A METHODOLOGICAL FRAMEWORK FOR INTEGRATION OF DATA MINING IN ORGANIZATIONS

Uroš Bole

Doctoral Dissertation
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia, July 2013

Evaluation Board:

Assist. Prof. Dr. Barbara Koroušić Seljak, Ph.D., Chairman, Jožef Stefan Institute, Ljubljana

Prof. Dr. Jurij Jaklič, Member, Faculty of Economics, University of Ljubljana

Assoc. Prof. Dr. Marko Bajec, Member, Faculty of Computer and Information Science, University of Ljubljana

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL

Uroš Bole

A METHODOLOGICAL FRAMEWORK FOR INTEGRATION OF DATA MINING IN ORGANIZATIONS

Doctoral Dissertation

METODOLOŠKI OKVIR UVAJANJA PODATKOVNEGA RUDARJENJA V ORGANIZACIJE

Doktorska disertacija

Supervisor: Assist. Prof. Dr. Gregor Papa

Ljubljana, Slovenia, July 2013

Index

Abstract

In this thesis, we address the problems and challenges faced by data mining (DM) practitioners in the initial stages of DM technology integration in organizations. The findings (with minor modifications) may also be applied in other analytics domains, i.e. forecasting/extrapolation, modeling, experimental design, simulation, and optimization.

While it is evident that DM now represents a significant technology for strategic applications, there appears to be a dearth of empirical studies that consider in detail the initial (embryonic) stages in DM management to enable an appropriate foundation for its later successful implementation. Most extant theory either fails to consider the distinctive context and aim of the embryonic DM process or focuses on large-scale DM implementation. Yet, in the great majority of organizations, the embryonic DM process is a *sine qua non* of enterprise-wide DM integration.

Our research therefore aimed to propose a methodological framework – a system of principles, practices, and procedures – to guide practitioner decision making through the embryonic DM process. We hypothesize that the application of the methodological framework increases the likelihood of success of embryonic DM. Due to the nature of the artifact, we applied a design science research methodology. Embedded within the design process we also applied a structured-case framework to identify best practices of embryonic DM. Primary data was principally collected through semi-structured interviews with DM practitioners. The proposed formulation of a methodological framework was validated and reported through a series of case studies.

Our findings indicate a significant range of considerations and reveal additional issues for applied decision making in the context of DM requirements and process success. Addressing best practices of embryonic DM a critical success factors framework is proposed. It suggests four success measures and seven success factors which, if managed well, lead to success. Moreover, a process model for carrying out embryonic DM is designed. The findings extend extant theory on DM implementation and can therefore be used for comparative studies and the development of cumulative knowledge.

Povzetek

V disertaciji obravnavamo izzive in probleme, s katerimi se srečujejo deležniki uvajanja podatkovnega rudarjenja (ang. data mining) v organizacije. Izsledki se ob manjših spremembah lahko uporabijo tudi v drugih domenah analitike (napovedovanje/ekstrapolacija, modeliranje, eksperimentalno načrtovanje in optimizacija).

Dandanes ni več dvoma, da je podatkovno rudarjenje potencialno strateško pomembna tehnologija. Zato preseneča dejstvo, da ni empiričnih raziskav, ki bi podrobno obravnavale zgodnje (motivacijske) faze uvajanja oz. pobud podatkovnega rudarjenja, na katerih temelji uspeh kasnejših faz integracije. Večina literature ne upošteva posebnosti konteksta in ciljev motivacijskega procesa. Preostala dela pa obravnavajo poznejšo integracijo podatkovnega rudarjenja, čeprav ugotavljajo, da se velika večina organizacij motivacijskim pobudam ne more izogniti.

Namen našega raziskovalnega dela je bil predlagati metodološki okvir – sistem načel, praks in postopkov – za pomoč pri izvajanju motivacijske faze uvajanja podatkovnega rudarjenja. Raziskovalna hipoteza pravi, da tak metodološki okvir poveča verjetnost uspešne izvedbe motivacijske faze podatkovnega rudarjenja. Glede na naravo artefakta smo raziskave izvajali po metodologiji »design science«, ki določa znanstveni postopek, ko je namen raziskave oblikovanje novega artefakta. Identifikacija dobrih praks – ključnih dejavnikov uspeha – pa je temeljila na metodi »structured-case«. Primarne podatke smo večinoma zbirali po metodi pol-strukturiranih intervjujev. Predlagani metodološki okvir smo ovrednotili z uporabo metode študije primera.

