications and consumer expectations, is presented by Kahn et al. (2002). The researchers developed a conceptual model named product and service performance for information quality (PSP/IQ) model (**Table 10**) for mapping DQ dimensions according to their roles in product and service quality. Kahn et al. (2002) extend previous research on managing information as product to incorporate the service characteristics of information delivery based on quality definition as conformance to specification and as exceeding consumer expectations. The IQ dimension mapping is according to whether they can be achieved by conformance to specifications or by considering the changing expectations of consumers performing organizational tasks, and whether each IQ dimension is primarily an aspect of product quality or service quality.

Table 10. Mapping the IQ dimensions into the PSP/IQ model

IQ dimensions	Conforms to specifications	Meets or exceeds consumer expectations
Product Quality	**Sound Information** • Free-of-Error • Concise Representation • Completeness • Consistent Representation	**Useful Information** • Appropriate Amount • Relevancy • Understandability • Interpretability • Objectivity
Service Quality	**Dependable Information** • Timeliness • Security	**Usable Information** • Believability • Accessibility • Ease of Operation • Reputation

[Source: Kahn et al. (2002]

The four quadrants in PSP/IQ model i.e. sound, dependable, useful, and usable information provide a basis for assessing how well organizations develop sound and useful information products and deliver dependable and usable information services to information consumers. The model has two key contributions: such an assessment provides a baseline for determining what improvements should be made. It also provides a way to compare information quality across organizations, and to develop IQ benchmarks.

The value of the PSP/IQ model expanded by Lee et al. (2002) to AIMQ methodology with MIT's TDQM program is in the possibility of forming a basis for information quality assessment and benchmarking. The methodology is useful in identifying IQ problems, prioritizing areas for IQ improvement, and monitoring IQ improvements over time. The methodology is illustrated through its application in five major organizations. The methodology consists of three components: the PSP/IQ model of IQ, a questionnaire for measuring IQ, and IQ gap analysis techniques for interpreting the IQ measures. The researchers develop and validate the questionnaire and use it to collect data on the status of organizational IQ. These data are used to assess and benchmark IQ for four quadrants of the model. The AIMQ methodology provides a practical IQ tool to organizations. It has been applied in various organizational settings, such as the financial, healthcare, and manufacturing industries.

Pipino et al. (2002) distinguish between subjective and objective information quality assessments. Subjective data quality assessment is basically a qualitative approach and reflects the needs and experiences of stakeholders: the collectors, custodians, and consumers of data products. An objective assessment is basically a quantitative approach and it can be task-independent or task-dependent. Task-independent metrics reflect states of the data without the contextual knowledge of the application, and can be applied to any data set, regardless of the tasks at hand (Ballou et al., 1998; Wang, 1998; Pipino et al. 2002). The researchers also present an approach that combines the subjective and objective assessments of data quality and illustrate how it has been used in practice. They emphasize that companies must deal with both the subjective perceptions of the individuals involved with the data and the objective measurements based on the data set in question.

Price & Shanks (2005) present an IQ framework for understanding and defining information quality to support the decision-making process based on semiotic theory, the linguistic theory of sign-based communication, to describe the form, meaning, and use-related aspects of information and can serve as a theoretical framework to integrate the different approaches required to define quality criteria for each of these different information aspects. The focus of their paper is on information quality properties (i.e. quality categories and criteria), with additional consideration given to the quality goals and assessment techniques suitable for each quality category where the specific aim is to provide a theoretical foundation for the later development of practical quality assessment tools and guidelines.

Other researchers (e.g. Ballou & Pazer, 1985; Even & Shankaranarayana, 2007) present models to assess the impact of data and process quality upon the outputs of multi-user information-decision systems and revise DQ metrics and measurement techniques to incorporate and better reflect contextual assessment. The metrics are driven by data utility, a conceptual measure of the business value that is associated with the data within a specific usage context by addressing several dimensions of DQ at the collection, input, processing and output stages.

Nicolaou & McKnight (2006) examined the role of information quality in the success of initial phase inter-organizational (I-O) data exchanges. They proposed perceived information quality (PIQ) as a factor of perceived risk and trusting beliefs, which will directly affect intention to use the exchange.

Tee et al. (2007) examined factors influencing the level of DQ within a target organization. The researchers investigated and tracked DQ initiatives undertaken by the participating organization. The results confirmed that to ensure and maintain DQ, commitment to continuous DQ improvement is necessary. Most importantly, the research found that sustaining DQ gains requires mutual understanding by operations personnel, management, and funding sources as well as the provision of adequate incentives and modifications to institutional constraints. However, changing work processes and establishing a DQ awareness culture are required to motivate further improvements to DQ.

Frank (2008) investigated DQ and how it influences the quality of a decision. His research suggests an analysis to determine the effects of DQ on the quality of decisions provides criteria whether to invest in DQ improvement. He uses an example of an environmental engineering decision to demonstrate a general method to assess the influence of DQ on the decision in a GIS environment. This research shows that the uncertainty in aspects, which are poorly known, dominate the uncertainty of many decisions. Efforts to collect more or better data to improve the DQ of those stored in a GIS would not reduce uncertainty in the decision significantly. This result seems to be consistent with results from other studies for this very large class of decisions. The article gives a general method to assess whether collecting better data improves a decision or not.

Trčka et al. (2009) provided an analysis approach to discover dataflow errors in workflows (WFs) models. They presented a systematic classification of possible flaws and formulated these errors as data-flow "Anti-patterns" (e.g. missing data), expressed in terms of a temporal logic to formalize our anti-patterns. The goal of anti-patterns is to formally describe repeated mistakes such that they can be recognized and repaired. Typical errors include accessing a data element that is not yet available or updating a data element while it may be read in a parallel place. Since the anti-patterns are expressed in terms of temporal logic, the well-known, stable, adaptable, and effective model-checking techniques can be used to discover data-flow errors. Moreover, the approach enables a seamless integration of control-flow and data-flow verification. This study was a first step towards a unifying framework for the integral analysis of workflows taking into account both control and data flow. However, this approach is limited in principle to processes and models in WF systems and not to processes in general (Trčka et al., 2009).

Shariat Panahy et al., (2013) presented a framework to support process activities in information systems based on analyzing the relationships among four major and critical quality dimensions; accuracy, completeness, consistency, and timeliness for process improvement. The relationships are represented by related hypothesis developed to determine dependencies among the independent variables and dependent variable including their dependencies with control variable, which is improvement process. The researchers conducted a qualitative approach using a questionnaire and the responses were assessed to measure reliability and validity of the survey. It was validated through different statistical techniques on the data gathered from the survey to interpret the results. The results show that the items of each DQ dimension and improvement process are reliable and valid and there is a relationship between DQ dimensions and improvement process. Moreover, this framework can be adopted and used effectively to evaluate DQ in information systems within organizations and industrials to evaluate relationships among DQ dimensions to improve the involved process.

Belhiah et al. (2016) presented a model to clearly identify the opportunities for increased monetary and non-monetary benefits from improved DQ within an Enterprise Architecture context. The model aims to measure, in a quantitative manner, how key business processes help to execute an organization’s strategy and then to qualify the benefits as well as the complexity of improving data, that are consumed and produced by these processes. These findings will allow selecting data quality improvement projects, based on the latter's benefits to the organization and their costs of implementation (Belhiah et al. 2016).

In the health care domain, the DQ problem is particularly complex due to the wide variety of information sources, systems and users, combined with the sometimes urgent nature of health care decisions (Bai et al., 2018). The effectiveness, efficiency and reliability of health services depends on the quality of data in health information systems used by health care providers, managers and decision-makers. For this case, Bai et al. (2018) proposed in research-in-progress, a DQ framework and set of tools to help an information management team at a health care institution to monitor DQ as part of their business intelligence process. The proposed tool has the capacity to capture and visualize different stakeholder perceptions of DQ depending on the user role and the task they are responsible for. This study uses an action case method to develop a prototype of the tool and provide an illustration of its implementation and usage by multiple stakeholders.

Yang et al. (2017) investigated the data integration process in line with the TPC-DI Benchmark[3], which is the first and well-known industry data integration (DI) benchmark. It is designed to benchmark the data integration and serve as a standardisation to evaluate the ETL tools performance. They found a set of typical DQ problems that can occur in the data integration process. For each DQ problem, they have defined the problem and provided examples to demonstrate the problem trigger and possible effects and measured it effectively and meaningfully, based on general quality approaches (e.g. DMAIC and benchmarking) and DQ dimensions. In order to facilitate the DQ management in data integration, they have classified the DQ problems into different DQ dimensions. This result also indicates which DQ dimensions are important in data integration. These important dimensions can help researchers and practitioners to set the focus in DQ management and reduce the unnecessary cost and time.

Heinrich et al. (2018) proposed a set of five requirements for DQ metrics that aim to support both decision-making under uncertainty and an economically oriented management of data quality, based on a decision-oriented framework. These clearly defined requirements

[3] Developed by TPC - a non-profit corporation founded to define transaction processing and database benchmarks. For more details about TPC-DI you can find at: http://www.tpc.org/tpcdi/

are relevant to identify inadequate metrics, which may lead to wrong decisions and economic losses. Moreover, the researchers demonstrated the applicability and efficacy of these requirements by evaluating five DQ metrics for different DQ dimensions. They analysed the existing literature and justified this set of requirements based on a decision-oriented framework. In addition, they discussed on practical implications when applying the presented requirements. However, these DQ requirements are measurement-oriented to support decision-making process.

Zhang et al. (2019) developed an approach for discovering DQ problems in unexplored and repurposed datasets. Such repurposed datasets can be found in government open data portals, data markets and several publicly available data repositories that are perceived to have relevance and potential to create value for end users. In such scenarios, applying top-down DQ checking approaches is not feasible, as the consumers of the data have no control over its creation and governance. The approach, named LANG, based on bottom-up manner by using generic exploratory and/or analytical methods that can be effectively applied in settings where data creation and use is separated. It was developed applying a design science approach foundation on the basis of semiotics theory and DQ dimensions. Moreover, the approach is refined based on focus groups sessions with data custodians and based on repeatability evaluation results.

2.3.5.3. Data quality problems - Summary and main conclusions

The amount of various approaches, methods and models dealing with DQ topic over the years certainly indicates the interesting and high importance that researchers and scientific literature ascribe to this topic and to the variety of business' problems involved in it.

From an in-depth look at the above approaches, methods and models it can be seen that some of them have similar emphases and some have differences but in practice complement each other. In general, the majority of the approaches and models presented above concern user-centric and DQ requirements are determined in a top-down manner following organizational structures and data governance frameworks. In fact, many of these approaches and methods (e.g. Willshire & Meyen, 1997; Lee et al., 2002; Pipino et al., 2002; DeLone & McLean, 1992; 2003; Belhiah et al., 2016; Heinrich et al., 2018) are measurement-oriented and propose usable metrics in the context of DQ in practice and analyze the DQ dimensions indirectly mostly on databases and technical aspects based on applying their measures and assessment techniques, with the overall goal being the success of implementing the IS.

On the other hand, the other group of the approaches and models presented here (e.g. Price & Shanks, 2005; Even & Shankaranarayana, 2007; Frank, 2008; Bai et al., 2018) focused on the effectiveness and efficiency of organization by determining the effects of data quality on decision making processes and its improvement for the benefit of the organizations. Another sub-group of the above approaches and models (e.g. Wang & Strong, 1996; Wand & Wang, 1996; Dvir & Evans, 1997; Wang, 1998; Kahn et al., 2002; Yang et al., 2017) assumed that an analogy exists between quality issues in product or service manufacturing and those in information manufacturing and the quality of information can be examined through approaches and techniques taken from industrial quality domain (e.g. TQM, DMAIC) to meet consumer expectations.

The main advantages of the approaches and models presented here are the reference general quality approaches and the analysis of information quality based on DQ dimensions at the organization level especially by quantitative assessment of DQ with usable metrics. However, quantitative assessment of DQ and its technical aspects are not included in the scope of this study. Furthermore, these approaches and perspectives are limited, since they do not focus on process requirements at the design stage, and none of them links it to the defined process goals and their achievement in high quality by the process.

In general, each of the approaches and models presented here has a unique contribution to the importance and usage of DQ dimensions in general, and for DQ improvement in particular. However, none of them have found a deep and specific reference to the impact of the DQ dimensions and their relationship to the existing dependence of the information items on the design of BPs as a preliminary quality product for the analysis and design of the new IS.

To cover these gaps and as main conclusions, we need to design and develop a new method that complements the limitations and business needs in the approaches and methods presented above. In fact, the new expected method is more closely and based on some foundations and assumptions presented by the approach of Wand & Wang (1996). Furthermore, we are going to extend it to the analysis and design phases of BPs as the basis for the analysis and design phase of the IS. In addition, in the new method we are going to deepen the use of DQ dimensions and identify and analyze the dependencies that exist between data items in the processes and their mutual impacts and effects.

2.3.6. Data quality problems and business process design

The importance of high data quality and the need to consider data quality in the context of business processes are well acknowledged (Glowalla & Sunyaev , 2013a; 2014a). Generally, models and methods are also used in the wider area of ISs to represent BPs of organizations; processes are described in terms of activities, their inputs and outputs, causal relationships between them, and functional/non-functional requirements but the notion of DQ within BP design methods has received relatively little attention. Such methods and models are needed in order to help the process analysts, designers or practitioners, to analyse and foresee process behaviour, measure performance, and design possible improvements (Batini & Scannapieca, 2016).

In this sense, we rely on an argument raised by Cappiello et al. (2013) that the analysis of the relevant literature has been done revealed two different points of view, which sometimes are dealt with jointly, about the usability of DQ: the first one related to assure quality of data as meeting requirements at process design phase, and on the other hand, the second one, related to the adequate design of BP to guarantee the adequate levels of DQ at execution time. In principle, we will focus this study on the first option, i.e. to ensure the DQ by setting DQ requirements for the process design phase. Moreover, it is a fundamental principle in software engineering that design errors should be prevented as early as possible. The later the errors are identified, the more cost and effort are needed to fix the errors. This principle also holds for designing and automating BPs (Sun & Zhao, 2013).

As already mentioned, BPs consume and produce information and are based on information flows (English, 2001; Schultz, 2006), and "many business processes leave their 'footprints' in transactional information systems" (Aalst, 2005, p.198). The ability of companies to react quickly is often due to their capacity to handle information and communication technologies in support of business evolution requirements (Daoudi & Nurcan, 2007). Data errors can be caused by both design and operational problems (Arachchi et al., 2015). Hence, the information flow in BP becomes a critical issue for process success and it should be examined as recommended by Soffer & Wand (2007).

In the past two decades many researchers have addressed the problem of DQ from different organizational perspectives. Xu et al. (2002) developed a framework for identifying DQ issues in implementing ERP systems. Other researchers e.g. (Fisher & Kingma, 2001; Fisher et al., 2003; Jung et al., 2005; Slone, 2006; Raab, 2008) investigated the impact of DQ on the performance of organizational units (including individuals), evaluated the costs and benefits of DQ initiatives, and assessed the impact of DQ on operations and decision making.

Another approach to cope with this issue is to improve, redevelop, and revise the DQ perspective of BPs affected, as a major basis for IS design, since BPs are supported, controlled and/or monitored by ISs. For example, Shankaranarayanan et al. (2000) suggested managing information as an information product (IP). They presented a more elaborative approach which proposes the information product map (IP-MAP) as a method to systematically model the manufacture of an IP. It permits the specification of BPs by means of a conceptual map and a sort of activity diagrams. The IP-MAP is an extension of the Information Manufacturing System (IMS) model proposed earlier by Ballou et al. (1998). It offers several advantages including the ability to visualize the manufacture, implement continuous improvement and quality-at-source, and measure the quality of the IP using appropriate quality dimensions. Furthermore, IP-MAP serve as a foundation upon which a suite of quality dimensions may be identified and implemented for managing the IP quality.

DQ problems can also appear in enterprise resource planning (ERP) systems. Cao & Zhu (2013) investigated inevitable data quality problems resulting from the tight coupling effects and the complexity of ERP systems in their case study in China. The findings show that organizations that have successfully implemented ERP can still experience certain data quality problems and the efficient operation of ERP systems largely depends on DQ. They identified major data quality problems in data production, storage and maintenance, and utilization processes. The researchers also analyzed the causes of these DQ problems by linking them to certain characteristics of ERP systems within an organizational context.

Glowalla & Sunyaev (2013b; 2014b) examined in qualitative research how ERP systems are used within DQM to provide a DQ fit for tasks within the insurance sector. Basically, ERP systems are unlikely to fit completely to organizations' needs specifically concerned with the service sector. They examined the information-intensive insurance sector applying the task technology fit (TTF) theory in an explorative study. TTF theory was developed in order to assess linkage between IS use and individual performance, depending on the IS fit for tasks.

The researchers conducted semi-structured expert interviews as a focus group, with participants in IS/IT strategic decision making. The results show the need to apply a broad perspective on ERP systems to account for DQ impacts beyond the system itself. Their post-implementation perspective adds valuable sector-specific and general insights for practitioners and researchers, linking ERP systems' use to DQM in evolving task environments. However, due to the small number of organizations, a generalization of this study is limited to specific implications and rich insights, which need to be reevaluated when used in other contexts. In addition, the selection of participants focusses the study on IT strategic issues only. Other DQ issues may exist from a functional accounting or distribution perspective.

Application data are an integral part of business processes. Data can be created, modified, and deleted during the execution of business processes (Weske, 2012; 2019). Since BPs consist of a set of activities that are related, these activities operate on an integrated set of application data. At the process level, data dependencies between process activities exists and is typically described by data flow (Weske, 2012; 2019).

Sun et al. (2006), and later Sun and Zhao (2013) in their continued study, emphasized the importance of data-flow perspective in workflow analysis. According to Sun et al. (2006) workflow technology has become a standard solution for managing increasingly complex BPs. Successful BP management depends on effective workflow modeling and analysis. They argue that given a correct process sequence, errors can still occur during workflow execution due to incorrect data-flow specifications, and no formal methodologies are available for systematically discovering data-flow errors in a workflow model. The researchers presented a data-flow analysis framework for detecting data-flow anomalies. Their paper takes a minimalist approach and defines three basic data-flow anomalies - missing data, redundant data, and conflicting data, based on concepts of data dependencies in processes. Also, they have developed algorithm that can be used as a road map for the implementation of the data-flow perspective. The data-flow framework includes two basic components: data-flow specification and data-flow analysis; these components add more analytical rigor to business process management (Sun et al., 2006).

Basically, Sun et al. (2006) research includes two important advantages: first, it helps workflow practitioners in discovering data-flow errors in a workflow model and systems. In addition, it adds more analytical rigor to business process management at design stage, since there have been only little treatments of the data-flow perspective in the literature and no formal methodologies are available for systematically discovering data-flow errors in a workflow model. The second important advantage is the integration idea of workflow model with the concepts of data dependencies in processes. However, there are some limitations here: first, this research focuses on workflow model and is commercial workflow systems-oriented, lying on the assumption that the quality of requirements specification fully exists. Second, the approach is applicable when a dataflow specification is well connected, complete and concise. Furthermore, the researchers did not address the approach or method that defines how they collected the data or information flows in the processes and whether the requirements specification of the data or information flows in the processes, and if it correctly represents the data flow transformation in the process. In addition, they did not relate to the approach or method of how to deal with incomplete or partial data or information that is used in processes.

Soffer (2010) explores the inaccuracies of data as the situation where the IS does not truly reflect the state of a domain where a process takes place. In the mentioned research, the potential consequences of data inaccuracy concepts are discussed and formalized based on ontology foundations. The research has a main contribution by providing the foundation to support the design of robust processes that avoid problems related to data inaccuracy. However, this research is limited since the discussion is about inaccuracy dimension aspects only and the author ignored the data items dependency concept.

Ofner et al. (2012) suggested an approach that conceptualizes DQ for business process modelling and introduces DQ as further criteria in a multi-criteria decision-making process during BP re-designs. The approach is based on key concepts and metrics from the DQ management domain and supports decision-making in the process of re-design projects. Furthermore, the researchers used a case study and expert interviews to evaluate the suggested approach and showed that the DQ oriented process modelling approach facilitates and improves managerial decision-making in the context of process re-design and DQ is considered as a success factor for BPs and is conceptualized using a rule-based approach. The modelling method represents one of the first comprehensive approaches for integrating a DQ perspective into BPM.

Caro et al. (2012) presented the 'Business Process including Data Quality view point' (BPiDQ) methodology (see **Figure 23**) that is oriented to support the modelling and design of data quality-aware business process and the generation of DQ requirements for the software development. BPiDQ allows business people to include DQ needs in business process modeling using DQ Flags. Then, for each one of these DQ Flags, BPiDQ enables workers to specify DQ requirements that will drive improvements over the original BP model in order to guarantee the DQ level required. Furthermore, the methodology supports the specification of use cases for the data quality-aware software development.

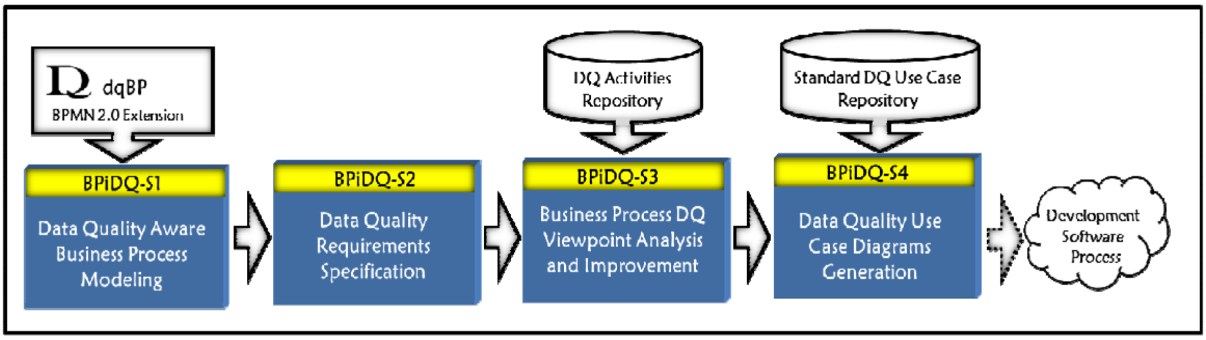


Figure 23. BPiDQ methodology to design DQ-aware business processes

[Source: Caro et al., 2012]

As shown in **Figure 23**, BPiDQ is composed of four stages. The first stage (BPiDQ-S1) in the BPMN Descriptive level, starts by introducing high-level DQ requirements into the BP model. In the second stage (BPiDQ-S2) the high-level DQ requirements will be refined in order to generate low-level DQ requirements based on a set of relevant DQ dimensions. In the third stage, (BPiDQ-S3) in the BPMN Analytic level, the DQ requirements will guide the data quality aware BP improvement that will imply the addition of new activities or the modification of the process flow. Finally, the fourth stage (BPiDQ-S4) supports the generation of use case diagrams to specify DQ software requirements. The main advantage of suggested methodology is the linkage between DQ requirements to business process using DQ Flags and is based on a set of relevant DQ dimensions. However, the main drawback of this method is that it does not take into account the dependencies and the dependency relation types between the data items.

Falge et al. (2012; 2013) examined collaborative BPs, which exchange data between enterprises, and where information is created and modified. They identified, by means of a qualitative content analysis on Business Networking case studies, DQ requirements of collaborative BPs in business networking and basically, to support DQM experts in organizations. Business Networking is defined as the organization and management of IS/IT supported business relationships with internal and external business partners. The research results showed which combinations of data classes (e.g. order data, forecast data) and quality dimensions (e.g. business rule conformity) are crucial for the different collaborative business processes in business networks. They used a set of DQ dimensions already known in the literature, and also defined a newly set of relevant DQ dimensions. These DQ requirements can be used by BP analysts and practitioners to identify needs for DQM initiatives more efficiently and structure the BP analysis.

Furthermore, in their follow-up study, Falge et al. (2013) argue that large and multidivisional enterprises need to establish Corporate Data Quality Management (CDQM) to achieve data of high quality in order to meet a number of strategic business requirements, such as enterprise-wide process harmonization, integrated customer management or compliance. They presented a new method for developing and implementing a CDQM strategy based on analysis outputs of focus group, expert interviews and participative case studies. On the one hand the method provides guidance to a CDQM team. On the other hand, for corporate executives' management the method ensures that the CDQM strategy is derived from their business objectives and that their requirements are systematically taken into account and fulfilled. However, new technologies supporting enterprise collaboration (e.g. social networking) with respect to additional requirements on DQ have not been examined. Also, the DQ requirements ignored data dependency concept.

Cappiello et al. (2013) introduced a top-down data quality-aware BPs model to support BP designers. They presented a BPMN 2.0 extension, allowing business analysts to specify DQ requirements in BP models. This extension allows business analysts to be aware of DQ issues and provides a systematic method for DQ management. However, this research is limited to BP modeling and the authors ignored data items dependency considerations and their relation type.

Based on Caro et al. (2012) BPiDQ methodology, Cappiello et al. (2013) proposed a comprehensive methodology shown in **Figure 24**, to support BP designers in identifying DQ requirements using DQ Flags, and selecting the required actions to satisfy such requirements during the design of BPs. With respect to Caro et al. (2012) methodology, the proposed methodology provides more details about the actions that the process designer has to perform in order to detect the most suitable DQ improvement activities and consequently change the BP model (Cappiello et al., 2013). It is important to note that this methodology is suitable for both the redesign of existing processes and the design of new ones.

Figure 24. Methodology to model DQ-aware business processes

[Source: Cappiello et al., 2013]

The main limitation of the proposed approach is that it starts from the assumption that the involved actors have a good knowledge of the analyzed BP. In fact, the lack of information about the analyzed BP could affect the effectiveness of the presented methodology. In addition, their approach is not formally anchored and does not support a detailed and systematic analysis of potential DQ issues in a business process.

Glowalla & Sunyaev (2013a) examined the varying applications of process modelling languages for process-driven data quality management (PDDQM) methodologies and techniques (such as Workflow, DFD etc.) based on reviewing the extant literature. Additionally, they examined how to integrate DQ aspects into existing process models while simultaneously controlling model complexity. For these purposes, the researchers reviewed 1,555 articles from 1995 onwards. From this collection, they focused and examined 26 articles and 46 process models in detail regarding the varying application of process modeling languages within organizations for PDDQM. Furthermore, they applied the conceptual framework depicted in **Figure 25** to cope and answer their research questions by providing two options to integrate data quality into existing process models: within-model integration and across-model integration. Within-model integration allows to enhance existing process models with DQ information by integrating DQ checks. Across-model integration provides a new process model with an information product-centric perspective, linking it to existing models. In addition, the researchers also examined the integration approaches' impact on the models' complexity and patterns for complexity reduction. This way one can know which patterns for complexity reduction can be applied, taking into account the existing process model characteristics.

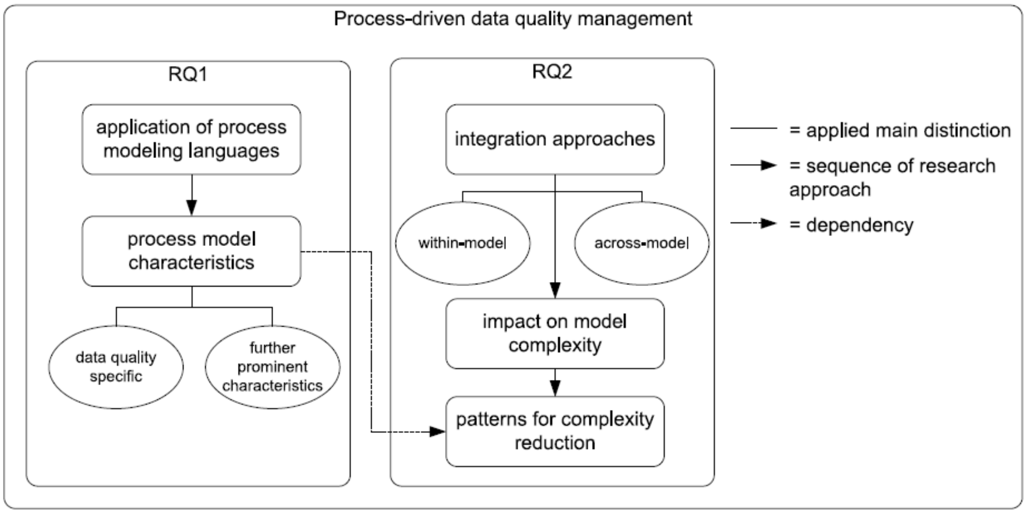


Figure 25. PDDQM - Conceptual research framework

[Source: Glowalla & Sunyaev, 2013a; 2014a]

Further into their research, Glowalla & Sunyaev (2014a) provided representational requirements for PDDQM that should be integrated within existing process modelling languages. The two main contributions of this research are: first, this synthesis provides practitioners examples of how organizations can apply well-known and mature process modelling languages, enhancing them with information about DQ that suits their particular organizational needs, instead of switching to new process modelling languages. Second, the integration of DQ into different process modelling languages and the context-specific application of these languages show existing needs and opportunities for further differentiated research in PDDQM.

However, some limitations exist in these researches: first, a specific selection by considered articles describing such an application of process-driven DQ led to the achievement of a small number of relevant articles. Hence, a generalization beyond the identified articles regarding the application of process modelling languages and methodologies is problematic. Second, the authors focus on the integration of DQ into process modelling only. Third, they ignored data items dependency considerations. Finally, they excluded processes that are inherent to IS/IT (e.g., the optimization of data warehouse internal processes).

Another research, made by Glowalla et al. (2014) provide a process-driven application of the combined conceptual life cycle (CCLC) model for process exploration and DQ improvement. They conducted an interpretive, in-depth case study in a medium-sized company, which launched a process optimization initiative to improve DQ in the certification process to evaluate and demonstrate its application. Therefore, the data quality dimensions applied in the CCLC model and their definitions merely provided a basis for identifying DQ requirements from data consumers' view. The dimensions applied allowed a holistic view on the identified relevant data quality requirements and issues from data consumers' view.

Zhang et al. (2014) presented an approach for discovering data quality issues using generic exploratory methods, which they derived through experimentation with a real data set based on public transport. The approach uses DQ requirements identification and can provide a generic set of guidelines for data-driven discovery of DQ requirements. Basically, DQ requirements are determined in a top-down manner (see **Figure 26**) following organizational structures and data governance frameworks. In such scenarios applying top-down approaches is not feasible. Users need to be empowered with data exploration capabilities that allow them to investigate and understand the quality of data sets and, subsequently, the implications for their use (Zhang et al., 2014).

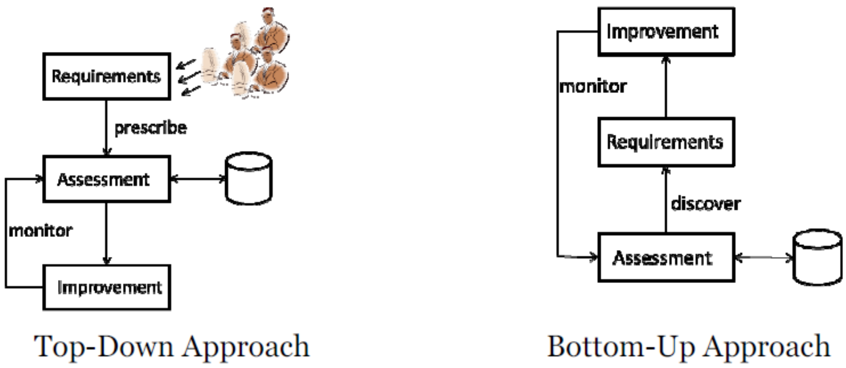


Figure 26. Two approaches for data quality requirements

[Source: Zhang et al., 2014]

In the top-down approach, derivation of user requirements comes first, and determines the metrics for the later stages of assessment and improvement. The bottom-up approach allows requirements to be dynamically discovered and adapted as the use and understanding of the data set expands, which can be instrumental in creating value from large and unfamiliar data sets. The question is to what extent can the quality of a data set be explored in a bottom up manner without access to well defined data quality measures. The researchers argue that a bottom-up approach has the potential to discover (at least part of) the DQ requirements using exploratory and/or analytical methods (Zhang et al., 2014).

The four stages of the approach provide guidelines for undertaking a systematic and comprehensive assessment of the syntactic and semantic DQ of an unfamiliar data set. By following these guidelines, the user can arrive at a concrete set of DQ requirements that can be used to perform data cleansing and DQ enforcement tasks. The researchers declared that they will conduct as a future work, an empirical study in the form of a focus group, to evaluate the validity of the results (DQ requirements) generated from the approach.

Gharib et al. (2018) discussed the importance of modelling and analysing IQ requirements during the early phases of the BPs design. They argue that such requirements should be analysed in their social and organizational context, which are very important, since BPs are mostly enacted by social actors. In particular, they proposed a multidimensional model as shown in **Figure 27**, based on their previous research (Gharib & Giorgini, 2015), and extend the framework for analysing IQ that considers social and organizational aspects and requirements while analysing the set of DQ dimensions (e.g. accessibility, accuracy, consistency etc.) and mapping these requirements into workflow nets with actors named WFA-net. WFA-net is a workflow modelling language, based on Petri nets for modelling and analysing control flow, information flow, and IQ requirements of the BP.

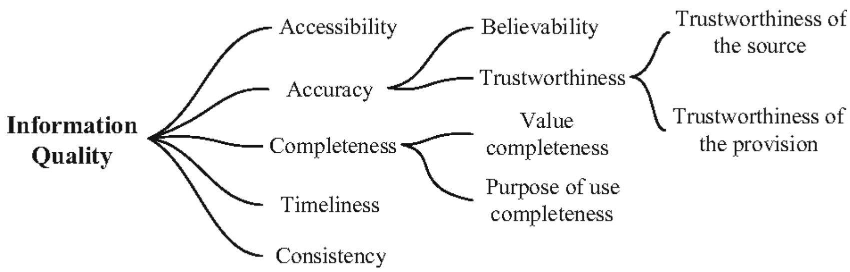


Figure 27. Multi-dimensional model for analyzing IQ for socio-technical systems

[Source: Gharib et al., 2018]

The researchers illustrated and evaluated their approach to demonstrate its utility and efficacy by developing and implementing a prototype tool and test its applicability on a simulation basis with artificial data, on two realistic scenarios abstracted from the US stock market system (e.g., New York Stock Exchange (NYSE), NASDAQ, etc.) as a case study (Gharib et al., 2018). The approach's process is shown in **Figure 28** and is composed of three main phases, namely: Modelling, Mapping, and Analysis.

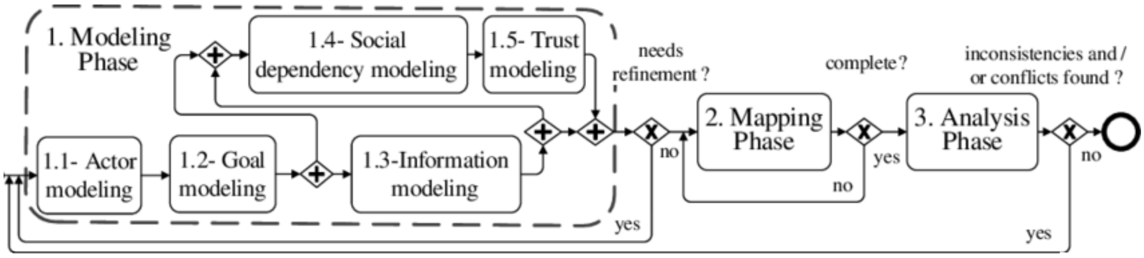


Figure 28. The process for modeling and reasoning about IQ requirements in BP

[Source: Gharib et al., 2018]

In **Figure 28** the modelling phase aims to model the IQ requirements of the system-to-be in their social and organizational context, where the BP is executed. This includes actor, goal, information, social dependency and trust modelling. When the modelling phase is completed, and there is no need for any further refinements, we can proceed to the mapping phase.

The mapping phase aims to map the IQ requirements model that has been produced in the previous phase into activities of the WFA-net to represent the intended process taking into consideration the actors who are responsible for achieving such goals, and information that such goals produce, read, modify and/or send. When the mapping phase is completed, then we proceed to the analysis phase. The analysis phase aims to verify the correctness and consistency of the BP model based on a set of properties of the design that can be used to verify the correctness and consistency of the control flow, information flow, and IQ requirements of the WFA net, i.e., the WFA-net is correct and consistent, if all of these properties possess.

However, this approach suffers from the following limitations: there is a binary requirement satisfaction i.e. the approach only deals with binary requirement satisfaction (e.g., a goal can be either satisfied or denied). In particular, the approach does not support qualitative requirements reasoning and only deals with binary DQ requirement satisfaction, i.e., information can be either accurate or inaccurate, believable or unbelievable, etc. In addition, all DQ dimensions have the same priority and importance to the system. For example, information accuracy has the same importance for the system as information consistency, completeness, etc.

Other limitations are: the approach cannot deal with more than one BP at the same time and it does not support customized analysis, it only supports verifying all the properties of the design, i.e., a user cannot choose which properties of the design to be verified. Also, the suggested tool installation is not user friendly, since it requires several applications to be installed on the host machine (e.g., Java, Sirius, Acceleo, etc.) to run appropriately (Gharib et al., 2018).

For the convenience of the reader, we summarized in **Table 11** the main methods and approaches as mentioned above, with a description of their main advantages and limitations, with an emphasis on approaches or methods that deal with aspects of DQ requirements in the design of BPs.

Table 11. Comparison of methods and approaches

The author(s)	Main advantages	Main limitations
Sun et al. (2006) Sun & Zhao (2013)	• Framework for detecting data-flow anomalies in a workflow model. • Helps workflow practitioners in discovering data-flow errors in a workflow model. • Focus on the concept of data dependency. • Some data quality dimensions took in account.	• The focus is on workflow model only and commercial workflow systems oriented. • Lying on the assumption that the quality of requirements specification exists fully. • The approach is applicable when a dataflow specification is well connected, complete and concise. • The focus on data-flow anomalies but the analysis does not deal with accuracy dimension as a significant part of DQ requirements for processes. • The researchers did not refer to the types of dependency relation.
Soffer (2010)	• Formalizing the data inaccuracy concept for incorporating data in-accuracy considerations in process models. • Based on ontological foundations.	• Focus on data inaccuracy only. • Workflow and process model oriented. • Ignoring from data dependency concept.
Ofner et al. (2012)	• An approach that conceptualizes DQ for business process modeling. • Introduces DQ as further criteria in a multi-criteria decision-making process during business process re-designs. • Practitioners may benefit from the modeling method that solves a relevant problem known in practice.	• DQ dimensions are not used in the approach. • The authors ignored data items dependency considerations.
Caro et al. (2012)	• The methodology oriented to support the modelling and design of data quality-aware business process and the generation of DQ requirements for the software development. • DQ requirements based on a set of relevant DQ dimensions. • BPiDQ include DQ needs in business process modeling using DQ Flags.	• The methodology limited to business process modeling. • The authors ignored data items dependency considerations. • The researchers did not refer to the dependency relation types.
Falge et al. (2012; 2013)	• The method helps BP analysts and practitioners to identify needs for DQM initiatives more efficiently and structure the BP analysis.	• New technologies supporting enterprise collaboration (e.g. social networking) with additional requirements on DQ have not been examined. • DQ dimensions are not used in the method.

	• The DQ requirements help prioritize data classes for a DQ oriented collaborative BP design, taking into consideration the data lifecycle. • The method provides guidance to a CDQM team and for corporate executives' management the method ensures that the CDQM strategy is derived from their business objectives and that their requirements are systematically taken into account and fulfilled.	• Ignoring the data dependency concept.
Cappiello et al. (2013)	• The approach allows business analysts to be aware of data quality issues and also provides a systematic method for data quality management. • The approach takes into account the cost of implementing actions and the benefits of doing so. • The approach conducts the most suitable data quality improvement activities. • The approach is suitable for both the redesign of existing processes and the design of new ones.	• The methodology is limited to business process modeling. • Based on the assumption that the involved actors have a good knowledge of the analyzed BP. • The authors ignored data items dependency considerations and their relation type. • Lack of information about the analyzed BP could affect the effectiveness of the methodology.
Glowalla & Sunyaev (2013a; 2014a)	• Data and data quality aspects are considered in the context of the process modelling languages. • The authors provided two options to integrate data quality into existing mature process models. • The approach provided representational requirements for PDDQM that should be integrated within existing process modelling languages.	• Focus on integration of data quality into process modelling languages only. • Specific selection by considered articles describing such an application of process-driven data quality led to a small number of relevant articles. • Requirements based on data quality dimensions are not provided. • The authors ignored data items dependency considerations and their relation type. • The authors explicitly excluded processes that are inherent to IS/IT systems.
Gharib et al. (2018)	• Proposed a multidimensional model and approach for analyzing IQ requirements in their social and organizational context. • The approach is based on a well-known Petri Nets foundation. • Emphasized the importance of modelling and analyzing IQ requirements during the early phases of the business processes design. • Advocate that IQ requirements should be analyzed in their social and organizational context. • DQ dimensions are used in the approach model.	• Focus on integration of data quality requirements into WFA-net modelling languages only. • The approach deals only with binary requirement satisfaction i.e., some requirement e.g. goal can be either satisfied or denied. • The approach deals only with binary DQ requirement satisfaction i.e., information can be either accurate or inaccurate, believable or unbelievable, etc. • All DQ dimensions have the same priority and importance to the system. • The approach cannot deal with more than one BP at the same time. • Ignoring from data dependency aspects and their relation type.

[Source: own study]

2.4. Discussion

Information quality is a key success factor for a successful company in a dynamic environment and it is considered a critical issue in organizations in order to achieve effective and efficient BP design and in supporting IS development. Many organizations are starting to assess and improve the quality of their data and information resources and most of them recognize their existing dependency on IQ for everyday business operations and for decisions making.

The accelerated growth of BPM field over the years has led many researchers to recognize its importance and focus on methods and tools for designing, enacting, and analyzing BPs. However, the existing literature regarding BPs design and validation lacks methods and models for dealing with quality of information flows in BPs and design phase errors. The existing design methods in BPM remain manual and require a lot of effort, and therefore result in inefficiency in design tasks and potential errors in BP design. Furthermore, the focus of research attention has been on combining data flow with activity flow i.e. data-aware process design and it has been undertaken already to some extent yet avoided design time errors. This issue is crucial when it comes to numerous instances of BPs models, which bring us to the world of Big Data analysis and the additional challenges emerging from this fact.

Successful accomplishment of IS projects is also a crucial challenge and remains a timeless goal for researchers and practitioners. Moreover, BP design and IS design in enterprises are often not well aligned and the complex interrelations between BPs and ISs are not adequately understood and currently considered to be in development.

In order to cover these gaps and as a response to these challenges, in the next section we are going to propose a new conceptual model and a new method that deals with these issues and the limitations in the variety of approaches and methods presented above. Recall, our research goals are mainly concerned with the BP design and validation activities to reflect the requirements from IS as represented in the real-world. This kind of conceptual model and method can provide support and enable us to identify and define relevant attributes and dimensions of DQ. These attributes and dimensions will later also serve as a basis for IS requirements. Furthermore, they can serve as complementary solutions for BP and IS designers and practitioners, help them to better perform the analysis and design tasks of BPs and to achieve better DQ requirements at this stage for business needs.

As already reviewed above, many researchers emphasized the challenge concerns providing approaches, methods and models that would support the quality of analysis and design of BP to be more robust, and avoid problems related to DQ aspects and their dimensions in order to ensure their correctness. In addition, DQ aspects and BPs design have not received particular attention in literature so far, and were not discussed enough jointly. There is a great challenge in dealing with DQ deficiencies derived from data dependency aspects and coping with their possible effects in order to achieve business defined goals on a high quality level.

Figure 29 is the illustration of the main idea and the research rationale as presented here.

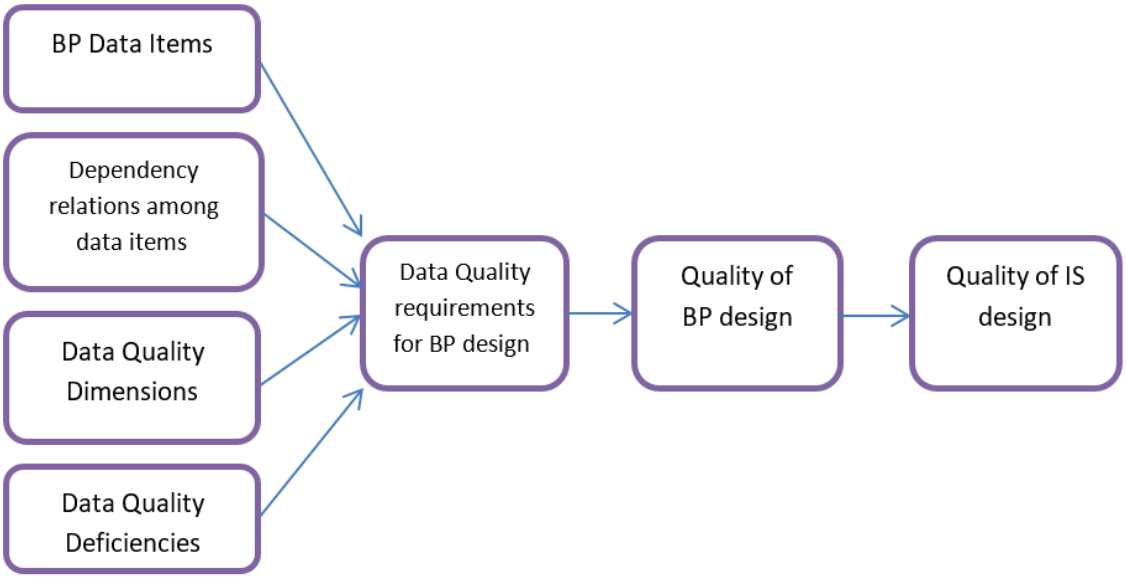


Figure 29. The research rationale

[Source: own study]

We can summarize it in two conclusions: firstly, it is required to analyze and plan the DQ requirements and design it with the expected outputs in BPR. Secondly, we need a new method which will help BP and IS analysts, designers and practitioners to better perform the analysis of BPs and allows them to examine and define implications and DQ requirements at this stage, in order to establish a new IS with higher quality. Both the above issues are challenging and suffer from lack of theoretical basis, studies and application methods.

2.5. Summary

The chapter provided a literature review and references about BP and DQ domains with an emphasis on the impacts on quality of BP design. In addition, as part of covering the first research goal, various aspects, constructs and problems of DQ are described and reviewed. Furthermore, the main challenge for BP and IS analysts, designers and practitioners, is identifying the requirements that impact on BP design quality. These requirements must be considered and examined during the early stages of designing the processes.

In addition, previous solutions, approaches, methods and models used today are presented above in order to identify gaps and limitations in these approaches, methods and models in comparison to the existing new needs and enhances them in accordance of BP and IS designers and practitioners expectations and requirements and to mitigate them in BP analysis and design phases to achieve high quality level of BPs. Special focus is given in our research to the relationship between the quality of BP design phase and the quality of design of a new IS and their effects in achieving successful accomplishment of IS projects as a crucial challenge and timeless goal for IS researchers and practitioners.

Part II: The Research Artifacts

3 Research Artifacts Development

Design research in information systems (IS) builds and evaluates artifacts that address business needs (Tremblay et al., 2010). The design of artifacts can be described as having two phases repeated in an iterative pattern: the development of the artifact and its evaluation (Tremblay et al., 2010). Design science research is inherently a problem-solving process. The fundamental principle is that knowledge and understanding of a design problem and its solution are acquired in the building and application of an artifact (Hevner & Chatterjee, 2015).

Part II of this dissertation addresses primarily the design and development process of research artifacts to improve BP's design and to achieve high quality and alignment between BP designs and IS design.

The first sub-chapter (3.1) introduces the background, the idea and foundations of real-world (RW) representation with BPs and ISs and the mutual impact between BPs and ISs representation. Then, in sub-chapter 3.2, the foundations, basic concepts, definitions and assumptions regarding the proposed conceptual model, as a meta-model for DQ problem are introduced and characterized. In sub-chapters 3.3 to 3.5, the constructs and requirements of our main research artifact are described. Finally, in sub-chapter 3.6, the main research artifact, **Data Quality Deficiencies Prediction (DQDP)** method and its steps and constructs are described. In addition, a summary of relevant conclusions and implications at the dissertation level are presented in sub-chapter 3.7, respectively concerning each of the two main artifacts.

3.1. The Real World System vs. Representation System

For decades, many researchers (e.g. Kent, 1978; Wand & Wang, 1996; Soffer & Wand, 2007; Soffer, 2010) emphasize the role of real-world (RW) in planning and design of BP and IS. These researchers make a clear distinction between objects behavior in the real-world system and their representation in process or in ISs.

Basically, an IS is considered a mean for representation of a real-world system as perceived by users and is built for use by the user whose view of the real-world system is captured in the design of the system. Moreover, the IS role is to be a reflection of what we have done in process without making any change and it does not make any operations and has no control on the changes. From our point of view, IS acting as a representation machine of what we have in the real-world and reflect later only the defined rules and controls for her.

The view of an IS as a representation of the real-world is not new. Kent (1978) for example, states that “an information system…is a model of a small, finite subset of the real world”. In addition, Wand & Wang (1996) indicate that the role of an IS is "to provide a representation of an application domain or real-world system as perceived by users". Furthermore, an observation of the real-world system shows that it is limited to a defined set of data items.

Basically, IS, as a means for real-world and process representation, can suffer from quality aspects such as poor level of representation, i.e. inconformity between the real-world system behavior and its representation in IS. In fact, not every inconformity has an impact on process performance quality and its outputs. Hence, we can find sometimes some cases that inconformity in real-world representation can disturb the anticipated systematic process, yet sometimes it can be unnoticeable.

A similar potential of quality aspects exists in BPs, i.e. inconformity between the real-world system behavior and its representation in BPs design. Business processes are defined and operate over a given domain and change its state. The design of many processes is done subject to the assumption that there are no problems related to the quality of the information and assumed perfect values used through data items in the process. Nevertheless, processes in practice, can suffer from quality aspects such as poor level of DQ along their activities or in communication between processes. Furthermore, process design without taking quality considerations into account is probably expected to fail as discussed above.

Because of a poor design, a process can fail and achieve one or more undesired and poor results in the process outcomes. These poor outcomes are a result of a low level of strictness about DQ dimensions, since there is a dependency over data values in process. Moreover, the data flow in processes and along their activities is the basis for data requirements represented in IS stage. Hence, if we want to ensure IS to work properly and presenting desired data values in high quality, we have to check these data items values represented by those data items before recording them to IS to ensure DQ and for presenting desired values.

For reader convenience, we adopt from Soffer (2010) a conceptual model (**Figure 30**) by using UML class diagram created by the OMG group (OMG, 2005) to summarize the terms and constructs and made a comparison of the two parallel levels of abstraction and their relationships.

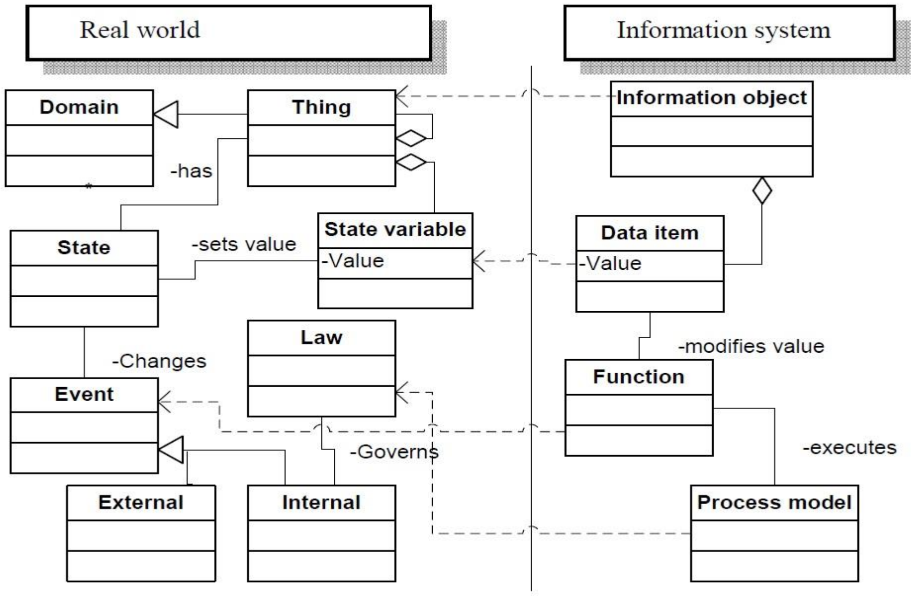


Figure 30. Real world/domain system vs. Representation

[Source: Soffer, 2010]

In **Figure 30**, the left side represents the domain or real-world system and the right side shows its representation i.e. IS. The domain or real-world system has state variables, i.e. data items whose values constitute the state of the domain or real-world system at a given time. These values are governed by the domain laws and are changed by external or internal events.

The right side is the representation of the domain or real-world system by data items and their values, where the value of a data item represents the value of state variable at a given domain. The law (or rule) refers to data items and determines whether or how their values change. These values are changed by activities that are performed by users. The users rely on representation values via IS to operate over the domain according to the law. Hence, the representation serves as a mediator to the law operations.

This kind of an approach will be used while developing the conceptual model presented in the next sub-chapters. We are going to relate it to business process representation based on the same analogy, where a business process representation is said to be a representation of a real-world system behavior if observation on the process representation at a given time enables us to infer the state of the real-world system (at the same or at a different time). Changes happening in the real-world should be reflected in data values appearing in ISs.

Hence, we can operate over the real-world and change its state based on its reflection. However, things sometimes do not work out smoothly and sometimes we cannot know exactly its real state, so we have to trust and base on its representation. Consequently, a process can fail and achieve undesired and poor results as outcomes.

The next sub-chapters will discuss and cope with this idea to develop a conceptual model and new method for specifying DQ in process design and its representation.

3.2. The Proposed Conceptual Model

Design-science research (DSR) efforts begin in principle, with simplified conceptualizations and representations of domain problems (Hevner et al., 2004). Within the information systems field, the task of conceptual modeling involves building a representation of selected phenomena in some domain. A conceptual model is an abstract framework for understanding significant relationships among the entities of some environment (Wand et al. 2000). High-quality conceptual modeling work is important because it facilitates early detection and correction of system development errors. It also plays an increasingly important role in activities like business process reengineering and documentation of best-practice data and process models in enterprise resource planning systems (Wand & Weber, 2002).

Based on the literature review discussed in the previous part and our own study, we built a conceptual model (depicted in **Figure 31**) for data quality assessment and business process design combined, to summarize the central constructs and terms discussed and their relationships by using UML class diagram (OMG, 2005). The developed conceptual model (also known as meta-model) will be applied for identification of relations of different data constructs within the process. This kind of conceptual model can also support and enable us to identify and define relevant dimensions for DQ requirements underlying the proposed method. These dimensions will later be a basis for the IS requirements. The description of the figure follows.

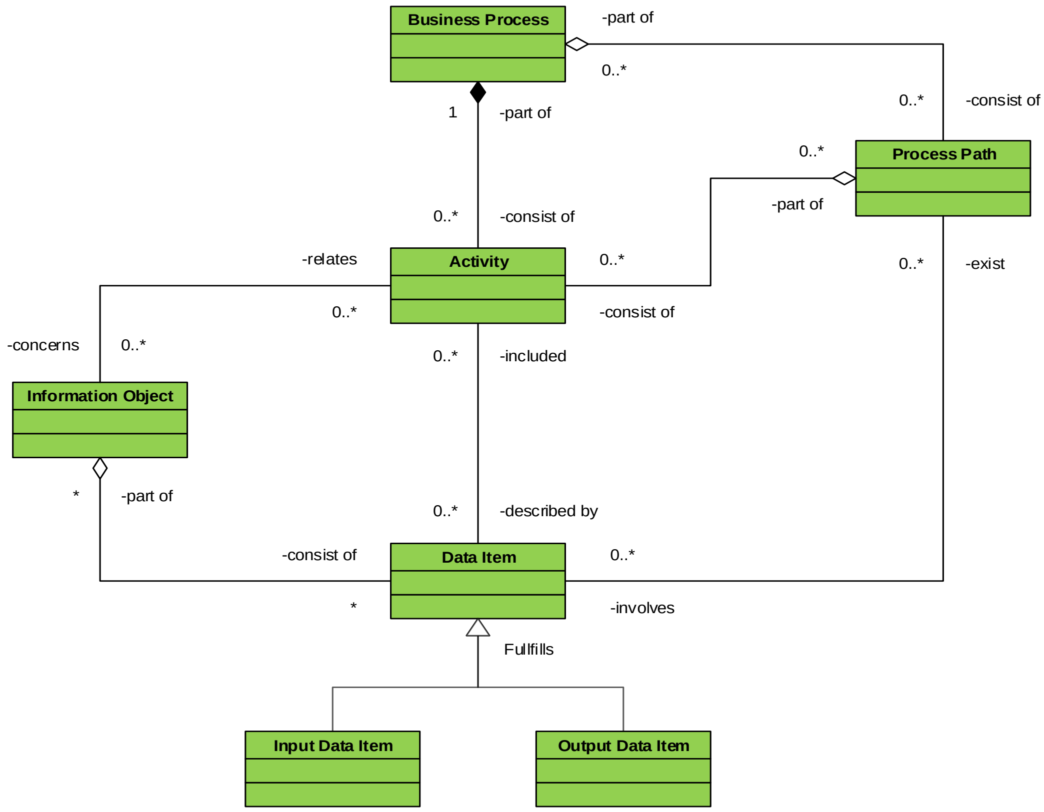


Figure 31. The conceptual model (meta-model) constructs

[Source: own study]

3.2.1 The conceptual model: Definitions and assumptions

Based on foundations as described above in detail, we extend it and formulate below a set of definitions and assumptions.

Definition 1: *Business Process* (named further as a *Process)* refers to a set of activities.

Process specifies which steps are required and in what order they should be executed. It is also known as: routing definition, procedure and workflow script (Aalst, 2004), e.g. purchase order, tax declarations and insurance claims process. $\boldsymbol{p}$ symbolizes a process.

Definition 2: *Activity* ($\boldsymbol{a_i}$) (also named task, step, process element, or work-item) is a logical step or unit of work in a business process, e.g. typing a letter, stamping a document, checking personal data, etc. (Aalst, 2004; Sun et al., 2006).

$\boldsymbol{a_i}$ is symbolizes a single activity and $\boldsymbol{A_p}$ represents a set of activities that define the process tasks domain, where: $\boldsymbol{A_p} = \{a_1, a_2, \dots, a_m\}$.

In general, an activity contains one or more data items and each activity can perform different processing and operations on a data item.

Assumption 1: The process should be valid in perspective of any modelling notation techniques.

Assumption 2: The represented structure of activities in process matches the structure in the real-world domain.

Definition 3: Basically, business process is a template for *process instance*s (also named cases or job), which are handled and need to be processed following the process definition, e.g. a customer order, a job application, an insurance claim, a building permit, etc. in general, each instance has a unique identity (Aalst, 2004).

Definition 4: *Data Item* $(\boldsymbol{d_j})$ is data element in process which contains value as a raw data, i.e. collection of characters, and can be given any value e.g. text, string, numeric, date, Boolean value, etc.

$\boldsymbol{d_j}$ is symbolizes a single data item in process and $\boldsymbol{D_p}$ represents a set of data items that define a process domain, where: $D_p = \{d_1, d_2, \dots, d_n\}$.

Assumption 3: The correctness of data items flow should be valid in perspective of process modelling notation.

Definition 5: *Information Object* is a general name for any data item or set of data items can be given as a collection of data values (e.g. form, report, formal document, application, certificate, entity, etc.).

Information object has a collection of data values in a business process which is stored in a database that can be classified and have stated relationships to other entities.

Assumption 4: Data Item in a given process can be considered also as an *Information Object*.

In practice, any data item value could be store in a field of database and used as an input for an activity or an output of an activity.

Assumption 5: Each activity is associated with two sets of data items - input and output data items set.

In general, each activity $\boldsymbol{a_i}$ takes in a set of input data items $\boldsymbol{D_{in}}$ and produces a set of output data items $\boldsymbol{D_{out}}$.

$\boldsymbol{D_{in}}(a_i)$ - represent set of input data items for activity a_i.

$\boldsymbol{D_{out}}(a_i)$ - represents set of output data items from activity a_i.

Where: $\{\boldsymbol{D_{in}}, \boldsymbol{D_{out}}\} \in D_p$.

The next construct in our model is the paths or trajectories in process.

Definition 6: A *Process path* is one possible trajectory or route within a business process that includes a set of activities in an order route from the initial activity to the final activity that have dependencies between them.

In fact, a trajectory in a business process describes the sequence of data processing path generated by the order and dependencies of activities and by process rules. Each trajectory describes the sequence of dependencies between the different data items.

Business processes are executed in order to achieve some pre-specified outcome(s) (Soffer & Wand, 2007; Dumas et al., 2018). The desired outcome of a process is a process goal. The outcomes of the process are derived from its goals. Furthermore, process goals are the basis for the DQ requirements definition.

Definition 7: *Process goal* is the desired outcome(s) to achieve by the process.

Assumption 6: Each process has at least one goal to achieve.

$\boldsymbol{g_k}$ is symbolizes a single goal of the process and G_p represents a set of goals to achieve by the process domain, where: $\boldsymbol{G_p} = \{g_1, g_2, \dots, g_n\}$.

$\boldsymbol{D_g}$ $-$ represents the set of data items values that define the process goals to achieve by the process, where: $D_g \subseteq D_p$.

Normal conclusion of the process means achieving a final set of data items values, defined by process goals and its rules. In practice, completion of the process in high quality means also to achieve the set of data values belonging to the defined process goal(s) in fitting to the expected outcome(s) by process manager(s), which can be under subjective perspectives. An examination of the compliance of the process with this subjective expected outcome(s) is beyond the scope of our research.

3.3. Requirements for New Proposed Method

3.3.1. Data quality requirements

BP design is based on a set of requirements which are collected and analyzed by professionals' process analysts and designers. As it is well known, the collection of these requirements constitutes a significant basis for the stages of analysis and design of a new IS and these requirements are crucial later when developing an IS. However, many empirical studies (e.g. Nasir & Sahibuddin, 2011; Nwakanma et al., 2013; Sadiq et al., 2013; The Standish Group, 2014; 2015; Komai et al., 2017) show that more than half of the errors which occur during ISs development stage are the result of inaccurate or incomplete requirements at the analysis and design stages. Furthermore, according to Meta Group, 60% – 80% of IS project failures can be attributed directly to poor requirements gathering, analysis, and management (Kaur & Sengupta, 2011).

BPs in practice can suffer from quality aspects such as poor design or poor level of DQ within their activities or in communication between different processes. Furthermore, designing a process without taking quality considerations into account is probably expected to fail as discussed above. Consequently, a process can fail and achieve undesired and poor results as outcomes. One of the most critical aspects of process design and analysis is specification's quality of data requirements and data flow in process activities (Sadiq et al., 2003) known as a non-functional requirements. Recall, functional requirements specify *what* an IS does, whereas quality requirements describe *how well* those functions are accomplished (Blaine & Cleland-Huang, 2008). A significant part of these quality requirements and in fact the main goal of this research, relates to DQ requirements.

Definition 8: A *data quality requirements (DQR)* are set of quality requirements concerning data items values grouped by DQ dimension.

Data quality requirements (DQR) are increasing as a wider range of data becomes available and the technology to mine the data shows the value of data that is 'fit for use' (Kerr & Norris, 2004) and without such a structure for understanding organizational information requirements, information cannot be managed (Auster & Choo, 1996). Both researchers and practitioners have widely recognized the need for an IS to deliver real quality to its stakeholders (Blaine & Cleland-Huang, 2008).

3.3.2. Using focus group (FG) method for collecting DQ problems and requirements

Our main challenge, derived from the first research goal, is to get to know the research domain in depth for gathering DQ problems aspects and collecting deficiencies and requirements from the real environment of the studied field in order to improve the processes design. For this purpose, we found the *'focus group'* method as a useful and effective means in obtaining inputs from practitioners and end users and elicit requirements for the development the new DQDP method (Kontio et al., 2004; Gregor, 2006; Belanger, 2012). Later, we will evaluate it, inter alia, on case study in company from forwarding and moving industry as a real-world environment, since it is considered to be rich in processes and in data and information items.

Focus groups are a popular qualitative research method often applied to different areas (Brandtner et al., 2015). Focus group (FG) is "a special qualitative research technique in which people are informally “interviewed” in a group discussion setting" (Neuman, 2014, p.471) and "a way of collecting qualitative data" (Onwuegbuzie et al., 2009, p.2). Focus groups are carefully planned discussions designed to elicit group interaction and obtain perceptions on a defined area of interest in a permissive, comfortable and nonthreatening environment (Kontio et al., 2004; Krueger & Casey, 2014). In general, the procedure is that a researcher chooses an expert domain group (typically between 5-12 participants in group) for interviewing and for open discussion by all group members.

Focus group research has rapidly grown in the past 20 years in qualitative research (Neuman, 2014) and is now one of the most widely used research tools to investigate new ideas in many research fields. The use of focus groups in design science research suggests interesting opportunities and challenges (Stewart et al., 2007; Tremblay et al., 2010; Brandtner et al., 2015).

Focus groups have been helpful in developing and maintaining quality improvement efforts, test monitoring procedures or solution ideas, and generally understand issues relating to quality (Krueger & Casey, 2014). Furthermore, the focus group method can be used to collect characterizing information about specific research topic (Nili et al., 2017), current practices, experience, collecting lessons learned from recommendations; and identifying potential root causes of problems or phenomena to IS researchers as they allow researchers to get deeper into the topic of interest (Belanger, 2012). The focus group method is a fast and cost-effective method to obtain experiences from practitioners and users. It can provide content rich, qualitative information and reveal insights that are difficult or expensive to capture with other methods (Kontio et al., 2004; Nili et al., 2017).

In addition, we have chosen to use a *semi-structured interviews* method during the focus groups sessions. A *semi-structured interview* is a method of research used most often in qualitative research (Edwards & Holland, 2013). While a structured interview has a rigorous set of questions which does not allow one to divert, a semi-structured interview is open, allowing new ideas to be brought up during the interview as a result of what the interviewee says. The interviewer in a semi-structured interview generally has a framework of themes to be explored, for identification of insights into an issue from the perspective of participants or end-users, involving in qualitative textual data, including observational data (Kugler, 2002; Edwards & Holland, 2013). Using this method enables us a combination of group and personal discussion with the ability to incorporate tools such as open-ended questions. Furthermore, it can uncover rich descriptive data on the personal experiences of participants, and it can be used to develop a preliminary hypothesis, explain relationships and create a foundation for further research. Also, this method is used during both the early and late stages of exploring the research domain (Creswell, 2014).

3.3.2.1. Focus groups types: Exploratory vs. Confirmatory

The focus group method is particularly useful for two types of focus groups: exploratory focus groups (EFG), which are used for the design and refinement of an artifact and when little is known about the phenomenon but also can be used as confirmatory focus groups (CFG), which are used to test hypotheses and for evaluation and the confirmatory proof of an artifact's utility in the application field (Stewart et al., 2007; Tremblay et al., 2010).

Design science researchers seek to design an artifact. The two design research goals for using focus groups are the incremental improvement of the design of the artifact and the demonstration of the utility of the design. For this reason, we have suggested the different focus group types of EFG and CFG. While the objectives of the two group types are very different, the methods of analyzing the focus group data from both EFG and CFG can be similar (Tremblay et al., 2010).

Figure 32 illustrates the positioning of the two types of focus groups, exploratory and confirmatory, in the design research process.

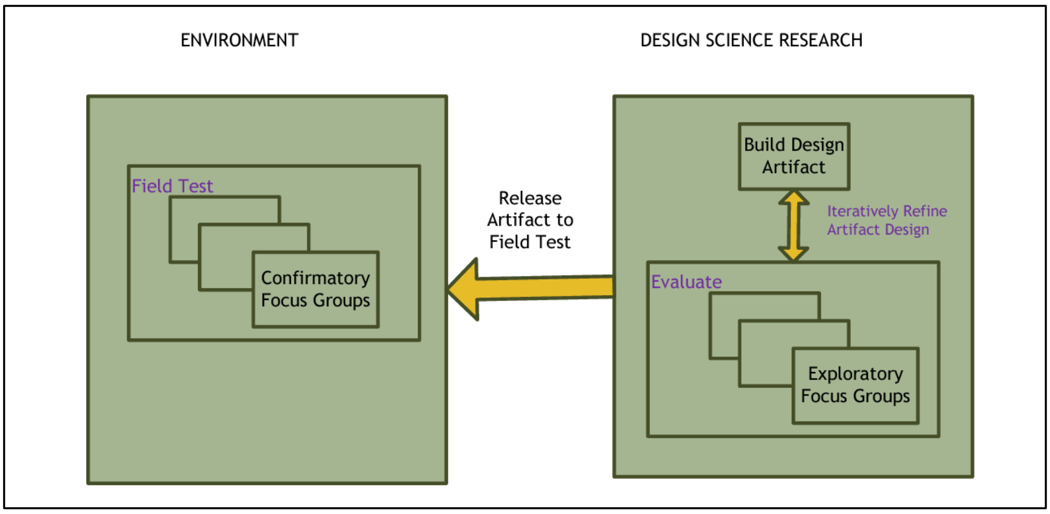


Figure 32. Focus groups in design science research (DSR)

[Source: Tremblay et al., 2010]

The cycle of building and evaluating in design research using EFGs continues until the artifact is released for a field test in the application environment. Then, the field test of the design artifact may employ confirmatory focus groups (CFGs) to establish and demonstrate the utility of the artifact design in the application field use. Basically, we analyze each type along this research, taking into consideration the two primary goals of design research: (1) design and refinement (EFG) and (2) evaluation (CFG) of a design artifact. In addition, for the refinement of an artifact design, focus groups can be applied to study the artifact in order to propose improvements. Once the artifact is released for field tests in the application domain, focus groups can be applied to establish its utility (Tremblay et al., 2010; Brandtner et al., 2015).

3.3.2.2. Focus group session steps

Over the years, various researchers on several academic sources and studies about focus groups research (e.g. Krueger, 2002; Kontio et al., 2004; Tremblay et al., 2010; Belanger, 2012; Krueger & Casey, 2014; Brandtner et al., 2015; Stewart & Shamdasani, 2015; Nyumba et al., 2018), have suggested approaches and outlined key steps required for proper planning and conducting of successful focus groups. Each of these steps is critical to the overall success of focus group meeting.

In this research, we adopted the approach of Tremblay et al. (2010), which developed as a procedural model (**Figure 33**) on how to apply and adopt focus groups especially in a DSR context by adapting traditional focus group techniques for use in design research projects. The model summarizes the basic steps to conduct a focus group in an effective manner. In fact, this procedural model would be applicable for any research-oriented use of focus groups.

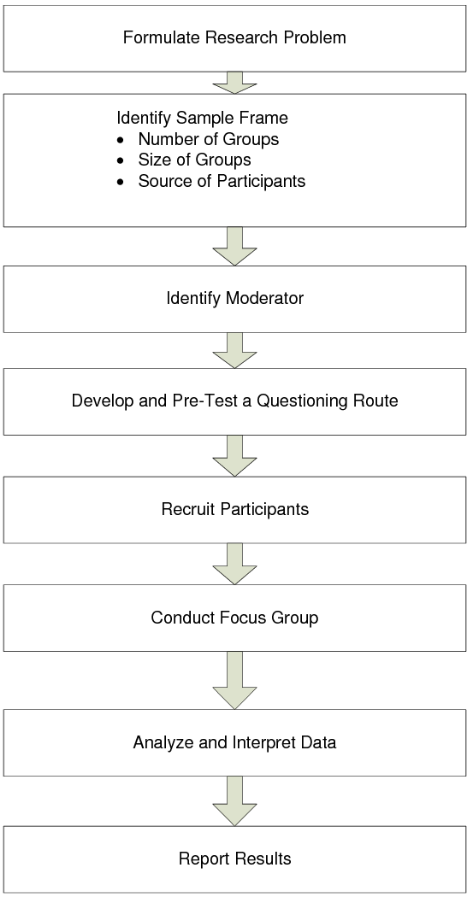


Figure 33. Focus groups steps

[Source: Tremblay et al., 2010]

For the convenience of the reader, we have described the method steps of the above process in Appendix I. The explanation of each step in the above method process is based on Tremblay et al. (2010) with several additions from other several academic sources and studies about focus groups research (e.g. Krueger, 2002; Kontio et al., 2004; Belanger, 2012; Krueger & Casey, 2014; Nili et al., 2017; Nyumba et al., 2018).

3.3.3. Establishing and conducting exploratory focus groups (EFGs) sessions

In order to explore the research domain in depth and to identify the information needs, DQ deficiencies, problems, requirements and expectations of all those employees and other parties involved in the domain processes, we established three exploratory focus groups (EFGs) with 24 participants in total, representatives from ten different organizations operating in a variety of areas and divided into two categories in general.

The aim of these focus groups study is obtaining feedback and opinions from participants i.e. domain experts, BP and IS analysts, designers and practitioners about the need, problems, requirements and expectations for designing a new method to cope with DQ aspects and deficiencies at process design.

The first two focus groups were established at Ocean Group Ltd., which is considered Israel's leading company and consists of domain experts from various occupations and senior practitioners' representatives who represent, in principle, end-user requirements for any new IS project. It is important to understand the nature of the work environment because its characteristics will affect the flow and use of information. Employees in specific environments make assumptions based on their work experience about what constitutes problem resolution and what makes information useful and valuable in their contexts (Hibberd & Evatt, 2004).

The third EFG consisted of a variety of BP and IS professionals and experts from various organizations and they were considered as BP & IS practitioners with extensive experience[4]. These information professionals possess competencies to understand how information is used, how it should be organized and structured, and how it is best sourced and presented – all key to mapping information flows (Hibberd & Evatt, 2004).

Generally, the EFGs meetings addressed the following main objectives:

- Exposing participants to the researched problem and research motivation, from scientific and business needs aspects.
- Collecting IQ aspects, problems and requirements from real-world business environment.
- Identify possible characteristics and components for designing the new method.

[4] More details about the focus groups participants and their profiles are included in Appendix C.

The following table (**Table 12**) presents the list of EFGs and the main characteristics of each of them.

Table 12. List of exploratory focus groups (EFGs)

EFG #	Date	Place	Number of participants	Type of participants	The main purpose / topic
EFG1	2.3.15	Ocean Group	8	Domain experts & IS practitioners	• Exposing participants to the researched problem and research motivation, from scientific and business needs' perspective. • Collection of IQ problems, aspects and requirements from business environment.
EFG2	13.5.15	Ocean Group	5	Domain experts & IS practitioners	• Exposing participants to the researched problem and research motivation, from scientific and business needs' perspective. • Collection of IQ problems, aspects and requirements from real world business environment.
EFG3	10.2.16	Zefat Academic College	11	BP & IS experts	• Exposing participants to the researched problem and research motivation, from scientific and business needs' perspective. • Discussion about IQ problems, aspects and requirements collection. • Definitions of the artifact's constructs. • Presenting a draft of the new artifact concept for feedbacks.

[Source: own study]

The following two charts show the distribution of EFGs participants profile according to profession and seniority years of professional experience. We can see in **Figure 34** that half of the EFGs participants (12) are domain experts and senior IS users (first category), considered as seniors IS end users and the second half of the EFGs participants are IS and BP professionals (second category). In addition, **Figure 35** shows that about 80 percent of participants have over 6 years of seniority experience.

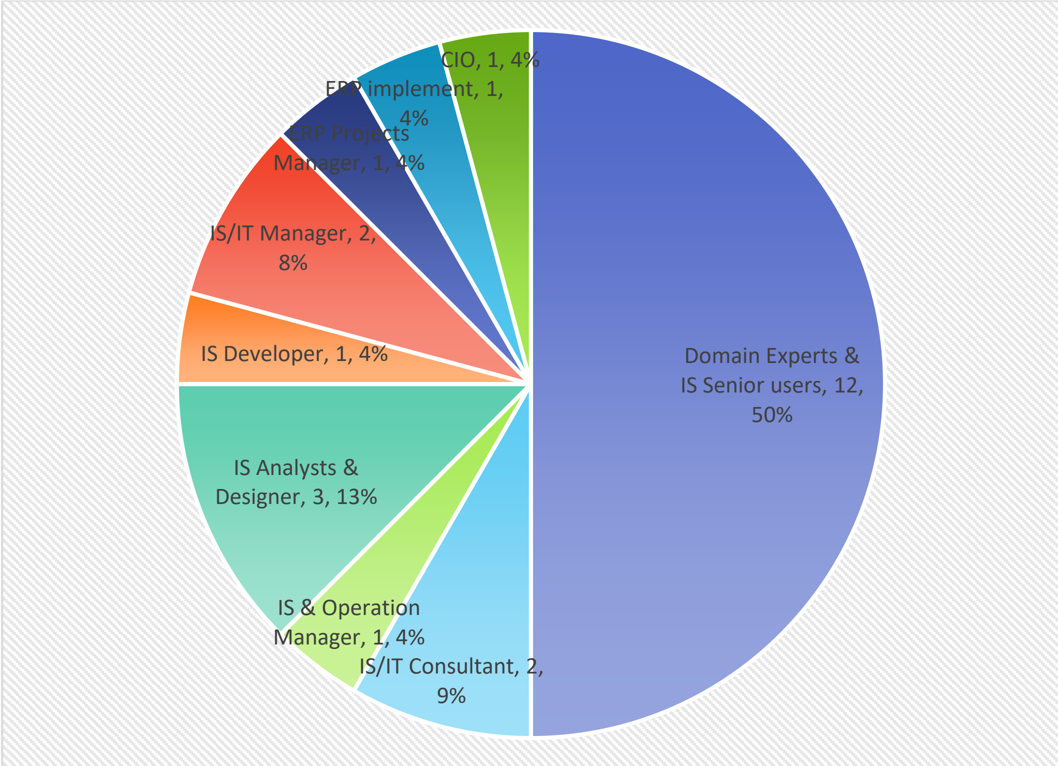


Figure 34. Distribution of EFGs participants by profession

[Source: own study]

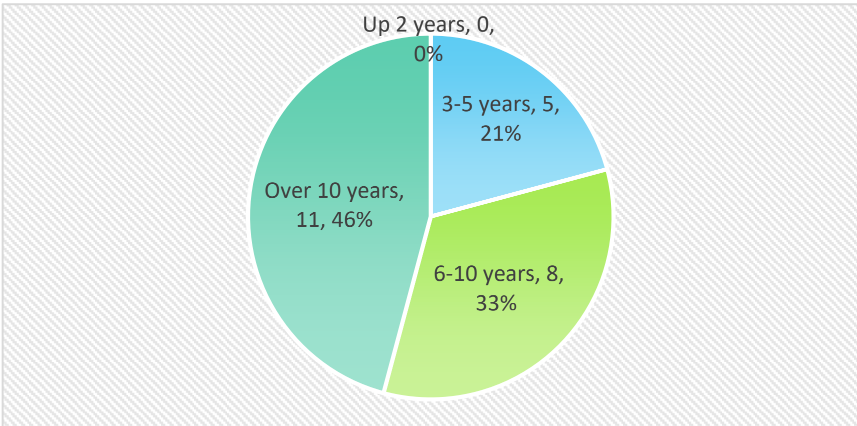


Figure 35. Distribution of EFGs participants by seniority years

[Source: own study]

3.3.3.1. The exploratory focus groups (EFGs) at Ocean Group Ltd.

The first two focus groups were established at Ocean Group Ltd. as formerly mentioned. Ocean Group is chosen since it is considered Israel's leading company and veteran in relocation and international forwarding and moving industry. Furthermore, this industry is generally considered to be rich in data and information and complexity of its processes, making it a challenging business. The company maintains an independent quality management system (QMS) and has international ISO 9001 certification, including maintaining a set of procedures, work instructions and internal forms. The company's management invests many efforts and resources to achieve high quality and continuous improvement in all its activities and processes.

In consultation with the company's CEO and president, we carefully defined the focus group's participants list, which is composed of thirteen managers and professional employees from various occupations with many years of experience. The participants were divided into two focus sub-groups, EFG1 and EFG2, according to the two main divisions at Ocean Group, i.e. relocation and forwarding division and operation and logistic division with 8 and 5 participants respectively per each sub focus group, which directly related to all the company's sales export services, relocation, forwarding, logistics, and storage processes.

The focus groups' participants were well experienced with the export service domain processes and well familiar with problems and failures that usually accompany the processes and can point out directions for solutions and recommendations for improvement. Furthermore, these focus groups participants were able to assess the potential of the improvement proposed by the new method we are developing.

As part of the preparation for the EFG session, participants were asked to prepare themselves for FG meeting in advance based on a preliminary questionnaire with open ended questions (see Appendix A) with some examples of processes from their workspace and to map their information flows. These questions directed participants to think about problems related to the quality of information they encounter in their daily work environment and their needs in this regard.

Generally, the FG meeting includes in principle a semi-structured interview with the participants. For this purpose, we also used interviews with the FG participants to elicit and gather a set of DQ deficiencies and flaws from FG participants' experience, per each chosen dimensions of data quality, as a basis for identifying DQ needs and requirements in process. The EFG discussion was documented in principle using FG script template (see an example in Appendix D).

The following **Table 13** presents a summary of the EFG 1 script as an example which took place at Ocean Group Ltd[5].

Table 13. EFG 1 script summary

#	Component	Description
1	Place	Ocean Group Ltd. / Relocation & Forwarding Division
2	Date	2.3.15 (at 10:00-12:00 AM)
3	Number of Participants	8 Participants - Managers and professional employees with many years of experience from various occupations.
4	The Topics / Purposes	• Exposing participants to the researched problem and research motivation, from scientific and business needs' perspective. • Collecting IQ aspects, problems and requirements from real world business environment. • Identifying possible characteristics and components for designing the new method.
5	The Main Activities	• Practical study and understanding of the problems, aspects and instances of poor quality of information quality dimensions in relocation, international forwarding and moving processes industry. • Practical study and understanding of the importance, problems and aspects of the idea of dependency between data items in relocation, international forwarding and moving processes industry. • Identifying potential process failures, as a result of IQ problems and existing dependency aspects among values of data items. • Receiving suggestions to improve the analysis and design of business processes. • Receiving expectations, characteristics and features needed for a new method.
6	The Main EFG Discussion Results, Feedbacks & Outcomes	**The FG discussion focused on two cases levels:** **1. The DQ problems and outcomes at process level:** ○ DQ problems of any data item values in process level and their impact on decision making in process. ○ DQ problems of any data item values in process and their impact on process performance and its quality. **2. The dependency idea and DQ problems at process level:** ○ The dependency between data items in a given process and its relation to process' goals. ○ The dependency in a given process on information items obtained from a third-party source. **The FG discussion main outcomes & results:** **1. The DQ problem and outcomes at process level:** ▪ **DQ problems of any data item values in process and their impact on decisions making in process:** ○ DQ problems may lead to or result in the wrong decision in process. ○ Good decisions are critical to process success and are based on high quality information. ○ There is a need to identify potential failures related to DQ problem in advance to achieve high quality in process design.

[5] The full script document of EFG1 is available in Appendix D

		▪ **<u>DQ problems of any data item values in process and their impact on process performance and its quality:</u>** o These kinds of DQ problems can lead to a waste of resources and unnecessary expenses in the process. o These kinds of DQ problems can lead to achieve certain outcomes that are not in the process' goals. o DQ problems in any data item values in process can cause the process to achieve undesired results in the process. o DQ problems of any data item values in process can cause a waste of time and valuable resources and to perform unnecessary tasks. o DQ problems of any data item values in process can lead the process to get stuck (i.e. be in deadlock situation). **<u>2. The dependency idea and DQ problems at process level:</u>** ▪ **<u>The dependency between data items in a given process and its relation to process' goals:</u>** o There is a great need to identify data items dependencies which relate to process' goals and their potential failures related to DQ problem in advance. o The dependency between different data items in the process is critical to achieve the process' goals. o Being able to identify different paths of dependencies relationships between data items in process in advance. ▪ **<u>The dependency in a given process on information obtained from a third-party source:</u>** o Can be a delay in receiving information from an internal or external source. o There is no full control in receiving of information from a third-party source. The information from a third-party source can also suffer also from a low level of DQ. o This kind of dependency is critical to achieve the process' goals and its outcomes. o This kind of dependency can cause the process to get stuck (i.e. be in a deadlock situation).

[Source: own study]

The following paragraphs summarize the main conclusions based on the above script table.

Data item values with DQ problems in process have a significant impact **on decisions making and performance quality** in process as mentioned above. In fact, DQ problems may lead to or result in the wrong decision in process. Many FG participants described some examples (e.g. current consumption data of various raw materials for inventory planning and forecasting or customer's real status and its visa type for customs issues etc.) where they made a wrong decision in the process because of DQ problems and emphasized that accuracy and completeness of data items are the significant dimensions of causes or results of wrong decision in process. Good decisions are critical to process success and based on high quality information. There is a great need to ensure the accuracy and completeness of all data items used in process to achieve high quality decisions.

Furthermore, DQ problems of any data item values in process have a significant impact **on process performance and its quality**. The FG participants described some situations and examples where they were faced with the kind of DQ problems that led them to perform unnecessary tasks or waste time, resources and unnecessary expenses in the process and thus, achieved undesired outcomes in the end of process. In some cases, due to a lack of information or a delay in receiving it from a third party caused the process to get stuck (i.e. be in deadlock situation) without achieving the process' goals. In this context, many FG participants emphasized the need to identify potential failures related to DQ problems and prevent them in advance.

The second topic of FG session focused on the **dependency idea between data items** in a given process and its relationship to DQ problems at process level. Many FG participants stated that dependencies between different data items in the process and their potential failures related to DQ problems should be identified in advance since they are critical to achieve the process' goals. They described some situations and examples where they faced DQ problems with kind of dependency between data items in given process (e.g. due to error or missing values of data items) and their results due to the dependency relationship between data items[6]. Furthermore, DQ problem can be created due to a delay in receiving a data item from internal or external source (i.e. missing values of data items is created). The dependency mostly concerns external third party to the organization which can cause the process to get stuck (i.e. be in deadlock situation) since there is no full control in process on information from the third-party source.

3.3.3.2. The exploratory focus group (EFG) of BP & IS experts

The third exploratory focus group (EFG 3) consists of participants who are considered BP and IS experts and are well experienced in BP and IS analysis and design[7].

The analysis of the problems raised by the domain experts' groups and the findings of the discussions in the previous EFGs that emerged from them, clearly show that our research motivation and problem and their aspects still exist and are relevant. Furthermore, with this focus group, we can confront and validate the various issues, problems and claims about DQ

---

[6] More examples of DQ problems are available in Appendix H.

[7] More details about the focus groups participants and their profiles appears in Appendix C.

deficiencies and flaws, raised by the professional managers and domain experts in previous focus groups. Additionally, we can examine the degree of relevance and the added value of all the improvement suggestions for BP design raised by them and compare it with the knowledge and professional experience of IS and BP experts.

To establish the presented problems and needs for an applied method, we created an open ended questionnaire dedicated to BP and IS experts (see Appendix B). The participants were asked at the beginning and as a preparation for the FG meeting to fill the questionnaires for FG discussion. Basically, this questionnaire contains two sections. In the first section, participants had to rate a series of statements and claims relating to the state of IQ and the dependence between data items and what their effects on the quality of the processes and the performance of the organization are in general, based on their professional and extensive experience. These statements and claims were defined in the orientation of the problems and aspects related to information quality and dependencies on data items in BP design requiring improvement as raised by previous FG participants.

The second section contains open-ended questions related to the use of existing models and methods and the extent to which they relate to the analysis of IQ and dependence on information items at process design stage and problems created there. In addition, the participants were asked to indicate whether these methods meet their needs and their recommendations for improvement and their expectations from a new method.

Generally, the rating is done based on a *Likert scale* (Likert, 1932). A Likert scale which is widely used in survey research is ideal for capturing attitudes, opinions, emotions and feedback from the respondent on virtually any topic (McLeod, 2008; Joshi et al., 2015), allows the individual to express his opinion and is an ever-present method of collecting attitudinal data (Subedi, 2016). A Likert scale (often called summative scale) is the sum of responses on several Likert items, where a Likert item is simply a statement, claim or a question that the respondent is asked to evaluate by giving it a quantitative value on any kind of subjective or objective statement or dimension, with the level of agreement/disagreement being the dimension most commonly used (McLeod, 2008; Neuman, 2014; Subedi, 2016).

The respondents were asked to indicate how much they agree or disagree with the following statements and claims. The respondents indicate their level of agreement or disagreement for each of the statements on a format of a typical five-level Likert item i.e.

range from a score of 1 to 5 scales, where the highest score, 5, represents strongly agree and lowest score 1 represents strongly disagree. One must recall that Likert-type data is ordinal data, i.e. we can only say that one score is higher than another, not the distance between the points.

Table 14 shows a summary of agreement level results of respondents' attitudes to a series of statements and claims presented in questionnaire (see Appendix B). The analysis is based on descriptive statistics, using mode, median and average measures, which are considered the most suitable and most commonly used for Likert scale analysis (McLeod, 2008), and presents the calculated scores of raw data provided by filled questionnaires responses of EFG3 participants per each of statements and claims. The data it is coded as follows: 5-Strongly agree; 4-Agree; 3-Moderately agree; 2-Disagree, and 1-Strongly disagree.

Table 14. Summary of agreement level - Mode, median and average scores

#	The statement / claim	Mode	Median	Average
1	At the analysis stage of business process and / or information system, there is little attention to the quality of information in the requirements definition.	3	3	3.5
2	In process analysis and specification there is little (if any) use of tools, models or methods for analyzing the quality of information in defining end user requirements.	5	5	3.9
3	There is a lack of tools, models and methods for BP & IS analysts and designers to achieve high quality analysis with information quality requirements.	5	4	3.9
4	Identifying the dependencies between data items in process is essential to obtaining high-quality information system specification outcomes.	5	5	4.7
5	BP & IS analysts and designers neglect the idea of the dependencies between various data items in BP & IS specification.	4	4	3.5
6	I need a method helping me to identify in earlier the dependencies between the data items and potential failures expected from the dependencies.	5	5	4.5
7	Poor quality information causes or may cause to achieve undesired outcomes and decisions and poor quality of process performance.	5	5	4.3
8	Poor information quality products can cause an organization's process(s) to get stuck.	5	4	3.6
9	Poor-level information quality outcomes can cause unnecessary delays in the process and failure to meet the planned time for the information system implementation.	5	5	4.8
10	Poor-level of information quality outcomes often result in unnecessary work, repairs and wasted valuable work time.	5	5	4.9

[Source: own study]

Figure 36 shows the above calculated scores data by distribution bar graph of EFG participants' agreement level per each statement or claim by scores of mode, median and average measures.

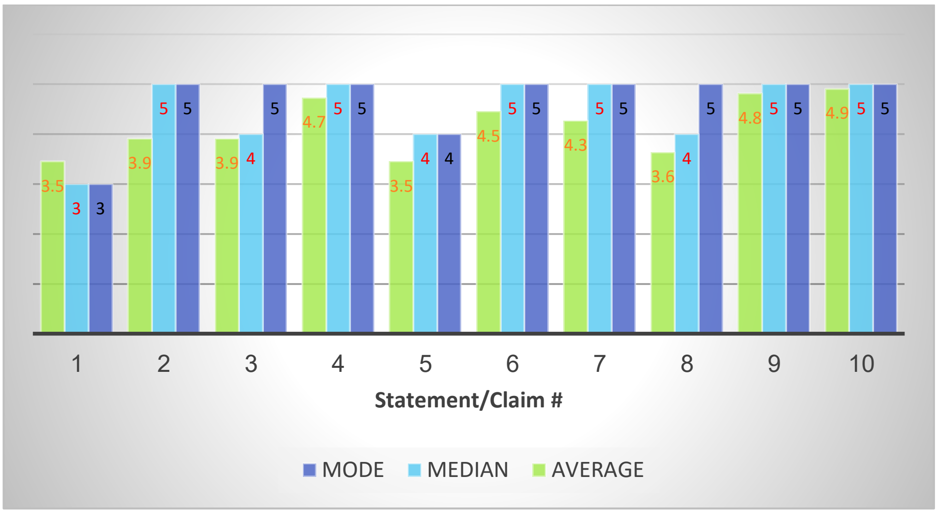


Figure 36. Distribution of agreement level per each statement by mode, median and average scores

[Source: own study]

Table 14 and **Figure 36** clearly show that according to the mode and median statistics measures calculated scores, the majority of EFG participants, i.e. BP and IS experts, rated agree level (4) to a strongly agree level (5) in nine out of ten statements and claims presented to them regarding the state of IQ in organizations today, except for claim 1 where the opinions were slightly different and the extent of the respondents' agreement with this claim was in general calculation at the moderately agree level. The average calculation scores of BP and IS experts' ratings show a similar relative result.

The EFG discussion about the open questions section followed by interviews and analysis with the participants about IQ problems and aspects collection confirms these results clearly. All EFG participants stated that they were almost constantly dealing with IQ issues and problems arising from dependencies between data items in the analysis and specification of BPs and the design of a new information system for the organization. In addition, they stated that, in general, they did not use specific methods for analyzing the IQ aspects and for identifying dependencies between the data items in the processes and their resulting problems. They stated that in principle, they depend mainly on the interviews outcomes and information obtained from domain experts and leading end users and through analysis and investigating of the information and knowledge that were obtained from them.

Furthermore, half of participants stated that a requirements specification document that relies solely on domain experts' interviews and questioning, provides a partial picture and basically covers partial knowledge and is expected to cause a variety of IQ problems later on and especially at the IS development stage. Another third noted that the lack of knowledge and incomplete information or IQ deficiencies is only apparent during the development stage of the IS, because end users are not always aware of all information needs and the need to disclose it in advance despite their extensive experience. The overall result at the enterprise level is that it wastes valuable time in completing the lack of information and knowledge, investing extra resources, and finally leads to a delay in the system development process for these end users.

The second part of the EFG discussion with BP and IS experts dealt with questions about how they address IQ issues and problems in light of the lack of appropriate models and methods and their expectations and requirements of a new method. The next graph (**Figure 37**) shows the distribution of answers to the question of how BP and IS experts are coping with IQ issues and requirements and dependency on data items with IQ aspects and problems at BP analysis and design stages.

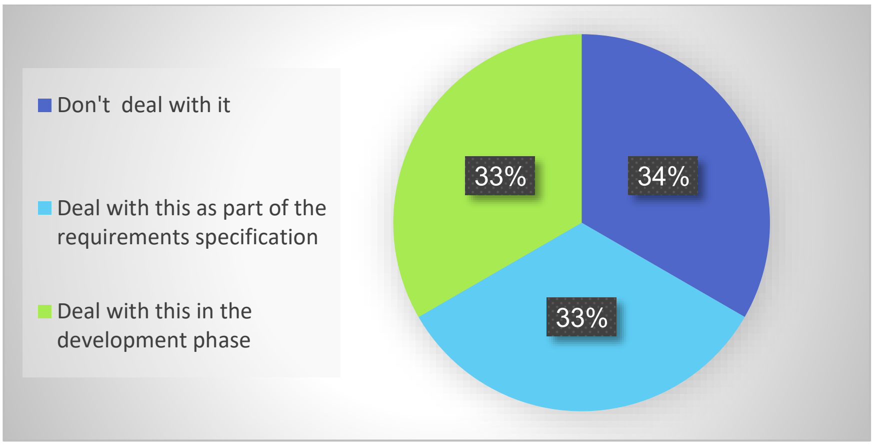


Figure 37. Distribution of BP and IS experts dealing with IQ issues

[Source: own study]

One third stated that they do not deal with them individually or deal with them very little in the analysis and specification phase of the processes and the information system. Another third stated that because of the complexity of the IQ subject they leave IQ issues to be addressed later at the development stage. Only a third stated that they addressed the issue as a part of the general specification of the requirements; however, not in a structured and systematic way.

A very interesting fact that emerged from this discussion is that none of the participants indicated that he or she addresses this issue individually or uniquely as part of the process and design phase of the BP or IS. Apparently, this is due to a lack of tools and methods focused on this topic.

The last part of the EFG discussion with BP and IS experts dealt with their expectations and requirements in the development of a new method for handling IQ aspects and dependencies between data items and to address IQ issues and problems in the context of appropriate models and methods.

The majority of participants (80%) noted the need and importance of a new method for the betterment of IQ and the specification's outcomes of the analysis and design of BP. Moreover, such a method can assist them in effectively addressing and defining IQ requirements in the early stages of BP i.e. at the analysis and design stages rather than the development stage of the IS itself. In addition, most of the participants expect that the new method will be able to direct them and help them identify relationships between various data items in the process and their dependencies and identify in advance potential deficiencies related to the DQ, based on the DQ dimensions, in the analysis and design phases.

The EFGs are useful in indicating changes needed in the design of the method while conducting each of the EFGs (Tremblay et al., 2010). After conducting each of the EFGs, significant changes were made to both the design artifacts (the expected method) and to the focus group scripts outputs for the next EFG.

Table 15 presents example design changes introduced after the three EFGs. These changes were required to be implemented later in the new suggested method as an artifact.

Table 15. Example design changes made after EFGs

EFG #	Component	Design change	Reason(s)
EFG 1	Research goal(s) and outcomes	They should be presented in terms of expected artifact and to whom it is intended in principle.	This definition is important since FG includes several domain experts and end users and some of them did not understand what their role is in the process and how they are affected by the research output.
EFG 1	Process goal(s)	The process goal(s) should be presented in terms of data items in the process and its (their) values that need to be achieved.	This will allow the progress of the process to be monitored through the dependency paths between the data items throughout the process and ensure that we have reached all the pre-defined data items as representations of the process goals.

EFG 1	Dependency relation types	Addition of dependency relation type of data or information item, received from an external entity i.e. third party.	Since the organization does not have full control over a third party, failure in receiving data or information item from it or delay in its delivery may leads to delay in the process and/or get stuck in some cases.
EFG 2	Dependency between data items	All dependencies paths between the data items can be identified in advance.	In this way, it is possible to identify the progress paths in the process in advance through the sequence of dependencies that exist between the information items.
EFG 3	Design the desired artifact	The design output should be a sequential step-by-step method as an artifact that guides the BP & IS analysts and designers what to do and what outcomes are needed at each stage.	Such a method will help BP & IS analysts and designers to systematically focus on the IQ and the dependencies aspects between the data items, so that no important outcomes will be missed, and the final specification will be of the highest quality.
EFG 3	Representation of dependency relation types	The proposed method should map any existing dependencies between data items with the appropriate dependency type.	In this way, the potential problems or deficiencies of any existing dependency can be identified in advance.
EFG 3	Matrices for all existing dependencies and for their relation types	Combined representation of all existing dependencies and their type on the same platform of matrix.	To reduce the overload in the specification work in order to mitigate the BP and IS analysts and designers work.

[Source: own study]

3.4. The Proposed Method - Components and Formalization

3.4.1. Identify the dependency relations concept among data items

Now we formalize the notion of dependency relations among data items values by adopting the idea of data dependency based on Sun et al. (2006) and Sun and Zhao (2013) in the following definitions below, and then we connect them to DQ dimensions and to process design outcomes in terms of deficiencies in process design representation**.**

Definition 9: *Data item dependency* is a situation in a process domain where value of data item $d_{in}(a_j)$ is affected by the value of data item $d_{out}(a_i)$ *via* of connective set of values. We define data item $d_{in}(a_j)$ as a dependent on data item $d_{out}(a_i)$.

In general, each activity $\boldsymbol{a_i}$ takes in a set of input data items $\boldsymbol{D_{in}}$ and produces a set of output data items $\boldsymbol{D_{out}}$ and $\{D_{in}, D_{out}\} \in D_P$.

Given two data items, $\boldsymbol{d_{out}(a_i)}$ and $\boldsymbol{d_{in}(a_j)}$, Where:

$\boldsymbol{d_{out}(a_i)}$ - represents the output data item of activity $\boldsymbol{a_i}$.

$\boldsymbol{d_{in}(a_j)}$ - represents the input data item of activity $\boldsymbol{a_j}$.

$\boldsymbol{d_{in}(a_j)}$ is dependent on $\boldsymbol{d_{out}(a_i)}$, denoted as $\boldsymbol{d_{out}(a_i)} \Rightarrow \boldsymbol{d_{in}(a_j)}$, if:

(1) $\boldsymbol{d_{in}} \in D_{in}$ and $\boldsymbol{d_{out}} \in D_{out}$ and $\{D_{in}, D_{out}\} \in D_p$, and (2) $\boldsymbol{a_i} \Rightarrow \boldsymbol{a_j}$ where $\{\boldsymbol{a_i}, \boldsymbol{a_j}\} \in \boldsymbol{A_p}$

.

Dependency event is with respect to a sequence of data items processing by activities where values of data items are determined based on dependency relationships between them. For example, before a customer sales order is approved, the customer's credit limit (data item $d_{out}(a_i)$) should be known in order to decide whether to accept the customer's order (data item $d_{in}(a_j)$).

Definition 10: *Dependencies Matrix*, denoted as *DM*, summarizes all dependency states that exist between various data items and their values in the selected process.

The matrix is a two-dimensional table that records each existing state dependency between various data items in a process domain. Each square on matrix, contain exist state dependency between two data values.

More formally, dependencies matrix can be defined as follows:

Assumption 7: Given the total number of data items ***n*** in a process ***p***, data items dependencies matrix *DM* is a *n* by *n* table where the cell element (r_i, c_j) shows where dependency exist on the data item d_j . If there is a dependency between data items it is identified as 1; if no dependency exists it is marked as a blank gray cell. Places with impossible dependency marked as blank white boxes.