Izsledki izpostavijo številne napotke za izvajanje začetnih pobud uvajanja podatkovnega rudarjenja. Sistem načel je izražen v okviru ključnih dejavnikov uspeha motivacijske faze podatkovnega rudarjenja in predlaga štiri merila uspešnosti in sedem dejavnikov uspeha, ki, ob ustrezni izvedbi, vodijo do uspeha. Predlagamo tudi proces modela (postopek) za izvajanje motivacijskega podatkovnega rudarjenja. Poleg tega so mnogotere ugotovitve nov prispevek k znanosti in se zato lahko uporabijo za primerjalne študije in razvoj kumulativnega znanja.

Abbreviations

ADR	=	Action Design Research
BI	=	business intelligence
BIE	=	Building, Intervention, Evaluation
CF	=	conceptual framework
CRISP-DM	=	Cross Industry Standard Process for Data Mining
CRM	=	customer relationship management
CSF	=	critical success factor
D3M	=	domain-driven data mining
DM	=	data mining
ERP	=	enterprise resource planning
IP	=	infrastructural projects
IS	=	Information Systems
IT	=	information technology
MSGPS	=	the model of the stages of group problem solving
OLAP	=	online analytical processing
PDSA	=	plan, do, study, act
ROI	=	return on investment

1. Introduction

There is significant evidence reported throughout the last decade of the increasing demand for data mining (DM) among organizations (Gartner, 2011; LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011; Luftman & Ben-Zvi, 2010). This trend is a reflection of the search for new sources of competitive advantage (Viaene & Van Den Bunder, 2011) and promises important benefits for industry leaders (Davenport & Harris, 2007). Their competitors are subsequently forced to follow suit (Davenport, Harris, & Morison, 2010). However, different surveys show a growing lag between the demand for DM and its implementation (Eckerson, 2007; Rexer, 2011, 2012). DM researchers and practitioners consistently suggest that this is a problem that must be addressed by the DM research community (Cao, 2010; Elder, 2007).

It is also apparent that embryonic DM initiatives are the main challenge because they are not driven by senior managers. Davenport and Harris (2007, p. p.106) distinguish between two paths to becoming an analytical competitor (see Figure 1). The "full-steam-ahead" track counts on top management support, unlike the "prove-it" route, which we denominate as the *embryonic* DM initiative. Promoted by a mid-level manager, it tends to be a series of projects that form a cyclical process. It consists of the identification of a business problem that can benefit from DM, the implementation of a localized project to show DM benefits, and the propagation of benefits generated, until enough success has been built to secure the sponsorship of executives (p.117). This embryonic DM process is particularly complex to manage because it must be carried out in the face of the resistance of an established organization, without top management support, and with fewer resources (p. 115, 117). The embryonic process may take up to three years and when unsuccessful, broader DM integration may be delayed, or even abandoned, with subsequent loss of competitiveness (p.116).

From the technical point of view there is no difference between embryonic DM and enterprise wide DM. Embryonic DM is singular because a new relationship must be built between the business stakeholders and DM experts (Davenport & Harris, 2007). This also implies that a DM expert is likely to be new to the domain, while for the business people DM is an unknown technology. In order to jointly solve problems both need to develop some understanding of the other domain.

Embryonic DM process is also important for DM researchers in their mission to develop new techniques and algorithms. DM is fundamentally an applicative science, i.e. the development of new techniques and algorithms greatly depends on scientists' exposure to real-world problems (Wu et al., 2003). Therefore scientific contribution and relevance of their work depends on researchers' ability to obtain and carry out embryonic, applicative projects (Pechenizkiy, Puuronen, & Tsymbal, 2008). In these initiatives

researchers often face the same challenges as do non-academic practitioners (Cao, 2010).