This matrix can help BP or IS analysts, designers and practitioners in many aspects: first, to identify all existing dependencies relation between data items values. Second, to identify all trajectories or paths in process. Third, it serves as a means of control that we have not missed any dependence between data items. Fourth, to identify the most important data items that are highly dependent. Finally, it can be a tool for BP or IS analysts, designers and practitioners, which help to identify the cases where the process is expected to be in a deadlock situation or achieve low level of outcomes, i.e. data values not in the process goals.

3.4.2. The dependency relation types

Now we formalize the notion of dependency relation types between any data items dependencies in the process domain.

We identified and elicited a set of dependency relation types *(RT)* between any data items dependencies based on the initial definitions of Sun et al. (2006) and Sun and Zhao (2013) and based on our in-depth preliminary study and generalized them into high level categorization of dependency relation types. Basically, we identified six dependency relation types which are presented in **Table 16**. Each of dependency relation type received an appropriate code represent by $\boldsymbol{rt_i}$ variable code ($\boldsymbol{t_i}$ in short).

The following assumptions and definitions describe the formalization of these dependency relation types with respect to type and characters of data item dependencies.

Assumption 8: Each of data item dependency in process representation, which have been identified in matrix *DM*, can be categorized into and are represented by one or more of six dependencies relation types.

Table 16. Types of dependencies relation between data items

Dependencies relation type code	Relation type description	Example(s)
rt_1	Aggregation (part of)	Customer or supplier details e.g. first name, last name, address, phone, etc.
rt_2	Sequence of appearance (mandatory)	Before approving the customer's order, his credit limit is checked, and it should be in good standing.
rt_3	Shared information	Some of customer's data items are shared with customer's order and used as a supplementary detail.
rt_4	Formula calculation	The calculation of the amount premium for payment for elementary insurance is a multiplication of the total sum insured by a given percentage.
rt_5	External information event	Receiving information about schedule of future flights from airline companies or pending purchase order for bids from various suppliers.
rt_6	Conditional information	Customer order confirmation by supervisor manager although there is an exception on customer credit.

[Source: own study]

Definition 11: *Aggregation* is used to express the 'part-of' / 'has-a' relationship between data items.

In our model, the reference is to an *information object* (also known as an entity) containing set of data items e.g. customer, supplier, service, quotation, purchasing order, etc. Each of them contains information details i.e. set of data items that is an integral part of them.

Assumption 9: If an information object is deleted, all of its composite parts i.e. data item values, are "normally" deleted with it.

For example, each business customer, supplier or even employee has several data items (e.g. first name, last name, address, phone, etc.). If such customer, supplier or employee is deleted then all data in respect thereof will be deleted.

Definition 12: *Sequence of appearance* (also known as mandatory dependency) describes the appearing order of processing data items by activities in process path.

In our case, the reference is to *mandatory dependency relation* between data items in a defined order of appearance, where value of data item $d_{in}(a_j)$ is affected by value of data item $d_{out}(a_i)$ and must use it, so the value of data item $d_{out}(a_i)$ must be known. If one value of the data items in a sequence is unknown, then the process will not be properly executed, and it is expected to become stuck and to be in a deadlock situation.

For example, in caring on customer order process from any customer, if we want to approve the customer' order, we need to know the 'inventory level' and 'unit price' of each ordered item in advance, which are on *sequence of appearance* and as a mandatory value. Without these data items values, we cannot confirm the customer' order and the process at this point is expected to be in delay or in a deadlock situation.

Definition 13: *Shared information* represents a subset of data items that are further shared and used in other information objects in different places in the process, but originally, they are a part of a given information object (or entity).

For example, the data items 'Customer name' and 'Customer address' of customer, are used in 'Quotation to customer' or in 'Customer order' as shared data items. Originally, these two data items are *'part of'* Customer entity as an *information object*. If the 'Quotation to customer' or 'Customer order' will cancel or delete, they remain in our repositories and vice versa, i.e. if an information object is deleted, all its composite parts i.e. data item values, are "normally" deleted with it.

Definition 14: *Formula calculation* describes a given data item which is composed of several other data items values in process, and that performs any quantitative calculation.

For example, purchasing order process from any supplier is composed of (among others) and based on the multiplication of 'quantity' and 'unit price' of each ordered item for the calculation of 'total price' value of the purchasing order. If one of these data components is missing or wrong, then the 'total price' value of the purchasing order cannot be calculated or achieve wrong calculation and the process at this point is expected to achieve wrong outcomes or to be stuck i.e. in a deadlock situation.

Another example is calculation of body mass index (BMI) which is based on value of a patient's weight and height. Recording incorrect value of a patient's weight or height can lead to an incorrect BMI calculation result and may lead to obtain incorrect medical conclusions by the attending doctor.

Definition 15: An *external information event* describes a situation and result of actions of objects outside the process domain where a dependency relation exists on data item or information that must be received from third party, mostly reflect an external one.

Assumption 10: External information is outside of the process control.

Basically, this kind of this data item or information can activate the process domain as a trigger when the process is ready to start or should be arrived along the process and it may fail and not be received. For example, when booking a flight ticket in a travel agency, all information about future flights from all airlines should be available and up to date at the travel agency. If any airline delays the transmission of its future flight data, then the image of the future flight status of the travel agency is not up-to-date and can delay the process of ordering flight tickets from customers and the process is expected to become stuck.

Definition 16: *Conditional information* describes a situation of dependency among data items values along process representation, where the value of data item $d_{in}(a_j)$ is affected by value of data item $d_{out}(a_i)$ and uses it under some conditions i.e. conditionally depends on.

In fact, the process may be executed without using this value; hence, data item $d_{out}(a_i)$ in representation can be unknown under some conditions.

For example, in a mortgage approval process from a bank, approving the 'final mortgage loan amount' for customer depends on 'Manager approval' which depends upon the 'signature of the guarantors' by the borrower, conditioned by the 'customer’s risk status' which is another data item. If the status of customer’s risk level is low and there is no 'signature of the guarantors', i.e. in the status representation is unknown yet or null, the process can still be in progress since the customer risk level is low.

Assumption 11*:* Given the total number of data items *n* in a process *p*, the matrix of dependencies relation types, denoted as *MoRs,* is a *n* by *n* table where the cell element (r_i, c_j) represents the dependency relation type(s) where data items dependency exist on the data item $\boldsymbol{d_j}$.

If there is a dependency between data items, it is represented by the symbol $rt_{\boldsymbol{i}}$ (t_i in short) – which is one or more of six dependency relation types (from **Table 16**); if no dependency exists it is marked as a blank gray cell. Places with impossible dependency marked as blank white boxes.

The *MoR*s matrix can helps BP and IS designer or practitioners in many aspects: first, to identify the classification types of dependencies relation between data items values. Second, it helps to identify the data items with many dependencies relation types between values and can indicate the need to give them more attention when designing the process as important and highly dependent data items. Finally, this matrix can also help us to identify the type of potential failure dependencies between all data items and predict the cases where the process is expected to achieve outcomes not in the process goals set or to be in a deadlock situation, i.e. obtain problematic data items values with respect to various data quality dimensions, generated by Wang & Strong (1996), e.g. accuracy, completeness, consistency, timeliness, etc. In general, we name all these kinds of potential failures and problems as *data quality deficiencies* and the next section will discuss it in more detail.

3.4.3. The data quality deficiencies characters

Now we introduce and formalize the notion of *data quality deficiencies* in process design representation in terms of real-world (RW) reflection.

The next definition is based on foundations and assumptions of the approach presented by Wand & Wang (1996) and it aligns with Heinrich's et al. (2018) definition of data quality as presented earlier.

Definition 17: A *data quality deficiency* is an inconformity between the view of the real-world system that can be inferred from a process representation and the view that can be obtained by directly observing the real-world system.

We followed by Wand & Strong (1996) and Wand & Wang (1996) classifications of DQ categories and dimensions (see **Table 6** and **Table 7**) and made a distinction between DQ categories and dimensions that are relevant to IS at design stage and DQ categories and dimensions that are relevant to IS at the run-time and operation stage. We found it as relevant to our case from several reasons: first, Wand & Wang (1996) classification is general and refers to any IS. Second, we adopt this classification for dealing with an information system for BPM, whose role is representation of a domain state or real-world system as already discussed above in details in sub-chapter **3.1** (p. 97).

Since our focus in this research is on the analysis and design stage of BPs, the DQ deficiencies at BP design stage refers to a set of dimensions derived from intrinsic or contextual data quality categories. Furthermore, there is a matching with Wand & Wang IS concept, since an IS for BP representation is a special case of a general IS. Consequently, we found Wand & Wang classification of DQ dimensions as valid and useful to our case. Moreover, DQ dimensions are a central notion to assessing DQ and modelling DQ requirements (Zhang et al., 2019). Still, we adopt these DQ dimensions with some changes and additions.

Basically, our research focuses on four DQ dimensions: accuracy, completeness, consistency and timeliness that are considered to be the most important and critical dimensions for DQ assessment (Scannapieco et al., 2005; Batini & Scannapieco, 2006; Shariat Panahy et al., 2013; Jaya et al., 2017; Yang et al., 2017; Gharib et al., 2018; KPMG, 2018) and cope with the kind of DQ deficiencies derived from the real-world domain state or in process representation with regard to BP design stage.

For analyzing and recognizing the source of deficiency in each dimension, we distinguish between data item value deficiencies and structural data item deficiencies, since there are cases of deficiencies which concern and were originated by values and there are other cases of deficiencies which concern and were originated by structure and format representation:

Definition 18: *Data item value deficiency* concerns shortcomings in the representation of data item value in process or IS (e.g. on database).

Definition 19: *Structural data item deficiency* concerns shortcomings in the way of data items are representing the domain (e.g. in different format or mapping to the wrong data item in RW domain by data items represented in process or IS).

Definition 20: *Dependency deficiency* is a failure in a process representation which reflects indirect deficiency that is created due to the defined dependency among values of data items in process representation.

This deficiency is based on the process domain rules and involves or relates to one or more DQ dimensions at design stage (e.g. accuracy, completeness, etc.). We assume that this dependency deficiency in process representation can lead the process to execute improperly and fail without achieving its goals and can lead to some potential process failure outcomes:

(1) Achieve data items values not in the process goals set.
(2) Can lead the process to be in a deadlock situation.
(3) The expected external information event fails to occur.

The implications of unidentified potential failures in the general level of process could be for example:

- Receiving a wrong decision(s) in process.
- A waste of valuable time, unnecessary expense or unnecessary work as part of the process.

Due to the complexity and scope and for simplicity's sake, our research will focus its efforts on eliminating data quality problems derived by the above three options. In addition, we consider an expected external information event failing to occur as a completeness deficiency, i.e. missing data item value at representation that can lead the process to be delayed or in a deadlock situation.

3.4.4. Data quality dimensions and dependency deficiencies

Now we formalize the notion of data quality deficiencies in BP design representation in terms of DQ dimensions and their dependencies. Basically, the focus here is on four DQ dimensions; accuracy, completeness, consistency, and timeliness, which are the most important and critical dimensions for DQ assessment as mentioned above.

According to the above definitions with integration to DQ dimensions which are described in detail above (see sub-chapter **2.3.4;** p. 63), and based on Wang & Strong (1996) and Pipino et al. (2002) definitions, we formulate the next definitions.

Definition 21: *Accuracy* is the extent to which data is correct, reliable, and certified free of error.

Definition 22: *Accuracy deficiency* is a wrong or incorrect representation of the RW domain state in process or IS representation, i.e. the values of data items in process or IS representation are different than values of data items which they represent.

Given dependency between data items $\boldsymbol{d_1}$ and $\boldsymbol{d_2}$, a wrong value in the first data item representation (d_1) may lead to a wrong value in the dependent data item representation (d_2), and eventually can lead the process goals to become unreachable.

Accuracy deficiency, in principle, deals with data item value shortcomings, namely, wrong value may exist in some data item which represent value of data item in RW domain but it can be related to structural data item shortcomings as well, namely, mapping the wrong data item by data items represented in process or IS. Basically, this kind of shortcoming is out of our scope based on assumption 2, since we assume IS reflects what we have done in process.

Definition 23: *Completeness* is the extent to which data is not missing and is of sufficient breadth, depth, and scope for the task at hand.

Definition 24: *Completeness deficiency* is a missing or incomplete representation of RW domain state in process or IS representation, i.e. some values of data items which should represent existing values of data items are missing in process or IS representation.

Missing or incomplete value in one data item may lead the process to be in a deadlock situation and missing or incomplete values in others dependable data items in process representation, which can eventually lead the process goal to become unreachable.

Completeness deficiency, in principle, deals with value shortcomings but it can be related to structural shortcomings as well, namely, not all data items are represented by data items in the IS. Basically, these kind of shortcomings are out of our scope based on assumptions 2 and 3, since we assume the correctness of data items flow and should be valid in perspective of process modelling notation and since IS is a reflection of what we have done in process. We assume then, that there are no structural deficiencies exist and it's correct, since it's a fundamental assumption in our discussion here and we can't deal with values deficiencies in completeness without this assumption.

Definition 25**:** *Consistency* is the extent to which data is presented in the same format and compatible with previous data.

Definition 26: *Consistency deficiency* refers to irrelevant, ambiguous, redundant or meaningless values of data items in process or IS representation of data items value of RW domain, i.e. there are data items in the process or IS representation with no mapping back to relevant data item in the RW domain.

Consistency dimension deals with values' deficiencies as a result of changes in the RW without ensuring the necessary changes in which mean structural deficiencies as mentioned above. Inconsistent, ambiguous, redundant or irrelevant values in one data item representation may lead to achieve irrelevant or redundant values in process representation or wrong values or missing values in other dependent data items and can eventually lead the process goals to become unreachable or in a deadlock situation.

An example of inconsistent data may be depicted as if the process representation includes data items' values about old technology e.g. floppy disk assumed to be used as data storage medium or mapping to historical standards (e.g. ISO 9002, DOD 2167, etc.). It may have been used historically and still exists in the process representation even though it does not exist and is not used anymore in the RW domain. Other example, if the representation of customer details in a process includes personal data items (e.g. his marital status, education or hobbies, etc.), that are irrelevant and useless during the process even though these data items are correct, they are considered as redundant data items.

Sometimes, a data item can be produced as an output data by one activity but there is no need for any other activities or a need to use it as input data and it is not part of the final output data of process goal(s) such as the name of the department or the name of the sales person handling the customer's order service.

In principle, we are differentiating here between two options of potential deficiencies. The first option deals with potential deficiencies in value representation, namely, the same value of a data item in process representation is mapped to several values of data item e.g. a customer order can be approved as long as their customer credit level is not exceeded or the inventory level represented by the IS is above the order point or safety stock value even though these values are independent and discrete in the process domain state.

The second option is vice versa and related to structural deficiencies, i.e. several values of data items in process or IS representation are mapped to one value of data item in RW. For example, customer address at the RW domain can be represented in more than one format for different purposes such as for marketing needs to locate the customer and for finance needs to send him an invoice.

Another perspective of this option related to decomposition of value of data item in RW into several values of data items in process or IS representation but without mapping or matching the exact value to the appropriate data items e.g. customer address which decomposes into more than one value of data items in the representation level such as street name, house number, zone, city, zip code, etc. but without matching between them. From a structural point of view, if the same data item is represented more than once (duplicates) and the values of these duplicates are consistent then it's not a problem. However, they may be inconsistent.

Based on the above two options and their examples we can conclude that this dimension deals with both structural and values deficiencies. Basically, these two options depend on the decision of granularity level in representation. However, structural deficiencies can lead to inconsistency in values and values deficiencies if the granularity level is unsuitable to the business requirements and to its needs, like in the granularity of inventory level representation.

The decision about granularity level in process or IS representation as compared to granularity level of the RW depends on the granularity level of the existing rules in reference to the representation as a result of business requirements. The change later in the decision of granularity level is a result of business requirements problems or business need.

A good example for this idea is the year 2000 bug problem (also known as the *Y2K bug*). Y2K bug was a ticking time bomb for all major computer applications. During 1960s to late 80s there was a widespread practice in all computer software to use two digits for representing a year on date rather than using 4 digits (e.g. 99 instead of 1999). This was done to save computer disk storage and memory space because these resources were relatively expensive in those times. As the year 1990's approached their end, experts began to realize this major shortcoming in the computer application software. In the year 2000, the computer systems could interpret 00 as 1900 and mess up all the computing businesses work. In this case, businesses required to change it by using 4 digits, namely change the requirement, to prevent themselves from chaos.

Definition 27: *Timeliness* is the extent to which data is sufficiently up-to-date and the age of the data must be appropriate for the task at hand.

Definition 28: *Timeliness deficiency* is the time delay between the change in RW domain state and the required modification in process or IS representation.

A time delay in appearance of data item value may lead to delay between the change in RW domain state and the required modification in process or IS representation. This time delay of values can lead also to delay in others dependable data items in the process representation, which means they can be not up-to-date or with missing values and eventually can lead the process to a deadlock situation. For example, daily output of customer bank account balance based on non-up-to-date information from the day before, due to working in batch computing mode environment, which means it does not exist in the RW.

Timeliness deficiencies results can appear in two possible formats. The data event whose representation is delayed may initiate or update the value of a data item. If value initiation is delayed, then the representation has a missing value i.e. completeness deficiency. If a value update is delayed, then the representation has wrong value i.e. an accuracy deficiency. These deficiencies will exist until the time delay is over.

Based on the above definitions, in the next chapter we identify some possible deficiencies in process representation that can occur as a result of a poor level of process design and along data values' production. We argue that a specific deficiency can cause or lead to a deficiency chain that influences each other, as a result of existing dependency aspects among values of data items which participate in the process through their representation. These kind of deficiencies and potential process failures in its outcomes can lead to failures in IS development stage and may lead the developed IS to be imperfect, without faithfully reflecting the domain, since the data flow requirements in IS are dependent upon and derived from data flow and dependencies in processes (Recker et al., 2006; Soffer & Wand, 2007). Failures in IS in terms of DQ dimensions are:

(1) Provide a wrong data value reflection in process or IS representation.
(2) Provide a partial representation or missing representation of the data items values which may be required by the rules of the process.
(3) Fail to reflect or delay in reflection of an external or internal information event with the expected data item value.

3.4.5. Collecting DQ deficiencies and requirements

Based on the literature review, the EFGs discussions results, participants' feedbacks, and following the formalization step and its definitions and assumptions, we identified and summarized a set of information-related problems in BPs design. Moreover, we generalized them by common features and characteristics and created sets of potential DQ deficiencies and flaws, grouped by these DQ dimensions and with respect to the identified dependencies relation types between data items (see **Table 16;** p. 124).

Basically, this analysis and summary focused on four DQ dimensions i.e. accuracy, completeness, consistency and timeliness, which are considered most critical and significant dimensions for DQ analysis and were selected carefully based on results of focus groups discussions and based on some references from literature as already mentioned above.

These four dimensions have been marked as ACR, CMP, CNS and TML respectively as an abbreviated identifier and for reader convenience. Each of identified DQ deficiency marked as

DQ deficiency code start with the letters DEF as a prefix with a sequential number separately per each dimension.

In the next following four tables (**Table 17** to **Table 20**) we summarized sets of types of DQ potential deficiencies and flaws and categorized them into each of DQ dimension. Later, all these DQ deficiencies and flaws issues will be transformed into formal DQ requirements. Moreover, these DQ deficiencies issues and set of DQ requirements will be used as a basis material for the new proposed method and later for validation phase based on case study in the next part. Additionally, this kind of auxiliary tables can significantly relieve BP and IS analysts and designers in detecting DQ failures and shorten the execution time of the analysis and design specification work, as a number of FG participants have raised a concern that kind of this method can cause them to overload in the specification work and they estimate that it will also affect the overall budget framework of IS project and they are afraid that it is expected to grow.

It is important to emphasize here that we have made the best efforts we could to produce these DQ deficiencies and DQ requirements lists and get to the depth of things regarding the process quality information issues. To the best of our knowledge, these DQ deficiencies tables cover the range of identified problems.

Table 17. DQ potential deficiencies types for accuracy dimension

DQ dimension	DQ deficiency code	The deficiency description	The implications and potential deficiencies at the process level	Examples
Accuracy (ACR)	DEF1-ACR	Wrong values or error in any data item values in process representation.	Wrong value in one data item can lead to achieve wrong values in other dependable data items in process and can eventually lead the process goal(s) to become unreachable or lead the process to be in a deadlock situation.	• Wrong value in data items such as 'quantity' or 'unit price' of purchasing order can lead to wrong value in 'total price' value as a dependable data item.
	DEF2-ACR	Wrong values or error in data items value which reflect the RW in process representation.	Wrong value in list of data items can lead to achieve missing values in others dependable data items in process and eventually can lead the process goal to become unreachable or to be in a deadlock situation.	• Wrong values in list of bill of materials (BOM) of final product. • Wrong value in list of data items on packing list to customer. • Wrong value about current date/time value can lead to incorrect values in employee work hours report.

	DEF3-ACR	Wrong values or error in data items which are part of any information object, list, form, etc. in process representation.	Wrong value in one data item can lead to achieve wrong values in others dependable data items in process and eventually can lead the process goal to become unreachable or to be in deadlock situation.	• Wrong value in data item e.g. phone number or address in employee or customer application.
	DEF4-ACR	Wrong values or error in sequence of appearance of data items in process representation.	Wrong value in sequence and order of data items appearance can lead to achieve wrong or missing values in others dependable data items in process dependency paths and eventually can lead the process goal to become unreachable or to be in a deadlock situation.	• Choosing wrong value about requested flying destination can lead to the shipment of luggage to wrong destination.
	DEF5-ACR	Wrong values or error in shared data items of any information object in process representation.	• Achieving data items not in the process goals set. • The process could be in deadlock situation.	• Wrong value about customer details e.g. customer address can lead to wrong customer address in customer delivering details.
	DEF6-ACR	Wrong values or error recorded in data items which are part of any formula calculation in process representation.	Wrong value in received and recorded data items which are part of the formula can lead to achieve wrong values in others dependent data items.	• An incorrect value of a patient's weight or height can lead to an incorrect Body Mass Index (BMI) calculation and may lead to incorrect medical conclusions. • BMI Formula: BMI = weight (kg) ÷ height^2 (m^2)
	DEF7-ACR	Wrong value or error in external data or information item received by third party.	Wrong value or error in received external data items can lead to achieve wrong values in others dependable data items.	• Wrong value about flying dates from airlines companies can lead to delayed arrival of the customer at the destination country.
	DEF8-ACR	Wrong or error in conditional data items values which are part of decision situation	Can lead to achieve wrong values in others dependable data items values.	• Wrong value about customer line of credit level in obtaining a loan can lead to obtain manager approval.

[Source: own study]

Table 18. DQ potential deficiencies types for completeness dimension

DQ dimension	DQ deficiency code	The deficiency description	The implications and potential deficiencies at the process level	Examples
Completeness (CMP)	DEF1-CMP	Missing any data item values in process representation.	Missing value in one data item can lead to achieve missing values in others dependable data items in process and eventually can lead the process goal(s) to become unreachable or lead the process to be in deadlock situation.	• Missing value in data items such as 'quantity' or 'unit price' of purchasing order can lead to missing value in 'total price' value as a dependable data item.
	DEF2-CMP	Missing data item value which reflect the RW in process representation.	Can lead to achieve missing or wrong values in others dependable data items and achieving data item values not in the process goals set.	• The real status of a foreign worker in a given country is missing. • Missing of data in daily/ monthly employee attendance report.
	DEF3-CMP	Missing value in data items which are part of any information object, list, application or form in process representation.	Can lead to achieve missing values in others dependable data items and can lead the process to be in a deadlock situation and eventually can also lead to process goal to become unreachable.	• Missing value about foreign worker details e.g. visa type, copy of passport, etc. or missing value about ordered service can lead a problem in obtaining a work permit for the caregiver.
	DEF4-CMP	Missing value in sequence of appearance of data items in process representation.	Can lead to achieve missing or wrong values in others dependable data items and can lead the process to be in deadlock situation and eventually can also lead the process goal to become unreachable.	• Missing value about requested destination lead to missing value about the port at destination.
	DEF5-CMP	Missing value in shared data items of any information object.	Can lead to dilemma and delay in process progress or lead the process to be in a deadlock situation.	• Missing value about prices of ordered items can lead to delay in delivery of quotation to the customer.
	DEF6-CMP	Missing data items or value in received and recorded data items which are part of any formula calculation.	Can lead to achieve missing or wrong values in others dependable data items and can lead the process to be in a deadlock situation.	• Missing items in cargo items report. • Missing value about requested destination.
	DEF7-CMP	Missing data items or values	Can lead to dilemma and delay in process progress	• Missing of products prices list from supplier.

		in received external data or information.	or can lead the process to be in a deadlock situation.	• The expected advanced payment from customer not received. • No customer approved for quotation.
	DEF8-CMP	Missing data items or values which are of conditional situation and decision.	Can lead to achieve missing or wrong values in others dependable data items and achieving data item values not in the process goals set.	• Missing an update of customer loan payment details.

[Source: own study]

Table 19. DQ potential deficiencies types for consistency dimension

DQ dimension	DQ deficiency code	The deficiency description	The implications and potential deficiencies at the process level	Examples
Consistency (CNS)	DEF1-CNS	Irrelevant, useless or redundant of any data items values in process representation.	Can lead to achieve mistakes in other data items or to cause waste of time and resources or delay in process progress or can lead to achieve data item values not in the process goals set.	• Collecting redundant or irrelevant data items values from customer which are useless in process (e.g. Marital status, customer occupation, number of children, etc.).
	DEF2-CNS	Inconsistency, unambiguous or meaningless of data item value of RW in process representation and vice versa.	Can lead to achieve mistakes or wrong values in other dependable data items or to cause waste of time and resources or delay in process progress or can lead to achieve data item values not in the process goals set.	• Two different customers relate to the same customer application or order. • Different values about contacts in customer application. • Representation of historical objects e.g. ISO9002, Floppy disk, etc. which is invalid in RW representation.
	DEF3-CNS	Inconsistency in data items values which are part of any information object in the process representation.	Can lead to achieve mistakes or wrong values in other dependable data items or to delay in process progress or can lead to achieve data item values not in the process goals set.	• Different values about the telephone number (e.g. office, home or mobile) set to the same telephone number in customer application. • Customer real status (Tourist, new immigrant, foreign resident, etc.) vs. his formal record in government authorities for customs regulation needs.
	DEF4-CNS	Inconsistency, unambiguous or meaningless in data items values in sequence of appearance of data items in	Can lead to achieve mistakes or wrong values in others dependable data items or to delay in process progress or can lead to achieving data item	• Inconsistency, unambiguous or meaningless in value about requested destination can lead to missing value about the port at destination.

		process representation.	values not in the process goals set.	
	DEF5-CNS	Inconsistency, unambiguous or meaningless in shared data items of any information object.	Can lead to achieve mistakes, wrong decision or wrong values in other dependable data items or delay in process progress or can lead the process to be in a deadlock situation.	• Different values about the customer address (e.g. office, home or mobile) set to the same address in customer application. • Inconsistency in storage of customer's items location in warehouse.
	DEF6-CNS	Inconsistency, unambiguous or meaningless on data items that are part of the formula are consistency and there is no unambiguous or meaningless data items.	Can lead to achieve mistakes, wrong decision or wrong values in others dependable data items or delay in process progress or can lead the process to be in a deadlock situation.	• Two different customer sites (e.g. main office or warehouse) but only one of them is relevant to the supply of the ordered products.
	DEF7-CNS	Inconsistency, unambiguous or meaningless on external data items are consistency and there is no unambiguous or meaningless data items.	Can lead to achieve mistakes, wrong decision or wrong values in others dependable data items or delay in process progress or can lead the process to be in a deadlock situation.	• Inconsistency between ordered items list of customers to the packing list report of items delivered by supplier.
	DEF8-CNS	Irrelevant, useless or redundant data items values exist in any data item, which are of conditional situation and decision.	Can lead to achieve mistakes in other data items or to cause waste of time and resources or delay in process progress or can lead to achieve data item values not in the process goals set.	• Collecting redundant or irrelevant data items values from customer which are useless in process (e.g. Marital status, customer occupation, number of children, etc.), can lead to mistakes or wrong decisions in process.

[Source: own study]

Table 20. DQ potential deficiencies types for timeliness dimension

DQ dimension	DQ deficiency code	The deficiency description	The implications and potential deficiencies at the process level	Examples
Timeliness (TML)	DEF1-TML	A time delay in appearance of any data item value in process representation.	Can lead to dilemma and to achieve missing values in others dependable data items and delay in process progress or can make the process to be in a deadlock situation.	• A time delay in receiving blood test results can lead to delay in medical care that the patient should receive.

	DEF2-TML	Change in data item value of RW representation is not reflected in real time by equivalent data item value in process representation.	Can lead to achieve wrong values in others dependable data items and achieving data item values not in the process goals set.	• Changes in flying schedule plan from airlines companies not reported in real time to travel agency i.e. the change is not reflected in process.
	DEF3-TML	The expected data item from third party does not appear on time or delay in receiving it as planned according to the sequence of appearance in process.	Can lead to achieve missing or wrong values in other dependable data items or delay in process progress or can make the process to be in a deadlock situation.	• The expected approval for quotation from customer might not arrive within a given time. • A time delay in acceptance of schedules plan from airlines companies can lead to delay of ticket issuing for customer to his destination country.

[Source: own study]

Accordingly, we created by generalization a set of information quality requirements and guidelines (**Table 21**) with respect to those dimensions of IQ. These requirements were elicited and selected based on analysis and results of focus groups discussions and based on the literature review, with respect to the above set of tables of DQ deficiencies and flaws, and the identified dependencies relation types between data items (**Table 16;** p. 124).

Table 21 depicts a generic catalogue with variety of DQ requirements and guidelines grouped by DQ dimensions. This table can assist BP and IS designers and practitioners in identifying the possible and appropriate set of requirements and guidelines that should be taken into account at the design stage and help them to cope with DQ deficiencies and flaws grouped by dependency relations types between data items and DQ dimensions. To the best of our knowledge, this requirements and guidelines list covers the range of identified problems.

Table 21. Generic catalogue of DQ requirements and guidelines

Data quality dimension	Requirement code	The requirement & guidelines description
Accuracy (ACR)	RQ1-ACR	Ensure a proper representation, reliable and free of error of any data items values in a given process.
	RQ2-ACR	Identify whether data item value from RW map to a wrong data item value in process representation and fix it.
	RQ3-ACR	Ensure a proper representation, reliable and free of error of data items values which are part of any information object in the process representation.
	RQ4-ACR	Ensure a proper representation, reliable and free of error in sequence of appearance, mandatory of data items or in others dependent data items based on process paths in process representation.
	RQ5-ACR	Ensure a proper representation, reliable and free of error in shared data items values of any information object.
	RQ6-ACR	Ensure reliable and free of error in recorded data items values which are part of any formula calculation in process.
	RQ7-ACR	Prevent accepting wrong or unreliable data item value from external and internal third party.
	RQ8-ACR	Ensure accuracy in conditional data items values which are part of decision situation.
Completeness (CMP)	RQ1-CMP	Ensure the completeness of any data item values in process representation.
	RQ2- CMP	Ensure the completeness of RW data item values in process representation and complete the missing data item value where it's necessary in the process.
	RQ3- CMP	Ensure the completeness of data item values which are part of any information object in the process representation.
	RQ4- CMP	Ensure the completeness of data item values in sequence of appearance, mandatory of data items or in others dependent data items based on process paths in process representation.
	RQ5-CMP	Ensure the completeness in shared data items values of any information object.

	RQ6-CMP	Ensure the completeness in recorded data items values which are part of any formula calculation in process.
	RQ7-CMP	Ensure the completeness of data item value received from external or internal third party.
	RQ8-CMP	Ensure the completeness in conditional data items values which are part of decision situation.
Consistency (CNS)	RQ1-CNS	Ambiguous, meaningless, redundant or inconsistency in data items values in process representation should be eliminated.
	RQ2-CNS	Identify ambiguity, lack of sense or inconsistency between RW representation to data item values in process representation i.e. different data item values in RW mapping to the same data item in process or data item values in process without mapping to RW and vice versa.
	RQ3-CNS	Identify ambiguity, meaningless or inconsistent data item values which are part of any information object in the process representation.
	RQ4-CNS	Identify ambiguous, meaningless or inconsistent data items in sequence of appearance, mandatory of data items or in others dependent data items based on process paths in process representation.
	RQ5-CNS	Identify ambiguity, meaningless or inconsistent in shared data items values of any information object.
	RQ6-CNS	Identify ambiguity, meaningless or inconsistent in recorded data items values which are part of any formula calculation in process.
	RQ7-CNS	Identify ambiguity, meaningless or inconsistent data item value received from external or internal third party.
	RQ8-CNS	Identify ambiguity, meaningless or inconsistent in conditional data items values which are part of decision situation.
Timeliness (TML)	RQ1-TML	Ensure that the expected data item value will appear on time as it planned or define its frequency according to the process representation.
	RQ2-TML	Change in data item value of RW representation should be reflected in real time by equivalent data item value in process representation.
	RQ3-TML	Ensure that the expected data item values which are part of information object will appear on time as it planned in the process representation.

	RQ4-TML	Ensure that the expected data item value will appear in real time in sequence of appearance, mandatory of data items or in others dependent data items based on process paths in process representation.
	RQ5-TML	Ensure that the expected data item values which are shared data items values of any information object will appear on time as it planned in the process representation.
	RQ6-TML	Ensure that the expected data item values which are which are part of any formula calculation in process will appear on time as it planned in the process representation.
	RQ7-TML	Prevent time delay in appearance of expected data item value in process from external and internal third party. Set a time monitoring to define whether data item value is to be accepted or changed within a given time in process representation.
	RQ8-TML	Prevent time delay in appearance of expected data item value in conditional data items values which are a part of decisive situation in process representation.

[Source: own study]

3.5. Summary of Main Challenges and Requirements for the New Method

3.5.1. Challenges to cope and address

In sub-chapter **1.3**, the "Problem Description" (p. 10), we reviewed the background and circumstances for the failure of IS/IT projects and the formation of failures related to the quality of the information at the process design stage. The main point was the exclusion of quality requirements at the analysis and design stage of the BPs, which can lead to non-compliance with the requirements and needs of the end users and errors that occur during ISs development. In fact, these are the result of inaccurate or incomplete requirements and errors in BPs specifications. Moreover, the data flow in BPs and along their activities is the basis for representation of data requirements in IS stage. Hence, if we want to ensure IS to work properly and present desired data values in high quality, we have to check values of these data items before recording them to IS to ensure DQ and the presentation of desired values.

Additionally, the existing literature contains many studies and academic efforts to deal with this by different approaches, methods and models as presented above (in sub-chapter **2.3.5**; p. 69). Although these approaches and methods relate to the DQ aspects and to the quality requirements derived from them, the issues of dependency relation types on data items values, the DQ deficiencies and their impact on the quality of process design in particular are neglected. In addition, the aspects of data items dependencies and their requirements of IQ based on the DQ dimensions presented above are critical to understanding the structure of the BPs and for defining additional IQ requirements. In addition, we used different relation types of dependencies between data items to show the possible effects and consequences thereof on the quality of the process in general.

BP models can be used to make the process explicit as-is. However, a gap is created between the requirements process to-be and the as-is process. To overcome this gap and to acquire an optimal fit between the designed process and its planned outputs, and finally to achieve soundness of processes used for designing of future IS, there is a great need for a new structured method to lead and direct the BP and IS analysts, designers and practitioners.

The new proposed method should take into account all the factors mentioned above to help BP and IS analysts, designers and practitioners to predict potential failures in the early stages of BPs, especially at analysis and design phases, and ultimately contributes to improving the quality of design and development phases of the new IS. Moreover, in IS, by definition, the artifact was created to address an important organizational problem. Hence, a design artifact is complete and effective when it satisfies the requirements and constraints of the problem it was meant to solve (Hevner et al., 2004).

3.5.2. Summary of main requirements for the new method

Artifacts should be evaluated based on the requirements of the context of their respective application and implementation environment (Peffers et al., 2012). Moreover, according to the FEDS approach it is necessary to choose the properties for artifact evaluation i.e. it entails determining what to evaluate with regards to the artifact, and it include its goal(s), general features, and of course, practical requirements of the artifact that are to be subject to evaluation (Venable et al., 2016).

Based on the above literature review and following the results of EFGs sessions, the participants' feedbacks and the implications of DQ potential deficiencies types discussed in above sub-chapters **3.3** and **3.4**, we identified and summarized in **Table 22**, a set of main requirements for a new method as an artifact. This list reflects the BP and IS analysts', designers' and practitioners' expectations for the new method and it is oriented to the quality of information in the BP design phase.

Table 22. A summary of requirements for the new method

#	The requirements
1	The ability of mapping and managing the entire information flows in a process and to see the "big picture" in advance.
2	The ability to reflect all dependencies that exist between information items in a process.
3	Be able to identify each type of each dependency relation and their impact in a process.
4	Be able to identify potential DQ failures related to dependency and its type between data items in a process.
5	Be able to identify the potential DQ problems in a process grouped by DQ dimensions in advance.
6	Be able to identify in advance the impact of potential DQ problems in a process on its performance and quality in advance.
7	The new method should enable BP and IS analysts, designers and practitioners to identify relevant DQ requirements for each potential DQ failure.
8	A new method should help BP and IS analysts, designers and practitioners to improve the quality of analysis and design outputs of BPs and IS.

[Source: own study]

3.6. The Proposed Data Quality Deficiencies Prediction (DQDP) Method

In this chapter, we present our new and original method, namely the **Data Quality Deficiencies Prediction (DQDP) method** or **DQDP method** in short, which is one of the main purposes of our research and in fact its main goal. Our third goal was to design and develop a new method for DQ assessment and to predict DQ deficiencies and potential failures in BPs design and prevent them in advance. In addition, the new method provides the answer for our second research question i.e. achieving the ability for DQ assessment and predicting DQ deficiencies and potential failures in BPs design. In fact, the new method is also designed to eliminate DQ deficiencies and flaws while examining the effect of dependencies among different data item values based on DQ dimensions on DQ and process failures.

The planning and construction of the new method is based on our own study, the conceptual model constructs, the analyzing and summarizing of the EFGs outputs and participants' feedbacks, the formalism outputs (definitions and assumptions) and summary of main requirements and expectations as discussed in details in the previous section. Moreover, the above set of DQ potential deficiencies and the DQ requirements collections, is also taken into account in building the proposed method.

3.6.1. The proposed DQDP method

The Data Quality Deficiencies Prediction (DQDP) method is depicted in **Table 23** on a high level. It consists of three layers: the process layer, the data layer and DQ requirements layer.

Table 23. The data quality deficiencies prediction (DQDP) method - Layers and steps

The layers	The step	The step description
Process layer	**Step 1**	Choose a valid process model.
Data layer	**Step 2**	Transform process goal(s) into a set of data items.
	Step 3	Identify set of inputs and outputs data items per each activity.
	Step 4	Build matrices of all dependencies between data items and relation types.
Data Quality Requirements layer	**Step 5**	Map and assign per each dependency relation type the appropriate of DQ requirement(s) from catalogue with respect to DQ dimensions.

[Source: own study]

Table 24 presents the new proposed DQDP method in more detail, including the scope of each layer i.e. the activities of each step and sub-steps (of steps 4 and 5), their guidelines and expected outcome(s) for implementation by BP and IS analysts, designers and practitioners.

Table 24. The DQDP method - steps, sub-steps components & outputs

The layers	The step	The step description	The step expected outcome(s)
Process layer	**Step 1**	Choose a valid process model.	• A valid process model is given • Process goals (G_p) are given. • A_p - Set of process activities. • D_p - Set of process data items.
Data layer	**Step 2**	Transform process goal(s) into a set of data items.	• D_g - Set of data items that define the process goals.
	Step 3	Identify set of inputs and outputs data items per each activity.	• List of inputs and outputs data items per each activity. • Set of all possible paths i.e. trajectories along the process.
	Step 4	Build matrices of all dependencies between data items and relation types.	
	Step 4.1	Identify all data items dependencies and build a *DM* matrix.	• A DM matrix of all existing data items dependencies.
	Step 4.2	Identify independent data items and dependencies which are not possible and mark them.	• Marked independent and impossible dependencies areas.
	Step 4.3	Map each of data items dependencies into relevant dependency relation types by transforming DM matrix to MoRs matrix.	• A new MoRs matrix of all data items dependencies with relevant dependency relation type's codes.
Data Quality Requirements layer	**Step 5**	Map and assign per each dependency relation type the appropriate of DQ requirement(s) from catalogue with respect to DQ dimensions.	
	Step 5.1	Assign each of the identified DQ deficiencies types into the relevant dependency relation type grouped by DQ dimensions.	• An assistant table with all identified DQ deficiencies types grouped by relevant dependency relation types and DQ dimensions.
	Step 5.2	Map and assign for each data items dependency the appropriate requirement(s) from catalogue grouped by dependency relation types and DQ dimensions.	• A new assistant table with set of appropriate requirements & guidelines should taking in account grouped by dependency relations between data items and DQ dimensions.

[Source: own study]

For the reader's convenience, the next sub-chapter (**3.6.2**) presents a description of each step and sub-step of the proposed DQDP method including an illustration of the proposed DQDP method with examples. Moreover, **Figure 38** presents the above detailed DQDP method steps illustrated as a Use Case diagram (OMG, 2005) for the reader's convenience and for BP and IS analysts, designers and practitioners.

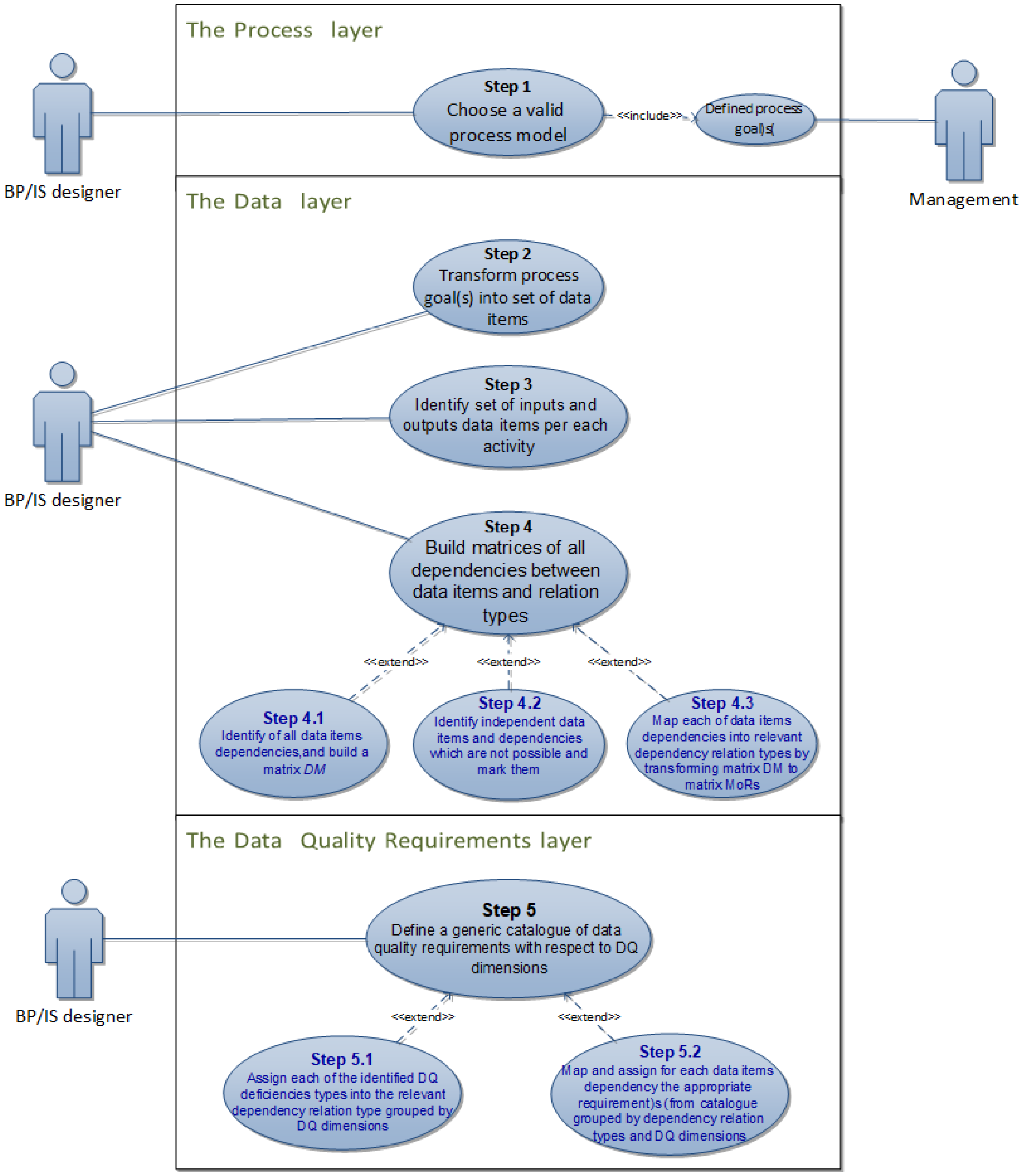


Figure 38. The DQDP method illustrated by Use Case diagram

[Source: own study]

3.6.2. Description and illustration of steps of the proposed DQDP method

This sub-chapter presents a description and a supplementary explanation of each step and sub-step of the proposed DQDP method and describes its activities and expected outcome(s) to be achieved with an illustration of the method based on a small example of business process named ***'Ordering a book on website'***.

The example process includes 4 activities with 7 data items. The 4 activities are: (1) Select a book on website (a_1); (2) Order this book (a_2); (3) Record customer details (a_3); and, (4) Deliver the book to customer address (a_4). The 7 data items are: (1) Book title (d_1); (2) Book price (d_2); (3) Customer name (d_3); (4) Customer address (d_4); (5) Order number (d_5); (6) Payment status (d_6) and, (7) Delivery status (d_7).

- ***Step 1: Choose a valid process model:***

In this step there is a need to choose a valid process model presented by any process modeling notation (e.g. BPMN diagram, Activity diagram, Petri nets, etc.) for the given process. As we already assumed in the above assumptions, the business process model and its data items flow should be correct and valid in every notation. In fact, we consider here the notion of soundness which was originally defined by Aalst (2013) as a preliminary requirement, where a process model is considered sound if it cannot get stuck before reaching the end (termination is always possible) and all parts of the process can be activated. Basically, soundness was originally defined for workflow nets (WF-nets) but it applies to all modeling techniques (Aalst, 2013).

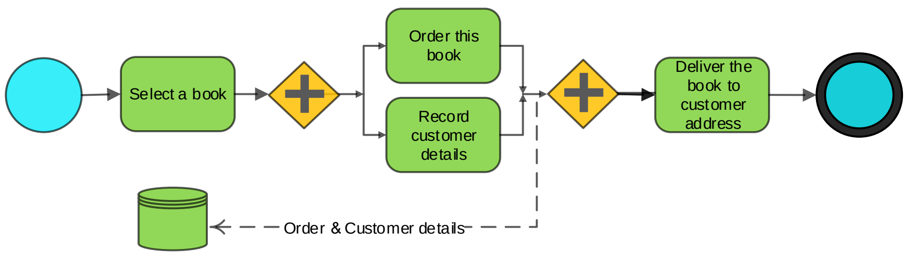


Figure 39. A process model for ordering a book on web site - an example

[Source: own study]

The example of a process model which is presented in **Figure 39** was created using Business Process Modelling Notation (BPMN) , since it is an up to date and very wide-spread modeling notation for BPs and widely recognized as de facto standard for BP modeling (Abramowicz et al., 2007; Heinrich & Paech, 2010; Harmon & Wolf, 2011; OMG, 2011). Moreover, BPMN is readily understandable and acceptable by all business users and aims to bridge the gap between the BP design and the process implementation (Abramowicz et al., 2007; Heinrich & Paech, 2010).

- ***Step 2: Transform process goal(s) into a set of data items:***

The outcomes of the process are derived from its goals. At this step there is a need to transform the process goal(s) (G_p) into a set of data items (D_g), which is a sub-group of all process data items. In fact, in our analysis, these are the values of the data items that represent the outcomes of the process. The idea is to be focused on achieving the process outcomes derived by process goals as defined by process' manager or by organization's management.

The process goals (G_p) of the above illustration example are 'Sell a book to the customer' and 'Deliver the ordered book to the customer' once its order is paid. Hence, the set of data items that represent the process goals (D_g), and actually represent its outcomes are order number (d_5), payment status (d_6) as 'Paid' and delivery status (d_7) that should be 'Sent' (to customer). Then, the outcomes of the process are considered as achieved.

- ***Step 3: Identify set of inputs and outputs data items per each activity:***

In this step, based on the chosen process model notation, all inputs and outputs data per each activity and along the process should be identified and presented in an appropriate table.

Table 25 represents all data items in manner of input and output data per each activity along the illustration process.

Table 25. List of input & output data items per each activity in example process

Activity ID	Activity description	Input data item ID $(D_{in}(a_i))$	Output data item ID $(D_{out}(a_i))$	Remarks
a_1	Select a book on web site		d_1, d_2	
a_2	Order this book	d_1, d_2	d_5	
a_3	Record customer details	d_5	d_3, d_4	
a_4	Deliver the book to customer address	d_3, d_4, d_5, d_6	d_7	d_6=Paid and d_7=Sent

[Source: own study]

- ***Step 4: Build matrices of all dependencies between data items and relation types:***

  Basically, this step has three sub-steps (as shown in **Table 24**).

- ***Sub-step 4.1: Identify of all data items dependencies and build a matrix DM***

In ***sub-step 4.1***, there is a need to identify all data items dependencies which exist in a given process and build a dependencies matrix, namely *DM matrix*. The size of the DM matrix depends on the number of data items along and within the process i.e. 7x7 (based on assumption 7; see p. 123). **Figure 40** shows the *DM matrix* with all exists data items dependencies in our example process.

Data item description	ID#	d_1	d_2	d_3	d_4	d_5	d_6	d_7
Book title	d_1							
Book price	d_2	1						
Customer name	d_3							
Customer address	d_4			1				
Order number	d_5	1	1	1	1			
Payment status	d_6					1		
Delivery status	d_7			1	1	1	1	

An area of impossible dependencies

Legend:	
1	Dependency exist
	No dependency exist
	Impossible dependency

Figure 40. Data items dependency matrix (DM) for the example process

[Source: own study]

For example, in this example process, a dependency of data items exists between 'Book price' (d_2) and 'Book title' (d_1), i.e. the value in 'Book price' (d_2) depend on the value in 'Book title' (d_1). Another example, a dependency of data items exists between 'Delivery status' (d_7) and 'Payment status' (d_6). Recording a wrong value about 'Payment status' (d_6) can lead to process deficiencies and choose wrong value in 'Delivery status' (d_7) and finally cause delay in delivering the ordered book to customer.