Despite their documented importance and relevance, embryonic DM initiatives are still sparsely addressed in the literature. Beyond singling out and defining embryonic DM initiatives, Davenport and Harris's (2007) book focuses on enterprise-wide DM implementation as do other reports, e.g. (Davenport et al., 2010; Kiron & Shockley, 2011; LaValle et al., 2011). Most of the remaining theory does not discriminate between DM integration management before and after key decision-maker support has been obtained, e.g. (Nemati & Barko, 2003; Sim, 2003). This dissertation reports on our research effort aimed at improving our understanding of embryonic DM initiatives. Our study builds on and complements existing research with an exploration of the managerial considerations in the embryonic DM process of gaining key decision-maker support.

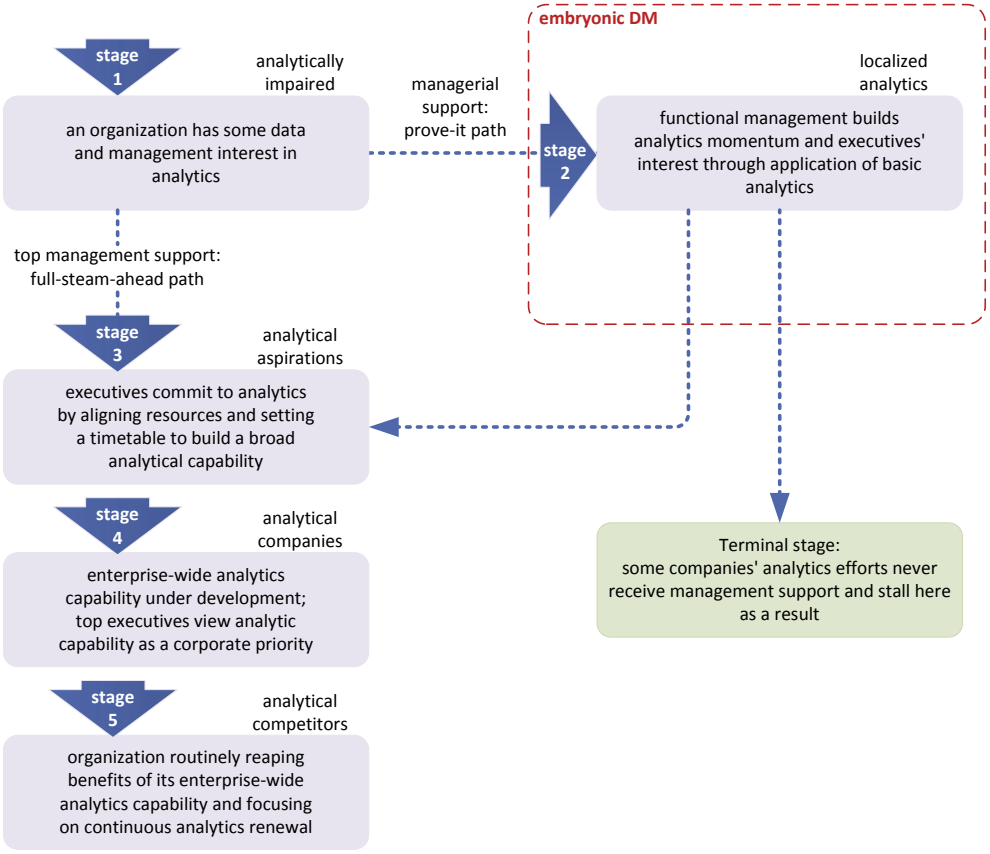

Figure 1: *Embryonic DM, i.e. analytics, (stage 2) is a singular step within the enterprise-wide analytics implementation process. Adopted from (Davenport & Harris, 2007).*

Our aim was to help increase the probability of success of embryonic DM initiatives by designing a methodological framework, i.e. a system of principles, practices, and procedures (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007) to guide practitioner decision making. We therefore applied a design science research methodology (Hevner, March, Park, & Ram, 2004; Peffers et al., 2007) as a general framework. Design science proposes a scientific procedure for research whose aim is design of a new artifact, e.g. a methodological framework (Hevner et al., 2004). Given the absence of extant theory on the embryonic DM process, and contradicting findings on the best practices of DM in

general, we carried out an embedded study to identify the Critical Success Factors (CSFs) of embryonic DM. Since CSFs are arguably not an artifact, a different research methodology was adopted. Our research applied a structured-case methodology (Carroll & Swatman, 2000) through a combination of secondary (reports) and primary data collection (interviews with practitioners) in an attempt to determine the best practices, i.e. the CSFs, of embryonic DM integration. The outcome is the proposed formulation of a methodological framework to guide practitioner adoption which was further validated and reported through a series of case studies.