- ***Sub-step 4.2: Identify independent data items and dependencies which are not possible and mark them***

In ***sub-step 4.2***, there is a need to identify and marked independent data items and impossible dependencies between data items as is done and shown in **Figure 40**. In addition, all set of possible paths i.e. trajectories along the process should be identified, based on a sequence of data items flows and activities for progressing from start to the end of process. The identified trajectory in our sample process is: d_1 - d_5 - d_3 - d_6 - d_7 . It represents the existing sequence of transitions between the various data items along the activities in process, starting from the first data item i.e. 'Book title' (d_1) to the last data item in the route i.e.

'Delivery status' (d_7). Furthermore, these trajectories not only show us the sequences of dependencies that exist between the various data items in the process, but also illustrate to the BP & IS analysts, designers and practitioners what are the data items that participate in many dependent paths and their effect on the progress of the process, and thus they also become more sensitive and important data items.

- ***Sub-step 4.3: Map each of data items dependencies into relevant dependency relation types by transforming DM matrix to MoRs matrix***

Then in ***sub-step 4.3***, there is a need to identify and assign the appropriate of dependency relation type between any data item dependencies. Each of data items dependency in matrix *DM* (**Figure 40**) is to transform to appropriate dependency relation type code based on **Table 16** (p. 124) and will be represented in an updated new matrix (**Figure 41**), namely *Matrix of Relations or MoRs* in short (based on assumption 11; see p. 124). Recall, all the dependency relation types described in details in sub-chapter **3.4.2** (p. 123) and depicted in **Table 16** (p. 124). Each of dependency relation type received an appropriate code represent by rt_i variable code (or t_i in short).

Figure 41 shows the new *MoRs matrix,* with transforming all exists dependencies of data items values in our example process into appropriate dependency relation type codes, denoted as rt_i (t_i in short) for each exist dependency.

Data item description	ID#	d1	d2	d3	d4	d5	d6	d7
Book title	d1							
Book price	d2	t1						
Customer name	d3							
Customer address	d4			t1				
Order number	d5	t3	t3/t4	t2/t3	t3			
Payment status	d6					t2		
Delivery status	d7			t3	t3	t3	t2	

An area of impossible dependencies

Legend:	
ti	Dependency relation type
(grey)	No dependency exist
(white)	Impossible dependency

Figure 41. MoRs matrix of data items dependency relation type for example process

[Source: own study]

For example, a dependency of data items exists between 'Book price' (d_2) and 'Book title' (d_1), i.e. the value in 'Book price' (d_2) depends on the vale in 'Book title' (d_1) as described in dependency matrix (*DM matrix*) (**Figure 40**) at former step. Since 'Book price' (d_2) is part of 'Book title' (d_1) as an information object, it assigned in *MoRs* matrix (**Figure 41**) as rt_1 dependency relation category (according to **Table 16**). Another example, a dependency of data items exists between Order number (d_5) and 'Book price' (d_2). In this case, the dependency on data item 'Book price' (d_2) has two roles respectively in this representation: firstly, it used as a shared data item, denoted as rt_3 since it is part of 'Book title' (d_1) as an information object, and secondly it also represented as rt_4 dependency relation category i.e. formula calculation, since it part of calculation the amount of customer's order.

- ***Step 5: Map and assign per each dependency relation type the appropriate of DQ requirement(s) from generic catalogue with respect to DQ dimensions:***

  Basically, this step has <u>two sub-steps</u> (as shown in **Table 24**).

- ***Sub-step 5.1: Assign each of the identified DQ deficiencies types into the relevant dependency relation type grouped by DQ dimensions***

In **sub-step 5.1**, there is a need to categorize and assign each of the DQ deficiencies types which are identified in sub-chapter **3.4.5** (see **Table 17** to **Table 20**; pp. 134-138) with the relevant dependency relation type which are identified in **Figure 41**, grouped by DQ dimensions and then presented in an appropriate table (**Table 26**). Later, each association is translated into a set of DQ requirements and guidelines.

Table 26. Types of DQ deficiencies grouped by dependency relation types and DQ dimensions

Dependency relation type	rt_1	rt_2	rt_3	rt_4	rt_5	rt_6
DQ dimension	Aggregation (part of)	Sequence of appearance (mandatory)	Shared information	Formula calculation	External information event	Conditional information
Accuracy (ACR)	DEF1-ACR DEF2-ACR DEF3-ACR	DEF1-ACR DEF2-ACR DEF4-ACR	DEF1-ACR DEF2-ACR DEF5-ACR	DEF1-ACR DEF2-ACR DEF6-ACR		
Completeness (CMP)	DEF1-CMP DEF2-CMP DEF3-CMP	DEF1-CMP DEF2-CMP DEF4-CMP	DEF1-CMP DEF2-CMP DEF5-CMP	DEF1-CMP DEF2-CMP DEF6-CMP		
Consistency (CNS)	DEF1-CNS DEF2-CNS DEF3-CNS	DEF1-CNS DEF2-CNS DEF4-CNS	DEF1-CNS DEF2-CNS DEF5-CNS	DEF1-CNS DEF2-CNS DEF6-CNS		
Timeliness (TML)	DEF1-TML DEF2-TML	DEF1-TML DEF2-TML	DEF1-TML DEF2-TML	DEF1-TML DEF2-TML		

[Source: own study]

Each cell of the table of dependency relation types contains the relevant DQ deficiencies type's codes in order, which was defined in the previous tables (see sub-chapter **3.4.5**, **Table**

17 to **Table 20**; pp. 134-138) grouped by DQ dimensions. Each code starts with the letters DEF as a prefix with a sequential number separately per each dimension (marked as ACR, CMP, CNS and TML respectively). In addition, the table shows that columns rt_5 and rt_6 are empty, since these dependency relation types have not been identified in the example process i.e. in *MoRs* matrix (**Figure 41**). Such a platform allows BP and IS analysts and designers to map the relevant IQ requirements according to each type of DQ problem identified in this table.

For example, as described in dependency matrix (*DM matrix*) (**Figure 40**) at a former step, a dependency of data items exists between 'Book price' (d_2) and 'Book title' (d_1), i.e. the value in 'Book price' (d_2) depends on the vale in 'Book title' (d_1). Since 'Book price' (d_2) is *part of* 'Book title' (d_1) as an information object, it is assigned in *MoRs* matrix (**Figure 41**) as rt_1 dependency relation category (according to **Table 16**). In this case, the types of DQ problems identified and grouped by accuracy dimension in this table are: DEF1-ACR; DEF2-ACR and DEF3-ACR. Based on **Table 17**, DEF1-ACR represents wrong values or an error in any data item values in process representation; DEF2-ACR represents wrong values or error in data items value which reflect the RW in process representation and DEF3-ACR represents wrong values or error in data items which are part of any information object, list, form, etc. in a process representation respectively. In practice, BP and IS analysts and designers are required to examine and identify the potential errors and defects in a given process representation associated with DQ in relation to the type of DQ problem identified.

Another example, a dependency of data items exists between 'Order number' (d_5) and 'Customer address' (d_4). The dependency on data item 'Customer address' (d_4) assigned in *MoRs* matrix (**Figure 41**) as a *shared data item*, denoted as rt_3 dependency relation category (according to **Table 16**), since 'Customer address' (d_4) is *part of* 'Customer name' (d_3) as an information object, and it is used here as a *shared data item* in both information objects. In this case, the types of DQ problems identified and grouped by timeliness are: DEF1-TML; DEF2-TML and DEF3-TML. Based on **Table 20**, DEF1-TML represents a time delay in appearance of any data item value in process representation; DEF2-TML represents a change in data item value of RW representation that is not reflected in real time by equivalent data item value in process representation and DEF3-TML represents situation that the expected data item does not appear on time as planned according to the sequence of appearance in process representation respectively.

- ***Sub-step 5.2: Map and assign for each data items' dependency the appropriate requirement(s) from catalogue grouped by relation dependency types and DQ dimensions***

Finally, in **sub-step 5.2**, there is a need to map and assign for each types of DQ problems identified in **Table 26** at a former step, the appropriate DQ requirements and guidelines from the generic catalogue (**Table 21**) grouped by dependency relation types and DQ dimensions.

Table 27 describes the sets of DQ requirements which are mapped and assigned per each type of DQ deficiencies based on the generic catalogue of DQ requirements in our example process.

Table 27. Mapping DQ requirements per each dependency relation type grouped by DQ dimensions

Dependency relation type / DQ dimension	rt_1 Aggregation (part of)	rt_2 Sequence of appearance (mandatory)	rt_3 Shared information	rt_4 Formula calculation	rt_5 External information event	rt_6 Conditional information
Accuracy (ACR)	RQ1-ACR RQ2-ACR RQ3-ACR	RQ1-ACR RQ2-ACR RQ4-ACR	RQ1-ACR RQ2-ACR RQ5-ACR	RQ1-ACR RQ2-ACR RQ6-ACR		
Completeness (CMP)	RQ1-CMP RQ2-CMP RQ3-CMP	RQ1-CMP RQ2-CMP RQ4-CMP	RQ1-CMP RQ2-CMP RQ5-CMP	RQ1-CMP RQ2-CMP RQ6-CMP		
Consistency (CNS)	RQ1-CNS RQ2-CNS RQ3-CNS	RQ1-CNS RQ2-CNS RQ4-CNS	RQ1-CNS RQ2-CNS RQ5-CNS	RQ1-CNS RQ2-CNS RQ6-CNS		
Timeliness (TML)	RQ1-TML RQ2-TML RQ3-TML	RQ1-TML RQ2-TML RQ4-TML	RQ1-TML RQ2-TML RQ5-TML	RQ1-TML RQ2-TML RQ6-TML		

[Source: own study]

The above table, provides us with a formal set of requirements and guidelines for BP and IS analysts, designers and practitioners, which required to improve the IQ in the process, thus helping us achieving higher quality in the process design phase. Each of table cell contains the requirements and guidelines needed for the identified DQ problems of each dependency relation type, based on a generic catalogue of DQ requirements which created in **Table 21**. In addition, **Table 27** shows also that columns rt_5 and rt_6 are empty, since types of DQ problems of these dependency relation types have not been identified in the example process i.e. in *MoRs* matrix (**Figure 41**).

Based on the example and **Table 26**, dependency matrix (*DM*) (**Figure 40**) and *MoRs* matrix (**Figure 41**) presenting the types of DQ problems identified in **Table 26,** grouped by accuracy dimension are: DEF1-ACR, which represent wrong values or error in any data item values in process representation; DEF2-ACR, which represent wrong values or error in data items value which reflect the RW in process representation, and DEF3-ACR, which represent wrong values or error in data items which are a part of any information object, list, form, etc. in process representation. The identified requirements of information quality for these types of DQ problems and deficiencies in this dimension based on are RQ1-ACR i.e. ensure a proper representation, reliable and free of error of any data items values in a given process; RQ2-ACR i.e. identify whether data item value from RW map to a wrong data item value in process representation and fix it and RQ3-ACR i.e. ensure a proper representation, reliable and free of error of data items values which are part of any information object in the process representation respectively.

In practice, BP and IS analysts and designers are required to examine and identify the appropriated requirements based on **Table 21** for these potential errors and defects in a given process representation associated with DQ in relation to the type of DQ problem identified.

3.7. Summary

In this chapter, our research artifacts for DQ improving in BP design are presented. In the first sub-chapter (3.1) we introduced the background, the idea and foundations of real-world representation with BPs and ISs and the mutual impact between BPs and ISs representation. Then, in sub-chapter 3.2, we presented a conceptual model as a meta-model for DQ problem and as an abstract framework for understanding relationships of some concepts and entities within a particular problem domain. Then, in sub-chapters 3.3 to 3.5, the constructs and requirements of our main research artifact i.e. the new method are described. Finally, in sub-chapter 3.6, the main research artifact i.e. the Data Quality Deficiencies Prediction (DQDP) method is presented in detail including describing its constructs and steps with illustrations.

The proposed DQDP method built with reference to other existing methods, approaches and other solutions used today in order to enhances its capabilities relative to the other methods in such a way as to meet BP and IS analysts, designers and practitioners expectations and to alleviating them in their BP design mission to achieve a high quality business process.

In fact, these two artifacts are the outcomes for achieving the second and third research goals and an answer for first and second research questions.

Within the next chapter the validation of the developed DQDP method is presented.

Part III: Research Validation

The validation part in research describes the evaluation activity of design artifacts. Evaluation of design artifacts is a central key activity and crucial in design science research (DSR), as it provides a feedback for further development and (if done correctly) assures the rigor of the research (March & Smith, 1995; Hevner et al., 2004; Peffers et al., 2008; Vaishnavi & Kuechler, 2004; 2015; Wieringa, 2014; Brandtner et al., 2016; Venable et al., 2016). Furthermore, it requires researchers to rigorously demonstrate the utility, quality, and efficacy of a design artifact using well-executed evaluation methods (Hevner et al, 2004; Venable et al., 2016).

The validation of the proposed DQDP method represents an important addition to the theoretical part and it can help in the identification of open issues to be challenged in future research.

4 Evaluation of Proposed DQDP Method

The evaluation phase provides essential feedback to the construction phase as to the quality of the design process and the design product under development (Hevner et al., 2004). The main purpose of the evaluation chapter is to measure to what extent the artifact solves the problem that has been defined at the beginning of the research (Peffers et al., 2008). Basically, this chapter is the proof for achieving the research goals and the answer for all research questions. Furthermore, within a DSR process, evaluation aims to verify if and how well the design objectives or requirements specified beforehand were accomplished. This can be obtained by logical proof or empirical evidence, such as performance indicators (quantitative) or client feedbacks (qualitative) like in our case. The results of the evaluation show whether the objectives are fulfilled, or if further adjustment is necessary (Peffers et al., 2008). As discussed more fully in Hevner (2007), two forms of artifact evaluation are performed in a design research project - the evaluation of the artifact to refine its design in the design science build/evaluate cycle and the field testing of the released artifact in the application environment (Tremblay et al., 2010).

The business environment establishes the requirements upon which the evaluation of the artifact is based. DSR artifacts i.e. constructs, models, methods, and instantiations must be exercised within appropriate domain environments and appropriate subject groups must be obtained for such studies. Furthermore, in DSR artifact evaluation within a specific environment is of crucial importance (Peffers et al., 2012). The evaluation of designed artifacts typically uses methodologies available in the knowledge base (Hevner et al., 2004; Peffers et al., 2012).

This chapter deals with the evaluation of our research artifact i.e. the new proposed DQDP method, based on case study analysis, focus group sessions and comparison with other existing methods in order to validate its utility and usability in depth. The first sub-chapter 4.1 introduces the criteria for evaluation of our proposed DQDP method. Then in sub-chapter 4.2 the background and foundations of case study analysis and the case study domain is described. Moreover, in sub-chapter 4.2.3, the implementation and evaluation of our proposed DQDP method along its steps in a selected case study for checking its completeness and quality is presented and discussed. In addition, sub-chapter 4.3 presents the evaluation of proposed DQDP method based on confirmatory focus groups (CFGs) sessions for checking and demonstrating its utility and efficacy. Then in sub-chapter 4.4 the evaluation of proposed DQDP method based on comparison with other methods is described. The evaluation review in term of research thesis and its degree of suitability are discussed and described in sub-chapter 4.6. Finally, the summary of evaluation results and chapter summary are described in sub-chapters 4.5 and 4.7 respectively.

4.1. Evaluation Criteria of Proposed DQDP Method

Over the years, various researchers (e.g. March & Smith, 1995; Hevner et al., 2004; Peffers et al., 2012; Prat et al., 2014; Brandtner et al., 2016; Venable et al., 2016) have presented different criteria for DSR and for IS artifact evaluation in particular. Sometimes these researchers use synonyms or have the same meaning for the same criteria. Moreover, evaluation criteria in DSR in general, lacks a systematic list of evaluation criteria for artifacts and an associated set of evaluation methods and the literature presents evaluation criteria in a fragmented manner (Prat et al., 2014).

Nevertheless, there are three most common and well-known criteria in DSR, stated by Hevner et al. (2004). The three are *utility*, *efficacy*, and *quality* of a design artifact and repetitive as basic criteria in the DSR literature, where researchers are required to examine them within the context of their artifact evaluation. These significant and important criteria must be rigorously demonstrated in artifact evaluation via well-executed evaluation methods (Peffers et al., 2012).

According to Prat et al. (2014), the *utility* criterion measures the quality of the artifact in practical use and its benefits. In addition, the term utility is often used synonymously to the term usefulness in literature (Prat et al., 2014) and utility has often been assessed through perceived usefulness (Brandtner et al., 2016). Utility is considered the ultimate goal of DSR and it is common to people and organizations (Prat et al., 2014).

The *efficacy* criterion is the degree to which the artifact produces its desired effect (i.e. achieves its goal). The *quality* criterion according ISO 9001 standard, is the degree to which a set of inherent characteristics of an object or product fulfils a set of requirements and its ability to satisfy stated and implied needs (ISO 9000:2015). In fact, the term object or product refers to our artifact, i.e. the proposed DQDP method. According to Hevner et al. (2004), quality of IS/IT artifacts can be evaluated in terms of functionality, completeness, consistency, accuracy, performance, reliability, usability, fitting the organization, and other relevant quality attributes.

Venable et al. (2016) share the same criteria raised by Hevner et al. (2004) and stated in addition, that properties of the method to be evaluated also included its *applicability*, *usability* and *effectiveness*. Venable et al., (2016) define *effectiveness* as the utility aspect of the artifact in the environment. The *Effectiveness* is sometimes distinguished from efficacy whereas Hevner et al. (2004) use these two terms interchangeably. *Effectiveness* means the artifact instantiation works in a real situation (Prat et al., 2014).

March & Smith (1995) in their comprehensive study, identified a set of evaluation criteria for different artifacts. They identified the following additional criteria for evaluation of method: *operationality* (i.e. the ability to perform the intended task or the ability of humans to effectively use the method), *efficiency*, *generality*, and *ease of use*.

Usability (or functionality) means the product's suitability to its use. The official ISO 9241-11 (2018) standard defines the term *usability* in a broader manner: “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use”. Basically, functionality and usability are two interchangeable terms. The artifact *generality* is goal generality i.e. the broader the goal addressed by the artifact, the more general the artifact (Prat et al., 2014). *Generality* is mentioned by March and Smith (1995) for methods.

In this context, Venable et al. (2016) presents a set of recommended frameworks with properties and features for an artifact evaluation. Furthermore, they present inter alia, the properties framework of Smithson & Hirschheim (1998) and stated that it is recommended as a means for artifact evaluation in case the artifact is part of design and/or an artifact for new IS. Since our main research artifact i.e. the DQDP method, is an IS artifact and is considered as part of the new IS design, we found this framework as an appropriate additional framework for evaluation of our artifact and we adopt their recommendation.

The properties framework of Smithson & Hirschheim (1998) adapting both, *rationality and understanding properties* for artifact evaluation. Rationality properties include *efficiency* and *effectiveness* dimensions i.e. aspects of quality assurance, *cost-benefit*, *user satisfaction* and *resource utilization*. Understanding properties include social action and cognitive psychology aspects. Our focus here is on the rationality properties and less on the social action and cognitive psychological aspects.

Based on FEDS approach the chosen criteria should be aligned with our main evaluation strategy i.e. *Human Risk & Effectiveness* as mentioned above, since the major design risk here is human or socially oriented and/or with real users in their real-world context and/or it is a critical goal of the evaluation is to rigorously establish that the utility/benefit will continue in real situations and over the long run. Furthermore, it is also oriented to functional purpose of the evaluation i.e. *summative evaluations* to determine the extent that our DQDP method (the artifact) match requirements and expectations and compatible with naturalistic evaluation strategy which explores the performance of the artifact in its real-world environment with real people and real systems.

In summary, the chosen evaluation criteria include the basic and the most common and recommended criteria in DSR literature, i.e. *utility*, *efficacy*, *quality* and *usability* of a design artifact while the attributes of quality assurance include *efficiency, generality* and *user satisfaction*.

4.2. Evaluation of Proposed DQDP Method Based with Case study

4.2.1. Case study analysis

Case study is an empirical research method that uses a mix of quantitative and qualitative evidence (mostly qualitative design) to examine in depth a contemporary phenomenon, program, event, activity, process, or one or more individuals in its real-life context (Teegavarapu et al., 2008; Creswell, 2014; Yin, 2014; Zhu et al., 2014). Moreover, it is a method used to narrow down a very broad field of research into one easily researchable topic (Shuttleworth, 2008) and study artifacts in depth in a business environment to justify and evaluate research (Hevner et al., 2004). The case(s) are bounded by time and activity, and researchers collect detailed information using a variety of data collection procedures over a sustained period (Creswell, 2014). However, study of a small number of cases can limit the reliability or generality of findings (Sidi et al., 2009).

Case study research is the most common qualitative method used in information systems (Sidi et al., 2009) and it is very popular among researchers since it utilizes naturally existing information sources such as people and interactions between people within the scope of the case (Karlsson, 2016). Additionally, case study research allows the exploration and excels in the understanding of a complex issue, can extend experience or add strength to what is already known through previous research and is considered a robust research method (Zainal, 2007; Sidi et al., 2009).

The case study research design is especially useful for testing whether scientific theories, to generate hypotheses, and to validate methods and models actually work in the real-world and is often recommended as a suitable research method for software engineering researchers and practitioners for improvement purposes of the studied phenomenon (Kontio, 2004; Shuttleworth, 2008; Teegavarapu et al., 2008; Runeson & Höst, 2009). Case studies implement the artifact in a real-world situation to evaluate not only its utility, but also its effect on its environment (Peffers et al., 2012). Furthermore, **this method** is widely used in data quality research domain (Zhu et al., 2014). Case study is also popular with practitioners as a tool for evaluation and organizational learning (Baškarada, 2014) and it is an ideal method when the aim of research is to find answers to "How" and "Why" types of research questions (Teegavarapu et al., 2008; Baškarada, 2014; Yin, 2014).

Development of a case study for validation purpose is needed for achieving the fourth research goal and answering our third research question. Moreover, it aims to provide a proof-of concept for the viability of the DQDP method and is used as a fundament for the method. In addition, this part shows how the proposed DQDP method can meet requirements outlined in sub-chapter 3.5 (p. 143).

4.2.2. The case study: Sales of sea export service

4.2.2.1. About international forwarding and moving industry

The case study which will validate the new proposed method is based on the central process of the international forwarding and moving industry. The global freight forwarding and moving industry is vast, both in terms of market size and huge numbers of people employed and it is considered one of the most important industries today and the most influential on logistics supply chain management (Manners-Bell J., 2014).

According to the European Freight Forwarders Association (EFFA), the term "freight forwarding industry" refers to a set of global logistics services for the exporter and importer in moving cargo to an overseas destination, which include services of any kind relating to the carriage, storage, handling, packing, etc., using modern information and communication technology. Basically, management and operation of the process is done by an international freight forwarder (Manners-Bell J. 2014).

An international freight forwarder is an agent for the exporter and importer in moving cargo to an overseas destination. These agents are familiar with the export and import rules and regulations of foreign countries, the methods of shipping, and the documents related to the foreign trade. The global service includes sub services such as carriage and inland moving, storage, warehousing, freight consolidation, cargo insurance and more. Once the order is ready for shipment, freight forwarders should review all documents to ensure that everything is in order. This is of special importance with regards to letters of credit payment terms. They may also prepare the bill of lading and any special required documentation. After shipment, freight forwarders can route the documents to the seller, the buyer, or to a paying bank. Freight forwarders can also make arrangements with customs brokers overseas to ensure that the goods comply with customs export documentation regulations. A customs broker is an individual or a company that is licensed to transact customs business on behalf of others.

4.2.2.2. Sales process of sea export service

Our case study analysis is based on sales process of the export services. Export services are one of two popular and favorite services in the international forwarding industry. To demonstrate our idea, we decided to focus on sales process domain of sea export services. The main reasons for choosing this process as a case study are because it is considered one of the most complex processes in the international forwarding and moving domain and is considered as a rich process in data items. It should be noted that the process has numerous instances and is even simplified, has many different data items associated. Moreover, in this process there are many aspects of dependencies between data items that are critical to the process success as a whole and therefore the importance of the components of the information quality in such a process is very high.

4.2.3. Implementation of proposed DQDP method on case study

In this research, we designed and developed a DQDP method, which implements our DQ quality approach and now, based on the above chosen case study, we further demonstrate its validity. This implementation has two core objectives: first, it helps in demonstrating the feasibility of the method. The second objective is related to the validation of the method as explained in the introduction to this part and chapter.

This section describes the implementation process of our proposed DQDP Method and its defined layers and steps as depicted in **Table 23** and **Table 24** in particular as described in above sub-chapter **3.6** (see pp. 145 -146).

4.2.3.1. Implementation steps of DQDP method

Step 1: Choose a valid process model

To demonstrate our idea, we decided to focus on sales process domain of sea export services. It should be noted that the process has numerous instances and is even simplified, has many different data items associated. The valid process model is created by using BPMN technique (OMG, 2011), as a graphical notation for describing BP models, which is widely recognized as de-facto standard for BPs modeling as mentioned above (sub-chapter **3.6.2**; p. 148).

Figure 42 presents a valid process model of our case study process by using BPMN diagram notation. Basically, this process domain is triggered and started by an external information event, i.e. receiving customer application into the first activity, including a request for quotation for sea export service. The application form is received by the sales department representative who is responsible for handling it.

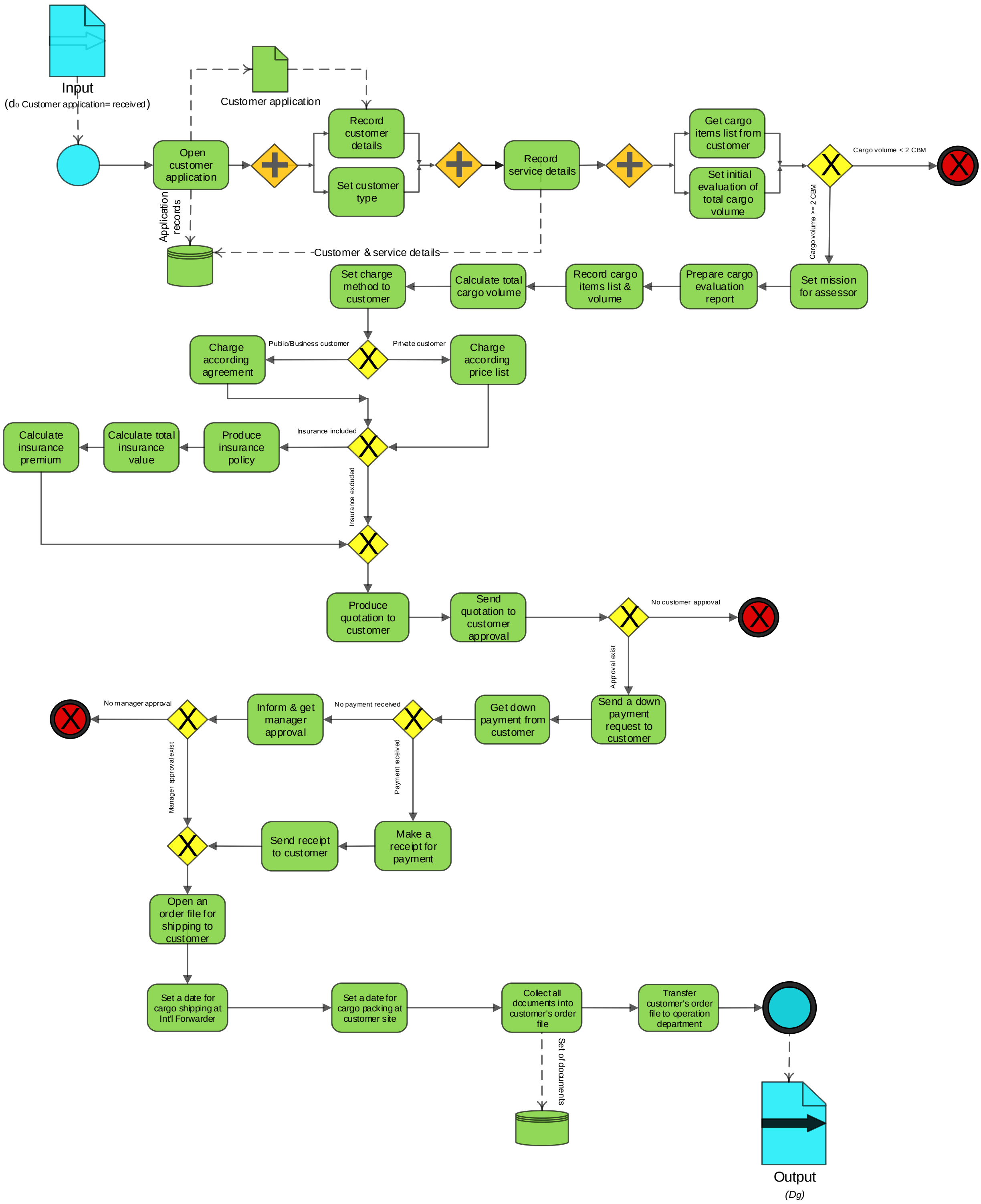


Figure 42. Sales process of sea export service (BPMN diagram)

[Source: own study]

Based on the process of our case study described above and its BPMN diagram, we identified a set of activities (A_p) and a set of data items (D_p) that exist in our case study process. **Table 28** presents the list and details of all the activities in the process of our case study, where: $A_p = \{a_1 \dots a_{30}\}$.

Table 28. List of activities in case study

Activity ID	Activity description
a_1	Open customer application
a_2	Record customer details
a_3	Set customer type & degree of risk
a_4	Record service details
a_5	Set responsible department
a_6	Get cargo items list from customer
a_7	Make initial evaluation of total cargo volume
a_8	Set mission for assessor (for cargo evaluation)
a_9	Prepare cargo evaluation report
a_{10}	Record cargo items list & volume
a_{11}	Calculate a total cargo volume
a_{12}	Set charge method to customer
a_{13}	Charge according to price list
a_{14}	Charge according to agreement
a_{15}	Produce insurance policy
a_{16}	Calculate total insurance value
a_{17}	Calculate insurance premium
a_{18}	Produce quotation to customer
a_{19}	Send quotation to customer approval
a_{20}	Send a down payment request to customer
a_{21}	Get down payment from customer
a_{22}	Inform & get manager approval
a_{23}	Cancel application or quotation
a_{24}	Make a receipt for payment
a_{25}	Send receipt to customer
a_{26}	Open an order file for shipping to customer
a_{27}	Set a date for cargo shipping at Int'l Forwarder
a_{28}	Set a date for cargo packing at customer site
a_{29}	Collect all documents into customer's order file
a_{30}	Transfer customer's order file to operation department

[Source: own study]

Table 29 presents the list and details of all data items in our process case study, where $D_p = \{d_0 \dots d_{32}\}$.

The trigger of the process is an external information event represent by data item d_0 i.e. Customer application status = receive

Table 29. List of data items in case study

Data item ID	Description	Data type	Examples of data values	Remarks
d_0	Customer application status	String	Received; Closed; Cancelled	External information event Limited List
d_1	Customer application	Form	Form details	
d_2	Customer application no.	String	CA0001; CA0002; …	
d_3	Application date/time	Date/Time	29/01/16; 08:45	
d_4	Customer name	String	Michael Vaknin	
d_5	Customer address	String	Arnon St. #22, TLV	
d_6	Customer type	String	Private; Business	Limited List
d_7	Degree of risk	Boolean	Low/High	
d_8	Service type	String	Export; Import; Storage	Limited List
d_9	Shipping destination (name of state)	String	Poland; Israel; …	Limited List of Countries
d_{10}	Port at destination	String	Gdansk; Haifa; …	Limited List of Ports
d_{11}	Shipping requested date	Date/Time	29/01/16; 08:45	
d_{12}	Is insurance included?	Boolean	Yes/No	
d_{13}	Responsible department	String	Export/ Import/ Storage	Limited List
d_{14}	Cargo items list from a customer	Report (of Items list)	Desk; Chairs; Closet; Freezer; Bed; Table; Cloths; …	An external information event
d_{15}	Initial evaluation of cargo volume	Number	4.5 CBM	Amount value in CBM (Cubic Meter)
d_{16}	Mission for assessor	String	Opened; Closed; Cancelled	Limited List
d_{17}	Assessor cargo evaluation report	Report (of Items list)	Items list report	External information event
d_{18}	Total cargo volume	Number	24.5 CBM	Amount value in CBM (Cubic Meter) Calculate by formula
d_{19}	Charge method (per CBM)	Boolean	Agreement / Price list	Charge according appropriate method
d_{20}	Charge cost/tariff (per CBM)	Number/ Currency	1,800 US $	Tariff in US $
d_{21}	Insurance policy	Form	123452	Document

d_{22}	Total insurance value	Number /Currency	1,800 US $	Amount value in US $ Calculate by formula
d_{23}	Insurance premium value	Number /Currency	750 US $	Amount value in US$ Calculate by formula
d_{24}	Quotation status	String	Opened; Sent; Approved; Closed; Cancelled	Limited List
d_{25}	Customer approval?	Boolean	Yes/No	External information event
d_{26}	Down payment request	Form	Form details + amount of down payment in US $	e.g. document Calculate by formula
d_{27}	Down payment received?	Boolean	Yes/No	Bank account transactions
d_{28}	Receipt for customer status	String	Opened; Sent; Closed; Cancelled	Limited List
d_{29}	Manager approval?	Boolean	Yes/No	
d_{30}	Customer order file status	String	Opened; Completed; Closed; Cancelled	Limited List
d_{31}	Shipping date at Int'l Forwarder	Date/Time	08/02/16; 09:45	External information event
d_{32}	Packing date at customer site	Date/Time	29/01/16; 08:00	

[Source: own study]

Step 2: Transform process goal(s) into a set of data items

The outcomes of the process are derived from its goals (based on definition 7; p. 103). Recall, the goal of the process (G_p) of our case study is *to supply sales process of export services for customers* i.e. exporters who want to ship their cargo to any destination abroad. In fact, at this step, this goal is transformed to a set of data items values, denoted as D_g, that represent the outcomes of the process as defined by the process manager or by the organization's management.

The selected set of data items is a sub-group and part of data items list in a given process, i.e. **Table 29** in our case study where: $D_g \subseteq D_p$.

The following set of data items is identified as a process goal representation (D_g) of process in our case study analysis:

$$D_g = \{d_4, d_8, d_9, d_{10}, d_{17}, d_{18}, d_{20}, d_{24}, d_{28}, d_{30}, d_{31}, d_{32}\}$$

Furthermore, each of the data items that belong to the process goals (D_g) eventually gets a final value at the end of the process. For example: d_4= John James; d_9= Poland; d_{10}= Gdansk; d_{18}= 24.5 CBM; d_{24}= Approved; d_{30}= Closed; d_{32}= 29/01/19; 08:00, and so on. If for example, one of them is incorrect or missing, then it can lead the process goal to become unreachable and/or to lead the process to be in a deadlock situation

Step 3: Identify set of inputs and outputs data items per each activity

In the next step we identified all inputs and outputs data items per each activity along the process of our case study and its notation as described above. Recall, we assumed (on Assumption 5; p. 102) that each activity a_i is associated with two sets of data items - input and output data items set, $\boldsymbol{D_{in}}(a_i)$ and $\boldsymbol{D_{out}}(a_i)$ respectively, and $\{\boldsymbol{D_{in}}, \boldsymbol{D_{out}}\} \in D_p$.

In this way, we can map in advance at the BP design stage, what data items are involved in each activity in the process and what their characteristics are in terms of IQ. **Table 30** includes all input and output data items per each activity along this process.

Table 30. List of input and output data items per each activity in case study

Activity ID ($\boldsymbol{a_i}$)	Activity description	Input data item ID ($\boldsymbol{D_{in}(a_i)}$)	Output data item ID ($\boldsymbol{D_{out}(a_i)}$)	Remarks
a_1	Open customer application	d_0	d_1, d_2, d_3	
a_2	Record customer details	d_1, d_2	d_4, d_5	
a_3	Set customer type & degree of risk	d_4	d_6, d_7	
a_4	Record service details	d_1, d_2	$d_8, d_9, d_{10}, d_{11}, d_{12}$	d_{12} *is conditional*
a_5	Set responsible department	d_2, d_8	d_{13}	
a_6	Get cargo items list from customer	d_2, d_4	d_{14}	
a_7	Make initial evaluation of total cargo volume	d_{14}	d_{15}	
a_8	Set mission for assessor (for cargo evaluation)	d_3, d_{15}	d_{16}	
a_9	Prepare cargo evaluation report	d_3, d_{16}	d_{17}	
a_{10}	Record cargo items list & volume	d_{17}	d_{18}	
a_{11}	Calculate total cargo volume	d_{18}	d_{20}	
a_{12}	Set charge method to customer	d_4, d_6	d_{19}	
a_{13}	Charge according to price list	d_{19}	d_{20}	*if* d_6*='Private'*
a_{14}	Charge according to agreement	d_{19}	d_{20}	*if* d_6*='Business'*
a_{15}	Produce insurance policy	d_4, d_{12}	d_{21}	d_{21} *is conditional if* d_{12}*='Yes'*

a_{16}	Calculate total insurance value	d_{17}, d_{21}	d_{22}	
a_{17}	Calculate insurance premium	d_{22}	d_{23}	
a_{18}	Produce quotation for a customer	$d_2, d_3, d_4, d_5, d_6, d_8, d_9, d_{10}, d_{11}, d_{18}, d_{20}, d_{23}$	d_{24}='Opened'	d_{23} *is conditional*
a_{19}	Send quotation to customer approval	d_{24}='Opened'	d_{25}	
a_{20}	Send a down payment request to customer	d_{25}	d_{26}	*if* d_{25}='Yes'
a_{21}	Get down payment from customer	d_{26}	d_{27}	
a_{22}	Inform & get manager approval	d_{27}='No'	d_{29}	d_{29} *is conditional if* d_{27}='No' *and* d_7='High'
a_{23}	Cancel application or quotation	d_{15}<2CBM, *or* d_{25}='No', *or* d_{29}='No'	d_1='Cancelled' *or* d_{24}='Cancelled'	
a_{24}	Make a receipt for payment	d_{27}='Yes'	d_{28}='Opened'	*if* d_{27}='Yes'
a_{25}	Send receipt to customer	d_{28}='Opened'	d_{28}='Sent'	
a_{26}	Open an order file for shipping to customer	d_{28}='Sent' *or* d_{29}='Yes'	d_{30}='Opened'	d_{29} *is conditional if* d_{29}='Yes'
a_{27}	Set a date for cargo shipping at Int'l Forwarder	$d_4, d_5, d_8, d_9, d_{10}, d_{11}, d_{30}$	d_{31}	
a_{28}	Set a date for cargo packing at customer site	$d_4, d_5, d_8, d_{11}, d_{31}$	d_{32}	
a_{29}	Collect all documents into customer's order file	d_{31}, d_{32}	d_{30}='Completed'	
a_{30}	Transfer customer's order file to operation department	d_{30}='Completed'	d_{30}='Closed'	

[Source: own study]

Step 4: Build matrices of all dependencies between data items and relation types

Basically, this step has three sub-steps:

- ***Step 4.1: Identify all data items dependencies and build a DM matrix***

In this sub-step, we identified all data items dependencies that exist in our case study.

Based on the BPMN Diagram of our case study and the idea of data dependencies described above in sub-chapter **3.4.1** (p. 122), we built a *dependency matrix,* denoted as *DM* (see **Figure 43**), which summarizes all dependencies that exist between various data items listed above and their values in the selected process. We then collect dependencies based on a study of possible process paths and assess whether an activity depends on a certain value or not. Each square on the matrix with the digit 1 contains an existing dependency state between two data item values. The size of the matrix DM depends on the number of data items along and within the process i.e. 32x32. **Figure 43** shows the *DM* matrix with all data items dependencies in our case study process.

Data item description	ID#	d0	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10	d11	d12	d13	d14	d15	d16	d17	d18	d19	d20	d21	d22	d23	d24	d25	d26	d27	d28	d29	d30	d31	d32
Customer application status	d0																																	
Customer application	d1	1																																
Customer application no.	d2		1																															
Application date/time	d3		1																															
Customer name	d4		1																															
Customer address	d5					1																												
Customer type	d6					1																												
Degree of risk	d7					1																												
Service type	d8		1																															
Shipping destination (name of state)	d9		1																															
Port at destination	d10										1																							
Shipping requested date	d11		1																															
Is insurance included?	d12		1																															
Responsible department	d13									1																								
Cargo items list from a customer	d14		1																															
Initial evaluation of cargo volume	d15															1																		
Mission for assessor	d16			1	1	1	1			1							1																	
Assessor cargo evaluation report	d17				1	1	1											1																
Total cargo volume (in CBM)	d18																		1															
Charge method (per CBM)	d19							1																										
Charge cost/tariff (per CBM)	d20																				1													
Insurance policy	d21					1								1																				
Total insurance value	d22																		1				1											
Insurance premium value	d23																						1	1										
Quotation status	d24			1	1	1	1			1	1	1	1		1					1	1	1			1									
Customer approval?	d25																									1								
Down payment request	d26					1																				1	1							
Down payment received?	d27					1																						1						
Receipt for customer status	d28					1	1								1											1			1					
Manager approval?	d29					1		1	1																				1					
Customer order file status	d30					1	1												1							1	1			1	1			
Shipping date at Int'l Forwarder	d31					1				1	1	1	1							1												1		
Packing date at customer site	d32					1	1						1							1													1	

Legend: 1 Dependency exist; No dependency exist; Impossible dependency

An area of impossible dependencies

Figure 43. Data items dependency matrix (DM) in our case study

[Source: own study]

Recall, the boxes with 1 digit represent the places where dependency between data items exists. The grey boxes represent places where dependency does not exist and finally, the white boxes represent places where dependency is impossible.

Finally, we identified all possible paths along the given process i.e. the trajectories of all processing activities and the sequences of dependencies of their data items flows sets based on progressing from start point to the end of our process case study:

1. d_0 - d_1 - d_4 - d_{24} - d_{25} - d_{26} - d_{27} - d_{28} - d_{30} - d_{31} - d_{32}
2. d_0 - d_1 - d_4 - d_{25} - d_{26} - d_{27} - d_{28} - d_{31} - d_{32}
3. d_0 - d_1 - d_2 - d_{15} - d_{16} - d_{17} - d_{18} - d_{25} - d_{26} - d_{27}- d_{28} - d_{31} - d_{32}
4. d_0 - d_1 - d_3 - d_{15} - d_{16} - d_{17} - d_{18} - d_{25} - d_{26} - d_{27}- d_{28} - d_{31} - d_{32}
5. d_0 - d_1 - d_3 - d_{25} - d_{26} - d_{27} - d_{28} - d_{31} - d_{32}

6. d_0 - d_1 - d_6 - d_{11} - d_{12} - d_{13} - d_{14}- d_{15} - d_{16} - d_{17} - d_{18} - d_{25} - d_{26} - d_{27}- d_{28} - d_{31} - d_{32}

7. d_0 - d_1 - d_7 - d_{10} - d_{24} - d_{25} - d_{26} - d_{27} - d_{28} - d_{31} - d_{32}

8. d_0 - d_1 - d_8 - d_{25} - d_{26} - d_{27} - d_{28} - d_{31} - d_{32}

9. d_0 - d_1 - d_9 - d_{25} - d_{26} - d_{27} - d_{28} - d_{31} - d_{32}

All the above trajectories actually represent the existing transitions between the various data items along the sequences of activities in the process of our case study and their dependencies, starting from the first data item i.e. 'Customer application status' (d_0) to the last data item in the route i.e. 'Packing date at customer site' (d_{32}). Furthermore, each trajectory describes the sequence of dependencies between the different data items. Furthermore, these trajectories not only show us the sequences of dependencies that exist between the various data items in the process, but also illustrate to the BP & IS analysts, designers and practitioners which data items participate in many dependent paths and their effect on the progress of the process, and thus also they become more sensitive and important data items. For example, it can be clearly seen that the data items d_0 , d_1 , d_{24} , d_{25} , d_{26} , d_{27} , d_{28} , d_{31} , d_{32} are repeated on all the above trajectories, making them also more sensitive and important data items.

For example, in this case study, a dependency of data items exists between 'Shipping destination' (d_9) and 'Port at destination' (d_{10}), i.e. the value in 'Port at destination' (d_{10}) depends on the value in 'Shipping destination' (d_9). Recording a wrong value about requested 'Shipping destination' (d_9) can lead to process deficiencies and choose wrong 'Port at destination' (d_{10}) and finally will cause to deliver the customer's cargo to the wrong country (as a destination). The potential results of these deficiencies at process level are: first, the customer's cargo will be delivered to a wrong destination, and second, wrong value about a requested destination can lead to calculating the wrong price and charge in 'Customer's quotation' (d_{24}), i.e. failures in achieving process goals and its outcomes, and the process is expected to be in a deadlock situation.

Another example of data items dependency exists between 'Assessor cargo evaluation report' (d_{17}), i.e. cargo items' list report and 'total cargo volume' (d_{18}). Missing items in cargo items' list can lead to deficiencies or failures in the process and the potential results of this deficiency example can be (1) a wrong value in a 'Total cargo volume' (d_{18}) and (2) wrong charge in 'Customer's quotation' (d_{24}) according to the total cargo volume. The potential result of these deficiencies at process level is failure in achieving process goals.

- ***Step 4.2: Mark independent data items and dependencies which are not possible***

Further to the previous sub-step 4.1 and on the same matrix we defined and marked the places where there is no dependence i.e. independent data items and impossible dependencies between data items. **Figure 43** shows the *DM* matrix with all existing data items dependencies in our case study, where the grey boxes represent places where dependency does not exist, and the white boxes represent places where dependency is impossible.

- ***Step 4.3: Map each of data items dependencies into a relevant dependency relation types by transforming DM matrix to MoRs matrix***

In this sub-step, we identified and assigned a set of dependency relation types between any data items dependencies as described in detail in sub-chapter **3.4.2**. Recall, all the dependency relation types depicted in **Table 16** (p. 124) and each of dependency relation type received an appropriate code represent by rt_i variable code. Then, we mapped and assigned for each of data items dependency the appropriate type of dependency relation dependency code represented by rt_i (t_i in short) in an updated matrix (**Figure 44**), namely *MoRs* matrix *i.e. Matrix of Relations.*

Data item description	ID#	d_0	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}	d_{16}	d_{17}	d_{18}	d_{19}	d_{20}	d_{21}	d_{22}	d_{23}	d_{24}	d_{25}	d_{26}	d_{27}	d_{28}	d_{29}	d_{30}	d_{31}	d_{32}
Customer application status	d_0																																	
Customer application	d_1	t_2,t_5																																
Customer application no.	d_2		t_1																															
Application date/time	d_3		t_1																															
Customer name	d_4		t_3																															
Customer address	d_5					t_1																												
Customer type	d_6					t_1																												
Degree of risk	d_7					t_1																												
Service type	d_8		t_3																															
Shipping destination (name of state)	d_9		t_1																															
Port at destination	d_{10}										t_1,t_2																							
Shipping requested date	d_{11}		t_1																															
Is insurance included?	d_{12}		t_1																															
Responsible department	d_{13}									t_1,t_2																								
Cargo items list from a customer	d_{14}		t_2,t_5																															
Initial evaluation of cargo volume	d_{15}															t_2,t_4,t_5																		
Mission for assessor	d_{16}			t_3	t_2,t_3	t_3	t_3			t_3							t_2,t_6																	
Assessor cargo evaluation report	d_{17}				t_3	t_3	t_3											t_2																
Total cargo volume (in CBM)	d_{18}																		t_2,t_4,t_5															
Charge method (per CBM)	d_{19}							t_2																										
Charge cost/tariff (per CBM)	d_{20}																				t_2													
Insurance policy	d_{21}					t_3								t_6																				
Total insurance value	d_{22}																		t_2,t_4,t_5				t_1,t_2											
Insurance premium value	d_{23}																						t_1	t_2,t_4										
Quotation status	d_{24}			t_3	t_2,t_3	t_3	t_3			t_3	t_3	t_3	t_3		t_3					t_1,t_4	t_3	t_1,t_4			t_3,t_6									
Customer approval?	d_{25}																									t_2								
Down payment request	d_{26}					t_3																				t_3,t_4	t_2,t_5							
Down payment received?	d_{27}					t_3																						t_2						
Receipt for customer status	d_{28}					t_3	t_3								t_3											t_3,t_4			t_2,t_6					
Manager approval?	d_{29}					t_3		t_6	t_6																				t_2,t_5,t_6					
Customer order file status	d_{30}					t_3	t_3												t_3							t_3	t_3			t_2,t_6	t_2,t_3,t_6			
Shipping date at Int'l Forwarder	d_{31}					t_3				t_3	t_3	t_3	t_2,t_3							t_3												t_2		
Packing date at customer site	d_{32}					t_3	t_3						t_2,t_3							t_3													t_2,t_3,t_5	

Legend:
t_i Dependency relation type
No dependency exist
Impossible dependency

An area of impossible dependencies

Figure 44. MoRs matrix of data items dependency relation type in case study

[Source: own study]

Figure 44 shows the new *MoRs matrix,* with all existing dependencies of data items values in our case study process with appropriate dependency relation type codes, denoted as rt_i for each exist dependency.

For example, in our case study (and in international export or import services in general), if we want to calculate the 'Insurance premium' (data item d_{23}) for shipping cargo, we need the 'Total insurance value' (data item d_{22}) as a mandatory value, which is on the *sequence of appearance* and assigned in *MoRs* matrix as t_2 dependency relation category according to **Table 16** (see **Figure 44**). Furthermore, the 'Total insurance value' (data item d_{22}) also depends on 'Assessor cargo evaluation report' (data item d_{17}) in process representation as a mandatory value, which is also in the *sequence of appearance* as t_2 dependency relation category, but it also categorized as t_4 i.e. *formula calculation* (for 'Total insurance value') and as t_5 relation category i.e. *external information event* dependency (received information from third party) respectively. Without this value i.e. 'Total cargo volume'(d_{18}), the process is expected to be in a deadlock situation. In other words, there is a mandatory dependency relation between data items along the sequence of appearance on process that can’t execute properly if any data item value is unknown or missing.

Another example in our case study concerns the *formula calculation*, denoted as rt_4 dependency relation category. In our case study, if we want to calculate the 'Insurance premium' value (data item d_{23}) for shipping cargo in an export service, we need the 'Total cargo volume' value (data item d_{18}) as a mandatory value. Without this value (i.e. 'Total cargo volume') the process is expected to be in a deadlock situation. In other words, there is at least one transformation in a process' laws that cannot execute properly, if data item d_{18} is unknown.

Basically, this matrix also helps us to identify the type of potential failure dependencies between all data items with respect to various data quality dimensions generated by Wang & Strong (1996) e.g. accuracy, completeness, timeliness, etc. We named these potential failures as *data quality deficiencies*, and as mentioned above, a data quality deficiency is an inconformity between the view of the real-world system that can be inferred from a process representation and the view that can be obtained by directly observing a real-world system. Identification of these data dependencies and generalization of dependencies relation types, may result in the development of rules enabling automatic checking of the business process design and the improvement of some of the explained problems.