Our findings extend current theory on DM implementation in several ways. The methodological framework indicates a significant range of CSF considerations and reveals additional issues for applied decision making in the context of embryonic DM requirements and process success. First, the differentiating characteristics of the embryonic DM initiatives are suggested. Second, a unique success measure within IS theory, *top management support*, is proposed. Third, it is shown that prior research on the CSFs of DM presents conflicting findings due to the lack of focus on either embryonic or enterprise-wide implementation. Fourth, a specific conceptual framework focusing on the success of embryonic DM initiatives is proposed, including a unique proposition within DM theory of two CSFs: *interdisciplinary learning*, and *process facilitation*. Finally, a new process model, *InterActive8*, for embryonic DM is proposed. By using the methodological framework, comparative studies can be made and cumulative knowledge developed. As the framework has a managerial perspective, it can also be used by practitioners since it reveals the issues that must be addressed when introducing DM.

Chapter 2 of this dissertation continues by identifying the specific properties of embryonic DM which motivate the problems and challenges faced by DM practitioners within this context. In Chapter 3, the research methodology is explicated. Next, in Chapter 4, best practices (CSFs) of embryonic DM are identified. Chapter 5 focuses on the design of the corresponding process model. In Chapter 6, this research is evaluated from the methodological point of view. Research contributions and practitioner relevance are discussed in Chapter 7. The dissertation concludes by summarizing the findings and discussing their limitations and the potential for future work.

2. Problem Identification

The integration of DM within an organization requires a specific managerial approach that is in some aspects different to that of a conventional Information Technology (IT) project or other Business Intelligence (BI) technologies. DM is considered a BI technology (Davenport & Harris, 2007; E. A. King & Rathburn, 2010; Moss & Atre, 2003; Wang & Wang, 2008). Yeoh and Koronious (2010) show that the implementation of BI requires a different implementation framework than other, conventional application-based IT projects such as operational or transactional systems. They argue that BI implementation is similar to other infrastructural projects such as enterprise resource planning (ERP) systems implementation. Hence a BI system implementation transcends "the purchase of software and hardware; rather, it is a complex undertaking requiring appropriate infrastructure and resources over a lengthy period (Yeoh & Koronios, 2010)." BI is therefore different in that it is arguably an IT infrastructure project whose capabilities support (or fail to support) important business processes (Moss & Atre, 2003; Wixom & Watson, 2001; Yeoh & Koronios, 2010).

DM is often viewed as an extension of a data warehouse (Eckerson, 2007) or as an optional final step of a BI system (Moss & Atre, 2003). However, this need not be the case. Often DM is an independent, stand-alone initiative that may precede and even motivate a data warehouse or a new BI system implementation (Davenport et al., 2010; Kohavi, Mason, Parekh, & Zheng, 2004; Lavrač et al., 2004). This may be the reason why most reports that focus on DM implementation treat it independently of a BI system, e.g. (Hilbert, 2005; Nemati & Barko, 2003; Sim, 2003). DM should therefore not be viewed as merely one aspect of a BI system.

Since DM is a BI technology (Davenport & Harris, 2007), its implementation is in many ways similar to implementations in other BI areas (data warehousing, OLAP, dashboards, scorecards, enterprise-wide analytics) and to other infrastructural projects (IP) such as ERP or CRM systems implementation. There is significant overlap between the CSFs of different types of IP (see Table 1), and embryonic DM implementation should, to some extent, draw from the IP implementation knowledgebase.

However, a dedicated study is necessary to account for the specifics of embryonic DM:

- local implementation;
- it leads to small, incremental changes to business processes;
- a credible business case may not be elaborated;
- top management support is not available.