Step 5: Map and assign per each dependency relation type the appropriate of DQ requirement(s) from catalogue with respect to DQ dimensions

- ***Step 5.1: Assign each of the identified DQ deficiencies types into the relevant dependency relation types grouped by DQ dimensions***

In this sub-step, we categorized and associated each types of DQ deficiencies and flaws which identified in sub-chapter **3.4.5** (p. 133) and detailed in **Table 17** to **Table 20** (pp.134-138), in accordance with the relevant dependency relation types which identified in **Figure 44**, i.e. *MoRs* matrix, grouped by DQ dimensions and accordingly we populated the following table (**Table 31**). Such platform allows BP and IS analysts and designers to map the relevant IQ requirements according to each type of DQ problem identified in this table and prevent them in advance.

Below is a set of examples illustrating some DQ problems related to information deficiencies in different cases presented in **Table 31** based on our case study with respect to DQ dimensions.

Table 31. Types of DQ deficiencies grouped by dependency RTs and DQ dimensions

Dependency relation type / DQ dimension	rt_1 Aggregation (part of)	rt_2 Sequence of appearance (mandatory)	rt_3 Shared information	rt_4 Formula calculation	rt_5 External information event	rt_6 Conditional information
Accuracy (ACR)	DEF1-ACR DEF2-ACR DEF3-ACR	DEF1-ACR DEF2-ACR DEF4-ACR	DEF1-ACR DEF2-ACR DEF5-ACR	DEF1-ACR DEF2-ACR DEF6-ACR	DEF1-ACR DEF2-ACR DEF7-ACR	DEF1-ACR DEF2-ACR DEF8-ACR
Completeness (CMP)	DEF1-CMP DEF2-CMP DEF3-CMP	DEF1-CMP DEF2-CMP DEF4-CMP	DEF1-CMP DEF2-CMP DEF5-CMP	DEF1-CMP DEF2-CMP DEF6-CMP	DEF1-CMP DEF2-CMP DEF7-CMP	DEF1-CMP DEF2-CMP DEF8-CMP
Consistency (CNS)	DEF1-CNS DEF2-CNS DEF3-CNS	DEF1-CNS DEF2-CNS DEF4-CNS	DEF1-CNS DEF2-CNS DEF5-CNS	DEF1-CNS DEF2-CNS DEF6-CNS	DEF1-CNS DEF2-CNS DEF7-CNS	DEF1-CNS DEF2-CNS DEF8-CNS
Timeliness (TML)	DEF1-TML DEF2-TML	DEF1-TML DEF2-TML	DEF1-TML DEF2-TML	DEF1-TML DEF2-TML	DEF1-TML DEF2-TML DEF3-TML	DEF1-TML DEF2-TML

[Source: own study]

Recall, each box of dependency relation types in **Table 31** contains the relevant DQ deficiencies type's codes in order, based on **Table 17** to **Table 20** (pp. 134-138) and grouped by each DQ dimensions. Each code starts with the letters DEF as a prefix with a sequential number represents the potentials types of DQ deficiencies and flaws separately per each dimension (marked as ACR, CMP, CNS and TML respectively). Later, each association types are translated into set of data quality requirements and guidelines. Moreover, the table shows that all columns have a content, since that all dependency relation types (rt_1 to rt_6) have been identified as a relevant in our case study i.e. appeared in *MoRs* matrix (**Figure 44**). Here are some examples to illustrate the role and benefits of **Table 31**.

The first example following the example as described at former step, and based on **Figure 43**, **Figure 44** and **Table 31.** In this case, a dependency of data items exists between 'Insurance premium' (d_{23}) and 'Total insurance value' (d_{22}) as a mandatory value, i.e. the value of 'Insurance premium' (d_{23}) depend on the value of 'Total insurance value' (d_{22}), and it assigned as *sequence of appearance* dependency relation category (rt_2) in *MoRs* matrix (**Figure 44**) according to **Table 16**. Since 'Insurance premium' (d_{23}) is calculated based on 'Total insurance value' (d_{22}), it assigned as *formula calculation* dependency relation category, denote as rt_4 in *MoRs* matrix (**Figure 44**) according to **Table 16**. Moreover, at the same time, 'Insurance premium' (d_{23}) is also *part of* 'Insurance policy' (d_{21}) as an information object, so it assigned as *aggregation* dependency relation category (rt_1) in *MoRs* matrix (**Figure 44**) according to **Table 16**.

In this case, the potential types of DQ problems identified in **Table 31**, grouped by accuracy dimension in this table are: DEF1-ACR, which represent wrong values or error in any data item values in process representation; DEF2-ACR, which represent wrong values or error in data items value which reflect the RW in process representation; DEF3-ACR, which represent wrong values or error in data items which are part of any information object, list, form, etc. in process representation; DEF4-ACR, which represent wrong values or error in sequence of appearance of data items in process representation, and DEF6-ACR, which represent wrong values or error recorded in data items which are part of any formula calculation in process representation. In fact, these deficiencies types can lead to making some mistakes, wrong decisions or wrong values in others dependable data items or delay in process progress or can lead the process to be in deadlock situation.

The second example based on the dependency of data items which exists between 'Mission for assessor' (d_{16}) and 'Customer address' (d_5). The dependency on data item 'Customer address' (d_5) assigned in *MoRs* matrix (**Figure 44**) as a *shared information*, denoted as rt_3 dependency relation category (according to **Table 16**), since 'Customer address' (d_5) is *part* *of* 'Customer name' (d_4) as an information object, and it used here as a shared data item in 'Mission for assessor' (d_{16}). In this case for example, there are inter alia, two potentials types of DQ problems, grouped by accuracy and consistency dimensions in **Table 31**. The first one is denoted as DEF1-ACR, which represents wrong value or error in 'Customer address' (d_5) value in process representation. This error can lead to achieve mistakes or wrong decision in the dependable data item in process i.e. 'Mission for assessor' (d_{16}), which in turn can lead the process to become unreachable or to be in deadlock situation. The second potential type of DQ problems relate to consistency dimension in **Table 31**, denoted as DEF5-CNS i.e. reflects type of inconsistency, unambiguous or meaningless in shared data items of any information object in the process representation. In fact, it represents situation where 'Customer address' contains two values e.g. home and office addresses but only one address is relevant to 'Mission for assessor' (d_{16}). Furthermore, this deficiency type can lead to making mistakes, wrong decisions or wrong values in others dependable data items or delay in process progress or can lead the process to be in deadlock situation.

The third example relate to *formula calculation* dependency relation, denote as rt_4, and describes a situation in which a data item is actually composed of several other data items values and that performs any quantitative calculation and others DQ problems. For example, the data item 'Quotation to customer' (d_{24}) is composed (among others) and based on the calculation of 'Total cargo volume' (d_{18}) multiplied by 'Charge cost/tariff' (d_{20}). In this case for example, there are inter alia, two potential types of DQ problems, grouped by accuracy and completeness dimensions in **Table 31**. If one of these data components is incorrect (accuracy) or is missing (completeness) i.e. potential type of DQ problems like DEF1-ACR and DEF1-CMP respectively, can appear. The expected results at the process level can be for example, that the final price value of the proposal is incorrect or cannot be calculated and the process is expected to achieve outputs and deliverables which are not in the process goals set and/or the process is expected to be stuck i.e. be in a deadlock situation in some cases.

The fourth example relates to *external information event* dependency relation, denote as rt_5, and describes the result of actions of objects outside the process domain where a dependency relation regarding data item that must be received from an external third party. This external information is required to activate the process domain when the process is ready to start or should be arrived along the process, and therefore are outside of the process control. In our case study for example, if the expected 'Advanced down payment' (d_{27}) from customer not received or the expected information about scheduling and shipping dates of vessels for cargo shipping at Int'l forwarder (d_{31}) fails to receive, then it relates to completeness and timeliness dimensions, i.e. the identified potential type of DQ problems are DEF1-CMP, DEF1-TML and DEF2-TML in respectively, can appears according to **Table 31**. Furthermore, in this case the process is also expected to be stuck i.e. be in a deadlock situation some if these data items do not arrive at the appropriate time.

The last example relates to *conditional information* dependency relation, denoted as rt_6, that describes a situation of dependency among values of data items along process representation, where the value of data item $d_{in}(a_j)$ is affected by value of data item $d_{out}(a_i)$ and uses it under some conditions i.e. conditionally depends on. The process may be executed without using this data item value; hence, data item $d_{out}(a_i)$ in representation can be unknown under some conditions.

In our case study for example, approving of opening 'Customer order file' (data item d_{30}) by 'Manager approval' (d_{29}) conditionally depends on receiving 'Down payment from customer' (data item d_{27}), conditioned by the 'Customer's risk level' (d_7) which is another data item. If the status of customer's risk level (d_7) is low and there is no down payment received, i.e. the down payment value (data item d_{27}) in representation is unknown yet or null, the process is still in progress since the 'Customer risk level' (d_7) is low.

- ***Step 5.2: Map and assign for each data items dependency the appropriate requirement(s) from catalogue grouped by relation dependency types and DQ dimensions***

Finally, in the last step, we mapped and assigned per each type of DQ deficiencies identified in **Table 31** in a former step , the appropriate DQ requirements and guidelines based on the generic catalogue of DQ requirements which described above (**Table 21**; p. 140), grouped by appropriate dependency relation type between data items and DQ dimensions into **Table 32**.

Table 32 describes the sets of DQ requirements which are mapped and assigned per each type of DQ deficiencies based on the generic catalogue of DQ requirements in our example process.

Table 32. Mapping DQ requirements per each dependency relation type

Dependency relation type / DQ dimension	rt_1 Aggregation (part of)	rt_2 Sequence of appearance (mandatory)	rt_3 Shared information	rt_4 Formula calculation	rt_5 External information event	rt_6 Conditional information
Accuracy (ACR)	RQ1-ACR RQ2-ACR RQ3-ACR	RQ1-ACR RQ2-ACR RQ4-ACR	RQ1-ACR RQ2-ACR3 RQ5-ACR4	RQ1-ACR RQ2-ACR RQ6-ACR	RQ1-ACR RQ2-ACR RQ7-ACR	RQ1-ACR RQ2-ACR RQ8-ACR
Completeness (CMP)	RQ1-CMP RQ2-CMP RQ3-CMP	RQ1-CMP RQ2-CMP RQ4-CMP	RQ1-CMP RQ2-CMP RQ5-CMP	RQ1-CMP RQ2-CMP RQ6-CMP	RQ1-CMP RQ2-CMP RQ7-CMP4	RQ1-CMP RQ2-CMP RQ8-CMP
Consistency (CNS)	RQ1-CNS RQ2-CNS RQ3-CNS	RQ1-CNS RQ2-CNS RQ4-CNS	RQ1-CNS RQ2-CNS RQ5-CNS	RQ1-CNS RQ2-CNS RQ6-CNS	RQ1-CNS RQ2-CNS RQ7-CNS	RQ1-CNS RQ2-CNS RQ8-CNS
Timeliness (TML)	RQ1-TML RQ2-TML RQ3-TML	RQ1-TML RQ2-TML RQ4-TML	RQ1-TML RQ2-TML RQ5-TML	RQ1-TML RQ2-TML RQ6-TML	RQ1-TML RQ2-TML RQ7-TML	RQ1-TML RQ2-TML RQ8-TML

[Source: own study]

The table above provides us with a formal set of requirements and guidelines for BP and IS analysts, designers and practitioners, which required to improve the information quality in the process, thus helping us achieve higher quality in the process design phase. Each of table cells contains the code number of requirements and guidelines needed for the identified DQ problems of each dependency relation type, based on generic catalogue of data quality requirements created in **Table 21**. Here are some examples to illustrate the role and benefits of **Table 32**.

The first example following the potential types of DQ problems as described at former step, grouped by accuracy dimension, i.e. DEF1-ACR, DEF2-ACR, DEF3-ACR, DEF4-ACR, and DEF6-ACR, based on **Table 31**. Respectively, the identified requirements of DQ for these types of DQ problems and deficiencies in this dimension, based on **Table 21**, are RQ1-ACR i.e. ensure a proper representation, reliable and free of error of any data items values in a given process; RQ2-ACR i.e. identify whether data item value from RW maps a wrong data item value in process representation and fix it and RQ3-ACR i.e. ensure a proper representation, reliable and free of error of data items values which are part of any information object in the process representation respectively; RQ4-ACR i.e. ensure a proper representation, reliable and free of error in sequence of appearance, mandatory of data items or in others dependent data items based on process paths in process representation, and RQ6-ACR i.e. ensure reliable and free of error in recorded data items values which are part of any formula calculation in process.

The next example following the potential types of DQ problems as described in a former step, relates to *formula calculation* dependency relation, denoted as rt_4. In this case, the identified potentials types of DQ problems are inter alia, DEF1-ACR and DEF1-CMP, grouped by accuracy and completeness dimensions as specified in **Table 31**. If one of data components is incorrect (accuracy) or is missing (completeness) these potential type of DQ problems, can appear. Respectively, the identified requirements of IQ for these types of DQ problems and deficiencies in accuracy dimension, based on **Table 21**, are RQ1-ACR i.e. ensure a proper representation, reliable and free of error of any data items values in a given process; RQ2-ACR i.e. identify whether data item value from RW maps to a wrong data item value in process representation and fix it, and RQ6-ACR i.e. ensure reliable and free of error in recorded data items values which are part of any formula calculation in process. In addition, the identified requirements of IQ for these types of DQ problems and deficiencies in completeness dimension, based on **Table 21**, are RQ1-CMP i.e. ensure the completeness of any data item values in process representation; RQ2-CMP i.e. ensure the completeness of RW data item values in process representation and complete the missing data item value where it's necessary in process, and RQ6-CMP i.e. ensure the completeness in recorded data items values which are part of any formula calculation in process.

4.2.4. Summary of evaluation results based on case study analysis

Our fourth goal was to develop a case study for validation based on BP from real-world environment to evaluate and demonstrate the completeness, utility and quality of our implemented DQDP method.

It is important to emphasize that during the implementation of our DQDP method on the above case study, we have identified a variety of DQ problems that will be presented below and that could have been avoided if our method had been implemented on this process. The main results and implications of case study validation in terms of the utility and efficacy of the implemented DQDP method are summarized in the following bullets.

(1) The evaluation clearly shows that processes in practice suffer from DQ problems as described in detail above in sub-chapter **3.4** (see p. 122) and there is a need to improve the quality of processes' outcomes. Basically, these DQ problems could float upwards after

the actual implementation of the process or IS. Such analysis at the design stage by using the DQDP method can help BP or IS analysts, designers and practitioners to predict potential failures in earlier stages, to fix these failures issues and prevent further DQ failures within the IS design process in advance. That's something that proves the utility of the DQDP method in real environment and its quality and efficacy, since we achieve effectiveness in the process execution and its outcomes.

(2) The role and importance of DQ dimensions in identifying DQ failures in the process was found to be significant here in improving the DQ. Along the process of our case study, the process' manager or operator is required in practice, to record a variety of data and information items required through a set of filled and uncontrolled forms or documents. In fact, some DQ deficiencies related to DQ dimensions like mistakes, missing data or inconsistencies in data recording were found between what appears in the original manual form in the process to what was recorded to the local information system. Furthermore, the process' manager has traced the causes of the problems that were preventing successful production of the required information.

For example, we sampled a customer's quotation (data item d_{24}) for export services. The original quotation (No. 40579) was issued to the customer on June 17, 2012, but no quotation confirmation by the customer (data item d_{25}) was found. This kind of DQ deficiency relates to incompleteness and timeliness DQ dimensions. An examination of this case revealed that the quotation status was not being monitored and there was no interaction with the customer, despite the fact that a sales person had spent time for preparing and probably the customer had chosen to perform the service through another company.

Another example relates to incompleteness and inaccuracy dimensions. In this case, the customer address (data item d_5) was fully specified in the manual form but in the reporting to the system it was partial (street and city only) and incorrect (workplace address), i.e., incompleteness and inaccuracy data. Another information problem was found when the total amount to be paid in the customer quotation appeared in US$, but it was recorded in the customer order in Israeli currency (NIS), i.e., inconsistency between the data in the original form and the stored data in local IS.

Another example relates to timeliness dimension. We sampled a customer's quotation (data item d_{24}) named Sigal Ben Ami (No. 18686) for export' customs brokerage and land transport services. The original quotation was issued to the customer on November 27, 2012 but the quotation was accepted by the customer (data item d_{25}) only a year later (on November 28, 2013). In fact, the new method creates continuous reporting of documents in the process and ensures full control over its implementation and progress.

(3) The role and importance of the dependency between data items in identifying DQ failures in the process and their relation types were also found to be significant in improving the information quality using our DQDP method.

For example, we sampled in a random a customer's order file. We checked the customer feedback form, designed to assess the level of customer satisfaction. We found that no customer details (data item d_4), customer order file number (data item d_{30}), and the date of the filled feedback form (data item d_{32}) were recorded and the form included a customer's signature only. The feedback form was filed in the customer file as is and was not reported in the system. If this customer decided to sue the company for negligence and/or poor level in the service she received, it would be difficult to prove her satisfaction with the service she received based on this form. Furthermore, closing a customer order file (data item d_{30}) depends on the existence of a complete customer feedback form with the customer details and the degree of satisfaction. Also, this kind of DQ deficiency relates to missing information item i.e. incompleteness dimension.

By the way, an examination of this case and the reason for this information failure revealed that the form was indeed attached to the customer file without being properly examined by the export coordinator due to lack of attention. Using our method could have prevented this in advance.

In another case, customer order file (No. 60678) for storage services was sampled. The customer's cargo was received in the company's warehouse on June 20, 2014 and until October 30, 2014, no invoice (data item d_{26}) was issued to the customer for down payments. Also, this kind of DQ deficiency relates to completeness and timeliness dimensions since the customer invoice was missing and was not issued to the customer within the time required by the organization's procedures.

In another case, we found another DQ problem related to accuracy dimension, of nonconformity in a customer quotation (No. 13168) for sea export services provided to the customer. The customer's original evaluation of cargo volume (data item d_{15}) was 11 CBM (cubic meters) and the customer received an initial quotation in the amount of $2600 and approved it. In practice, it was discovered that the original recorded volume was incorrect, and the correct volume measured was 17 CBM, i.e. a deviation or error of 6 CBM. The customer was required to pay an extra $850, but he refused to pay that extra, claiming he had already approved and signed the original quotation, and this error is the responsibility of the company salesman and was caused due to lack of attention.

Because the customer's cargo contents were already packed and stored at the company's warehouses and the customer refused to pay, the company had to delay the cargo shipment abroad. The result is that this customer's cargo did not arrive on time (at the original date set) at the destination abroad. Later, the customer filed a lawsuit against the company for negligence and for the damages he sustained, which caused an unnecessary financial expense.

This kind of DQ deficiency relates to accuracy dimension and to the dependency existing between cargo volume (data item d_{15}) and amount to pay derived from quotation (data item d_{24}). Since our DQDP method focuses on data items dependency considerations comprehensively and takes into account their relation type and their potential DQ problems, such a DQ problem could have been avoided by using the appropriates DQ aspects and requirements. The overall result is that the process did not achieve its original goals and we received other results from this process.

(4) The DQDP method includes important and useful matrices and set of reference tables as outputs, that enable BP or IS analysts, designers and practitioners to map the entire information flows in a given business process and to see a broader and more accurate "big picture" in advance and go down to the smallest details of data items flows in the business process. It enables identifying potential problems related to the DQ and enable us to predict the effects and impacts on the process level and adapt to DQ requirements in advance, thereby preventing failures.

Additionally, the method enables to identify the impact of those potential DQ deficiencies and problems in a business process in advance to improve its performances' quality. Failures in process design due to poor data quality can result, for example, to achievement of outputs and deliverables which are not part of the process goals and its objectives or can lead the process to be inefficient or in a deadlock situation in some cases.

(5) The DQDP method presents in detail organized potential DQ failures and requirements with respect to the most critical dimensions for DQ. These critical dimensions are part of our new method and used as a reference to identify and cope with their potential DQ failures to improve the quality of BP and IS design phase outputs. Furthermore, the implementation of DQDP method on the case study focuses on considerations and integration of DQ aspects and requirements into BP and IS analysis and design phases comprehensively, which are considered critical steps in the success of an IS project. Furthermore, the classifications of the dependency relation types identified between the various data items and their impacts on process outputs can help and direct the BP or IS experts for better focus on defining the potential failures of relevant IQ and requirements to address and prevent them in advance.

In summary, our new DQDP method focuses on data items dependency considerations comprehensively and takes into account their relation type and their potential DQ problems with respect to the most critical dimensions for DQ, which are considered critical steps in the success of BP and IS design. Based on the case study analysis described above and related examples presented, we can see the new contributions and capabilities of our new DQDP method in identifying IQ issues in a more systematic and organized manner and that all the requirements for the method were achieved. Furthermore, all of the above examples illustrate the usefulness and benefits of the DQDP method and the need to maintain quality in the process performance and its outcomes and by achieving effectiveness of process goal(s).

4.3. Evaluation of Proposed DQDP Method Based on Focus Groups

As part of a more comprehensive evaluation process of our new proposed DQDP method and to demonstrate its contribution, benefits and advantages over other existing methods, we also selected a *focus group* (FG) as a research method to evaluate and confirm the level of utility, efficacy and quality of our new DQDP method and to ensure its soundness, benefits and advantages.

The fifth goal of this dissertation was to evaluate and demonstrate the application of our new DQDP method in practice through a focus groups session and it is also an answer for our third research question.

Focus groups have the potential to provide great insights into phenomena of interest to IS researchers (Belanger, 2012). The focus group method is best suited to obtaining feedback from participants on new concepts, developing questionnaires, generating ideas, collecting or prioritizing potential problems, obtaining feedback on how models or methods of IS phenomena are presented, documented or are to be tested (Kontio et al. 2004; Belanger, 2012). The high level of interaction during a focus group study allows for deeper understanding on respondents' feedback, on the use of the artifact and on other issues influencing design (Brandtner et al., 2015). In addition, a focus group plays an important role in the design and development of IS artifact by showing a prototype to focus groups for feedback (Gregor, 2006; Belanger, 2012).

Furthermore, we made a distinction between two types of focus groups: **exploratory** and **confirmatory focus groups**, **EFG** and **CFG** in short respectively, in the design research process as illustrated in **Figure 32** (p. 107). We used the first type, i.e. EFG, for collecting and defining the artifacts requirements, and for the design and refinement of artifacts, while the second type, i.e. CFG, corresponds to the current phase, the evaluation stage i.e. to test and establish and demonstrate the utility of the artifact design in the application field use. For this purpose, we adopted here again, the steps of the procedural model of a focus group session (**Figure 33**), as described by Tremblay et al. (2010), and which explains how to apply and conduct focus groups especially in a design science research (DSR) context with respect to confirmatory focus groups (CFG) to evaluate the utility and usability of the new proposed DQDP method.

There is another aspect to consider here that is related to the nature of the evaluation and is based on the FEDS approach as explained in detail in sub-chapter **1.6.3** (p. 21). According to

Krueger & Casey (2014) focus group is considered the best and most appropriate method for gathering information for *formative* and *summative* evaluations. Since our evaluation is human oriented and its effect, it is necessary to examine the artifact in the real working environment with real users in their real-world context. Additionally, our research artifact was oriented to a functional purpose of the evaluation i.e. *summative* evaluation to determine the extent that the new artifact i.e. our DQDP method, match expectations and requirements and compatible with a naturalistic evaluation strategy which explores rigorously the performance of the artifact in its real-world environment with real people and real systems.

4.3.1. Establishing and conducting confirmatory focus groups (CFGs) sessions

Basically, the establishing and conducting confirmatory focus groups (CFGs) sessions are based on the procedural model steps of focus group session (**Figure 33**) presented by Tremblay et al. (2010) as mentioned above and extensively described in sub-chapter **3.3.2.2** (p. 107).

According to Tremblay et al. (2010), for design research, the FG participants should be from a population familiar with the application environment for which the artifact is designed so they can adequately inform the refinement and evaluation of the artifact. Care should be taken that the participant groups are from a similar pool for both EFGs and CFGs, so that CFGs are in fact confirming a final design (Tremblay et al., 2010).

The number of CFGs run in the field test depends on the consistency of results across the focus groups, the level of rigor required in the design research project, and the availability of resources needed to run additional CFGs (Tremblay et al., 2010). The unit of analysis is at the focus group and not the individual participants. Thus, it is crucial not to introduce any changes to the interview script and the artifact when multiple CFGs are conducted. This allows for the comparison of the results across CFGs to demonstrate and corroborate proof of utility of the artifact. Moreover, according to Tremblay et al. (2010) suggestion, at least two confirmatory focus groups (CFGs) should be run and conducted to achieve rigorous investigation and evaluation of the artifact with opportunities for quantitative and qualitative data collection and analyses across these CFGs.

For this purpose, we established four confirmatory focus groups (CFGs) with 29 participants in total, representatives from eleven different organizations operating in a variety of areas[8]. The participants from all organizations or companies did not know the other company participants beforehand. Furthermore, they were divided into two categories in general, to test and evaluate the utility and efficacy of our new DQDP method across different organizations. Two main focus groups consisted a mix of domain experts' participants from various occupations and senior practitioners' representatives who represent in principle, end-user requirements for any new IS project. The first one took place at company name *Elyakim Ben-Ari Ltd*, which considered Israel's leading company and veteran in civil engineering projects and quarries industry, and the second one took place at *Ocean group Ltd* as already described above (see sub-chapter **3.3.3**; p. 109), and which is considered Israel's leading company and veteran in international forwarding and moving industry.

The third and fourth CFGs took place in Zefat academic college and the Israel airports authority respectively. These two CFGs consisted a mix of BP and IS experts and data custodians' professionals from various organizations in various sectors of industry and service and considered as BP & IS with extensive experience.

Generally, the CFGs sessions were intended to achieve the following main objectives:

- Exposing CFGs participants to the new artifact i.e. our proposed DQDP method for professional feedbacks and its practical aspects.
- To demonstrate the structure and steps of proposed DQDP method from IQ aspects, problems and requirements.
- To demonstrate and evaluate the utility and usability of our artifact i.e. new proposed DQDP method in the application field by practical experience with the DQDP method by the FG participants and their feedbacks.
- To identify the challenges, obstacles and new directions that the organization is expected to face in adopting of our new method.

In principle, the CFG meeting included a distribution of a feedback form with an *open ended questions* for participants to fill at the end of session for evaluation purpose (see Appendix E).

[8] More details about the confirmatory focus groups (CFGs) participants and their profiles appear in Appendix F.

In addition, the CFG discussion was documented in principle using FG script template (see an example in Appendix G).

Results from the CFGs sessions can help us to achieve an objective assessment of the utility and efficacy of our new DQDP method, and accordingly identify opportunities for refining the method and facilitate a better understanding of its limitations. In fact, all the findings presented here and related to our new method are based on the feedbacks and opinions from variety of IS domain experts, e.g. BP and IS analysts and designers, software developers, designers, practitioners and experienced end users.

The following table (**Table 33**) presents the list of established CFGs and the characteristics of each of them.

Table 33. List of confirmatory focus groups (CFGs)

CFG #	Date	Place	Number of participants	Type of participants	The main purpose / topic
CFG1	13.3.19	Elyakim Ben-Ari Ltd	5	Domain experts & IS practitioners	• Exposing participants to the new artifact i.e. our proposed DQDP method as a solution to researched problem and research motivation, from scientific and business needs aspects. • To demonstrate the structure and steps of proposed DQDP method from IQ aspects, problems and requirements. • To demonstrate the utility of our artifact i.e. new proposed DQDP method. • To evaluate the utility and usability of the new proposed DQDP method in the application field. • The extent to which the new method fits into the organization. • The challenges that the organization is expected to face in adopting the new method.
CFG2	14.3.19	Ocean Group LTD	8	Domain experts & IS practitioners	
CFG3	20.3.19	Zefat Academic College	10	BP & IS experts	
CFG4	27.3.19	The Israel Airports Authority	6	BP & IS experts	

[Source: own study]

The following charts (**Figure 45**, **Figure 46** and **Figure 47**) present some descriptive statistics of the CFGs participants' profiles. **Figure 45** depicts the distribution of CFGs participants by profession. We can see in **Figure 45** that more than half of the CFGs participants (52%) are domain experts and IS practitioners (first category) and the second half of the CFGs participants are IS and BP professionals (second category). In addition, **Figure 46** shows that about 80 percent of participants possess over 6 years of experience.

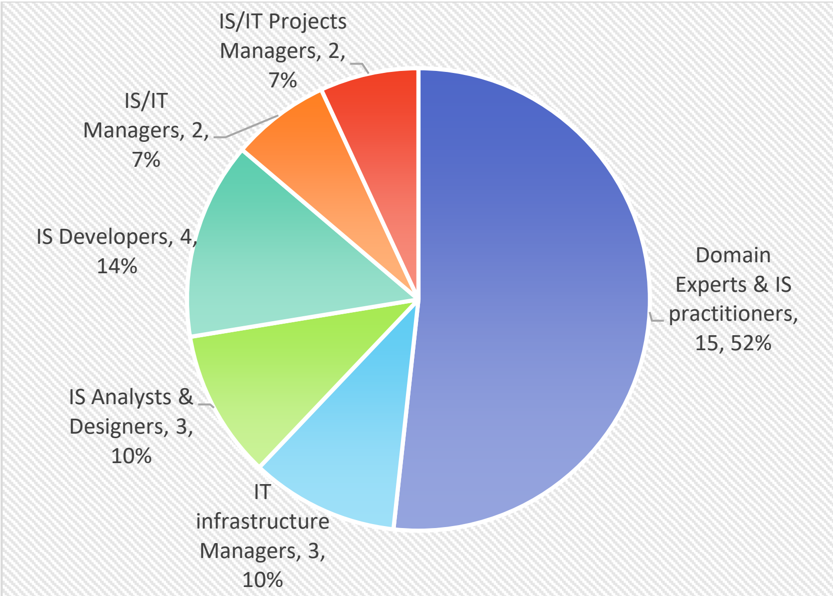


Figure 45. Distribution of CFGs participants by profession

[Source: own study]

Figure 46 depicts the distribution of CFGs participants by professional seniority.

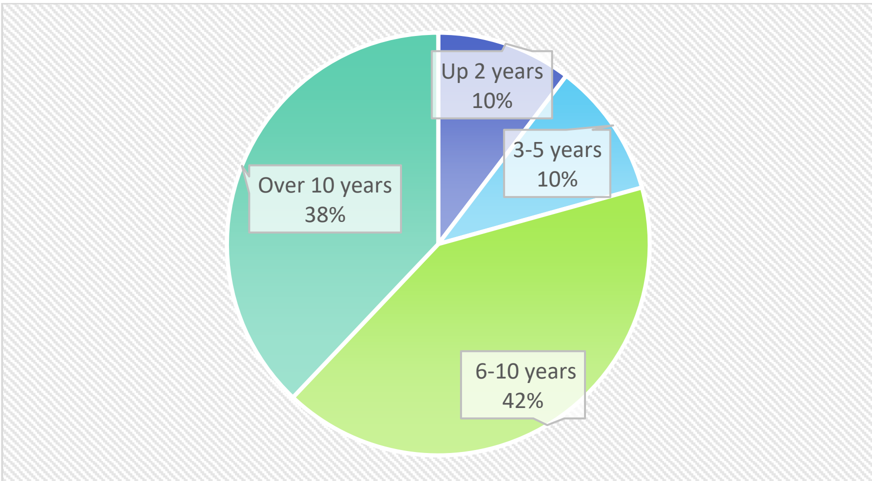


Figure 46. Distribution of CFGs participants by professional seniority

[Source: own study]

Figure 47 depicts the distribution of CFGs participants by organization sectors.

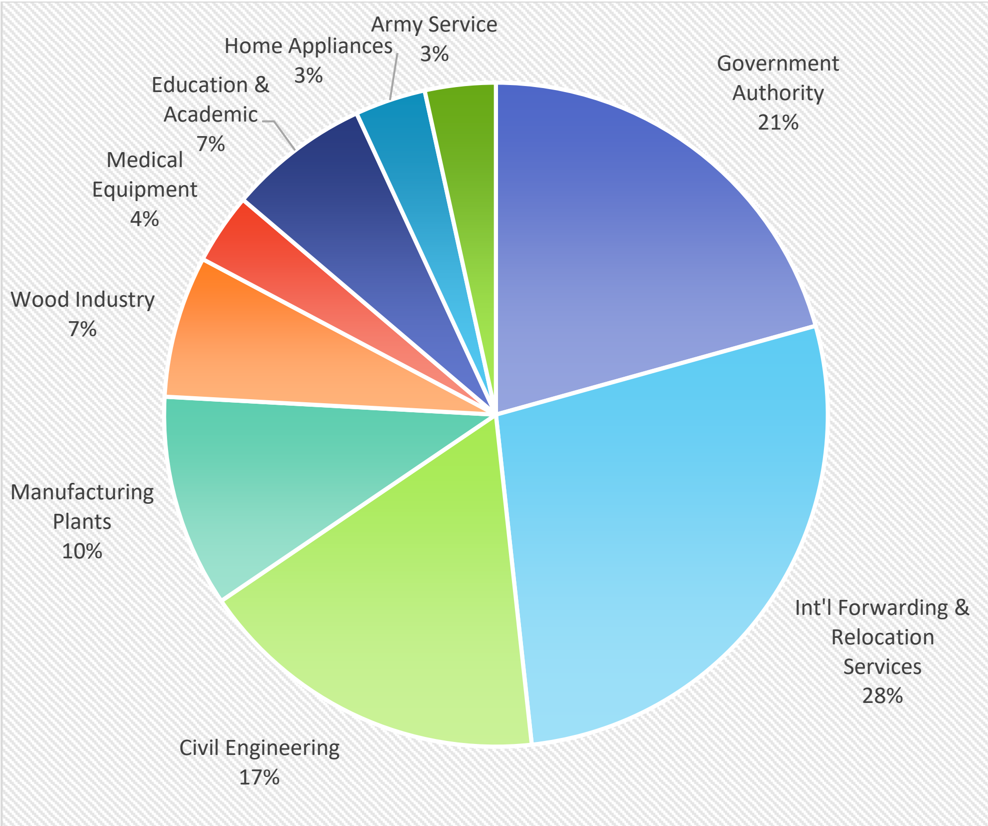


Figure 47. Distribution of CFGs participants by organization sectors

[Source: own study]

Table 34 as an example presents a summary of CFG script (CFG 3), which took place at Zefat Academic College[9]. Other focus groups followed a similar scenario.

Table 34. CFG 3 script summary

#	Component	Description
1	Place	Zefat Academic College
2	Date	20.3.19 (at 18:00-20:00 PM)
3	Number of Participants	10 Participants - IS and BP professionals (managers and employees) with many years of experience from various organizations
4	The CFG Topics / Purposes	• Exposing FG participants to the new artifact i.e. our proposed DQDP method for professional feedbacks and its practical aspects. • To demonstrate the structure and steps of proposed DQDP method from IQ aspects, problems and requirements. • To demonstrate the utility and usability of our artifact i.e. new proposed DQDP method in the application field. • To evaluate the utility and usability of the new proposed DQDP method through practical experience with the DQDP method by the FG participants and their feedbacks.

[9] The full script document of CFG 3 is available in Appendix G.

		• To identify the challenges and obstacles that the organization is expected to face in adopting of our new method. • To identify new directions and suggestions for development of the proposed method in the future.
5	The CFG Main Activities	**<u>The CFG discussion focused on four main activities:</u>** **1. Practical experience with the DQDP method by the FG participants:** o The FG participants were asked to analyze and characterize a mini process and define its requirements in IQ terms based on a preliminary case study. o The process specification should be in two stages: in the first stage, they were asked to define process specification according to any method they knew and without using the components and phases of the new method. In the second stage they were asked to create new specification for the same case study based on our new method steps. o The participants were asked to present their results' differences and their general opinion on the DQDP method. **2. General discussion about the utility of DQDP method through the following questions:** o The participants were asked to indicate how understandable and clear the new proposed DQDP method is. o The participants were asked to indicate their general impression from DQDP method. o Can the new method be suitable for use in your organization? o Would you recommend the new method to the relevant parties and people in the organization or to your professional colleagues? **3. General discussion about the main contributions and challenges in DQDP method adoption:** o The main contributions and benefits to the organization from implementing the new method. o The main challenges and obstacles that the organization is expected to face in adopting our new method. **4. General discussion about new directions and suggestions for development <u>the</u> proposed DQDP method in the future:** o Defining new directions for the development and improvement of the proposed method in the future.
6	The Main CFG Discussion Results, Feedbacks & Outcomes	**<u>The CFG discussion's main outcomes & results grouped by the above topics:</u>** **1. Practical experience with the DQDP method by the FG participants:** o The new method allows us to see the full picture of the process design in advance with emphasis on the information items and their dependencies. o There is a great need to identify potential failures related to DQ problem in advance to achieve high quality in process design. o The new method makes the analysis and design phase of a process more transparent, efficient and of higher quality. o DQ problems in any data item values in process (e.g. wrong or missing values) can cause the process to achieve undesired results in the process. o DQ problems may lead to or result in the wrong decisions in process. o Good decisions are based on high quality information and critical to process success. o The new method helps us to identify dependencies between different data items in advance which that are related and critical to achieve the process' goals and their potential failures in relation to DQ problem. o The new method enables to identify different paths of dependency relationships between data items in process in advance. o The new method helps us to identify potential failures related to a delay or missing pieces in receiving information items from internal or external source, which can lead the process to a deadlock situation since there is no full control in receiving the information from a third party source.

		o The new method helps us to identify in advance potential of a waste of time and valuable resources or to perform unnecessary tasks caused by DQ problems of any data item values in process. **2. General discussion about the utility of DQDP method through the following questions:** o The participants were asked to indicate how understandable and clear the new proposed DQDP method is. ✓ *The results clearly show that the majority of CFGs participants, a little more than 80%, indicated that the new proposed DQDP method is understandable to very understandable and clear (see also **Figure 48**).* o The participants were asked to indicate their general impression from DQDP method. ✓ *In principle, we received a wide range of answers and definitions for the proposed DQDP method, all of which are positive; e.g. very interesting, very comprehensive, perfect etc. (See details on page 193).* o The participants were asked to indicate if the new method can be suitable for use in their organizations and would they recommend it to the relevant parties in their organization or to other professional colleagues? ✓ *The results clearly show that the majority of CFGs participants, about 86%, indicated that the new method be suitable for use in their organizations and they would recommend to the relevant professional people and to other colleagues to implement the new DQDP method in their organization. In addition, some participants (14%) stated that they would recommend applying the new method under some conditions (see also Figure 50 on page 193).* **3. General discussion about the contributions, benefits, challenges and obstacles in DQDP method adoption:** o The main contributions and benefits to the organization from implementing the new DQDP method raised by the FG participants are: ✓ *The new method allows us to see the full picture of the process design in advance with emphasis on the data items and their dependencies.* ✓ *The new method allows us to identify potential failures related to DQ dimensions in advance to achieve high quality in process design.* ✓ *The new method makes the analysis and design phase of a process more transparent, efficient and of higher quality.* ✓ *The existing methods and techniques are inadequate and new methods or techniques to improve the process analysis and design are needed.* ✓ *The new method can help us to achieve good decisions based on high quality information and is critical to process success.* ✓ *The new method can help us in improving the information management and mapping the entire flows of information in the process and prevent duplication or shortages in data items in advance.* ✓ *The new method helps us to identify dependencies between different data items in advance which are related and critical to achieve the process' goals and their potential failures and problems related to DQ dimensions.* ✓ *The new method enables us to identify different paths of dependency relationships between data items in process in advance.* ✓ *The new method helps us to identify potential failures related to delay or missing information items upon receipt from internal or external source, which can lead the process to a deadlock situation since there is no full control in receiving of information from a third party source.* ✓ *The new method helps us to identify in advance potential of a waste of time and valuable resources or to perform unnecessary tasks caused by DQ problems of any data item values in process and to eliminate them.* ✓ *The new method allows us to see the direct connection and dependency between the analysis and design of BPs and IS design outputs.*

		○ The challenges and obstacles stated by the FG participants which the organization is expected to face in adopting our new method: ✓ *Appropriate training is required to deal with the complexity of the new method and to implement it successfully.* ✓ *It is necessary to incorporate in the implementation of the new method all the parties involved in the process while coordinating expectations between them.* ✓ *There is a need to redefine the existing processes and work procedures associated with the new method.* ✓ *There is a need for collaboration with relevant managers and employees and the ability to cope with employee objections to changes in work patterns.* ✓ *In some cases, it requires review and reassessment of existing information systems while addressing issues related to information quality and adapting to the requirements of the new method.* **4. General discussion about new directions and suggestions for development of the proposed DQDP method in the future:** ✓ *To automate the new proposed method in the future.* ✓ *Further to the method automation, receiving recommendations about dependencies, conflicts and potential failures related to the information quality in advance.*

[Source: own study]

After introducing the new method and giving the participants the opportunity to experiment with it, we asked all CFGs participants a set of general questions about the new method. First, we asked all the participants in the different CFGs to indicate how understandable and clear the new proposed DQDP method is?

The following chart (**Figure 48**) presents the distribution of the answers we received from all CFGs participants.

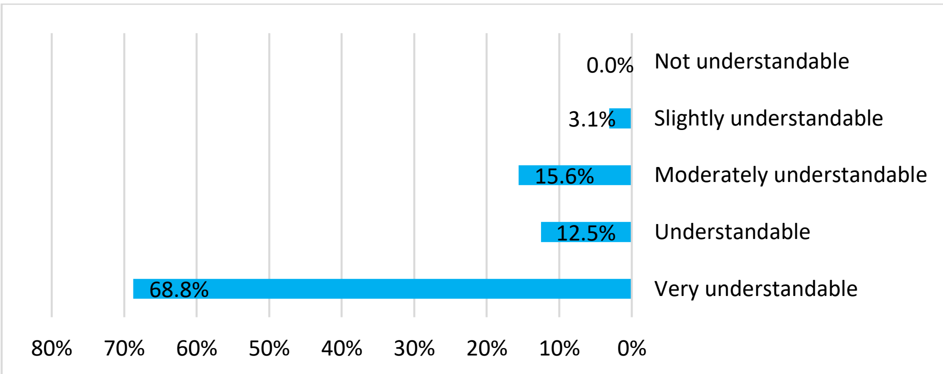


Figure 48. Distribution of CFGs participants' answers to the question "How understandable and clear the new proposed DQDP method is?"

[Source: own study]

Figure 48 clearly shows that the majority of CFGs participants, a little more than 80%, indicated that the new proposed DQDP method is understandable to very understandable and clear.

In addition, we asked all CFGs participants to indicate freely what do they think about the proposed method in general. **Figure 49** presents the distribution (in percentage), of the answers we received from all the CFGs participants to this question.

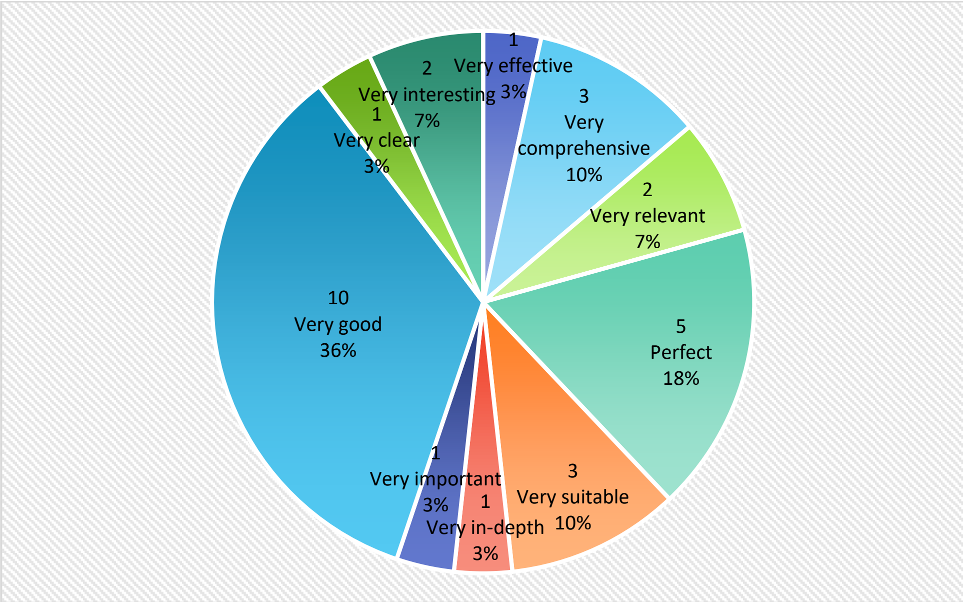


Figure 49. Distribution of CFGs participants' answers to the question "What do you think of the new method in general?"

[Source: own study]

Figure 49 clearly shows that all responses from the participants are positive regarding the use of the new proposed method without any exception and this is indicative of their higher degree of satisfaction.

Finally, we asked all CFGs participants to indicate whether they think the new method is recommended for implementation in their organization. The following chart (**Figure 50**) shows the distribution in percentage, of the CFGs participants' recommendation for implementing the method in their organization.

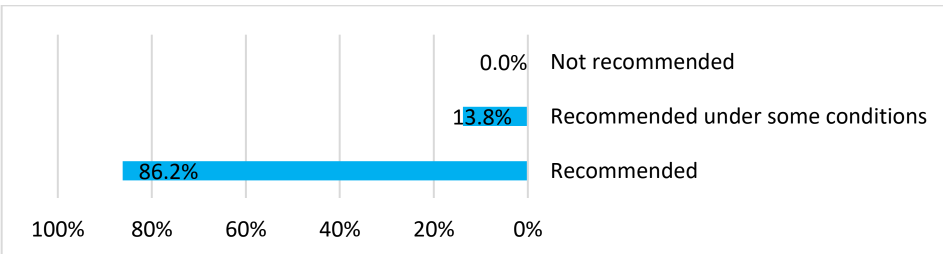


Figure 50. Distribution of CFGs participants' answers to the question "Would you recommend the new method?"

[Source: own study]

Figure 50 clearly shows that the majority of CFGs participants, about 86%, indicated that the new method be suitable for use in their organizations and they would recommend to the relevant professional people and to other colleagues to implement the new DQDP method in their organization. In addition, some participants (about 14%) stated that they would recommend applying the new method under some conditions. An examination of the matter reveals that in some organizations, and especially in the larger ones, there is a need to anchor this in the work procedures and / or to carry out pre-preparation activities among those involved in the organization in order to achieve a qualitative and complete implementation of the new method.

Furthermore, some CFGs participants noted that adopting the new method ensures on the one hand a higher quality of the analysis and design of the BP and the planned IS, but on the other hand the steps proposed in the new method are relatively long compared to the methods they are used to operate with and basically they are concerned that due to the pressure to advance the analysis and design of the new IS, they may not be given enough time to implement it fully.

4.3.2. Summary of evaluation results based on focus groups analysis

We selected a *focus group* as a research method to evaluate and demonstrate the utility, efficacy and quality of our new DQDP method and to ensure its soundness, benefits and advantages across a variety of organizations. Recall, our fifth goal was to evaluate and demonstrate the application of our new DQDP method in practice through focus groups sessions.

For this purpose, we established four confirmatory focus groups (CFGs) with a mix of domain experts' participants, BP and IS experts and data custodians' professionals and senior practitioners' representatives from various occupations, with extensive experience from variety of organizations who represent in principle, end-user requirements for any new IS project.

The collection of textual comments in addition to the quantitative Likert scales-based statements also proved to be helpful and provided us with further insight on our DQDP method and helped us understand participants' answer options and interpreting results. Furthermore, an in-depth analysis of the feedback forms and comments of the participants in all focus groups clearly and conspicuously highlighted the following main results.

(1) An overwhelming majority of confirmatory focus groups (CFGs) participants stated that the method presented was fully understood and considered to be an appropriate solution to problems (i.e. fitness for use) related to the IQ in analysis and design stages of BPs and information systems. In other words, this statement confirms the efficacy of the method since it produces its desired effect (i.e. achieved its goal).

(2) An analysis of feedback from participants of CFGs sessions shows that the majority of them (about 86%) have stated that the new DQDP method can be suitable for use in their organizations and they would recommend to the relevant professional people and authorities or to other colleagues to implement the new DQDP method in their organization as part of the work methodology in analyzing BPs and implementing IS/IT projects. In fact, these results demonstrate the utility that BP and IS analysts, designers and practitioners attribute to the new DQDP method and its usability i.e. the functionality and suitability to its use. However, some of participants recommend that it should be implemented in a gradual manner to reduce the resistance and uncertainty of senior designers, practitioners and project managers as a part of change management program in the organization.

(3) Most of the participants expressed a positive and sympathetic approach to the new proposed method, emphasizing in their feedback the degree of innovation and originality of the method for them. Moreover, they stated that BP and IS professionals can indeed benefit from the new method and improve the quality of analysis and design of BPs and ISs. In fact, these results demonstrate the potentials benefits and quality improvement that BP and IS analysts, designers and practitioners attribute to the new DQDP method implementation.

(4) Some of participants of BP and IS experts group stated that if this method was known and placed before them, they would have incorporated it into recent IS projects and completed them better and with a higher satisfaction of end-users.

(5) Additionally, some participants have stated that they are concerned that this kind of method can cause them to overload in the specification work and they estimate that it will also affect the overall budget framework of IS project and they are afraid that it is expected to grow. However, this is the framework of our proposed method and it may be worth considering in the future how to reduce its impact in this context.

4.4. Evaluation Based on Comparison with Other Existing Methods

Evaluation in DSR may also be concerned with a comparative evaluation of the new artifact with other artifacts to determine whether the new artifact is an improvement on the state of the art and also demonstrates more rigorous assessment of the effectiveness of the artifact (Prat et al., 2014; Venable et al., 2016).

In this section we implement this type of evaluation, i.e. a **comparative evaluation** of our proposed DQDP method in comparison with other exist methods. The goal is to compare our new method with the other existing ones, while reviewing the differences and focusing on their key limitations and weaknesses and to show the advantages and strengths of our DQDP method over the other existing ones.

Recall, in sub-chapter **2.3.6** we reviewed a variety of methods and approaches that emphasize and consider data quality aspects in the context of BPs design and summarized their advantages and limitations in **Table 11** (see p. 92). In the next table (**Table 35**) we present these approaches and methods with their main limitations and weaknesses, which are detailed in the above sub-chapter and mentioned table and compare them with the properties and advantages of our proposed method.

Table 35. Compare DQDP method with other methods and approaches - Summary

The author(s)	Main limitations and weaknesses	Advantages of our DQDP method
Sun et al. (2006) Sun & Zhao (2013)	• The focus is on workflow model only and commercial workflow systems oriented.	• The focus is on BP design phase in general and information systems oriented.
	• Lying on the assumption that the quality of requirements specification is fully exist.	• The new method helps to produce the relevant quality requirements.
	• The approach is applicable when a dataflow specification is well connected, complete and concise.	• The new method is applicable for BP design phase, helps to achieve completer and concise of dataflow specification.
	• The focus on data-flow anomalies but the analysis does not deal with accuracy dimension as a significant part of DQ requirements for processes.	• The new method focuses inter alia, on accuracy dimension as a significant part of DQ requirements for processes.
	• The researchers did not refer to the types of dependency relation.	• The new method focuses inter alia, on the types of dependency relation.
Soffer (2010)	• The research focuses on data inaccuracy problems only.	• The new method focuses on problems of four most important dimensions of information quality, inter alia, on accuracy dimension as a significant part of DQ requirements for BP design phase.