Below we further elaborate on each of the above points.

First, while IP are interdepartmental, large-scale integration technologies (S. F. King &

Burgess, 2008; Yeoh & Koronios, 2010), embryonic DM is local (Davenport & Harris, 2007). Second, embryonic DM does not lead to major changes that require large-scale change management (Davenport & Harris, 2007) as in IP implementations (S. F. King & Burgess, 2008; Yeoh & Koronios, 2010).

Table 1: *Infrastructural projects implementations show significant overlap.*

CSFs (non-technical)	*ERP*	*CRM*	*BI*	*Enterprise-wide Analytics*
top management support	*	*	*	*
championship related factors	*	*	*	*
clear vision & well established business case	*	*	*	*
team related factors	*	*	*	*
project management	*		*	
change management related factors	*	*	*	*
executive (top management) support	*	*	*	*
vendor support	*			
careful package selection	*			
knowledge management capabilities		*		
willingness to share data	*	*	*	
analysts				*
analytical culture				*
SOURCES:	ERP & CRM: (S. F. King & Burgess, 2008) BI: (Yeoh & Koronios, 2010) Enterprise-wide Analytics: (Davenport et al., 2010)			

Third, in embryonic DM a business case may not be elaborated (Davenport & Harris, 2007; Viaene & Van Den Bunder, 2011). The purpose of DM is to discover new, *previously unknown* knowledge (I. Bose, 2001; Moss & Atre, 2003; Spiegler, 2003). Not knowing what, if any, knowledge might be found in data implies that *prior* to carrying out a DM initiative it is *not possible* to present a credible *business case* (Davenport & Harris, 2007; Hermiz, 1999; E. A. King & Rathburn, 2010; Viaene & Van Den Bunder, 2011). Contradictory evidence is apparently presented by Yeoh and Koronios (Yeoh & Koronios, 2010), who imply that a *well-established business case* can be built when they demonstrate that it is a CSF of a BI system implementation. This paradox can only be explained if the BI implementations that were the subject of Yeoh and Koronios's study focused on BI technologies such as data warehousing, OLAP, dashboards, and scorecards. Hence, it may be concluded that DM differs from other BI technologies in that a business case identifying the proposed strategic benefits, resources, risks, costs, and timeline cannot be built.

The inability to elaborate a business case is therefore a unique characteristic of DM and leads to the fourth important difference in its implementation management: top management support is not available. The purpose of embryonic DM is precisely to secure executive support. A well-established business case is a CSF in BI systems implementations because it is the basis for another CSF – *top management support* (Wixom & Watson, 2001; Yeoh & Koronios, 2010). Similarly, executive sponsorship has been found to be critical to *enterprise-wide* DM implementation, which is the only way to realize DM's full benefits (Davenport et al., 2010; Davenport & Harris, 2007; Moss &

Atre, 2003). However, strong executive commitment to broad DM implementation at the outset is still extremely rare in today's organizations because most top managers have no experience with DM (Davenport & Harris, 2007). Therefore, in most organizations DM must first be implemented *locally* through an *embryonic DM initiative* whose principal aim is to obtain executive support (Davenport & Harris, 2007).

Davenport and Harris (2007) were the first to explicitly point out that the embryonic DM approach must be as different from the enterprise-wide approach as are the context and aim. An early DM initiative is pioneered *locally* by a *mid-level manager*. Without any support from the top, the proponent is up against the complex task of overcoming the resistance of an established organization: existing people, processes, data, technology, and culture. In addition, the DM champion must convince the executives of the technology's benefits and must do so with limited resources. Attracting top management attention and sponsorship requires a series of well-documented success stories. The process is cyclical and iterates between the search for a relevant business problem, the implementation of a localized project, and the propagation of the benefits (p. 117). If not successful, broader DM integration may be delayed indefinitely (p.116).

The embryonic DM process is under-researched. Most of the extant literature does not discriminate between DM integration management before and after key decision-maker support has been obtained. The few reports that do make the distinction focus on large-scale analytics at the enterprise level, e.g. (Davenport et al., 2010; Davenport & Harris, 2007; Kiron & Shockley, 2011; LaValle et al., 2011).