	• Workflow and process model oriented.	• The focus is on BP design phase in general and information systems oriented.
	• Ignoring from data dependency concept.	• The new method focuses on the dependency between data items and the potential DQ problems resulting from these dependencies.
Ofner et al. (2012)	• DQ dimensions are not used in the approach.	• DQ dimensions are part of our new method as a basis for DQ requirements at BP design phase.
	• The authors ignored from data items dependency considerations.	• The method focuses on the dependency between data items and the potential DQ problems resulting from these dependencies.
Caro et al. (2012)	• The methodology limited to business process modeling.	• The method focuses on BP design phase in general and information systems oriented.
	• The authors ignored from data items dependency considerations.	• The new method focuses on and takes into account the data items dependency considerations and their potential DQ problems.
	• The researchers did not refer to the dependency relation types.	• The new method focuses inter alia, on the types of dependency relation.
Falge et al. (2012; 2013)	• Ignoring from DQ dimensions and their potential problems in the method.	• DQ dimensions are part of our new method as a basis for DQ requirements at BP design phase.
	• Ignoring from data dependency concept and considerations.	• The new method focuses on and takes into account the data items dependency considerations and their potential DQ problems.
Cappiello et al. (2013)	• The methodology limited to business process modeling.	• The method focuses on BP design phase in general and information systems oriented.
	• Based on the assumption that the involved actors have a good knowledge of the analyzed BP.	• The method is for BP and IS analysts and designers.
	• The authors ignored from data items dependency considerations and their relation type.	• The new method focuses inter alia, on the types of dependency relation.
Glowalla & Sunyaev (2013a; 2014a)	• Focus on integration of data quality into process modelling languages only.	• The method focuses on integration of data quality aspects into BP and IS in general focused on design phase.
	• Requirements based on data quality dimensions not provided.	• The new method provided, inter alia requirements based on data quality dimensions.
	• The authors ignored from data items dependency considerations and their relation type.	• The new method focuses on and considers the data items dependency considerations, their relation type and their potential DQ problems.
	• The authors explicitly excluded processes that are inherent to IS/IT systems.	• The new method focuses on and takes into account considerations and requirements to cope with potential DQ problems with respect also to IS design.
Gharib et al. (2018)	• Focus on integration of data quality requirements into WFA-net modelling languages only.	• The method focuses on integration of data quality aspects into BP and IS in general focused on design phase.

	• The approach deals only with binary requirement satisfaction i.e., some requirement e.g. goal can be either satisfied or denied.	• The method presents the requirements in detail with reference to their potential failures.
	• The approach deals only with binary DQ requirement satisfaction i.e., information can be either accurate or inaccurate, believable or unbelievable, etc.	• The method presents the IQ requirements in detail with reference to their potential failures.
	• All DQ dimensions have the same priority and importance to the system.	• Our method does not address this issue.
	• The approach cannot deal with more than one BP at the same time.	• Our method does not address this issue.
	• Ignoring from data dependency aspects and their relation type.	• The new method focuses on and takes into account the data items dependency considerations, their relation type and their potential DQ problems.

[Source: own study]

4.4.1. Summary of evaluation results based on methods comparison

Based on the above comparison analysis table, we summarize the main advantages of our new DQDP method over other existing methods.

(1) DQDP method enables to map and manage the entire set of information flows in a business process. It includes important matrices and a set of tables as outputs, that enable us to see a broader and more accurate "big picture" in advance and go down to the smallest details of the BP, identifying potential problems related to the DQ and enable us to predict the effects and impacts on the process level and adapting DQ requirements in advance, thereby preventing failures.

(2) Furthermore, the new method focuses on considerations and integration of DQ aspects and requirements into BP and IS analysis and design phases comprehensively, which considered critical steps in the success of an IS project while most other methods and approaches have dealt with it from a BP Modeling perspective only.

(3) The DQDP method focuses on data items dependency considerations comprehensively and takes into account their relation type and their potential DQ problems while most other methods and approaches neglected the issue or deal with it in a limited or restricted way. Furthermore, the classifications of the dependencies' relation types identified between the various data items, their impact and the meanings derived from them in

terms of DQ deficiencies and requirements are innovative and unique to our method. In this way, BP or IS experts can focus on defining the potential failures of relevant DQ and requirements to address and prevent them in advance.

(4) The method enables to identify the impact of those potential DQ deficiencies and problems in a business process in advance to improve its performances' quality.

(5) The method presents in detail an organized list of potential DQ failures and requirements with respect to the most critical dimensions for DQ. These critical dimensions are a part of our new method and are used as a reference to identify and cope with their potential DQ failures to improve the quality of BP and IS design phase outputs.

(6) Finally, all the differences described above, including the weaknesses and limitations of the other methods and approaches versus the advantages and strengths of our new DQDP method, effectively illustrate the utility and benefits of the DQDP method over the other existing ones. Another thing, since the new method seeks to improve the quality of process' performance, its implementation is intended to debug process information quality failures to meet the highest quality of process goal(s) and its expected outcomes. Thus, it helps achieve a high level of process effectiveness and satisfaction of all process participants.

4.5. Summary of Evaluation Results

4.5.1. Evaluation results based on requirements fulfillment of DQDP method

Recall, in sub-chapters **3.3** to **3.5**, we discussed and identified a set of main requirements for fulfillment using a new DQDP method as an artifact and summarized it as a list in **Table 22** (p. 144). This list reflects the BP and IS analysts, designers and practitioners' expectations from the new DQDP method and it oriented to the quality of information in the BP design phase. For the convenience of the reader, we summarized in **Table 36** a status of requirements fulfillment of the DQDP method based on above three evaluation methods.

Table 36. A summary of requirements fulfillment status by the DQDP method

#	Evaluation method / The requirements	Case study	Confirmatory focus groups (CFGs)	Methods comparison
1	The ability of mapping and managing the entire of information flows in a process and to see the "big picture" in advance.	Achieved	Achieved	Achieved
2	The ability to reflect all dependencies that exist between information items in a process.	Achieved	Achieved	Achieved
3	Be able to identify the type of each dependency relation and their impact in a process.	Achieved	Achieved	Achieved
4	Be able to identify potential DQ failures related to dependency and its type between data items in a process.	Achieved	Achieved	Achieved
5	Be able to identify the potential DQ problems in a process grouped by DQ dimensions in advance.	Achieved	Achieved	Achieved
6	Be able to identify in advance the impact of potential DQ problems in a process on its performance and quality in advance.	Achieved	Achieved	Achieved
7	The new method should enable BP and IS analysts, designers and practitioners to identify relevant DQ requirements for each potential DQ failure.	Achieved	Achieved	Achieved
8	A new method should help BP and IS analysts, designers and practitioners to improve the quality of analysis and design outputs of BPs and IS.	Achieved	Achieved with some reservations	Achieved

[Source: own study]

Table 36 clearly shows that almost all the above stated requirements and expectations of BP and IS analysts, designers and practitioners for the new proposed DQDP method were achieved. However, there is one exception relating to the assessment of the expected improvement of the quality of analysis and design outputs of BP and IS following the implementation of the DQDP method. In this case, some reservations were raised by several CFGs participants as presented in the above findings.

4.6. Evaluation Results in Terms of Research Thesis

Recall, our research stated thesis assumed that the new suggested method improves the quality of BP design and helps BP and IS analysts and designers focusing on potential failures of data dependencies and their impact on quality requirements at an earlier stage of the IS design, than currently existing methods in the DQ domain.

The evaluation results and findings were based on three well-known research methods in DSR, taken from the existing knowledge base and proved beyond any doubt that our stated thesis was achieved. Furthermore, these results clearly show that our proposed method improves the quality of BP and IS design and helps BP and IS analysts, designers and practitioners focusing on potential failures of data dependencies and their impact on quality requirements at an earlier stage of the IS design, than existing methods in the DQ domain. The majority of the DQ deficiencies and problems presented along the research and as a result of existing dependencies between different data items in BPs can actually disappear or can be prevented in advance with the implementation of the new method in the early stages of analysis and design of BP and IS.

Additionally, our proposed method has been shown as better when compared to other existing methods. The comparative findings of our proposed method with respect to the other existing methods in our field clearly show that our method relates to new capabilities and components associated with DQ analysis that have been neglected or not extended by other existing methods.

Most of FG participants (especially those considered to be experts in BP and IS) showed a sympathetic and positive approach to this method and stated that they would recommend implementing it in their organizations. Moreover, regarding our original stated thesis, most participants' feedbacks supported it and indicated that our new DQDP method is novel as the thesis stated and considered more improved and wider than the other currently existing methods in DQ domain.

4.7. Summary

Evaluation of design artifacts is a central key activity and critical part of design science research (DSR), as it provides feedback for further development and requires researchers to rigorously demonstrate the utility, efficacy and quality of a design artifacts using well-executed evaluation methods.

Our research evaluation based on FEDS approach since it help to DSR researchers to design and improve their DSR evaluation activities. According FEDS approach there is two key goals of evaluation in DSR which are relevant here: (1) the utility aspect of the artifact in the environment i.e. its effectiveness, and (2) also the quality of the knowledge contributed by the construction of the artifact. Generally, designed artifacts are often components of a human-machine problem solving system as stated by Hevner et al. (2004). Furthermore, based on these evaluation goals, we found that *Human Risk and Effectiveness strategy* suggested by Venable et al. (2016), is more appropriate to our case and adopted it, since the major design risk is relates to people environment and user oriented i.e. the necessity to evaluate the new method rigorously with real users i.e. BP and IS analysts, designers and practitioners in their real domain's context to ensure that the utility/benefit of the new method will continue in real situations. Also, it concerns to knowledge sharing and corporate for collecting the problems, requirements and feedbacks from them.

Basically, the new DQDP method was evaluated summative as a functional purpose, i.e. it's done after the design artifact was developed to judge the extent that the outcomes match expectations. The evaluation was also based on a naturalistic paradigm since it conducted using a real process i.e. case study in a real organization and focus groups sessions, facing real problems to explore the performance of the DQDP method in a real environment within an organization. Furthermore, the evaluation process and its results and findings were based on three well-known research methods in DSR, taken from the existing knowledge base: case study analysis, CFGs sessions and a comparative evaluation of our proposed DQDP method in comparison with other previously existing methods.

As part of evaluation process of the new proposed DQDP method, the implementation of the method along its steps on chosen case study and its summary is described. Furthermore, the evaluation of proposed DQDP method based on CFGs sessions and a comparative evaluation to demonstrate its advantages over other existing methods also described.

Finally, all the above evaluation methods prove both the utility of the artifact in its environment and the transfer of knowledge to others and it is the proof for achieving the five research goals and the answer of the three research questions. In addition, they are intended to approve beyond any doubt that our stated thesis was achieved.

5 Discussion

DSR is inherently a problem-solving process. DSR implements IS artifacts within an application context (e.g., a business organization and/or process) for the purpose of improving the effectiveness and efficiency of that context (Hevner & Chatterjee, 2015). Furthermore, the correctness, effectiveness, and efficiency of the BPs supported by ISs are becoming vital to the organization. Poor DQ may be a cause of problems and losses related to organizational processes. Poor information quality may be a cause of problems and losses related to organizational processes and information systems design.

This dissertation proposed an original method, name Data Quality Deficiencies Prediction (DQDP) - a novel method uniquely suited to use in process and as a conceptual model for dealing with the DQ issue within BPs and ISs design.

We examined and evaluated the new DQDP method according to FEDS approach which is suitable for DSR research. Recall, FEDS approach focuses on the two key purposes of evaluation in DSR which are relevant here: (1) the utility aspect of the artifact in the environment i.e. effectiveness, and (2) also the quality of the knowledge contributed by the construction of the artifact. The evaluation strategy is fit to *Human Risk and Effectiveness* strategy category since the major design risk is user oriented i.e. concern to knowledge sharing and corporate for collecting the problems, requirements and feedbacks from BP and IS analysts, designers and practitioners within their real application context.

In fact, both purposes mentioned above were achieved: the utility aspect of the method has been checked in the real environment. The evaluation entails choosing the goals and set of requirements of the method that are to be subject to evaluation. In our case, the method goals and requirements have been checked in manner of set of criteria such as utility, efficacy, usability and quality, and its potential improvements in manner of DQ dimensions and their impacts on overall quality of BP and IS design. Basically, our analysis focuses on four DQ dimensions; accuracy, completeness, timeliness and consistency that are considered in the DQ literature to be the most critical and important dimensions for DQ and cope with kind of DQ deficiencies derived from the RW domain state or in process representation with regard to BP and IS design stage.

Rigorous in DSR is achieved by appropriately applying existing foundations and methods. Hence, the evaluation process and its results and findings were based on three well-known research methods in DSR, taken from the existing knowledge base: case study analysis, confirmatory focus groups (CFGs) sessions and a comparative evaluation of our proposed DQDP method in comparison with other previously existing methods.

The first method we used to evaluate our DQDP method is case study analysis which is widely used in DQ research domain. Case study implemented the artifact in a real-world situation to evaluate not only its utility, but also its efficacy and quality effects on its business environment. We conducted a case study to evaluate the applicability of the DQDP method for process exploration and improvement focusing on DQ aspects based on central process of international forwarding and moving industry. The results of evaluation based on case study analysis clearly show that processes in practice suffer from DQ problems. Basically, these DQ problems could have arisen after the implementation of the process or the information system. Such analysis at the design stage can help BP or IS analysts and designers to predict potential failures in earlier stages, to fix these failure issues and prevent further DQ failures within the IS design process.

Additionally, our DQDP method includes a set of matrices and important tables as outputs, that enable us to see a broader and more accurate picture and go down to the smallest details of the BP, identifying potential problems related to the information quality and enabling us to predict the effects and impacts on process level and adapting DQ requirements in advance, thereby preventing failures.

The second method we used to evaluate our DQDP method is a focus group (FG) which is considered as the best and most appropriate method for gathering information for summative evaluation. We conducted four confirmatory focus groups (CFGs) sessions with participants from variety of organizations and functions in various specialties.

An analysis of feedback from participants in the focus groups sessions shows that most of them have stated that they see a significant improvement in the process in terms of process quality, response speed and information quality in the process. An overwhelming majority of confirmatory focus groups (CFGs) participants stated that the method presented was fully understood and considered to be an appropriate solution to problems related to the information quality in analysis and design stages of BPs and ISs.

The majority of CFGs participants (about 86%) have stated that the new DQDP method can be suitable for use in their organizations and they would recommend it to the relevant professional people and authorities or to other colleagues to implement the new DQDP method in their organization as part of the work methodology in analyzing BPs and implementing IS/IT projects. Furthermore, based on the analysis of FG and feedbacks results, the new method is more improved, wider and has new capabilities than the other existing approaches and methods in DQ domain, since it covers DQ dimensions considerations' by adding DQ requirements based on data dependency and by using identified dependency relation types at the earlier stage in process design and/or IS implementation project.

However, the analysis of feedback from participants also shows that some of them have noted that the method is relatively broad and complex, and they will need to implement it in a gradual manner. Other participants indicated that there is a small risk of non-cooperation from professionals and senior users, and expressed concern that the new method needs to change their work patterns and procedures. For this reason, and in order to mitigate the BP and IS analysts' and designers' work, we have built a set of tables as a complementary foundation to our DQDP method (see the above **Table 17** to **Table 20** ; pp. 134-138) with a list of potential problems related to information quality grouped by the critical dimensions of DQ. This is so that they will be able to more quickly identify the potential problems in the candidate process design and choose from them only the relevant problems for this process. Furthermore, we combined the outputs of both sub-steps, 4.1 and 4.2 (of the proposed method) into one output in the dependency matrix i.e. *DM matrix* (see **Figure 43**; p. 170), thereby saving unnecessary work time. These results reflect the usability of our DQDP method i.e. the ability to perform the intended tasks and the ability of BP and IS analysts, designers and practitioners to effectively use the method.

The third method of evaluation is based on the analysis of methods comparison. Generally, there are other methods and approaches presented in literature, which attempt to deal with and improve DQ aspects and figure out how to detect and ensure DQ aspects into BP analysis and design and along IS design stages. However, these methods often do not address the original source of these problems and most of them ignored data dependency aspects and their important role in achieving BPs design that is more robust.

The advantages of the DQDP method can be clearly seen when comparing it to existing methods and approaches. DQDP method has several main advantages: first, it focuses on integration of DQ requirements into BPs analysis and design outcomes and takes in account data dependency considerations into DQ requirements in BP design for BP and IS designers and practitioners. Second, the ability to identify potential problems in advance related to IQ in the various data items by identifying the dependencies between them and the relation types of dependence is a significant added value in this method that cannot be found to the best of our knowledge in other methods.

Third, while the existing methods solve part of the problems related to the information quality, our new method expands and complements them in other additional information quality aspects, since it covers DQ dimensions considerations' by adding DQ requirements. Furthermore, our research expanded the idea of data dependencies by analyzing and developing new dependency relation categories. Based on data dependency and by using identified dependency relation types at the earlier stages of BP design and/or IS projects to achieve high quality results for all parties in the end.

In principle, we can see clearly that all the stated requirements for the method were achieved (summarized in **Table 36)** based on the well-known methods for evaluation with some exceptions as analyzed and described in former chapters. The proposed DQDP method is novel and is considered to be improved, wider, and has new capabilities than the other existing methods in the data quality domain, since it cover DQ dimensions' considerations by adding DQ requirements based on data dependency and by using identified dependency relation types at the earlier stage in BP design and/or IS implementation project.

The majority of the problems mentioned above in the context of information quality and as a result of existing dependencies between different data items can actually disappear or can be prevented in advance with the implementation of the DQDP method in the early stages of analysis and design of BP and IS.

If our new method is not implemented or will be delayed in its implementation, it is expected that the amount of problems can be extremely high in the context of the information quality and as a result of existing dependencies between the various data items. Moreover, it can cause unnecessary financial expenses and directly or indirectly affect the quality of performance and/or time delay of process design and information systems projects.

Part IV: Research Conclusions and Outlook

6 Conclusions Implications and Contributions

The final part summarizes the dissertation and provides the main conclusions and outlook for this research. In addition, it describes the research limitations and challenges while reviewing the most pressing open issues that should be tackled for future researches.

6.1. Research Conclusions

This research focused on the relationship between business process design and information quality management. Specifically, it was concerned with designing a new method for integrating DQ requirements into BPs design phase based on DQ dimensions and data dependency aspects. Such a method is important because it can assist BP and IS analysts, designers and practitioners in achieving high quality of IS/IT implementation projects by improving the BPs analysis and design stages as a source of requirements and surface for new IS design and development.

The main research problem, and in fact its main motivation and challenge, is that BPs and IS projects in practice can suffer from a poor level of DQ along their design activities and they are expected to fail and achieve undesired and poor outputs. Furthermore, poor DQ has negative impacts on the performance of an organization. Most of the organizations depend on the quality of data and information resources for everyday business operations and for decision making and they are aware of IQ problems and their costs, but in fact, they do not do enough to eliminate them from the organization's environment and reduce its influence on them. Hence, successful accomplishment of IS projects is a crucial challenge and remains a timeless challenge for BP and IS researchers and practitioners.

Basically, this dissertation dealt with five main research goals and three research questions. In principle, we can declare that all the goals of the dissertation were achieved, and all research questions were covered respectively. Now we will expand it with a proven explanation about each of them.

The first research goal was to identify a set of possible IQ aspects, problems and requirements that impact BP design quality. In Chapter 2 we reviewed many sources and studies from the literature that dealt with various aspects and constructs related to information quality and its related dimensions in general, and their role and contribution in improving BP analysis and design stages in particular. In sub-chapter **3.4** we were more deeply involved in the problem by identifying and collecting many practical problems from the real-world, based on the background we reviewed in Chapter 2, and through exploratory focus groups (EFGs) sessions. For this purpose, we established three EFGs with 24 professional and experienced participants in total, representatives of ten different organizations operating in a variety of areas. The EFGs aimed to explore the research domain in depth as well as collect DQ problems and deficiencies and to identify the information needs, DQ requirements and expectations of all those participants involved in the domain processes. The focus groups sessions included an open discussion, based on open ended questions and semi-structured interviews that enabled a combination of group and personal discussion. Then, we translated them into a set of DQ potential deficiencies and set of requirements related to information quality, grouped by main DQ dimensions in order to prepare an infrastructure for developing our new method from the existing knowledge base that adds value to BP and IS analysts and designers.

The second research goal was to design and present a conceptual model for data quality assessment and BP design combined, to summarize the central constructs and terms and their relationships (by using UML class diagram). In Chapter 3 we discussed and designed a conceptual model, its constructs and their relationships. The conceptual model (also known as meta-model) was aimed for identifying the relations of different data constructs within the BP. This kind of conceptual model also enabled and supported us later in identifying and defining the relevant aspects and constructs of our new method based on DQ dimensions for creating the DQ requirements underlying our proposed method.

The third goal was to design and develop a new method for information quality assessment and to predict data quality deficiencies and potential failures in BPs design and prevent them in advance. In Chapter 3 we developed and presented an original and novel method, named Data Quality Deficiencies Prediction (DQDP) for information quality assessment and predicting data quality deficiencies and potential failures in BPs design and prevent them in advance. The new developed DQDP method is designed for BP and IS analysts, designers and practitioners, takes into account data quality deficiencies and flaws, and issues from the real-world and mix it with data dependency issue and generate from a formal catalogue a set of requirements and guidelines (see **Table 21**; p. 140). Furthermore, we evaluated the new DQDP method based on FEDS approach. The evaluation process and its results and findings were based on three well-known research methods in DSR, taken from the existing knowledge base: case study analysis, confirmatory focus groups (CFGs) sessions and a comparative evaluation, of our proposed DQDP method in comparison with other previously existing methods.

The fourth goal was *to* develop a case study from the real world and collect a set of information quality requirements to verify and validate the utility of the implemented new DQDP method based on validity criteria in practice. In sub-chapter **4.2** we developed and presented the case study analysis for implementing and evaluating our proposed new method. The developed case study based on sales process of the sea export service in real-world company operating in the international forwarding and moving industry. Then, we applied our new DQDP method on the case study, to demonstrate and evaluate its applicability and generality in organization, as well as its ability to improve DQ problems in BP and IS design in practice.

The main reasons for choosing this process as a case study are because it is considered one of the most complex processes in the international forwarding and moving domain. Also, it is considered a rich BP in data items since it is a critical and popular service in the international forwarding industry. Moreover, this process contains a large collection of dependencies between information items and many aspects and considerations of dependencies must be taken into account since they are critical and significant to the process success as a whole, and therefore the importance of the components of the information quality in such a process is very high.

The fifth and last goal was to evaluate and demonstrate the application of the new DQDP method in practice through confirmatory focus groups (CFGs) sessions. In sub-chapter **4.3** we established and conducted four CFGs sessions for evaluation of our proposed new method with IS/IT professional and experienced participants from eleven different organizations operating in a variety of areas and evaluate the validity criteria in practice. The focus groups sessions included open discussion, based on open ended questions and semi-structured interviews that enabled a combination of group and personal discussion.

In this case, an overwhelming majority of participants stated that the presented method is an appropriate solution to problems related to the information quality in analysis and design stages of BPs and ISs. Furthermore, most of the participants emphasized in their feedbacks the degree of innovation and originality of the method for them and stated that the new DQDP method can improve the quality of analysis and design of BPs and ISs. They have stated also that the new DQDP method can be suitable for use in their organizations and they would make a recommendation to relevant professional people and authorities or to other colleagues to implement the new DQDP method in their organization. Furthermore, they also stated that if the DQDP method is not implemented or will be delayed in its implementation, it is expected that the amount of problems in the context of the information quality and as a result of existing dependencies between the various data items can be extremely high. Moreover, this delay may cause unnecessary financial expenses and directly or indirectly affect the quality of performance and/or time delay of process design and information systems projects.

However, part of the participants stated that they are afraid that this will cause them to overload in the specification work and they estimate that it will also affect the overall budget framework of IS project and they afraid that it is expected to grow. Another part of participants recommend that it should be implemented in a gradual manner to reduce the resistance and uncertainty of senior designers, practitioners and project managers as a part of a change management program in the organization.

We also demonstrated the utility and efficacy of our new DQDP method based on the analysis of methods comparison. All the differences described above, including the weaknesses and limitations of the other methods and approaches versus the advantages and strengths of our new DQDP method, effectively illustrate the usefulness and benefits of the DQDP method over the other existing ones. Basically, our new method expands and complements other existing methods and approaches in other aspects of information quality, i.e. it covers DQ dimensions considerations' based on the idea of data dependency and by categorizing them into identified dependency relation types and generating DQ requirements at the earlier stages of BPs design and/or IS projects to achieve high quality results for all parties in the end.

Finally, all the above evaluation methods, their steps and actions taken are designed to prove both the utility, efficacy and quality of our artifact in its environment and the transfer of knowledge to others. In addition, they are intended to prove beyond any doubt that our stated thesis was achieved. Recall, our thesis *was "The new suggested method improves the quality of business process design and helps BP analysts and designers focusing on potential* *failures of data dependencies and their impact on quality requirements at an earlier stage of the information systems design than currently existing methods in the data quality domain.*

Moreover, based on the above discussion, it is clearly demonstrated that almost all the stated requirements and expectations of BP and IS analysts, designers and practitioners from the new proposed DQDP method were fully achieved. However, there is one exception relating to the assessment of the expected improvement of the quality of analysis and design outputs of BP and IS following the implementation of the DQDP method. In this case, some reservations were raised by several CFGs participants as presented in the above findings.

6.2. Summary of Research Deliverables

This research explored the linkage between business process design and data quality requirements. It aimed to focus on the importance of considering information quality aspects into business process design in analysis and design stage. To the best of our knowledge, these two issues have not yet been discussed sufficiently, and still there is a lack of systematic approaches, methods and models with regard to data quality.

Furthermore, this research aimed to develop a conceptual model and a new method (DQDP) as an artifact for BP and IS analysts, designers and practitioners, based on analyzing the impacts of information quality dimensions on quality of BPs design. In addition, it aimed to generate a set of information quality requirements for BP design success at analysis and design stage and therefore the above-mentioned method is considered a novelty.

The conceptual model aimed to describe the general idea and the relationships among constructs of data quality requirements with business processes design in order to achieve the defined process goals. The main artifact in our research is the DQDP method which guided BP and IS analysts, designers and practitioners how to take into account and enter data quality requirements to BPs design stage in order to check and validate them, to achieve high level of quality and their defined goals. Furthermore, the novel DQDP method is considered more improved and wider than the other existing methods in data quality domain as the thesis suggests. While the application of the existing methods solves part of the problems related to the information quality, our new method actually expands and complements them in other aspects of the information quality, since it covers DQ dimensions considerations' by adding DQ requirements based on data dependency and by using identified dependency relation types at the earlier stages of BPs design and/or IS projects to achieve high quality results for all parties in the end.

Table 37 summarizes the artifacts that result from this dissertation.

Table 37. Artifacts resulting from this dissertation - Summary

Artifact	Description
Constructs	As the general background and concept formation, Chapters 1, 2 and 3 provided the basis and foundations of DQ domain include *vocabulary, attributes, terminology and principles* etc., used in this dissertation. They provided also clear definitions, classifications etc., and discussions of established concepts and principles, i.e. quality aspects of BPs design, DQ dimensions i.e. accuracy, completeness, consistency and timeliness, which are considered most critical, DQ models and methods for BP design which detailed in sub-chapter 2.3.5. etc.
Model	Based on the introduced constructs, Chapter 3 described a *conceptual model*, a high-level description of research problem scope, expressing relationships among BP and DQ constructs, i.e. entities and their relationships and constraints by using UML class diagram, to summarize the terms and constructs discussed. The conceptual model used for identification of relations of different data constructs within BP domain. In addition, it draws the general picture, defines assumptions and prepared the foundations for the *DQDP* method as a main artifact.
Method	Chapter 3 presented the *Data Quality Deficiencies Prediction (DQDP)* method developed in this dissertation as a main research artifact. The method is developed based on analysis of DQ deficiencies and set of requirements which detailed in sub-chapter 3.4, collected from real-world instantiations and exploratory focus groups (EFGs) sessions with IS and domain experts. Finally, it presented formally as a set of ordered steps to be implement by BP and IS analysts, designers and practitioners.
Instantiations	In order to verify whether the proposed new method is able to solve the proposed research problem and its fitness for use, sub-chapter 3.4 and Chapter 4 present the performed evaluation of the DQDP method, using set of instantiations, an examples and scenarios from real-world, with case study and focus group sessions evaluation.

[Source: own study]

6.3. Contributions and Implications of the Dissertation

Effective design-science research must provide clear contributions in the areas of the design artifact. According to Hevner et al. (2004) and their DSR guidelines, it holds the potential for three types of research contributions: The design artifact itself i.e., enable the solution of unsolved problems in the environment; foundations i.e., extend and improve the existing knowledge base of the domain, and/or design evaluation knowledge methodologies i.e., based on using evaluation methods. One or more of these contributions must be found in a given research (Hevner et al., 2004).

The following contributions and implications of the dissertation are both, theoretical and practical.

6.3.1. Theoretical contribution and extension for BP and IS literature

The alignment aspects between BPs and their support ISs is a central component of Information systems analysis and design (ISAD) domain, and remains a challenging ambition for many researchers, focusing on how improve it by seeking and developing new models and methods for this purpose. Furthermore, the correctness, effectiveness, and efficiency of the BPs supported by ISs are becoming vital to the organizations and considered as an ongoing research challenge as well. However, the literature of BP and IS design domains lacks new studies, methodologies, methods and models for dealing with quality of information flows on BPs, as already discussed in detail above, especially regarding the theoretical framework and methods for gathering DQ requirements to improve DQ within BP and IS design phases.

For these purposes, there are the two most significant theoretical extensions and contributions for BP and IS literature and for these domains' knowledge base from our research. First, is the concentration on the linkage between IQ and BP design combined and the importance of IQ requirements on BPs design success; later it will be a basis for IS requirements. These requirements might help us deal with information flow quality in terms of process design and validation activities and reflect the quality requirements from IS. Consequently, the contribution is by creating a theoretical foundations and conceptual framework for IQ, in order to analyze and validate processes in an organization, based on evaluation.

The second significant theoretical contribution is by developing and proposing an original and new artifact, a method name Data Quality Deficiencies Prediction (DQDP) to fill the lack of methods for dealing with DQ aspects and deficiencies within BPs and ISs design projects. The new DQDP method takes into account DQ deficiencies and flaws issues from the real-world and mixes them with data dependency issues to generate a formal catalogue of requirements and guidelines based on DQ dimensions for BP and IS analysts, designers and practitioners. Moreover, it helps them to achieve high quality performance and deliverables at the analysis and design stages, which is considered a significant milestone in BPs and ISs implementation projects.

The third significant theoretical contribution of our research concerns its rigor. The rigor is derived from the effective use of the knowledge base i.e., theoretical foundations and research methodologies. Design-science research (DSR) relies upon the application of rigorous methods in both the construction and evaluation of the design artifact. The construction of the conceptual model and the new DQDP method are based on aspects, methods and instantiations of information quality derived from the literature review and based on exploratory focus groups (EFGs) sessions.

The evaluation phase and its results were based on three well-known research methods in DSR: case study analysis, confirmatory focus groups (CFGs) sessions and a comparative evaluation of our proposed DQDP method in comparison with existing methods. These methods are considered rigorous evaluation methods taken from the existing general knowledge base. Additionally, the evaluation process was anchored throughout the components of FEDS approach, a novel evaluation framework uniquely suited to use and improve the DSR evaluation to demonstrate and confirm the utility, efficacy, usability and quality of our new DQDP method.

6.3.2. Practical contributions and implications

Generally, the practical contributions and implications of the dissertation are followed by our research motivation and problem, and are mostly targeted for BP and IS analysts, designers and practitioners, but also for organizations in general.

The following bullets summarize the most significance practical contributions and implications that our research achieved:

- The Data Quality Deficiencies Prediction (DQDP) method developed for BP and IS analysts, designers and practitioners to improve the analysis and design stages of BPs as a source of requirements and surface for IS design, and to achieve high quality in IS/IT development projects implementation in an organization.
- In addition to the previous point, organizations often need to self-examine and adapt their BPs and their structure in response to changes, trends and developments in the business environment. Poor design can lead the process to fail and achieve undesired and poor outputs not in the process goals set in terms of data items perspective, or result in a deadlock situation. In addition, poor DQ has a negative impact on the performance of BPs and thereby the success of organizations. The new method helps to reduce the negative impacts and achieve a robustness and validation of BPs design.
- Information is a critical asset and essential resource for business success in the information age and enables organizations to create competitive advantages. Hence,

BP and IS analysts, designers and practitioners, can use the new DQDP method to eliminate DQ deficiencies in BPs and for predicting potential failures at the process design phase, while examining the effect of dependencies among different data values on these failures based on DQ dimensions in the design of a BP.

- ❖ DQ is one of the most important elements of the data governance. Moreover, data quality is critical to organizational success, for achieving strategic and operational business goals and for improved decision-making and it is the responsibility of the organizations' managers to ensure this over time. The new method helps them to achieve these goals.
- ❖ Many organizations are suffering from poor data quality along their BPs and ISs as already discussed above, and they are starting to realize that poor DQ is hurting them. Furthermore, poor DQ increases operational cost in organizations because time and other valuable resources are spent detecting and correcting errors in BPs and ISs. Hence, the most significant practical implication and contribution of this research for BP and IS analysts, designers and practitioners, and for organizations in general, is to give them the ability to design new BP and IS at a high-quality level in an effectively and efficiently manner.
- ❖ IS are pervasive today in all forms of business organizations and provide operational and managerial information to managers and end users to support their decision-making processes. Hence, the information flows in BPs and ISs become a critical issue and they should be examined. Moreover, achieving high-quality data is an important competitive factor and persistent challenge for enterprises and managers and the new method helps them to cope with this goal.
- ❖ IS projects are complex and risky projects for managers and are inclined to be challenged or failed inter alia by a low level of DQ. Successful accomplishment of IS projects is a crucial and persistent challenge for BP and IS analysts, designers and practitioners. Hence, the practical implications and contributions of this research are giving BP and IS analysts, designers and practitioners, a usable method, intended to identify potential failures in BPs as a surface for IS at analysis and design stages and create DQ requirements based on DQ dimensions to conduct IS projects at a high quality level. In this context, another practical contribution is achieving higher quality

in IS requirements specification by using qualified process design, features and components.

6.4. Research Limitations and Recommendations for Future Researches

Despite the proposed DQDP method of this research, it is innovative and adds new capabilities over existing methods for BP and IS practitioners and professionals, however, some limitations are derived from this research but there are also some challenges and recommendations for future researches. In this context, it is appropriate to quote here from Gregor & Hevner (2013; p. 1), who claim that "…the development of a particularly novel artifact with high utility will be seen as a contribution to knowledge, even if the full understanding of why the artifact works is partial and incomplete".

Challenges concerning supporting the robustness of BPs and ISs design still exist. In this context, we have several recommendations for further research and challenges. First, our research analysis focused on four main dimensions of DQ in identification of DQ deficiencies and requirements and in an attempt to explore other deficiencies arising from an analysis of other DQ dimensions. Second, our research and the proposed method focused more on the process at the design phase. Further studies can use the ideas and method suggested here to examine BPs also at their run-time phase.

The third recommendation concern to evaluation of our proposed DQDP method: this method evaluated inter alia, based on one case study which implies that the empirical results of our study should be generalized with caution. Hence, our recommendation for further research is to implement the new method in more domains i.e. validate and evaluate it in other domain case studies to provide a broader basis for scientific conclusions and generalization. Fourth, in principle our analysis and validation focused on qualitative research aspects using qualitative research methods i.e. based on a case study and focus groups analysis. Hence, our recommendation for further research is to implement the new method and analyze and validate it using quantitative research methods.

Furthermore, one significant challenge that emerged from the CFGs discussions is the need to simplify the complexity of the method to make it easier for BP and IS to use it. The proposed recommendation in the future is to automate the new proposed method through the construction of a tool, computer program or suitable algorithm and make the process easier. Such a tool or computer program should present to the end user structures of dependencies between data items in the process, the matrices, identify conflicts and receive recommendations about potential failures related to the DQ in advance etc.

Further research directions could be adopting our DQDP method into information technology environments such as data warehouse (DWH), data mining, on line analytical processing (OLAP), extract, transform and load (ETL) tools, business intelligence (BI) etc. which are considered today as important sources of data, information and knowledge needs high quality data to mine, but data of sufficient quality is often lacking. In the Big Data era, DQ faces many challenges and high-quality data are the precondition for analyzing and using big data and for guaranteeing the value of the data so it can be another interesting area to enhance our research foundations and method.

Another new research direction relevant today concerns the integration of DQ issue with the requirements of global data protection regulation (GDPR). Any organization with a single customer, employee, or other party living in the EU is required to ensure high-quality data in its organizational repositories and databases. Further researches, new methods or models to support and ensure accuracy, completeness or consistency of its data are still required to help them in activities and explanation of what data they have; where they locate it; how to correct it; explain where they got it; and, if requested, delete it, otherwise, they are exposed to lawsuits or face potentially huge fines. Our method can help also in this case, as it is concerned, inter alia, to ensure the accuracy and completeness of the data, while closely controlling and monitoring of data repositories and databases.

References

1. Aalst, W.M.P, van der, (2004). "Discovering Coordination Patterns using Process Mining", In: Bocchi, L., & Ciancarini, P., (Eds), ***First International Workshop on Coordination and Petri Nets (PNC 2004)***. STAR, Servizio Tipografico Area della Ricerca, CNR Pisa, Italy, 2004, pp. 49-64.

2. Aalst, W.M.P, van der, (2005). "Business alignment: Using process mining as a tool for Delta analysis and conformance testing", ***Requirements Engineering***, Vol. 10, pp. 198-211.

3. Aalst, W.M.P, van der, (2008). "Challenges in business process analysis", ***Enterprise Information Systems: 9th International Conference, ICEIS 2007***, Funchal, Madeira, June 12-16, 2007. Berlin: Springer. (Lecture Notes in Business Information Processing; 12). pp. 27-42.

4. Aalst, W.M.P, van der, (2009). "Process-Aware Information Systems: Design, Enactment and Analysis", ***Wiley Encyclopedia of Computer Science and Engineering***, John Wiley & Sons, pp. 2221-2233.

5. Aalst, W.M.P, van der, (2013). "Business Process Management: A Comprehensive Survey", ***ISRN Software Engineering***, Vol. 2013, Article ID 507984, pp. 1-66.

6. Aalst, W.M.P, van der, Hofstede, A.H.M. ter, & Weske, M., (2003). "Business process management: a survey", in: ***Proceedings of the International Conference on Business Process Management (BPM 2003)***, 26-27 June, Eindhoven, The Netherlands.

7. Abate, M.L., Diegert, K.V., & Allen, H.W., (1998). "A Hierarchical Approach to Improving Data Quality", ***Data Quality***, Vol. 4, No. 1, Sep. 1998. Available at: http://www.dataquality.com (14.3.14).

8. Abbasi, A., Sarker, S., & Chiang, R.H.L., (2016). "Big Data Research in Information Systems: Toward an Inclusive Research Agenda", ***Journal of the Association for Information Systems (JAIS)***, Vol. 17, No. 2, Feb. 2016, pp. 1-32.

9. Abdullah, N., Ismail, S.A., Sophiayati, S., Sam, S.M., (2015). "Data Quality in Big Data: A Review", ***Int. Journal of Advance Soft Computing*** *and Its* ***Applications (IJASCA)***, Vol. 7, No. 3, Nov. 2015, pp. 16-27.

10. Abramowicz, W., Filipowska, A., Kaczmarek, M., & Kaczmarek, T., (2007). "Semantically enhanced Business Process Modelling Notation", In: Hepp, M., Hinkelmann, K., Karagiannis, D., Klein, R., Stojanovic, N., (eds.): Semantic Business Process and Product Lifecycle

Management. in: ***Proceedings of the Workshop SBPM 2007***, Innsbruck, April 7, 2007, CEUR Workshop Proceedings, pp.1-4.

11. Ackoff, R., (1989). "From data to wisdom", ***Journal of Applied Systems Analysis***, Vol. 16, pp. 3-9.
12. Al-Mashari, M., & Zairi, M., (1999). "BPR implementation process: an analysis of key success and failure factors", ***Business Process Management Journal***, Vol. 5, No. 1, pp. 87-112.
13. Alotaibi, Y., & Liu, F., (2017). "Survey of business process management: challenges and solutions", ***Enterprise Information Systems Journal***, Vol. 11, No. 8, pp. 1119–1153.
14. Alshazly, A.A., Elfatatry, A.M., & Abougabal, M.S., (2014). "Detecting defects in software requirements Specification", ***Alexandria Engineering Journal***, Vol. 53 (2014), pp. 513–527.
15. Alter, S., (2002). ***Information Systems***, Pearson Education International, Prentice Hall, New Jersey.
16. Anupindi, R., Chopra, S., Deshmukh, S., Mieghem, J., & Zemel, E. (1999). ***Managing Business Process Flows***, Prentice Hall.
17. Auster, E., & Choo, C.W., (1996). ***Managing information for the competitive edge***, Neal-Schuman Publishers Inc., New York, NY.
18. Arachchi, S.M., Chong, S.C., & Madhushani, A.G.I., (2015). "Quality Assurance and Quality Control in ERP Systems Implementation", ***American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)***, Vol. 11, No 1, pp. 70-83.
19. Bagchi, S., Bai, X., & Kalagnanam, J., (2006). "Data Quality Management using Business Process Modeling", in: ***Proceedings of the IEEE International Conference on Services Computing***, pp. 398-405.
20. Bai, L., Meredith, R., & Burstein, F., (2018). "A data quality framework, method and tools for managing data quality in a health care setting: an action case study, ***Journal of Decision Systems***, Vol. 27, No. 1, pp. 144-154.
21. Ballou, D.P., & Pazer, H.L., (1985). "Modeling Data and Process Quality in Multi-Input, Multi-Output Information Systems", ***Management Science Journal***, Vol. 31, No. 2, INFORMS, pp. 150-162.

22. Ballou, D.P., Wang, R.Y., Pazer, H., & Tayi, G.K., (1998). "Modeling information manufacturing systems to determine information product quality", ***Management Science Journal***, Vol. 44, No. 4, pp. 462–484.

23. Baškarada, S., (2014). "Qualitative Case Study Guidelines", ***The Qualitative Report***, Vol. 19, No. 40, pp. 1-18. Available online at: https://nsuworks.nova.edu/tqr/vol19/iss40/3 (5.3.19).

24. Baškarada, S., & Koronios, A., (2014). "A Critical Success Factor Framework for Information Quality Management", ***Information Systems Management***, Vol. 31, Issue 4, pp. 276-295.

25. Batini, C., Francalanci, C., Cappiello, C., & Maurino, A., (2009). "Methodologies for Data Quality Assessment and Improvement", ***ACM computing surveys***, Vol. 41, No. 3, pp. 1- 52.

26. Batini, C., & Scannapieca, M., (2006). ***Data Quality- Concepts, Methodologies and Techniques***, Springer-Verlag Berlin Heidelberg.

27. Batini, C., & Scannapieca, M., (2016). ***Data and Information Quality, Data-Centric Systems and Applications,*** Springer International Publishing, Switzerland.

28. Belanger, F., (2012). "Theorizing In Information Systems Research Using Focus Groups", ***Australasian Journal of Information Systems***, Vol. 17, No. 2, pp. 109-96.

29. Bevan, N., (1999). "Quality in use: meeting user needs for quality", ***Journal of Systems and Software***, Vol. 49, No. 1, pp. 89-96.

30. Bevan, N., & Bogomolni, I., (2000). "Incorporating user quality requirements in the software development process", Serco Ltd and IAI.

31. Bhatt, G.D., (2000). "Exploring the relationship between information technology, infrastructure and business process re-engineering", ***Business Process Management Journal***, Vol. 6, No. 2, pp. 139-163.

32. Bhatt, G.D., & Troutt, M.D., (2005). "Examining the relationship between business process improvement initiatives, information systems integration and customer focus: an empirical study", ***Business Process Management Journal***, Vol. 11, Issue 5, pp. 532-558.

33. Bhatt, G.D., & Emdad, A.F., (2010) "An empirical examination of the relationship between information technology (IT) infrastructure, customer focus, and business advantages", ***Journal of Systems and Information Technology***, Vol. 12, No. 1, pp. 4-16.

34. Biehl, M., (2007). "Success factors for implementing global information systems", ***Communications of the ACM***, New York, Jan 2007, Vol. 50, Issue 1, pp. 52-58.

35. Blaine, J.D., & Cleland-Huang, J., (2008). "Software Quality Requirements: How to Balance Competing Priorities", ***IEEE Software***, March/April 2008, pp. 22-24.

36. Blake, R., & Mangiameli, P., (2011). "The Effects and Interactions of Data Quality and Problem Complexity on Classification", ***Journal of Data and Information Quality (JDIQ)***, February 2011, Vol. 2, Issue 2, pp. 1-28.

37. Bocij, P., Greasley, A., & Hickie, S., (2015). ***Business Information Systems: Technology, Development and Management for the E-Business***, 5th edition, Pearson Education International, Prentice Hall, England.

38. Brandtner, P., Helfert, M., Auinger, A., & Gaubinger, K., (2015). "Conducting focus group research in a design science project: Application in developing a process model for the front end of innovation", ***Systems, Signs & Actions: An International Journal on Information Technology, Action, Communication and Work practices***, Vol. 9, No. 1, pp. 26-55.

39. Brandtner, P., Helfert, M., Auinger, A., & Gaubinger, K., (2016). "Evaluation of Design Artifacts - more than prototypes and case studies". ***AIS SIGPRAG Pre-ICIS Workshop 2016***. pp. 1-12.

40. Cai, L., & Zhu, Y., (2015). "The Challenges of Data Quality and Data Quality Assessment in the Big Data Era", ***Data Science Journal***, Vol. 14, No. 2, pp. 1-10,

41. Cao, L., & Zhu, H., (2013). "Normal accident: Data quality problems in ERP-enabled manufacturing", ***ACM Journal of Data and Information Quality***, Vol. 4, Issue 3, Article 11.

42. Cappiello, C., Francalanci, C., & Pernici, B., (2004). “Data quality assessment from the user’s perspective”, in: ***Proceedings of the 2004 international workshop on Information quality in information systems***, Paris, France, ACM Press, pp. 68-73.

43. Cappiello, C., Caro, A., Rodríguez, A., & Caballero, I., (2013). "An approach to design business processes addressing data quality issues", in: ***Proceedings of the 21st European Conference on Information Systems (ECIS2013)***, June 5-8, 2013, Utrecht, The Netherlands, pp. 1-12.

44. Cappiello, C., Pernici, B., & Villani, L., (2014). "Strategies for Data Quality Monitoring in Business Processes", in: ***Proceedings of the 15th International Workshops on Web Information Systems (WISE 2014)***, October 12 - 14, 2014, Thessaloniki, Greece, In: Benatallah B. et al. (eds). Lecture Notes in Computer Science, Springer, Cham, Vol. 9051, pp. 226-238.

45. Caro, A., Rodríguez, A., Cappiello, C., & Caballero, I., (2012). "Designing Business Processes able to satisfy Data Quality Requirements", in: ***Proceedings of the 17th International Conference on Information Quality (ICIQ)***, Paris, France. Nov. 16-17, 2012, pp. 31-45.

46. Cecez-Kecmanovic, D., Kautz, K., & Abrahall, R., (2014). "Reframing Success and Failure of Information Systems: a Performative Perspective", ***MIS Quarterly***, Vol. 38, No. 2, pp. 561–588.

47. Chandrasekaran, S., Gudlavalleti, S., & Kaniyar, S., (2014). "Achieving success in large complex software projects", ***McKinsey & Company***, July 2014, pp. 1-5.

48. Chintalapati, S.S., Prasad, Ch.G.V.N., Vineela, R., & Sowjanya, J., (2014). "Process mining: Business Intelligence (BI) and Business Process Management (BPM)". ***International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE)***, Research Paper, Vol. 4, No. 5, May 2014, pp. 774-780. Available online at: www.ijarcsse.com (10.4.16).

49. Chung, L., Nixon, B., You, E., & Mylopoulos, J., (1999). ***Non-Functional Requirements in Software Engineering***, Kluwer Academic Publishers, Boston.

50. Ćirić, Z., & Raković, L., (2010). "Change Management in Information System Development and Implementation Projects", ***Management Information Systems***, Vol. 5, No. 2, pp. 23-28.

51. Creswell, J.W., (2014). ***Research design: qualitative, quantitative, and mixed methods approaches***, 4th ed., SAGE Publications Inc., California, USA.

52. Davenport, T. H., & Prusak, L., (1998). ***Working knowledge: how organizations manage what they know***, Boston, MA: Harvard Business School Press.

53. Davenport, T.H., & Short, J.E., (1990). "The New Industrial Engineering: Information Technology and Business Process Redesign", ***Sloan Management Review***, Vol. 31, No. 4, pp. 11-27.

54. DAMA Dictionary, (2017). ***The Dictionary of Data Management***, 2nd Edition, DAMA International.

55. DAMA-DMBOK2 Framework - Guide Knowledge, (2014). Available at: https://www.dama.org/sites/default/files/download/DAMA-DMBOK2-Framework-V2-20140317-FINAL.pdf (5.3.19).

56. Daoudi, F., & Nurcan, S., (2007). "A benchmarking framework for methods to design flexible business processes", ***Software Process: Improvement and Practice***, Vol. 12, No. 1, pp. 51-63.

57. DeLone, W.H., & McLean, E.R., (1992). "Information systems success: The quest for the dependent variable", ***Information Systems Research***, Vol. 3, No. 1, pp. 60-95.

58. 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***, Spring 2003, Vol. 19, No. 4, pp. 9–30.

59. Donzelli, P., & Bresciani, P., (2004). "Improving Requirements Engineering by Quality Modeling – A Quality-Based Requirements Engineering Framework", ***Journal of Research and Practice in Information Technology***, Vol. 36, No. 4, November 2004, pp. 277-294.

60. Dumas M., La Rosa, M., Mendling, J., & Reijers H.A., (2018). ***Fundamentals of Business Process Management***, 2nd Ed., Springer, Berlin, Heidelberg.

61. Dumbleton, J., & Munro, D., (2015). "Global Data Quality Research - Discussion Paper 2015", ***Experian Data Quality***, pp. 1-20.

62. Dzepina, A., & Lehner, F., (2018). "Business Process Model Quality – Results from A Structured Literature Review", in: ***Proceedings of 26th European Conference on Information Systems (ECIS 2018)***, Portsmouth, UK, pp. 1-12.

63. Earl, M., (1997). "Information equity: getting value out of information", Executive Reports, ***Centre for Research in Information Management***, London Business School, London.

64. Edwards, R., & Holland, J., (2013). ***What is qualitative interviewing?*** , Bloomsbury Academic. pp. 2-3.

65. English, L.P., (2000). ***Achieving enterprise wide data quality***, Trillium Software System.

66. English, L.P., (2001). "Information Quality Management: The Next Frontier", ***ASQ - Annual Quality Congress Proceedings***, Charlotte, NC. Vol. 55, No. 0, May 2001, pp. 529-533.

67. Eppler, M.J., & Helfert, M., (2004). "A classification and analysis of data quality costs", in: ***Proceedings of the 9th International Conference on Information Quality (MIT ICIQ-04)***, Cambridge, Massachusetts. November 5-7, 2004.

68. European Freight Forwarders Association (EFFA), (2015). Available at: https://www.effa.com (24.11.15).

69. Even, A., & Shankaranarayanan, G., (2007). "Utility-Driven Assessment of Data Quality, ***ACM SIGMIS - The DATABASE for Advances in Information Systems Journal***, Vol. 38, No. 2, pp. 75-93.

70. Ewusi-Mensah, K., (1997). "Critical Issues in Abandoned Information Systems Development Projects", ***Communications of the ACM***, September 1997, Vol. 40, No. 9, pp. 74-80.

71. Experian, (2015). "The data quality benchmark report," ***Experian Data Quality***, Tech. Report, January 2015.

72. Experian, (2016). "Building a business case for data quality", ***Experian Data Quality***, Tech. Report, Available at: https://www.edq.com/globalassets/white-papers/building-a-business-case-for-data-quality-report.pdf (20.12.18).

73. Falge, C., Otto, B., & Österle, H., (2012). "Data Quality Requirements of Collaborative Business Processes", ***45th Hawaii International Conference on Systems Science (HICSS-45 2012)***, Grand Wailea, Maui, HI, USA, IEEE Computer Society, pp. 4316 - 4325.

74. Falge, C., Otto, B., & Österle, H., (2013). "Towards a Strategy Design Method for Corporate Data Quality Management", in: ***11th International Conference on Wirtschaftsinformatik***, 27th February – 01st March 2013, Leipzig, Germany, pp. 801-815.

75. Falk, T., Griesberger, P., Johannsen, F., & Leist, S., (2013). "Patterns for business process improvement-a first approach", in: ***Proceedings of the 21st European Conference on Information Systems (ECIS 2013).*** Utrecht, The Netherlands, pp. 151-163.

76. Filipowska, A., Kaczmarek, M., Kowalkiewicz, M., Zhou, X., & Born, M., (2009). "Procedure and guidelines for evaluation of BPM methodologies", ***Business Process Management Journal***, Vol. 15, No. 3, June 2009, pp. 336-357.

77. Firican, G., (2018). "What is Data Governance?", *in* DATA GOVERNANCE, JAN 29, 2018. Available at: http://www.lightsondata.com/what-is-data-governance/ (5.3.19).

78. Fisher, C., Chengular-Smith, I., & Ballou, D., (2003). "The impact of experience and time on the use of data quality information in decision making", ***Journal of Information Systems Research***, Vol. 14, No. 2, pp. 170–188.

79. Fisher, C., & Kingma, B., (2001). "Criticality of data quality as exemplified in two disasters", ***Journal of Information Management***, Vol. 39, pp. 109–116.

80. Foidl, H., & Felderer, M., (2019). "Risk-based data validation in machine learning-based software systems". In: ***Proceedings of the 3rd ACM SIGSOFT International Workshop on Machine Learning Techniques for Software Quality Evaluation (MaLTeSQuE '19)***, August 27, 2019, Tallinn, Estonia, pp. 1-6.

81. Forbes, (2017). "Poor-quality data imposes costs and risks on businesses". Available at: https://www.forbes.com/sites/forbespr/2017/05/31/poor-quality-data-imposes-costs-and-risks-on-businesses-says-new-forbes-insights-report (22.10.18).

82. Forster, F., (2006). "The Idea behind Business Process Improvement", ***BP Trends***. Available at: http://www.bptrends.com/ (26.12.15).

83. Francalanci, C., & Pernici, B., (2004). "Data quality assessment from the user's perspective", in: ***Proceedings of the 2004 international workshop on Information quality in information systems***, Paris, France, Communications of the ACM, pp. 68 – 73.

84. Frank, A.U., (2008). "Analysis of dependence of decision quality on data quality", ***Journal of Geography Systems***, Vol. 10, pp. 71–88.

85. Gaikema, M., Donkersloot, M., Johnson, J., & Mulder, H., (2019). "Increase the success of Governmental IT-projects", Systemics, Cybernetics and Informatics, Vol. 17, No. 1, pp. 97–105.

86. Gartner Research Group, (2006). "Gartner's Position on Business Process Management", ***Gartner, Inc.,*** Stamford, Connecticut.

87. Gartner Research Group, (2008). "Data Stewardship Case Studies Show Increased Organizational Focus on Data Quality and Information Governance", ***Gartner, Inc.,*** Stamford, Connecticut.

88. Gartner Research Group, (2009). "Key Issues for Business Process Management", ***Gartner, Inc.,*** Stamford, Connecticut.

89. Gartner Research Group, (2010). "Leading in Times of Transition: The 2010 CIO Agenda", EXP Premier Report Jan. 2010, ***Gartner, Inc.,*** Stamford, CT.

90. Gartner Research Group, (2011). "Measuring the Business Value of Data Quality", Oct. 2011, ***Gartner, Inc.***, Stamford, CT.,

91. Gartner Research Group, (2013a). "Successful business and IT transformation". ***Gartner, Inc.***, Stamford, CT.

92. Gartner Research Group, (2013b). "The State of Data Quality: Current Practices and Evolving Trends". ***Gartner, Inc.***, Stamford, CT.

93. Gartner Research Group, (2016). "Magic Quadrant for Data Quality Tools", Nov. 2016, ***Gartner, Inc***., Stamford, CT.

94. Gharib, M., & Giorgini, P., (2015). Modeling and reasoning about information quality requirements in business processes. In: ***Enterprise, business-process and information systems modeling***. Springer, Berlin, pp. 231-245.

95. Gharib, M., Giorgini, P., & Mylopoulos, J., (2018). "Analysis of information quality requirements in business processes, revisited", ***Requirements Engineering***, June 2018, Vol. 23, Issue 2, pp. 227-249.

96. Giaglis, G.M., (2001). "A Taxonomy of Business Process Modelling and Information Systems Modelling Techniques", ***International Journal of Flexible Manufacturing Systems***, Vol. 13, No. 2, pp. 209-228.

97. Glinz, M., (2007). "On Non-Functional Requirements", in: ***Proceedings of the 15th IEEE International Requirements Engineering Conference***, Delhi, India. IEEE Xplore, Conference: Requirements Engineering Conference, 2007. RE '07. 15th IEEE International, pp. 21-26.

98. Glowalla, P., Balazy, P., Basten, D., & Sunyaev, A., (2014). "Process-driven Data Quality Management – An Application of the Combined Conceptual Life Cycle Model". in: ***Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS 2014)***, Big Island, Hawaii, USA, January 2014, IEEE, pp. 4700-4709.

99. Glowalla, P., & Sunyaev, A., (2013a). "Process-driven Data Quality Management through Integration of Data Quality into Existing Process Models", ***Business & Information Systems Engineering (BISE)***, Dec. 2013, Vol. 5, Issue 6, pp. 433-448.

100. Glowalla, P., & Sunyaev, A., (2013b). "Managing Data Quality with ERP Systems - Insights from the Insurance Sector", in: ***Proceedings of the European Conference on Information Systems (ECIS 2013)***, June 5-8, Utrecht, Completed Research, Vol. 184, pp. 1-12.

101. Glowalla, P., & Sunyaev, A., (2014a). "Process-driven data quality management: A critical review on the application of process modeling languages", ***ACM Journal of Data and Information Quality (ACM JDIQ)***, Vol. 5, Nos. 1–2, Article 7 (August 2014), pp. 1-30.

102.Glowalla, P., & Sunyaev, A., (2014b). "ERP system fit – an explorative task and data quality perspective", ***Journal of Enterprise Information Management (JEIM),*** Vol. 27, Issue 5, pp. 668–686.

103.Glowalla, P., & Sunyaev, A., (2015). "Influential Factors on IS Project Quality: A Total Quality Management Perspective", in: ***Proceedings of 36th International Conference on Information Systems (ICIS),*** Fort Worth, USA, December, pp. 13-16.

104.Goetz, M., Leganza, G., McGovern, S., & Lynch, D., (2015). "The Forrester Wave™: Data Quality Solutions", ***Forrester research Inc.***, Q4-2015. Available at: https://www.forrester.com (18.12.17).

105.González-Pérez, S., Alenina, E., & Grigorienko, V., (2014). "Business Process Management and New Technologies",***The 2nd annual conference В.И. Кравцовой: "Managing Competitiveness",*** Dec. 2013, Moscow, Russia, pp. 1-8.

106.Gregor, S., (2006). "The Nature of Theory in Information Systems," ***MIS Quarterly***, September 2006, Vol. 30, No. 3, pp. 611-642.

107.Gregor, S., & Hevner, A., (2013). "Positioning and Presenting Design Science Research for Maximum Impact", ***MIS Quarterly***, Vol. 37, No. 2, pp. 337-355.

108.Griesberger, P., Leist, S., & Zellner, G., (2011). "Analysis of Techniques for Business Process Improvement", in: ***Proceedings of European Conference on Information Systems (ECIS2011)***, No. 20.

109.Hammer, M., & Champy, J., (1993). ***Reengineering the Corporation – A manifesto for Business Revolution***, Nicholas Brealey Publishing, London.

110.Hamzah, M.A., AlQudah Siti, N., & Binti, S., (2014). “The role of Data Quality and internal control in raising the effectiveness of AIS in Jordan Companies.”, ***International Journal of Scientific and Technology Research (IJSTR)***, Vol. 3, No. 8, August 2014, pp. 298-303.

111.Harmon, P., (2004). "Evaluating an Organization's Business Process Maturity", ***BP Trends***, Vol. 2, No. 3, March 2004, pp. 1-11.

112.Harmon, P., (2007). ***Business Process Change: A Guide for Business Managers and BPM and Six Sigma Professionals***. 2nd edition. Morgan Kaufmann. Burlington (USA).

113.Harmon, P., (2014). ***Business Process Change: A Business Process Management Guide for Managers and Process Professionals***, 3rd Edition, Morgan-Kaufmann Publishers Inc., San Francisco, CA, USA.

114.Harmon, P., & Wolf, C., (2011). "Business Process Modeling Survey", ***BP Trends***. Available at: http://www.bptrends.com/ (26.12.15).

115.Haug, A., Zachariassen, F., & van Liempd, D., (2011). "The cost of poor data quality", ***Journal of Industrial Engineering and Management***, Vol. 4, No. 2, pp. 168-193.

116.Havey, M., (2005). ***Essential Business Process Modeling-Theory in Practice***, O'Reilly Media Inc.

117.Heilbronner, N.N., & Renzulli, J.S., (2016). ***Schoolwide Enrichment Model in Science***, Prufrock Press inc., Texas, USA.

118.Heinrich, B., Hristova, D., Klier, M., Schiller, A., & Szubartowicz, M., (2018). "Requirements for Data Quality Metrics", ***Journal of Data and Information Quality (JDIQ)***, Vol. 9, Issue 2, January 2018, Article No. 12, pp. 1-29.

119.Heinrich, R., Kappe, A., & Paech, B., (2011). "Modeling Quality Information within Business Process Models", in: ***Proceedings of the SQMB'11 Workshop***, TUM-I1104, Karlsruhe (Germany), February 21th. 2011, pp. 4-13.

120.Heinrich, R., Merkle, P., Henss, J., & Paech, B., (2017). "Integrating business process simulation and information system simulation for performance prediction", ***Software and Systems Modeling (SoSyM) Journal***, Vol. 16, Issue 1, Feb. 2017, pp. 257-277.

121.Heinrich, R., & Paech, B., (2010). "Defining the Quality of Business Processes", in: Engels, G. et al. (Eds.). in: ***Proceedings of Modellierung 2010***, GI, 2010. Bonner Köllen., LNI Vol. P-161, pp. 133–148

122.Heravizadeh, M., Mendling, J., & Rosemann, M., (2009). "Dimensions of Business Processes Quality (QoBP)", ***Business Process Management International Workshops08***, Milano, Italy, September 1-4, Vol. 17, Part 1, pp. 80-91.

123.Herrmann, A., & Paech, B., (2006). "MOQARE = Misuse-oriented Quality Requirements Engineering", ***Softwaretechnik-Trends***, Vol. 26, No. 1.

124.Hevner, A.R., (2007). "A Three Cycle View of Design Science Research", ***Scandinavian Journal of Information Systems (SJIS)***, Vol. 19, No. 2, pp. 87–92.

125.Hevner, A.R., & Chatterjee, S., (2015). "Design Science Research in Information Systems", in: ***Association for Information Systems (AIS)***, Reference Syllabi, Ed.: J. vom Brocke, Eduglopedia.org, 2015. Available at: http://eduglopedia.org/reference-syllabus/AIS_Reference_Syllabus_Design_Science_Research_in_IS.pdf (15.12.16).

126.Hevner, A.R, March, S.T., Park, J., & Ram, S., (2004). "Design Science in Information Systems Research", ***MIS Quarterly***, Minneapolis: March 2004, Vol. 28, No. 1, pp. 75-105.

127.Hibberd, B.J., & Evatt, A., (2004). "Mapping information flows: A practical guide", ***Information Management Journal***, Vol. 38, No. 1, Jan.-Feb. 2004, pp. 58-64.

128.Hiskey, M., (2018). "Data quality goes big-time with GDPR in 2018", ***InfoWorld***, Jan. 25, 2018. Available at: https://www.infoworld.com/article/3251072/data-management/data-quality-goes-big-time-with-gdpr-in-2018.html (10.9.18).

129.Hommes, B.J., & Dietz, J.L.G., (2001). "Business Process Modeling for the purpose of Applying Internet Technology", in: ***Proceeding of 6th CAiSE/IFIP8.1 -The international Workshop on Evaluation of Modeling Methods in Systems Analysis and Design (EMMSAD'01)***, 4-5 June, 2001, Interlaken, Switzerland, pp. 1-8.

130.Hoyle, D., (2009). ***ISO9000 Quality Systems Handbook***, 6th edition, Butterworth-Heinemann publication, Oxford, UK.

131.Huang, H.T., Lee, Y.W., & Wang, R.Y., (1999). ***Quality Information and Knowledge***, Prentice Hall, New Jersey.

132.IBM Big Data & Analytics Hub, (2016). "Extracting business value from the 4 V's of big data". Available at: http://www.ibmbigdatahub.com/infographic/ex-tracting-business-value-4-vs-big-data (20.12.18).

133.Ibrahim, R., Ayazi, E., Nasrmalek, S., & Nakha S., (2013). "An Investigation of Critical Failure Factors In Information Technology Projects", ***Journal of Business and Management***, Vol. 10, Issue 3 (May- Jun. 2013), pp. 87-92.

134.Imtiaz, M.A., Al-Mudhary, A.S., Mirhashemi, M.T., & Ibrahim, R., (2013). "Critical Success Factors in Information Technology Projects", ***International Journal of Social, Human Science and Engineering***, Vol. 7, No. 12, pp. 1913–1917.

135.Iivari, J., Parsons, J., & Wand, Y., (2006). "Research in Information Systems Analysis and Design: Introduction to the Special Issue", ***Journal of the Association for Information Systems***, Atlanta: Aug. 2006. Vol. 7, Issue 8, pp. 509-513.

136.Indulska, M., Recker, J., Roseman, M., & Green, P., (2009). "Process Modeling: Current Issues and Future Challenges", ***International Conference on Advanced Information Systems Engineering - CAiSE 2009***. Amsterdam: Netherlands, pp. 501-514.

137.Info-Tech Research Group, (2008). "Project Management: Back to Basics". Available at: http://www.infotech.com (12.5.14).

138.Ishikawa, K., (1985). ***What is Total Quality Control***, Prentice-Hall, Englewood Cliffs, NJ.

139.ISO 8000-62:2018. ***Data quality - Part 62: Data quality management***: Organizational process maturity assessment: Application of standards relating to process assessment.

140.ISO 9000:2015. ***Quality Management Systems – Fundamentals and Vocabulary***, 4th edition.

141.ISO/IEC 25012: Software engineering (2008) - Software product Quality Requirements and Evaluation (SQuaRE) — Data quality model, First edition.

142.ISO 9241-11:2018. Ergonomics of Human-System interaction, Part 11: Usability: Definitions and concepts. 2nd edition.

143.Javed, A., Maqsood, M., AshfaqQazi, K., & Ali, S.K., (2012). "How To Improve Software Quality Assurance", ***Advanced Computing: An International Journal (ACIJ)***, Vol. 3, No. 2, March 2012.

144.Jaya, M.I., Sidi, F., Ishak, I., Affendey, L.S., & Jabar, M.A., (2017). "A review of data quality research in achieving high data quality within organization". ***Journal of Theoretical and Applied Information Technology***, Vol. 95, No. 12, pp. 2647-2657.

145.Jayawardene, V., Sadiq, S., & Indulska, M., (2013) "An analysis of data quality dimensions". ***ITEE Technical Report No. 2013-01***. School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia.

146.Jeston, J., & Nelis, J., (2014). ***Business Process Management: Practical Guidelines to Successful Implementations***, 3rd Ed., Routledge -Taylor & Francis Group, New-York, NY.

147.Johnson, J., & Mulder, H., (2016). "CHAOS Chronicles, focusing on failures and possible improvements in IT projects", ***Systemics, Cybernetics and Informatics***, Vol. 14, No. 5, pp. 1-5.

148.Joshi, A., Kale, S., Chandel, S., & Pal, D., (2015). "Likert Scale: Explored and Explained", ***British Journal of Applied Science & Technology (BJAST)***, Vol. 7, No. 4, pp. 396-403.

149.Juárez, D., Schmidt, E.E., Stahl-Toyota, S., Ückert, F., & Lablans, M., (2019). “A Generic Method and Implementation to Evaluate and Improve Data Quality in Distributed Research Networks”, ***Methods of Information in Medicine***. Vol. 58, No. 2–3/2019, pp. 86–93.

150.Juddoo, S., & George, C., (2018). "Discovering the Most important Data Quality Dimensions in Health Big Data using Latent Semantic Analysis", in: ***Proceedings of IEEE International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD),*** Durban, South Africa, 6-7 August 2018, pp. 1-6.

151.Jugulum, R., (2016). "Importance of Data Quality for Analytics", in Sampaio P., and Saraiva P., (eds.) book, ***Quality in the 21st Century***, Springer International Publishing, Switzerland 2016.

152.Jung, W., (2004). "A review of research: an investigation of the impact of data quality on decision performance", in: ***Proceedings of the 2004 international symposium on Information and communication technologies (ISICT '04)***, Las Vegas, NV, June 16-18, 2004, pp. 166-171.

153.Jung, W., Olfman, L., Ryan, T., & Park, Y., (2005). "An experimental study of the effects of contextual data quality and task complexity on decision performance", in: ***Proceedings of IEEE International Conference on Information Reuse and Integration***, August 15-17, 2005, Las Vegas, NV, pp. 149–154.

154.Juran, J.M., & Godfrey, A.B., (1999). ***Juran's Quality Handbook***, 5th Edition, McGraw-Hill.

155.Kahn, B.K., Strong, D.M., & Wang, R.Y., (2002). "Information quality benchmarks: product and service performance", ***Communications of the ACM***, Vol. 45, Issue 4, April 2002, pp. 184-192.

156.Khan, Z., & Vorley, T., (2017). "Big Data Text Analytics an enabler of Knowledge Management", ***Journal of Knowledge Management***, Vol. 21, No. 1, pp. 18-34.

157.Kappelman, L., McKeeman, R., & Zhang, L., (2009). “Early Warning Signs of IT Project Failure: The Dangerous Dozen", ***EDPACS (EDP Audit, Control, and Security)***, December 2009, Vol. 40, No. 6, pp. 17-25.

*158.*Karel, R., & Richardson, C., (2010). "Avoid Process Data Headaches: Align Business Process and Data Governance Initiatives", ***Forrester Research Inc.***

159.Karlsson, M., (2016), "What is a case study?." Available at: http://www.diva-portal.org/smash/get/diva2:1051860/FULLTEXT01.pdf (15.12.18).

160.Karr, A.F., Sanil, A.P., & Banks, D.L., (2005). "Data quality: A statistical perspective", ***Statistical Methodology Journal***, Tech. Report No. 151, March 2005, Vol. 3, No. 1, pp. 137–173.

161.Kaur, R., & Sengupta, J., (2011). Software Process Models and Analysis on Failure of Software Development Projects, ***International Journal of Scientific & Engineering Research***, Vol. 2, Issue 2, February. pp. 1-5.

162.Kedad, Z., & Loucopoulos, P., (2011). "Considering quality factors for business processes during requirement engineering", in: ***Proceedings of 2011 5th International Conference on Research Challenges in Information Science (RCIS 2011)***, pp. 1-9.

163.Kenett, R.S., & Shmueli, G., (2016). ***Information Quality: The Potential of Data and Analytics to Generate Knowledge***, John Wiley & Sons, Ltd, Chichester, UK.

164.Kent, W., (1978), ***Data and Reality***, North Holland, New York.

165.Kerpedzhiev, G., König, U., Röglinger, M., & Rosemann, M., (2017). "Business Process Management in the Digital Age", ***BP Trends***, July 2017. Available at: www.bptrends.com (7.12.17).

166.Kerr, K., & Norris, T., (2004). "The Development of a Healthcare Data Quality Framework and Strategy", in: ***Proceedings of the 9th International Conference on Information Quality (IQ 2004)***, November 5-7, 2004, pp. 218-233.

167.Keskin, H., (2005). "The Relationships between Explicit and Tacit Oriented KM Strategy, and Firm Performance", ***Journal of American Academy of Business***, Cambridge, Vol. 7, No. 1, pp. 169–175..

168.Ketchen, D., & Short, J., (2012). ***Strategic Management: Evaluation and Execution***, Creative Commons Publishers, Digital Press.

169.Kirchmer, M., (2010). "Management of Process Excellence", in: Brocke J. vom & Rosemann M. (eds.), ***Handbook on Business Process Management 2***, Springer-Verlag Berlin Heidelberg, pp. 39-55.

170.Kirchmer, M., (2017). "Chapter 1: Business Process Management: What Is It and Why Do You Need It?", ***High Performance through Business Process Management: Strategy Execution in a Digital World***, Springer. pp. 1–28.

171.Knauer, T., Nikiforow, N., & Wagener, S., (2020). "Determinants of information system quality and data quality in management accounting", ***Journal of Management Control***, Vol. 31, pp. 97–121.

172.Knight, S.A., & Burn, J., (2005). "Developing a Framework for Assessing Information Quality on the World Wide Web", ***Informing Science Journal***, Vol. 8, pp. 159-172.

173.Knight, M., (2017). "What is Data Governance?", December 18, 2017. Available at: https://www.dataversity.net/what-is-data-governance/ (5.3.19).

174.Ko, R.K.L., Lee, S.S.G., & Lee, E.W., (2009). "Business Process Management (BPM) Standards: A Survey", ***Business Process Management Journal***, Vol. 15, No. 5, pp. 744-791.

175.Komai S., Saidi, H., & Nakanishi, H., (2017). "Study on Critical Success Factors Estimation in IT System Development", ***Sains Humanika Journal***, Penerbit UTM Press, Universiti Teknologi Malaysia. Vol. 9, No. 1-3 (2017), pp. 31–38.

176.Kontio, J., Lehtola, L., & Bragge, J., (2004). "Using the Focus Group Method in Software Engineering: Obtaining Practitioner and User Experiences", in: ***Proceedings of the International Symposium on Empirical Software Engineering (ISESE)***, Aug. 19-20, 2004, Redondo Beach, U.S.A. IEEE Computer Society, pp. 271- 280.

177.KPMG, (2016). "Now or never- 2016 Global CEO Outlook Survey". Available at: https://home.kpmg/content/dam/kpmg/pdf/2016/06/2016-global-ceo-outlook.pdf (31.7.17).

178.KPMG, (2018). "Data Quality Assessment". Available at: https://assets.kpmg/content/dam/kpmg/ch/pdf/data-quality-assessment.pdf (11.12.18).

179. Krasner, H., (2018). "The Cost of Poor Software Quality in the US: A 2018 Report", ***Consortium for IT Software Quality (CISQ)***, Sep. 26, 2018. Available at: https://www.it-cisq.org (11.2.20).

180.Krueger, R.A., (2002). "Designing and Conducting Focus Group Interviews", October 2002, pp. 1-18.

181.Krueger, R.A., & Casey, M.A., (2014). ***Focus Groups: A Practical Guide for Applied Research***, 5th, Sage Publications, USA.

182.Kurtz, C., Semmann, M., & Böhmann, T., (2018). "Privacy by Design to Comply with GDPR: A Review on Third-Party Data Processors", ***Twenty-fourth Americas Conference on Information Systems***, New Orleans, August 2018, pp. 1-10.

183.Lahajnar, S., & Rožanec, A., (2016). "The evaluation framework for business process management methodologies", ***Management: Journal of Contemporary Management Issues***, Vol. 21, No. 1, pp. 47-69.

184.Laranjeiro, L., Soydemir, S.N., & Bernardino, J., (2015). "A Survey on Data Quality: Classifying Poor Data", in: ***Conference: The 21st IEEE Pacific Rim International Symposium on Dependable Computing (PRDC 2015)***, At Zhangjiajie, China, pp. 179-188.

185.Laudon, K.C., & Laudon, J.P., (2018). ***Management Information Systems: Managing the Digital Firm***, 15th Edition, Pearson Education Limited, England.

186.Law, C.C.H., & Ngai, E.W.T., (2007) "Infrastructure IT Capabilities and business process improvement: association with IT governance characteristics". ***Information Resources Management Journal***, Vol. 20, No. 4, pp. 25–47.

187.Leavitt, H.J., (1965). "Applied organizational change in industry: Structural, technological and humanistic approaches". In: J.G. March (Ed.), ***Handbook of Organization***, pp. 1144-1170. Rand McNally and Company. Chicago, Illinois.

188.Lee, Y.W., Strong, D., Kahn, B., & Wang, R.Y., (2002). "AIMQ: A methodology for information quality assessment", ***Information & Management***, Vol. 40, No. 2, pp.133-146.

189.Leymann, F., & Altenhuber, W., (1994). "Managing Business Processes as an Information Resource", ***IBM systems journal***, Vol. 33, No. 2, pp. 326-348.

190.Likert, R., (1932). "A Technique for the Measurement of Attitudes". ***Archives of Psychology***, Vol. 140, pp. 1–55.

191.Liu, Q., Feng, G., Zhao, X., & Wang W., (2020). "Minimizing the data quality problem of information systems: A process-based method", ***Decision Support Systems***, Vol. 137, pp. 1-12.

192.Lohrmann, M., & Reichert, M., (2010). "Basic Considerations on Business Process Quality", ***Technical Report UIB-2010-03***, University of Ulm, Ulm, Germany.

193.Lohrmann, M., & Reichert, M., (2013). "Understanding Business Process Quality", In: Glykas, M., (eds), ***Business Process Management: Theory and Applications***, Studies in Computational Intelligence, Vol. 444, Springer, Berlin, Heidelberg, 2013, pp. 41-73.

194.Loucopoulos, P., & Heidari, F., (2012). "Evaluating quality of business processes", In: N. Seyff, and A. Koziolek (Eds.), ***Modelling and quality in requirements engineering***: essays dedicated to Martin Glinz on the occasion of his 60th birthday. Munster: MV-Wissenschaft, pp. 61-73.

195.Lu, R., Sadiq, S., & Governatori, G., (2009). "On managing business processes variants." ***Data & Knowledge Engineering***, Vol. 68, No. 7, pp. 642-664.

196.Loshin, D., (2011). "Evaluating the business impacts of poor data quality," ***Knowledge Integrity Incorporated, Business Intelligence Solutions***, Tech. Report, January 2011.

197.Madnick, S.E., Wang R.Y., Lee, Y.W., & Zhu, H., (2009). "Overview and Framework for Data and Information Quality Research", ***ACM Journal of Data and Information Quality***, Vol. 1, No. 1, Article 2, pp. 1-22.

198.Manners-Bell, J., (2014). ***Global Logistics Strategies: Delivering the Goods***, Kogan Page Publishers, London.

199.March, S.T., & Smith, G.F., (1995). "Design and natural science research on information technology", ***Decision Support Systems***, Vol. 15, No. 4, pp. 251-266.

200.March, S.T., & Storey, V.C., (2008). "Design Science in the Information Systems Discipline: An introduction to the special issue on design science research", ***MIS Quarterly***, Vol. 32, No. 4, pp. 725-730.

201.McLeod, S.A., (2008). "Likert scale". Available at: https://www.simplypsychology.org/likert-scale.html (10.3.19).

202.Meda, H.S., Sen, A.K., & Bagchi, A., (2010). "On Detecting Data Flow Errors in Workflows", ***Journal of Data and Information Quality (JDIQ)***, 07/2010, Vol. 2, No. 4, Issue 1, pp. 1-30.

203.Mondragón, M., Mora, M., Garza, L., Álvarez, F., Rodríguez, L., & Duran-Limon, H.A. (2013). "Toward a Well-structured Development Methodology for Business Process-oriented Software Systems based on Services", ***Procedia Technology***, Vol. 9, (2013), pp. 351-360.

204.Moody, D.L., Sindre, G., & Brasethvik, T., (2003). "Evaluating the Quality of Information Models: Empirical Testing of a Conceptual Model Quality Framework", ***IEEE***, pp. 295-305.

205.Moore, S., (2018). "How to Stop Data Quality Undermining Your Business", ***Gartner research Inc.***, Stamford, Connecticut. Available at: https://www.gartner.com/smarterwithgartner/how-to-stop-data-quality-undermining-your-business/ (20.12.18).

206.Mylopoulos, J., Chung, L., & Nixon, B., (1992). "Representing and Using Non-Functional Requirements: A Process-Oriented Approach", ***IEEE Transactions on Software Engineering***, Vol. 18, No. 6, June 1992, pp. 483-497.

207.Nagle, T., Redman, T., & Sammon, D., (2020). "Assessing data quality: A managerial call to action", ***Business Horizons***, Vol. 63, Issue 3, 2020, pp. 325-337.

208.Nasir, M.H.N., & Sahibuddin, S., (2011). "Critical success factors for software projects: A comparative study", ***Scientific Research and Essays***, Vol. 6, No. 10, pp. 2174-2186.

209.Neuman, W.L., (2014). ***Social Research Methods: Qualitative and Quantitative Approaches***, 7th Edition, Pearson Education ltd., Essex, UK.

210.Nicolaou, A.I., & McKnight, D.H., (2006). "Perceived Information Quality in Data Exchanges: Effects on Risk, Trust, and Intention to Use", ***Information Systems Research***, Vol. 17, No. 4, December 2006, pp. 332–351.

211.Nili, A., Tate, M., & Johnstone, D., (2017). "A Framework and Approach for Analysis of Focus Group Data in Information Systems Research". ***Communications of the Association for Information Systems***, Vol. 40, No. 1, pp. 1-21.

212.NISS Report, (2001). “Affiliates Workshop on Data Quality”, ***National Institute of Statistical Sciences***. Available at: http://www.niss.org/affiliates/dqworkshop/report/dq-report.pdf (12.5.14).

213.Nonaka, I., & Takeuchi, H., (1995). ***The Knowledge Creating Company***, Oxford University Press, New York.

214.Nüttgens, M., Feld, T., & Zimmermann, V., (1998). "Business Process Modeling with EPC and UML: Transformation or Integration?", in: Schader, M.; Korthaus, A. (Hrsg.): ***The Unified Modeling Language-Technical Aspects and Applications, Proceedings***, (Mannheim, October 1997), Heidelberg 1998, pp. 250-261.

215.Nwakanma, C.I., Asiegbu, B.C., Ogbonna, C.A., & Njoku Peter-Paul, C., (2013). "Factors Affecting Successful Implementation of Information Technology Projects: Experts' Perception", ***European Scientific Journal***, Sep. 2013 edition, Vol. 9, No. 27, pp. 128-137.

216.Nyumba, T.O., Wilson, K., Derrick, C.J., & Mukherjee, N., (2018). "The use of focus group discussion methodology: Insights from two decades of application in conservation", ***Methods in Ecology and Evolution***, 2018, No. 9, pp. 20-32.

217.Ofner, M., Otto, B., & Österle, H., (2012). "Integrating a data quality perspective into business process management", ***Business Process Management Journal***, Vol. 18, No. 6, pp. 1036-1067.

218.OMG-UML, (2005). ***The Unified Modeling Language (UML™)***, V2.0, Object Management Group. Available at: https://www.omg.org/spec/UML/2.0 (10.6.15).

219.OMG, (2011). ***Business Process Modeling Notation (BPMN™)***, V2.0, Object Management Group. Available at: http://www.omg.org/spec/BPMN/2.0 (20.6.17).

220.Onwuegbuzie, A.J., Dickinson, W.B., Leech, N.L., & Zoran, A.G., (2009). "A Qualitative Framework for Collecting and Analyzing Data in Focus Group Research", ***International Journal of Qualitative Methods (IJQM)***, Vol. 8, No. 3, pp. 1-21.

221.Otto, B., (2011). "Data governance", ***Business and Information Systems Engineering (BISE)***, Vol. 3, No. 4, pp. 241–244.

222.Otto, B., Hüner, K.M., & Österle, H., (2009). "Identification of Business Oriented Data Quality Metrics", in ***Proceedings of 14th International Conference on Information Quality (ICIQ 2009)***, Potsdam, Germany. pp. 122-134.

223.Otto, B., & Österle, H., (2015). ***Corporate Data Quality: Prerequisite for Successful Business Models*** (eBook), Publisher: epubli GmbH, Berlin, Germany.

224.Olson, J.E., (2003). ***Data Quality: The Accuracy Dimension***, The Morgan Kaufmann Publishers, Elsevier Science & Technology, San Francisco, CA.

225.Österle, H., & Otto, B., (2010). "Consortium Research: A Method for Researcher-Practitioner Collaboration in Design-Oriented IS Research", ***Business and Information Systems Engineering (BISE)***, Vol. 2, Issue 5, May 2010, pp. 283-293. Available at: http://aisel.aisnet.org/bise/vol2/iss5/3 (10.8.2016).

226.Patel, N.V., (2007). "Deferred Action: Theoretical model of process architecture design for emergent business processes", ***International Journal of Business Science and Applied Management***, Vol. 2, Issue 3, pp. 4-21.

227.Patel, N., & Hlupic, V., (2001). "Dynamic business process modelling (BPM) for business process change", ***International Journal of Simulation: Systems, Science & Technology***, Dec. 2001, Vol. 2 No. 2, pp. 51-64.

228.Peffers, K., Rothenberger, M., Tuunanen, T., & Vaezi, R., (2012). "Design science research evaluation", in: ***Proceeding of the 7th international conference on Design Science Research in Information Systems (DESRIST'*** 2012), May 2012, Las Vegas, NV, USA. pp. 398-410.

229.Peffers, K., Tuunanen, T., Rothenberger, M.A., & Chatterjee, S., (2008). "A Design Science Research Methodology for Information Systems Research", ***Journal of Management Information Systems***, Vol. 24, No. 3, Winter 2007-8, pp. 45-77.

230.Pierce, E., Yonke, C.L., Malik, P., & Nargaraj, C.K., (2012). "The state of information and data quality efforts in today's organizations", in ***Proceedings of the 17th International Conference on Information Quality (ICIQ)***, November 16-17, 2012, Paris, France. pp. 14-30.

231.Pipino, L.L., Lee, Y.W., & Wang, R.Y., (2002). "Data Quality Assessment", ***Communications of the ACM***, New York, April 2002, Vol. 45, No. 4, pp. 211-218.

232.Pohl, K., (1996). ***Process-Centered Requirements Engineering***, Research Studies Press Ltd., Taunton, Somerset, England.

233.Porter, M.E., (1985). ***Competitive Advantage***, The Free-Press, New York, USA.

234.Porter, M.E., & Millar, V.E., (1985). "How Information gives you Competitive Advantage", ***Harvard Business Review***, Vol. 64, No. 4, July- August 1985, pp. 149-160.

235.Prat, N., Comyn-Wattiau, I., & Akoka, J., (2014). "Artifact Evaluation in Information Systems Design Science Research - A Holistic View", In: ***Proceedings of 18th Pacific Asia Conference on Information Systems (PACIS 2014)***, Chengdu, China, Paper 23, pp. 1-16. Available at: http://aisel.aisnet.org/pacis2014/23 (20.9.19).

236.Price, R., & Shanks, G., (2005). "A Semiotic Information Quality Framework: Development and Comparative Analysis," ***Journal of Information Technology***, Vol. 20, No. 2, June 2005, pp. 88-102.

237.Raab, D.M., (2008). "Technical Measures for Data Quality Investments", ***DM Review***, June 2008, p. 10.

238.Rahimi, F., Møller, C., & Hvam, L., (2014). "Alignment between business process governance and IT governance", in: ***Proceedings of 20th Americas Conference on Information Systems***, Savannah, United States, 07/08/2014. pp. 1-12.

239.Rajkumar, G., & Alagarsamy, K., (2013). "Failure of Software Development Projects", ***The International Journal of Computer Science & Applications (TIJCSA)***, Vol. 1, No. 11, pp. 74-77, January 2013.

240.Recker, J., Mutschler, B., & Wieringa, R., (2011). "Empirical research in business process management: introduction to the special issue", ***Journal of Information Systems and e-Business Management*** , Vol. 9, Issue 3, pp. 303-306.

241.Recker, J., Rosemann, M., Indulska, M., & Green, P., (2006). "Business Process Modeling: A Maturing Discipline?", ***BPM Center Report NR***, BPM-06-20, BPMCenter.org

242.Redman, T.C., (1997). ***Data Quality for the Information Age***, Artech House Publishers Inc. Norwood, MA, USA.

243.Redman, T.C., (1998). "The Impact of Poor Data Quality on the Typical Enterprise", ***Communications of the ACM***, Vol. 41, No. 2, pp. 79-83.

244.Redman, T.C., (2004). "Data: An Unfolding Quality Disaster", ***DM Review Magazine,*** Aug. 2004. Available at: http://www.dmreview.com (4.6.14).

245.Redman, T.C., (2016). "Bad Data Costs the U.S. $3 Trillion per Year", ***Harvard Business Review***, digital article, Sep. 2016. Available at: https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year (20.12.18).

246.Regev, G., Soffer, P., & Bider, I., (2005). "Coordinated development of business processes and their support systems", ***Requirements Engineering***, Vol. 10, pp. 173-174.

247.Reijers, H.A., Mendling, J., & Recker, J., (2010). "Business Process Quality Management", in J. vom Brocke & M. Rosemann (eds.), ***Handbook on Business Process Management 1, International Handbooks on Information Systems***, Springer-Verlag Berlin Heidelberg.

248.Rosato, M., (2018). "Go Small for Project Success", ***PM World Journal***, Vol. 7, No. 5, May 2018, pp. 1-10. Available at: http://www.pmworldjournal.net/ (10.12.18).

249.Rosemann, M., & vom Brocke, J., (2015). "Six Core Elements of Business Process Management". In J. vom Brocke & M. Rosemann (Eds.), ***Handbook on Business Process Management: Introduction, Methods, and Information Systems*** (International Handbooks on Information Systems), 2nd ed., Vol. 1, Springer-Verlag Berlin Heidelberg, pp. 105-122.

250.Röthlin, M., (2004). "An Exploratory Study of Data Quality Management Practices in the ERP Software Systems Context". In: Dadam, P. and Reichert, M., (eds.), in: ***Proceedings of INFORMATIK 2004 – the 34th Annual Conference of the Gesellschaft fur Informatik***, Vol. 1, Bonn 2004, pp. 254-258.

251.Runeson, P., & Höst, M., (2009). "Guidelines for conducting and reporting case study research in software engineering", ***Empirical Software Engineering***, Vol. 14, No. 2, pp. 131-164.

252.Russell, N., Aalst, W.M.P., van der, & Hofstede, A.H.M., ter, (2016). ***Workflow Patterns: The Definitive Guide***, MIT Press, Cambridge, Massachusetts, London, England.

253.Sabherwal, R., & King, W., (1991). "Towards a theory of the strategic use of information resources: an inductive approach", ***Information and Management***, Vol. 20, No.3, pp. 191-212.

254.Sadiq, S., (ed.) (2013). ***Handbook of Data Quality: Research and Practice***, Springer-Verlag, Berlin Heidelberg, Berlin, Germany.

255.Sadiq, S., Jayawardene, V., & Indulska, M., (2011). "Research and industry synergies in data quality management", in: ***International conference on information quality (ICIQ2011)***, Adelaide, Australia, 18-20 November, 2011.

256.Sadiq, W., & Orlowska, M., (1999). "On Capturing Process Requirements of Workflow Based Information Systems", in: ***Proceedings of the 3rd Int'l Conference on Business Information Systems (BIS '99),*** Poznan, Poland, April 14-16, 1999, Springer-Verlag, pp. 195-209.

257.Sadiq, S., Orlowska, M., Sadiq, W., & Foulger, C., (2003). "Data flow and validation in workflow modeling", in: ***Proceedings of the 15th Australasian Database conference (ADC'2004)***, Dunedin, New Zealand, January 18-22, Vol. 27 of CRPIT, pp. 207-214.

258.Scannapieco, M., Missier, P., & Batini C., (2005). "Data quality at a glance," ***Datenbank-Spektrum***, Vol. 14, pp. 6-14.

259.Schultz, J.R., (2006). "Measuring Service Industry Performance: Some Basic Concepts", ***Performance Improvement***, April 2006, Vol. 45, No. 4, pp. 13-17.

260.Seiner, R.S., (2014). ***Non-Invasive Data Governance***, Technics Publications, Technics Publications, Basking Ridge, NJ, USA.

261.Serra, F., (2020). "Handling Context in Data Quality Management". In: Bellatreche L. et al. (eds) ADBIS, TPDL and EDA 2020 Common Workshops and Doctoral Consortium. TPDL 2020, ADBIS 2020. ***Communications in Computer and Information Science***, Vol. 1260. Springer, Cham. Available at: https://doi.org/10.1007/978-3-030-55814-7_32 (28.8.2020).

262.Shankaranarayanan, G., & Cai, Y., (2006). "Supporting data quality management in decision-making", ***Decision Support Systems***, Vol. 42, No. 1, pp. 302–317.

263.Shankaranarayanan, G., & Wang, R., (2007). "IPMAP: Current State and Perspectives". In: ***Proceedings of the 12th International Conference on Information Quality***, pp. 500–517.

264.Shankaranarayanan, G., Wang, R.Y., & Ziad, M., (2000). "IP-Map: Representing the manufacture of an information product", in: ***5th International Conference on Information Quality (ICIQ'2000)***, pp. 1-16.

265.Shariat Panahy, P.H., Sidi, F., Affendey, L., Jabar, M., Ibrahim, H., & Mustapha, A., (2013). "A Framework to Construct Data Quality Dimensions Relationships", ***Indian Journal of Science and Technology***, Vol. 6, No. 5, pp. 4422-4431.

266.Shaw, D.R., Holland, C.P., Kawalek, P., Snowdon, B., & Warboys, B., (2007). "Elements of a business process management system: theory and practice", ***Business Process Management Journal***, Vol. 13, No. 1, 2007, pp. 91-107.

267.Shuttleworth, M., (2008). "Case Study Research Design", Retrieved from ***Explorable.com***, Available at: https://explorable.com/case-study-research-design (23.8.2018).

268.Sidi, F., Jabar, M.A., Selamat, M.H., Ghani, A.A.A., & Ibrahim, H., (2009). "An Investigation into Methods and Concepts of Qualitative Research in Information System Research". ***Computer and Information Science***, Vol. 2, No. 4, pp. 109-115.

269.Slone, J.P., (2006). "Information quality strategy: An empirical investigation of the relationship between information quality improvements and organizational outcomes", ***Ph.D dissertation***, Capella University.

270.Smart, P.A., Maddern, H., & Maull, R.S., (2009). "Understanding business process management: implications for theory and practice", ***British Journal of Management***, Vol. 20, pp. 491-507.

271.Smithson, S., & Hirschheim, R., (1998). "Analyzing information systems evaluation: another look at an old problem", ***European Journal of Information Systems***, Vol. 7, No. 3, pp. 158–174.

272.Soffer, P., & Wand, Y., (2007). "Goal-driven multi-process analysis", ***Journal of the Association of Information Systems***, Vol. 8, Issue 3, pp. 175-203.

273.Soffer, P., (2010). "Mirror, mirror on the wall, can i count on you at all? Exploring data inaccuracy in business processes", Enterprise, Business-Process and Information Systems Modeling: in ***Proceedings of 11th International Workshop BPMDS 2010***, Vol. 50, pp. 14-25.

274.Souibgui, M., Atigui, F., Zammali, S., Cherfi, S., & Ben Yahia, S., (2019). "Data quality in ETL process: A preliminary study", ***Procedia Computer Science***, Vol. 159, pp. 676-687.

275.Stair, R., & Reynolds, G., (2010). ***Principles of information Systems***, 9th ed. MA. USA.

276.Standish Group, (2013). "CHAOS Manifesto 2013: Think Big, Act Small", ***The Standish Group International Inc.***, Available at: https://people.eecs.ku.edu/~hossein/811/Papers/standish-2013-report.pdf (20.4.14).

277.Standish Group, (2014). "The CHAOS Report 2014 – Project smart", ***The Standish Group International Inc.***, Available at: https://www.projectsmart.co.uk/white-papers/chaos-report.pdf (12.6.16).

278.Standish Group, (2015). "The CHAOS Report 2015", ***The Standish Group International Inc.***, Available at: https://www.standishgroup.com/sample_research_files/CHAOSReport2015-Final.pdf (12.6.16).

279.Standish Group, (2018). "Project Resolution Benchmark", ***The Standish Group International Inc.***, Available at: https://www.standishgroup.com/store/project-resolution-benchmark.html (11.2.20).

280.Storey, V.C., Dewan, R.M., & Freimer, M., (2012). "Data quality: Setting organizational policies", ***Decision Support Systems Journal***, Vol. 54, Issue 1, December 2012, pp. 434-442.

281.Strong, D.M., (2009). "Information Quality: Managing Information as a Product", In: LIU L., ÖZSU M.T. (eds) ***Encyclopedia of Database Systems***. Springer, Boston, MA.

282.Strong, D.M., Lee, Y.W., & Wang, R.Y., (1997). "Data Quality in Context", ***Communications of the ACM***, Vol. 40, No. 5, pp. 103-110.

283.Stewart, D.W., & Shamdasani, P.N., (2015). ***Focus Groups - Theory and Practice***, 3rd edition, Sage Publications, Newbury Park, CA.

284.Stewart, D.W., Shamdasani, P.N., & Rook, D.W., (2007). ***Focus Groups: Theory and Practice***, 2nd edition, Vol. 20, Sage Publications, Newbury Park, CA.

285.Subedi, B.P., (2016). "Using Likert Type Data in Social Science Research: Confusion, Issues and Challenges", ***International Journal of Contemporary Applied Sciences***, Vol. 3, No. 2, February 2016, pp. 36-49. Available at: http://www.ijcas.net/ (12.3.19).

286.Sun, S.X., Zhao, J.L., Nunamaker, J.F., & Liu Sheng O.R., (2006). "Formulating the Data-Flow Perspective for Business Process Management", ***Information Systems Research***, Vol. 17, No. 4, December 2006, pp. 374-391.

287.Sun, S.X., and & Zhao, J.L., (2013). "Formal workflow design analytics using data flow modeling", ***Journal of Decision Support Systems***, Vol. 55, Issue 1, April 2013, pp. 270-283.

288.Tayi, G.K., & Ballou, D.P., (1998). "Examining Data Quality", ***Communications of the ACM***, Feb. 1998, Vol. 41, No. 2, pp. 54-57.

289.Tee, S.W., Bowen P.L., Doyle P., & Rohde F.H., (2007). "Data quality initiatives: striving for continuous improvements", ***International Journal of Information Quality***, Vol. 1, No. 4, pp. 347-367.

290.Teegavarapu, S., Summers, J.D., & Mocko, G.M., (2008). "Case Study Method for Design Research: A Justification". In ***Proceedings of the ASME International Design Engineering Technical Conference (IDETC/DTM 2008)***, August 3-6, 2008, New York, NY, pp. 1-9.

291.Tepandi J., Lauk, M., Linros, J., Raspel, P., Piho, G., Pappel, I., & Draheim, D., (2017). "The Data Quality Framework for the Estonian Public Sector and Its Evaluation". In: ***Hameurlain A., Küng J., Wagner R., Sakr S., Razzak I., Riyad A. (eds) Transactions on Large-Scale Data- and Knowledge-Centered Systems XXXV. LNCS***, Vol. 10680. Springer, Berlin, Heidelberg, pp. 1-26.

292.Thomas, G., (2014). "Defining Data Governance", ***Data Governance Institute (DGI)***, Available at: http://www.datagovernance.com/adg_data_governance_definition (3.3.19).

293.Timmerman, Y., & Bronselaer, A., (2019). "Measuring data quality in information systems research", ***Journal of Decision Support Systems***, Vol. 126, pp. 1-7.

294.Tremblay, M.C., Hevner, A.R., & Berndt, D.J., (2010). "Focus Groups for Artifact Refinement and Evaluation in Design Research", ***Communications of the Association for Information Systems (CAIS)***, Vol. 26, Article 27, June 2010, pp. 599-618.

295.Trčka, N., Aalst, W.M.P., van der, & Sidorova N., (2009). "Data-Flow Anti-Patterns: Discovering Dataflow Errors in Workflows", in: ***Advanced Information Systems Engineering (CAiSE 2009)***, (LNCS 5565), Springer-Verlag, Berlin, pp. 425-439.

296.Ullah, A., & Lai, R., (2011). "Modeling business goal for business-IT alignment using requirements engineering", ***Journal of Computer Information Systems***, Spring 2011, pp.21-28.

297.Ulrich, D., & Smallwood, N., (2004). "Capitalizing on capabilities", ***Harvard Business Review***, June 2004, pp. 119-127.

298.Ulrich, F., (2014). "Multi-perspective enterprise modeling: foundational concepts, prospects and future research challenges", ***Software & System Modeling***, Vol. 13, pp. 941-962.

299.Umar, A., Karabatis, G., Ness, L., Horowitz, B. & Elmagardmid, A., (1999). "Enterprise Data Quality: A Pragmatic Approach", ***Information Systems Frontiers***, Vol. 1, No. 3, pp. 279-301.

300.Vaishnavi, V.K., & Kuechler, W., (2004). "Design Research in Information Systems", Available at: http://desrist.org/desrist/ (31.1.12).

301.Vaishnavi, V.K., & Kuechler, W., (2015). ***Design Science Research Methods and Patterns: Innovating Information and Communication Technology***, 2nd ed., CRC Press, An Auerbach Publications, New York.

302.Vaisman A. (2006). "Requirements Elicitation for Decision Support Systems: A Data Quality Approach". In: ***Proceedings of the Eighth International Conference on Enterprise Information Systems - ISAS***, pp. 316-321

303.Vaknin, M., & Filipowska, A., (2016). "Information Quality Framework for the Design and Validation of Data Flow within Business Processes - position paper", in: ***Proceedings of the 19th International Conference on Business Information Systems (BIS 2016),*** Workshop on Business and IT Alignment (BITA). 6-8 July 2016, Leipzig, Germany, pp. 158-168.

304.Vaknin, M., & Filipowska, A., (2017). "Modelling of data dependencies for executable business processes", ***STUDIA OECONOMICA POSNANIENSIA (SOEP)***, Vol. 5, No. 12, pp. 107-125.

305.Vanbrabant, L., Martin, N., Ramaekers, K., & Braekers, K., (2019). "Quality of input data in emergency department simulations: Framework and assessment techniques", ***Simulation Modelling Practice and Theory***, Vol. 91, pp. 83-101.

306.Venable, J., Pries-Heje, J., & Baskerville, R., (2016). "FEDS: a Framework for Evaluation in Design Science Research", ***European Journal of Information Systems (EJIS)***, Operational Research Society Ltd., January 2016, Vol. 25, Issue 1, pp. 77-89.

307.Venkatraman, N., (1994). "IT-Enabled Business Transformation: From Automation to Business Scope Redefinition", ***Sloan Management Review***, Winter 1994, Vol. 35, No. 2, pp. 73-87.

308.Verner, J.M., (2008). "Quality Software Development: What do we need to improve in the software development process?", ***WoSQ'08***, May 10, 2008, Leipzig, Germany.

309.Wand, Y., Storey, V.C., & Weber, R., (2000). "An Ontological Analysis of the Relationship Construct in Conceptual Modeling", ***ACM Transactions on Database Systems***, Vol. 24, No. 4, December 1999, pp. 494-528.

310.Wand, Y., & Wang, R.Y., (1996). "Anchoring data quality dimensions in ontological foundations", ***Communications of the ACM***, Vol. 39, No. 11, Nov. 1996, pp. 86-95.

311.Wand, Y., & Weber, R., (2002). "Research Commentary: Information Systems and Conceptual Modeling - A Research Agenda", ***Information Systems Research***, Vol. 13, No. 4, Dec. 2002, pp. 363-376.

312.Wang, R.Y., (1998). "A Product perspective on total Data Quality Management", ***Communications of the ACM***, Vol. 41, No. 2, pp. 58-65.

313.Wang, R.Y., Kon, H.B., & Madnick, S.E., (1993). "Data Quality Requirements Analysis and Modeling", in: ***The 9th International Conference of Data Engineering***, Vienna, Austria, April 1993.

314.Wang, R.Y., & Strong, D.M., (1996). "Beyond Accuracy: What Data Quality Means to Data Consumer", ***Journal of Management Information Systems***, Vol. 12, No. 4, pp. 5-34.

315.Wassén, O., (2017). "5 essential elements of Data Governance", ***NodeGraph***, December 8th, 2017. Available at: https://www.nodegraph.se/5-essential-elements-of-data-governance/ (5.3.2019).

316.Watts, S.G., & Shankaranarayanan, A.E., (2009). "Data quality assessment in context: A cognitive perspective", ***Decision Support Systems***, Vol. 48, pp. 202-211.

317.Weber, K., Otto, B., & Österle, H., (2009). "One Size Does Not Fit All- A Contingency Approach to Data Governance", ***ACM Journal of Data and Information Quality***, Vol. 1, No. 1, Article 4, June 2009, pp. 4:1-4:27.

318.Weske, M., (2012). ***Business Process Management: Concepts, Languages, Architectures***, 2nd Edition, Springer-Verlag Berlin Heidelberg, New York, Inc. Secaucus, NJ, USA.

319.Weske, M., (2019). ***Business Process Management: Concepts, Languages, Architectures***, 3rd Edition, Springer-Verlag Berlin Heidelberg, New York, Inc. Secaucus, NJ, USA.

320.Wetzstein, B., Ma, Z., Filipowska, A., Kaczmarek, M., Bhiri, S. , Losada, S., López-Cobo, J.M., & Cicurel, L., (2007). "Semantic Business Process Management: A Lifecycle Based Requirements Analysis" .In: Hepp, M., Hinkelmann, K., Karagiannis, D., Klein, R., Stojanovic, N., (eds.): Semantic Business Process and Product Lifecycle Management, ***Proceedings of the Workshop SBPM 2007***, Innsbruck, April 7, 2007, CEUR Workshop Proceedings, pp. 1-11.

321.Wieringa, R.J., (2014). ***Design science methodology for information systems and software engineering***. Springer-Verlag Berlin Heidelberg, London, UK.

322.Willshire, M.J. & Meyen, D., (1997). "A Process for Improving Data Quality", ***Data Quality Journal***, Vol. 3, No. 1, Sep. 1997. Available at: http://www.dataquality.com/997meyen.htm (22.5.2007).

323.Woodall, P., Borek, A., & Parlikad, A.K., (2013). "Data quality assessment: The Hybrid Approach", ***Information & Management Journal***, Vol. 50, Issue 7, Nov. 2013, pp. 369–382.

324.Xingsen, L., Lingling, Z., Peng, Z., & Yong, S., (2009). "Problems and systematic solutions in data quality", ***International Journal of Services Sciences***, Vol. 2, No. 1, pp. 53-69.

325.Xu, H., Nord, J.H., Brown, N., & Nord, G.G., (2002). "Data quality issues in implementing an ERP", ***Industrial Management & Data Systems***, Vol. 10, No. 2(1), pp. 47–58.

326.Yang, Q., Ge, M., & Helfert, M., (2017). "Guidelines of Data Quality Issues for Data Integration in the Context of the TPC-DI Benchmark", in: ***Proceedings of the 19th International Conference on Enterprise Information Systems (ICEIS 2017)***, Vol. 1, pp. 135-144.

327.Yin, R.K., (2014). ***Case Study Research: Design and Methods***, 5th edition, Sage Publications, Thousand Oaks, CA.

328.Ying, C., Hong, L., & Zhengchuan, X., (2004). "An Evaluation Framework for Inter-organizational Business Process Modeling Techniques", in: ***The 8th Pacific Asia Conference on Information Systems - PACIS 2004***, Shanghai, China.

329.Zainal, Z., (2007). "Case study as a research method", ***Journal Kemanusiaan***, Vol. 9, pp. 1-6.

330.Zairi, M., & Sinclair, D., (1995). "Business process re-engineering and process management: A survey of current practice and future trends in integrated management", ***Business Process Re-engineering & Management Journal***, Vol. 1, No. 1, pp. 8-30.

331.Zatyko, S., (2017). "Data quality vs. data governance", ***Experian Research***, June 7, 2017. Available at: http:// https://www.edq.com/blog/data-quality-vs-data-governance/ (2.3.19).

332.Zoet, M., Versendaal, J., Ravesteyn, P., & Welke, R., (2011). "Alignment of Business Process Management and Business Rules", ***ECIS 2011 Proceedings***, Paper 34. Available at: http://aisel.aisnet.org/ecis2011/34 (15.9.17).

333.Zhang, R., Indulska, M., & Sadiq, S., (2019). "Discovering Data Quality Problems - The Case of Repurposed Data". ***Business & Information Systems Engineering Journal (BISE)***, Vol. 61, Issue 5, pp. 575–593.

334.Zhang, R., Jayawardene, V., Indulska, M., Sadiq, S. & Zhou, X., (2014). "A data driven approach for discovering data quality requirements", in: ***Proceedings of the 35th International Conference on Information Systems (ICIS 2014)***, 14-17 Dec. 2014, Auckland, New Zealand.

335.Zhu, H., Madnick, S.E., Lee, Y.W., & Wang, R.Y., (2014). "Data and Information Quality Research: Its Evolution and Future", in: ***Computing Handbook: Information Systems and Information Technology***, 3rd Ed., Taylor & Francis Group, LLC., Chapman & Hall/CRC Press, Boca Raton, FL., pp. 16.1-16.20.

Appendixes

Appendix A: A preliminary questionnaire for exploratory focus group (EFG) of domain experts, BP and IS practitioners

Preliminary questionnaire for exploratory focus groups participants

Level of information quality and deficiencies in work processes

Dear Employee,

Please find as an attachment an open-ended questions that examines the level of information quality and information deficiencies in your work environment in relation to work processes you are in charge of the organization for collecting information quality requirements.

Information Quality (IQ), relates to the features, dimensions, constraints and problems of the information that the organization manages and produces as an outcome of business processes. Emphasis is placed here on dimensions such as accuracy, completeness, consistency and timeliness of information.

Enterprise information management can be performed using a computerized information system and / or manually. Information items can be given by any format e.g. text, numeric value, observations values, document, form, etc. that exist today in your organization's processes.

The questionnaire is for academic research needs in the field of information systems and contains relevant questions for research.

This questionnaire is drafted in a masculine form for convenience only, but it refers to male and female as one.

Thank you in advance for taking the time to review and completion of the questionnaire☺

Sincerely,
Michael Vaknin

Part A: Personal details

First & Last name:		Date:
E-mail Address:	Tel./Cell:	
Company:	Occupation:	
Seniority at work: **1.** Up 2 years **2.** 3-5 years **3.** 6-10 years **4.** Over 10 years		

Part B: Information Quality (IQ) aspects

The following questions relate to information items used in the process or processes you are in charge of in your organization's occupation scope.

1. Specify the information items which you think are the most important and sensitive in process:

 ----

2. Specify cases or examples where you think an error or lack of any data item in process may result in a wrong decision in the process:

 ----

3. Specify cases or examples where you think an error or lack of a data item might cause the process to achieve undesired results:

 ----

4. Specify cases or examples where you think an error or lack of a data item in the process, might cause a waste of time and performance of unnecessary work:

   ----

5. Specify cases or examples where you think an error or lack of a data item might cause the process to get stuck (be in a deadlock situation):

Part C: Dependency & data items aspects

The following two questions relate to dependencies between data items used in the process or processes you are in charge of in your organization's occupation scope.

Note: dependence between data items represents a relationship between the values of two different data items: the value of the data item dependent on it and other data item value that depends on it. An existing error or lack of one value may lead to error or lack of information about a dependent item.

6. Specify cases or examples where you think there is a dependency between different data items in the process, which is critical to achieving the process' goals and objectives:

   ----

7. Specify cases or examples where you think there is a dependency on data items obtained from a source, internal or external to the organization, which is critical to achieving the process' goals and objectives:

[Source: own study]

Appendix B: A preliminary questionnaire for exploratory focus group (EFG) of BP and IS designers & experts

Preliminary Questionnaire for Exploratory Focus Group Participants

Level of Information Quality and Deficiencies in Work Processes

Dear BP and IS Expert,

Please find as an attachment an open-ended questionnaire that examines the level of information quality in the analysis and design of work processes and information systems in organizations and projects, based on your personal experience.

Information Quality (IQ), relates to a range of features, dimensions, constraints and problems of the information resource that the organization manages and produces as an outcome of business processes. Emphasis is placed here on dimensions such as accuracy, completeness, consistency and timeliness of information. Another important emphasis is placed on the dependencies between data items in processes and information systems and the resulting problems.

Enterprise information management can be performed using a computerized information system and / or manually. Information items can be given in any format, e.g. text, numeric value, observations values, document, form, etc. that exist today in your organization's processes.

The questionnaire is for academic research needs in the field of information systems and contains relevant questions for research.

This questionnaire is drafted in masculine form for convenience purposes only, but it refers to male and female equally.

Thank you in advance for taking the time to review and completion of the questionnaire☺

Sincerely,

Michael Vaknin

Part A: Personal details

<table>
<tr><td colspan="2">First & Last name:</td><td>Date:</td></tr>
<tr><td>E-mail Address:</td><td colspan="2">Tel./Cell:</td></tr>
<tr><td>Company:</td><td colspan="2">Occupation:</td></tr>
<tr><td colspan="3">Seniority at work: 1. Up 2 years 2. 3-5 years 3. 6-10 years 4. Over 10 years</td></tr>
</table>

Part B: Statements about information quality (IQ) aspects

The following ten statements and claims relate to a variety of information quality aspects that have an impact on the quality of the information system's design outcomes. Please indicate how much you agree or disagree with each of the following statements or claims by marking X in the appropriate box and the value that most represents your opinion in accordance with your professional experience, where:

5- Strongly agree, 4- Agree, 3- Moderately agree, 2- Disagree, and 1- Strongly disagree.

#	The statement / claim	1	2	3	4	5
1	At the analysis stage of business process and / or information system, there is little attention paid to the quality of information in the requirements definition.					
2	In process analysis and specification there is little (if any) use of tools, models or methods for analyzing the quality of information in defining end user requirements.					
3	There is a lack of tools, models and methods for BP & IS analysts and designers to achieve high quality analysis with information quality requirements.					
4	Identifying the dependencies between data items in a process is essential to obtaining high-quality information system specification outcomes.					
5	BP & IS analysts and designers neglect the idea of the dependencies between various data items in BP & IS specification.					
6	I need a method helping me to identify earlier the dependencies between the data items and potential failures expected from the dependencies.					
7	Poor quality information causes or may cause the achievement of undesired outcomes and decisions and poor quality of process performance.					
8	Poor information quality products can cause an organization's process (s) to get stuck.					
9	Poor-level information quality outcomes can cause unnecessary delays in the process and failure to meet the planned time for the information system implementation.					
10	Poor-level of information quality outcomes often result in unnecessary work, repairs and are a waste of valuable work time.					

Part C: Process analysis and specification aspects

The following questions relate to the preliminary studying phase and the analysis of BPs requirements in the organization as a preliminary stage for the specification and design of the information system.

1. Specify the key tools, methods, or models that you currently use for analysis and documenting the process requirements and specification:

----

2. Going further into q. 1, indicate major weaknesses, problems, or limitations that you encounter when using them to define information quality requirements:

----

3. Indicate whether you identify and handle the aspects and dimensions of the information quality (as mentioned above) and the dependency that exists between the various data items separately during the analysis of BP. If so, indicate how:

----

4. What are your expectations and requirements for the development of a new method for handling information quality aspects and dependencies between data items and what do you think the main characteristics and / or components of such a method should be?

[Source: own study]

Appendix C: List of exploratory focus groups (EFGs) and their participants[10]

# EFG	# participant	Occupation	Work experience (in years)	Organization name & home page	The organization sector
1	1	VP Operation & Quality	25	**Ocean group Ltd – Relocation & Forwarding Division** http://www.ocean-il.co.il	Relocation, Int'l Forwarding, Moving and Storage services
	2	VP Sales & Marketing	8		
	3	Sales & Accounts Manager	15		
	4	Forwarding supply chain Manager	5		
	5	Export coordinator	15		
	6	Export coordinator	7		
	7	Export coordinator	6		
	8	Control Manager	17		
2	9	Logistic Purchasing & Storage Manager	12	**Ocean group Ltd – Operation & Logistic Division** http://www.ocean-il.co.il	Relocation, Int'l Forwarding, Moving and Storage services
	10	Transportation & Packing/ Unpacking crews Manager	6		
	11	VP Business Development	22		
	12	VP Relocation and Real Estate division Manager	8		
	13	IS/IT Manager	20		
3	14	CIO (in retirement)	35	**Bet Shemesh Engines Ltd** http://bsel.co.il	Developing and Manufacturing of Engine parts
	15	IS/IT & Operation Consultant	5	Freelance	Consulting & Implementing of IS & ERP Systems
	16	IS/IT & Operation Consultant	30	**Epilog Technology** http://www.epilog.co.il	Consulting & Implementing of IS & ERP Systems
	17	ERP Projects Manager	13	**Raan Engineering** http://www.raan.co.il	Consulting & Implementing of IS & ERP Systems and maintaining of IT infrastructures
	18	IS Analyst & Designer	5	**Israel Defense Forces** http://www.idf.il	Israeli Army

[10] The data appearing in the above table are valid to the date of filling out the questionnaires in the research.

	19	IS/IT Manager	20	**Emek - Regional High School** https://www.emek-school.org	Higher School Education
	20	IS Analyst & Designer	10	**Zefat Academic College** http://www.zefat.ac.il	Academic studies
	21	Software Development Manager	7	**Zefat Academic College** http://www.zefat.ac.il	Academic studies
	22	ERP Implement	4	**Medatech Ltd** http://medatech.com/	Consulting & Implementing of IS & ERP Systems and maintaining of IT infrastructures
	23	IS & BPM Analyst & Designer	8	**Shamir Optical Ltd** http://www.shamir.co.il	Manufacturer of bifocal lenses and molds Industry
	24	IS/IT & Operation Manager	5	**AVIVI Kitchens** http://www.avivi.com	Designing & Manufacturing Kitchens

[Source: own study]

Appendix D: EFG #1: Focus group script

1. **Date:** 2.3.15 at 10:00-12:00 AM

2. **Place:** OCEAN Group Ltd. (at conference room)

3. **Number of Participants:** 8 participants

4. **Type of Participants:** Domain experts and IS senior users (from a variety of positions with extensive professional experience).

5. **The EFG Topic / Purpose:**
   - Exposing FG participants to the researched problem and research motivation, from scientific and business needs aspects.
   - Collecting IQ aspects, problems and requirements from real world business environment.
   - Identifying possible characteristics and components for designing the new method.

6. **Preparation Tasks for FG Meeting:**
   - Coordination of the date of the meeting and the conference room with the administrative manager of the company.
   - Definition and invitation of participants for FG meeting: in consultation with the company's CEO and president, we defined the focus group participants, who are composed of eight managers and professional employees with many years of experience from various areas and activities at Ocean Group which directly related to all the company's sales export services, forwarding, logistics, and storage processes.
   - Preparation of PPT presentation.
   - Preparation of explanations of some important terms and various examples on the topics of information quality, dimensions of information quality and dependencies between information items.
   - Sending a reminder for participants to fill out an individual preliminary questionnaire (see Appendix A) as a preparation for the FG meeting.
   - Preparation and set up of projector, tape recorder, furniture, papers and writing tools for participants as well as a flip chart.

7. **The Course of the FG Discussion:**

7.1 Introduction

- The moderator introduced himself and his professional background to the participants and thanked them for participating in this meeting.
- The moderator presented the participants with an explanation of what the meaning and roles of the FG meeting are and how the group discussion will be conducted

with a discussion on how participants should discuss and act during the session and what the expectations from them are.

- The moderator explained to the participants that everything they say in this meeting remains confidential and will be used for academic purposes only and their real names will not be used in any report.
- The moderator explained to the participants why they are being asked to participate in this FG because: "Each of you works in a variety of positions with extensive professional experience and knowledge and we need your help in developing a new method to improve the way we are conducting business process design today ".
- The moderator explained to participants the researched problem's scope and research motivation from scientific and business needs aspects and clarified the research's practical challenges.
- The moderator explained to participants the ideas of data quality (DQ), DQ dimensions and the dependencies between data items in a process in general, and their potential errors, problems and failures.
- The moderator explained to participants the relationship and the important role of analysis and design of business processes and their quality in the organization for the design and quality improvement of an information system.
- The moderator introduced to participants the meeting objectives:
  - To understand how we can include your vast experience and knowledge about data quality problems into a new method to improve the quality of business processes design and how it will affect your processes' quality and your organizational information systems.
  - To get your suggestions on how you would like to improve these processes.
  - To get your opinion on the way these should be presented.

7.2 Discussion Topics and Cases

- The moderator presented to participants the planned cases for discussion in each team level based on their filled preliminary questionnaires.
- These are the 3 issues we need to consider here and cover today. The first topic includes six cases for discussion.

1. For each of the next four cases, specify significant outcomes of DQ problems based on cases or examples from your experience where:

<u>The DQ problems and outcomes at process level:</u>

1.1 DQ problems of any data item values in process and their impact on decision making in process.

1.2 DQ problems of any data item values in process and their impact on process performance and its quality.

<u>The dependency idea and DQ problems at process level:</u>

1.3 The dependency between data items in a given process and its relation to process' goals.

1.4 The dependency in a given process on information obtained from a third-party source.

2. General suggestions to improve the analysis and design of business processes.

3. Define their expectations and characteristics of a new method for improving the analyzing of information quality and the role of information dependencies between data items in processes design stage for a new information system.

- In order for us to cover all these topics and cases we have planned and to be more effective as well as meet the time planned for this meeting, we divided the group into three small teams, with each team having an internal discussion (in 25-30 minutes) for all topics and cases before discussing them as a group, and define a representative as a speaker for each team to present their outputs and what is agreed upon between them.

8. The EFG Discussion Main Results:

- The participants were asked to start the internal discussion about each of the above topics and cases before discussing them as a group.
- After 30 minutes, the participants were asked to summarize their discussion and to draw a list of final conclusions.
- Each team presented its own significant outcomes per each case from the internal discussion.
- All the results of each team were collected and written on the board for general group discussion and decision.

- The following is a summary of all significant outcomes and conclusions collected by all teams grouped by the above topics and cases:

1. **The DQ problem and outcomes at process level:**

1.1 DQ problems of any data item values in process and their impact on decisions making in process:

- DQ problems may lead to or result in the wrong decision in process.
- There is a great need to ensure the accuracy and completeness of all data items usage in process to achieve high quality decisions.
- Good decisions are critical to process success and are based on high quality information.
- The significant dimensions of causes or results of wrong decision in process are accuracy and completeness of data items.
- There is a need to identify potential failures related to DQ problem in advance to achieve high quality in process design.

1.2 DQ problems of any data item values in process and their impact on process performance and its quality:

- These kinds of this DQ problems can lead to a waste of resources and unnecessary expenses in the process.
- These kinds of this DQ problems can lead to achieve certain outcomes that are not in the process' goals.
- DQ problems in any data item values in process can cause the process to achieve undesired results in the process.
- DQ problems of any data item values in process can cause a waste of time and valuable resources and to perform unnecessary tasks.

- DQ problems of any data item values in process can lead the process to get stuck (i.e. be in deadlock situation).

<u>The dependency idea and DQ problem:</u>

1.3 The dependency between data items in a given process and its relation to process' goals:

- There is a great need to identify data items dependencies which relate to process' goals and their potential failures related to DQ problem in advance.
- The dependency between different data items in the process is critical to achieve the process' goals.
- Being able to identify different paths of dependencies relationships between data items in process in advance.

1.4 The dependency in a given process on information obtained from a third-party source:

- Can be a delay in receiving information from an internal or external source.
- There is no full control in receiving of information from a third-party source.
- The information from a third-party source can also suffer from a low level of DQ.
- Can be internal or external to the organization.
- The dependency is mostly on an external third party to the organization.
- This kind of dependency is critical to achieve the process' goals and its outcomes.
- This kind of dependency in process on information from a third-party source can lead the process to get stuck (i.e. be in deadlock situation).

2. <u>Suggestions to improve the analysis and design of business processes:</u>

- Investing more time and effort into early learning of the processes by BP and IS analysts and designers.
- Identifying the potential DQ problems and their impact on process performance and its quality in advance to improve the definitions of end-user's requirements.
- Analyze the process in-depth in full cooperation of BP and IS analysts and designers with the domain experts and end users.
- The existing methods are inadequate and new tools and/or methods for the process analysis are needed.
- To prevent duplication, shortages and deficiency in information management and provide the ability to cross various data items.
- To be able to identify potential failures related to dependency between data items.

3. <u>Requirements, characteristics, features and expectations needed for a new method:</u>

- A new method that can help us to see the "big picture" and fully analyze and define high quality requirements and to achieve success in IS projects.
- A new method that can help us in improving the information management and mapping the entire flows of information in the process in advance.
- Identify the potential DQ problems and their impact on process performance and its quality in advance.
- A new method that can help us to reflect all dependencies that exist between information items in process.
- Be able to identify different types of dependency relationships in each process.

○ Be able to identify in advance the impact of different types of dependency relationships in a process. ○ Be able to identify potential failures related to dependency between data items in a process. ○ The new method should be of added value to BP and IS analysts and designers. ○ A method that will allow us to see the direct connection between the analysis and design of processes and information system design products.
9. Summary of the EFG meeting: • The moderator summarized the FG meeting and presented its main results and outcomes and describes what is expected to be the next stage in the research process. • The moderator thanks all participants for participating in this meeting and their important contributions and invited the whole group to lunch☺

[Source: own study]

Appendix E: Feedback form for confirmatory focus group (CFG) participants

Feedback Form for Confirmatory Focus Group Participant

Dear Manager/Employee,

Please find as an attachment a feedback form with open-ended questions to get your opinion on a new method developed for business processes design phase in organization or business from every sector.

The purpose of the proposed new method is to improve the level of information quality in business processes design phase based on identifying the dependencies between different information items in the work processes at your responsibility and in your work environment in the organization.

Information item can be given any text, numeric value, document, form, etc.

Information Quality (IQ), relates to the features and dimensions of the information that the organization manages and produces as an outcome of business processes. Emphasis is placed here on dimensions such as accuracy, completeness, consistency and timeliness of information Where:

Dependency between information items represents a connection between the values of two different information items. An error or lack of value may result in an error or lack of the value of the dependent data item.

- ***Accuracy*** *- relates to the question of how important it is for an information item to be accurate and correct.*
- ***Completeness*** *- refers to the question of how important it is that an information item is represented fully and not partially or missing.*
- ***Consistency*** *- relates to the question of how important it is that an information item is relevant, unambiguous, and compatible with previous data items.*
- ***Timeliness*** *- refers to the frequency and acceptance on time of any information item from third party.*

Enterprise information management can be performed using a computerized information system and/or manually with respect to documents and forms that exist today in the organization.

The feedback form is for academic research needs in the field of information systems and contains relevant questions for research.

This questionnaire is drafted in masculine form for convenience only, but it refers to male and female equally.

Thank you in advance for taking the time to review and completion of the feedback form!

Sincerely,

Michael Vaknin

Part A: Personal details

First & Last name:		Date:
E-mail Address:	Tel./Cell:	
Company:	Occupation:	
Seniority at work: **1.** Up 2 years **2.** 3-5 years **3.** 6-10 years **4.** Over 10 years		

Part B

Here are questions about the new method you're currently exposed to within the Focus Group.

Please try to answer all questions in as much detail as you can about each one.

1. Is the description of the new method and its stages clear and fully understood?

   If not, please provide a specific description of the method and/or steps you did not understand:

   ----

2. Indicate your impression and your general opinion about the proposed new method:

   (Note: In your reply, please specify whether the method seems perfect or partial, appropriate/inappropriate for your organization, etc.)

   ----

3. Would you recommend immediately applying the new proposed method in your organization? Please specify:

   If so, please specify how improving the information quality through this method can contribute to the organization and to you in achieving the objectives of the process or area in your responsibility for better quality?

   ----

4. In your opinion, what do you think are the potential challenges (positives and negatives) that your organization is likely to face when deciding to implement the new method?

Please specify.

----

5. Specify cases or examples where you think the proposed method cannot support or alternatively can be an obstacle in achieving an improvement in the quality of the process performance:

----

6. Specify here any additional thing to what you have already mentioned, which comes to mind and should be taken into account and relate them to the proposed new method:

[Source: own study]

Appendix F: List of confirmatory focus groups (CFGs) and their participants[11]

# CFG	# participant	Occupation	Work experience (in years)	Organization name & home page	The organization sector
1	1	Accountant & IS practitioner	20	**Elyakim Ben-Ari Ltd** http://www.benari.co.il	Civil engineering & quarries
	2	Economist & IS practitioner	7		
	3	CPA & IS practitioner	7		
	4	Accountant & IS practitioner	10		
	5	Bookkeeper & IS practitioner	14		
2	6	President & CEO	35	Ocean group Ltd http://www.ocean-il.co.il	Relocation & Int'l forwarding services
	7	VP Business development	22		
	8	VP Relocation & real estate division manager	8		
	9	CRM & Digital media manager	7		
	10	VP Resources & infrastructures	2		
	11	Marketing & sales manager	2		
	12	Export coordinator	7		
	13	Customers service & Insurance claims manager	17		
3	14	Department manager & IS analyst	15	**ISCAR Ltd** http://www.iscar.co.il	Manufacturer of metalworking tools
	15	Software development manager	8		
	16	Technologist of computerized measurement processes	8		
	17	IS/BPM analyst & implementation	4	**Zefat Academic College** http://www.zefat.ac.il	Academic studies
	18	Operation & IS/IT manager	2	**AVIVI Kitchens** http://www.avivi.com	Designing & manufacturing kitchens
	19	Software design & development manager	7	**Shamir Optical Industry Ltd** http://www.shamir.co.il	Manufacturer of bifocal lenses & molds
	20	IT Manager	5	**Israel Defense Forces** http://www.idf.il/	Israeli army
	21	Software developer	7	**Zefat Academic College** http://www.zefat.ac.il	Academic studies

[11] The data appearing in the above table are valid to the date of filling out the questionnaires in the research.

	22	IS practitioner & Import division manager	8	**Mazonit I.S.R.L Ltd** http://www.mazonit.co.il/	Importing & marketing of wood surfaces and boards
	23	IS Practitioner & Import division Manager	10	**Tefal - Newpan Ltd** http://www.newpan.co.il	Home appliances
4	**24**	IS analyst & Projects manager	15	**The Israel Airports Authority** http://www.iaa.gov.il	Maintain, operate, develop & manage the airports in Israel
	25	Mechanical engineer & IS practitioner	12		
	26	IS development manager	10		
	27	Projects manager	15		
	28	Planning & Organizational engineer	15		
	29	IT infrastructures manager	9		

[Source: own study]

Appendix G: CFG #3: Focus group script

1. **Date:** 20.3.19 (at 18:00-20:00 PM)

2. **Place:** Zefat Academic College (at conference room)

3. **Number of Participants:** 10 participants

4. **Type of Participants:** IS and BP professionals (managers and employees) with many years of experience from various organizations.

5. **The CFG Topic / Purpose:**
   - Exposing FG participants to the new artifact i.e. our proposed DQDP method for professional feedbacks and its practical aspects.
   - To demonstrate the structure and steps of proposed DQDP method from IQ aspects, problems and requirements.
   - To demonstrate the utility and usability of our artifact i.e. new proposed DQDP method in the application field.
   - To evaluate the utility and usability of the new proposed DQDP method through practical experience with the DQDP method by the FG participants and their feedbacks.
   - To identify the challenges and obstacles that the organization is expected to face in adopting our new method.
   - To identify new directions for the development and improvement of the proposed method in the future.

6. **Preparation Tasks for CFG Meeting:**
   - Definition and invitation of participants for CFG meeting.
   - Coordination of the date of the meeting and the conference room with the participants.
   - Preparation of PPT presentation about the new proposed DQDP method.
   - Preparation of explanations of some important terms and various examples and use cases to illustrate the utility and usability of the new proposed DQDP method.
   - Preparation of feedback form for participants to fill out after the FG meeting (see Appendix E).
   - Preparation and set up of projector, tape recorder, furniture, flip chart, papers and writing tools for participants.

7. **The Course of the FG Discussion:**

7.1 Introduction

1. The moderator introduced himself and his professional background to the participants and thanked them for participating in the meeting.

2. The moderator presented the participants with an explanation of what the meaning of CFG meeting is and how the group discussion will be conducted.
3. The moderator explained to the participants that everything they say in this meeting is to remain confidential and used for academic purpose only and their real names will not be used in any report.
4. The moderator introduced the meeting objectives to the participants:
    - To get your opinion and feedback about the utility and usability of the new proposed DQDP method in the application field.
    - To identify the challenges and obstacles that the organization is expected to face in adopting our new method.
    - To identify new directions for the development of the proposed method and to improve it in the future.
5. The moderator explained to the participants why they are being asked to participate in this CFG because: "Each of you is working in a variety of positions in BP and IS domain with extensive professional experience and knowledge and we need your help in developing a new method in order to improve the way we conduct business process design today".
6. The moderator explained to participants briefly, the researched problem and research motivation from a scientific and business needs aspect.
7. The moderator explained to participants the ideas of data quality, DQ dimensions and the dependencies between data items in a given process in general, and their potential errors, problems and failures in particular.
8. The moderator explained to participants the relationship and the important role of analysis and design of business processes and their quality in the organization for the design and quality improvement of information system development.
9. The moderator presented to participants the concept of new proposed artifact i.e. DQDP method, as a solution to researched problem and research motivation, from scientific IQ aspects and business needs aspects.
10. The moderator demonstrated to participants the new DQDP method and its structure and each step in detail, based on examples and cases from real world about:
    - DQ problems of any data item values in process level and their impact on decision making in process.
    - DQ problems of any data item values in process level and their impact on process performance and its quality.

7.2 The Discussion Topics

<u>**The CFG discussion focused on four topics:**</u>

1. **Practical experience with the DQDP method by the FG participants:**
    - The FG participants were asked to analyze and characterize a mini process and define its requirements in IQ terms based on a preliminary case study.
    - The process specification should be in two stages: In the first stage, they were asked to define process specification according to any method they knew and without using the components and phases of the new method. In the second stage, they were asked to create new specification for the same case study based on our new method steps.

- The participants were asked to present their results' differences and their general opinion on the DQDP method.

2. **General discussion about the utility of DQDP method through the following questions:**
   - The participants were asked to indicate how understandable and clear the new proposed DQDP method is.
   - The participants were asked to indicate their general impression from DQDP method.
   - Can the new method be suitable for use in your organization?
   - Would you recommend the new method to the relevant parties and people in the organization or to your professional colleagues?
3. **General discussion about the main contributions and challenges in DQDP method adoption:**
   - The main contributions and benefits to the organization from implementing the new method.
   - The main challenges and obstacles that the organization is expected to face in adopting our new method.
4. **General discussion about new directions for the development and suggestions to improve the proposed DQDP method in the future:**
   - Defining new directions for the development and improvement of the proposed method in the future.

- In order for us to cover all these topics that we have planned, and to be more effective as well as meet the time planned for this meeting, we divided the participants into 5 couples, with each couple having an internal discussion (lasting 30 minutes) for all topics and cases before discussing them as a group, and to present their outputs and what is agreed upon between them.
- The participants were asked to start the internal discussion about each of the above topics and case study before discussing them as a group.
- After 30 minutes, the participants were asked to summarize their discussion and to draw a list of final conclusions in 10 minutes more.
- Each couple presented its own significant outcomes per each case study from the internal discussion.
- All the results of each couple were collected and written on the board for general group discussion and decision.

8. The CFG Discussion's Main Results:

- The following is a summary of all significant responses and conclusions collected by all participants grouped by the above topics:

1. **Practical experience with the DQDP method by the FG participants:**
   - The new method allows us to see the full picture of the process design in advance with emphasis on the information items and their dependencies.
   - There is a great need to identify potential failures related to DQ problem in advance to achieve high quality in process design.

- There is a significant difference when analyzing a process and identifying the potential failures related to the information quality in advance and the one dealing with it in real time and fixing it.
- The new method makes the analysis and design phase of a process more transparent, efficient and of higher quality.
- DQ problems in any data item values in process (e.g. wrong or missing values) can cause the process to achieve undesired results in the process.
- DQ problems may lead to or result in the wrong decisions in process.
- There is a great need to ensure the accuracy and completeness of all data items used in process to achieve high quality decisions.
- Good decisions are based on high quality information and critical to process success.
- The new method helps us to identify dependencies between different data items in advance that are related and critical to achieve the process' goals and their potential failures in relation to DQ problem.
- The new method enables to identify different paths of dependency relationships between data items in process in advance.
- The new method helps us to identify potential failures related to a delay or missing pieces in receiving information items from internal or external source, which can lead the process to a deadlock situation since there is no full control in receiving the information from a third party source.
- The new method helps us to identify in advance potential of a waste of time and valuable resources or to perform unnecessary tasks caused by DQ problems of any data item values in process.

2. **General discussion about the utility of DQDP method through the following questions:**
   - The participants were asked to indicate how understandable and clear the new proposed DQDP method is.
     - ✓ *The results clearly show that the majority of CFG participants, a little more than 80%, indicated that the new proposed DQDP method is understandable to very understandable and clear (see also **Figure 48**).*
   - The participants were asked to indicate their general impression from DQDP method.
     - ✓ *In principle, we received a wide range of answers and definitions for the proposed DQDP method, all of which are positive; e.g. very interesting, very comprehensive, perfect etc. (See details on page 193).*
   - The participants were asked to indicate if the new method can be suitable for use in their organizations and would they recommend it to the relevant parties in their organization or to other professional colleagues?
     - ✓ *The results clearly show that the majority of CFGs participants, about 86%, indicated that the new method be suitable for use in their organizations and they would recommend to the relevant professional people and to other colleagues to implement the new DQDP method in their organization. In addition, some participants (14%) stated that they would recommend applying the new method under some conditions (see also **Figure 50** on page 193).*

3. **General discussion about the contributions, challenges and obstacles in DQDP method adoption:**
   - The main contributions and benefits to the organization from implementing the new DQDP method raised by the CFGs participants are:
     - ✓ The new method allows us to see the full picture of the process design in advance with emphasis on the data items and their dependencies.
     - ✓ The new method allows us to identify potential failures related to DQ dimensions in advance to achieve high quality in process design.
     - ✓ The new method makes the analysis and design phase of a process more transparent, efficient and of higher quality.
     - ✓ The existing methods and techniques are inadequate and new methods or techniques to improve the process analysis and design are needed.
     - ✓ The new method can helps us to achieve good decisions based on high quality information and is critical to process success.
     - ✓ The new method can helps us in improving the information management and mapping the entire flows of information in the process and prevent duplication or shortages in data items in advance.
     - ✓ The new method helps us to identify dependencies between different data items in advance which are related and critical to achieve the process' goals and their potential failures and problems related to DQ dimensions.
     - ✓ The new method enables us to identify different paths of dependency relationships between data items in process in advance.
     - ✓ The new method helps us to identify potential failures related to delay or missing information items upon receipt from internal or external source, which can lead the process to a deadlock situation since there is no full control in receiving of information from a third party source.
     - ✓ The new method helps us to identify in advance potential of a waste of time and valuable resources or to perform unnecessary tasks caused by DQ problems of any data item values in process and to eliminate them.
     - ✓ The new method allows us to see the direct connection and dependency between the analysis and design of processes and information system design outputs.

   - The challenges and obstacles stated by the CFG participants which the organization is expected to face in adopting our new method:
     - ✓ Appropriate training is required to deal with the complexity of the new method and to implement it successfully.
     - ✓ It is necessary to incorporate in the implementation of the new method all the parties involved in the process while coordinating expectations between them.
     - ✓ There is a need to redefine the existing processes and work procedures associated with the new method.
     - ✓ There is a need for collaboration with relevant managers and employees and the ability to cope with employee objections to changes in work patterns.

✓ In some cases, it requires review and reassessment of existing information systems while addressing issues related to information quality and adapting to the requirements of the new method.
4. General discussion about new directions and suggestions for development of the proposed DQDP method in the future: ✓ To automate the new proposed method in the future. ✓ Further to the method automation, receiving recommendations about dependencies, conflicts and potential failures related to the information quality in advance.
9. Summary of the CFG meeting: • The moderator summarized the CFG meeting and presents the main results and outcomes and describes what is expected to be the next stage in the research process. • The moderator thanks all participants for participating in this meeting and emphasized their important contributions to the FG session☺

[Source: own study]

Appendix H: Examples of dependencies and DQ deficiencies based on case study

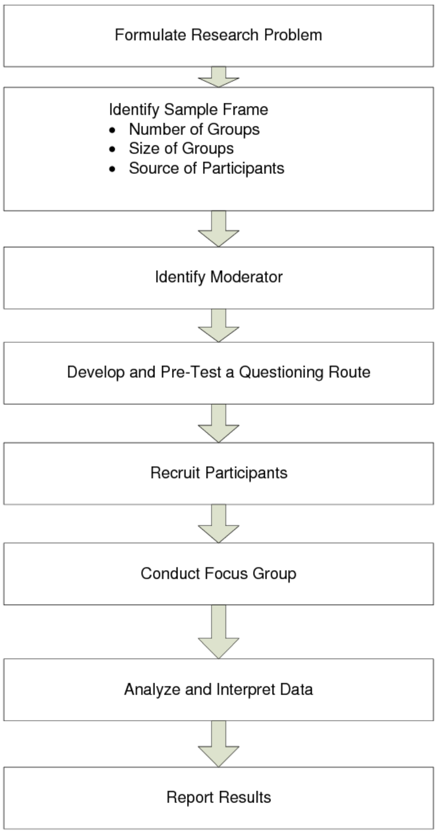
DQ dimension	Examples of dependencies	The potential deficiencies	The possible results of deficiencies	The implications of deficiencies at the process level	Examples of DQ requirements
Accuracy	Customer type ⇨ private customer or Public/Business customer $(d_6 \rightarrow d_{24})$	• Wrong value about customer type.	• Can lead to wrong charge in quotation according to customer type.	• Achieving data items not in the process goals set.	• Identify and define whether data item value is to be a mandatory or optional value based on the depending on the other data items values.
Accuracy	Shipping destination ⇨ Port at destination $(d_9 \rightarrow d_{10}/d_{24})$	• Wrong value about requested destination • E.g. Australia vs. Austria; Birmingham city in UK vs. Birmingham city in U.S. state of Alabama. • Oakland city in U.S. state of California vs. Oakland city in New Zealand.	• Customer cargo will delivering to a wrong destination. • Wrong value about requested destination can lead to issue a quotation to customer with wrong prices.	• Achieving data item values not in the process goals set.	• Check that the destination country exists in countries list. • Ensure that origin country ≠ destination country. • Create a link between country and possible ports.
Completeness	Customer type ⇨ Charge method $(d_6 \rightarrow d_{19}/d_{24})$	• Missing value about customer type. • E.g. Private vs. Business customer.	• Can lead to dilemma and missing charge in quotation according to customer type. • Can lead to wrong charge in quotation according to customer type.	• The process could be in a deadlock situation. • Achieving data items not in the process goals set.	• Define Customer type as a mandatory value. • Create a check box or selection option.
Completeness	Assessor cargo items list report ⇨ Total cargo volume evaluation $(d_{17} \rightarrow d_{18})$	• Missing items in cargo items list.	• Can lead to wrong value in total cargo volume evaluation. • Can lead to wrong charge in quotation according to total cargo volume.	• Achieving data items not in the process goals set.	• Create a data check box or selection option to ensure the completeness of all data items values. • Ensure the completeness of all input and output data items representation.
Consistency	Customer address ⇨Assessor evaluation mission $(d_5 \rightarrow d_{16})$	• Two different customer addresses (e.g. office or home) but only one of them is necessary to the evaluation mission by assessor.	• Irrelevant or redundant values can lead to choose the wrong address where the ordered service should be done.	• The process could be in a deadlock situation.	• Ensure the unambiguous of data items values representation in the process. Ambiguous or meaningless in data items values representation in process should be eliminated.
Consistency	Ordered service⇨ responsible department $(d_8 \rightarrow d_{13})$	• Two different values about ordered services in customer application (e.g. export & storage)	• Each ordered service in customer application (e.g. export & storage) belongs to one	• The process could be in a deadlock situation.	• Identify ambiguous or inconsistency representation of data item value in process i.e. different data item

DQ dimension	Examples of dependencies	The potential deficiencies	The possible results of deficiencies	The implications of deficiencies at the process level	Examples of DQ requirements
		set or map to the same responsible department without separation.	responsible department.		values in RW mapping to the same data item in process and vice versa.
Timeliness	Shipping date at Int'l Forwarder⇨ packing date of customer cargo $(d_{31} \rightarrow d_{32})$	• Shipping vessel schedule plan not arrived on time from int'l forwarder.	• Difficult to set the packing date of customer cargo.	• The process could be in a deadlock situation.	• Set a time monitoring to define whether data item value to be changed within a given time.
Timeliness	assessor cargo items list report ⇨ Total cargo volume evaluation $(d_{17} \rightarrow d_{18})$	• Assessor cargo evaluation report not arrived on time from assessor.	• Difficult to set the total cargo volume.	• The process could be in delay and/or a deadlock situation.	• Ensure that the expected data item will appear on time as planned according to the process.
Timeliness	Shipping requested date ⇨ Shipping date at Int'l Forwarder ⇨ Packing date of customer cargo $(d_{11} \rightarrow d_{31} \rightarrow d_{32})$	• shipping vessel schedule plan not arrived on time from int'l forwarder.	• Difficult to set the packing date of customer cargo. • Delay in shipping of customer's cargo to the destination.	• The process could be in a deadlock situation.	• Prevent time delay between the change in data item value of RW representation and the required modification of data item value representation in the process. • Set a reminder on weekly basis to get the vessel schedule plan from int'l forwarder.

[Source: own study]

Appendix I: Focus group session steps

This appendix summarizes the steps of the focus group method, which appear in the Figure below. As mentioned above (see sub-chapter **3.3.2**; p. 105), the explanation of each step is based on Tremblay et al. (2010) with several additions from other several academic sources and studies about focus groups research.



[Source: Tremblay et al., 2010]

- <u>**Step 1: Formulate a research problem**</u>

The process begins by identifying the main aim and defining the key research goals of the study. In order to effectively define the content and focus groups, the research goals and problem must be clearly identified. Based upon the research goals, a list of questions is prepared as guidance for each focus group discussion session.

The aim of these focus groups study is obtaining feedback and opinions from participants i.e. domain experts, BP and IS analysts, designers and practitioners about the need, problems, requirements and expectations for designing a new method to cope with DQ aspects and deficiencies at process design.

- **Step 2: Identify sample frame**

Three decisions are made in this step: (1) number of each type of focus group to run, (2) the desired number of participants in each group, and (3) what type of participant to recruit.

Tremblay et al. (2010) suggest conducting one pilot focus group, two explanatory focus groups and at least two confirmatory focus groups. The ideal number of participants depends on the objective of the focus group study: smaller groups require each participant to be more active while larger groups may lead to social loafing (Brandtner, 2015). According to Tremblay et al. (2010) larger focus groups exceeding six participants may be difficult to apply in a design science project since the subject matter in such projects is more complex than topics of traditional focus groups. Furthermore, for design research, the participants should be from a population familiar with the application environment for which the artifact is designed so they can adequately inform the refinement and evaluation of the artifact. Care should be taken that the participant groups are from a similar pool for both EFGs and CFGs, so that CFGs are in fact confirming a final design (Tremblay et al., 2010).

- **Step 3: Identify moderator**

The role of the moderator is critical in a focus group session. He is central to the discussion not only by managing existing relationships but also by creating a relaxed and comfortable environment for unfamiliar participants. Furthermore, the moderator should facilitate discussion but not allow his or her own opinions to influence the discussion and his main task is to listen and investigate deeper when necessary. It is often necessary to emphasize participant points to ensure that the contribution was correctly understood. For design research, the moderator needs to have a clear understanding of various aspects of the designed artifact. The moderator should also be familiar with the future artifact and be comfortable presenting its characters to focus group participants.

- **Step 4: Develop and pre-test a questioning route**

The questioning route is the agenda for the focus group. In the questioning route we set the direction for a group discussion and it should closely align with the research goals. The discussed topics to be ordered by importance, and within those topics, the questions are ordered from general to specific. For a designed artifact, this means beginning with an explanation of the motivation behind the design of this artifact, followed by a broad explanation of different scenarios on where and how the artifact could be utilized. The questions in a focus group session are mostly open ended. In principle, it is preferable to ask each participant to fill out an individual preliminary questionnaire with demographic information as a preparation for the focus group meeting and to base the discussion on it.

- **Step 5: Participant recruitment**

Participants' recruitment is a very important step and should be done very carefully (Nyumba et al., 2018). The recruitment of focus group participants is not a random selection, but rather is based on characteristics and backgrounds of participants in relation to the artifact that is being discussed. Furthermore, the value of the method is very sensitive to the experience and insight of the participants (Kontio et al., 2004). Design researchers prefer to recruit participants that are familiar with the application environment and would be potential users of the proposed artifact. A diversity of participants will potentially produce more creative ideas, but segregation of participants based on skills and knowledge may provide more in-depth tradeoffs in values and success measures. Care should be taken that the participant groups are from a similar pool for both EFGs and CFGs, so that CFGs are in fact confirming a final design.

Another important consideration is the number of participants to be invited for FG discussion. It is generally accepted that between 5-8 participants are enough for FG discussion. Phone calls and e-mails should be placed at least a month before the focus groups are planned. A few days before the focus groups the participants should be reminded.

- **Step 6: Conduct the focus group session**

Focus group sessions should be pleasant, enriching and stimulating event for both the participants and the moderator. Greeting the participants when they arrive is a good first step. The participants are generally seated in a U-shape arrangement to encourage collaboration and allow space for the moderator to demonstrate the presentation and details about artifact. A good approach is to get to know the participants before the questioning route begins.

Basically, the meeting should be documented using a protocol and discussion scripts and may be recorded if necessary, by video and/or audio medium. The moderator or someone on his behalf will take careful notes, noting any strong statements, ideas, declarations, reactions, etc. of participants. Moreover, time management is also important when conducting a focus group. The duration of the focus group discussion meetings should be about 1-2 hours. Participants are likely to suffer from fatigue when discussions are longer than two hours. A moderator should be able to recognize when all possible issues for a topic have been covered and move on to the next topic.

- **<u>Step 7: Analyze and interpret data</u>**

Focus group discussion usually yields both qualitative and observational data where analyses and interpretation of the focus group discussions can be challenging. Several techniques that are used for qualitative data analysis can be considered.

One possible approach is template analysis. Template analysis normally starts with at least a few predefined codes which help guide analysis. The first step in template analysis is to create an initial template by exploring the focus group transcripts, academic literature, the researchers' own experiences, anecdotal and informal evidence, and other types of exploratory research. The contents of the discussions are also examined for the meanings and implications for the research questions. Analysts will look for common themes and variations within the transcripts that would provide rich descriptions of the participants' reactions to design features. In template analysis, the initial template is applied in order to analyze the text but is revised between each EFG session. Once the final template is created after the final EFG, it is used to code the CFG sessions.

- **<u>Step 8: Report results</u>**

Focus group sessions produce mainly qualitative information about the objects of study (Kontio et al., 2004; Krueger & Casey, 2014). Qualitative results can be reported by creating an account structured around the main themes identified; drawing illustrative examples from each transcript as required. A similar approach can be taken when reporting focus group results. Short quotes are used to aid in the specific points of interpretation and longer passages of quotation are used to give a flavor of the original discussions. Summary tables can be very helpful, displaying both evidence and counter-evidence of the utility of the artifact by the focus group. Rich descriptions can further corroborate results by using quotes from the focus group participants. Furthermore, in this step, we collected the feedback forms and other comments of the participants in all focus groups and made an in-depth analysis of them.