Moreover, existing DM reports have been criticized for focusing on technical concerns such as data and algorithms (Fayyad, Piatetsky-shapiro, & Smyth, 1996). Practice-minded keynote speakers at DM conferences repeatedly insist that in addition to technical issues, the community also needs to address organizational concerns, e.g. (Elder, 2007; Fayyad, 2007; Fogelman Soulié, 2008). In light of these suggestions, some researchers promoted Domain-Driven DM (D3M), e.g. (Cao, 2008, 2009, 2010), in an attempt to propose guidelines and methodologies that put the domain expert at the center of the DM process, e.g. (Cao & Zhang, 2006, 2007). D3M provides valuable insights; however, focusing on the domain expertise addresses only part of the organizational and managerial complexity of a DM initiative. CRISP-DM (P. Chapman et al., 2000), being the best known and most used DM process methodology (KDnuggets, 2007a), also displays similar weaknesses. While it is a good process model from the technical viewpoint of a DM expert, it fails to address the process from the business perspective, nor does it attempt to integrate the two (Khabaza, 2007).

In summary, a focused examination of the embryonic stages of DM integration management is therefore justified because of the following:

- organizations are showing increasing interest in DM,
- embryonic DM initiatives are specific, complex and inevitable,
- empirical research on the topic is scarce and unsuitable, and
- the business implications of wider DM adoption are potentially significant.

To address our research question, we propose to design a methodological framework, i.e. a system of principles, practices, and procedures (Peffers et al., 2007). Such a framework should guide practitioner decision making. Moreover, it should facilitate comparative research studies and the development of cumulative knowledge. For embryonic DM a methodological framework would include three elements: conceptual principles to define what is meant by embryonic DM, best practice rules manifested as the CSFs of embryonic DM, and a process model for carrying out embryonic DM. As shown earlier, the first element of the framework, i.e. embryonic DM, has already been defined by Davenport and Harris (2007). On the other hand, the second element, i.e. practice rules, needs further examination, and the third, i.e. a process model, must be designed (see Figure 2).

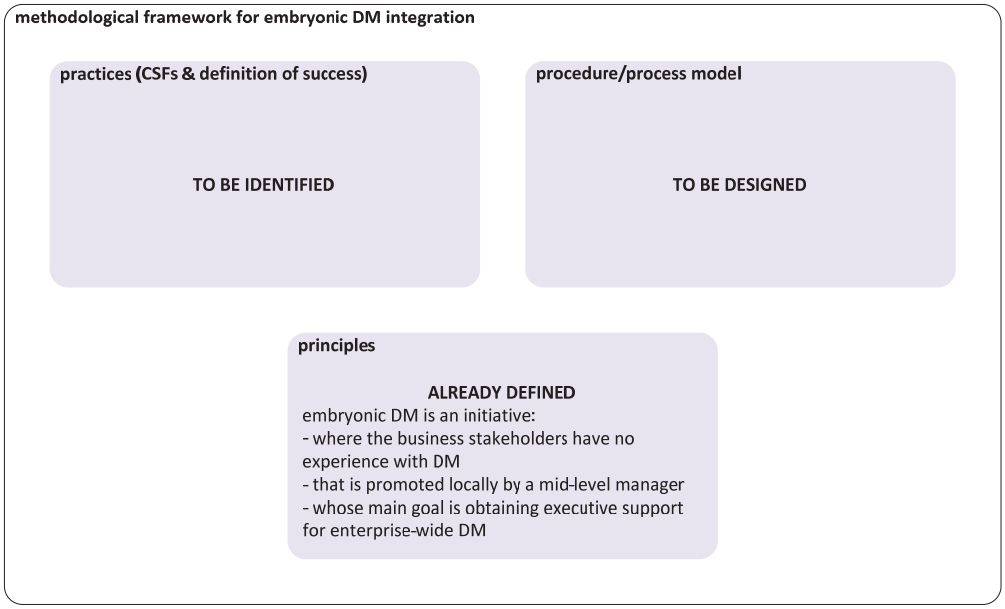

Figure 2: *Objective of this research: design of a methodological framework; the components missing in extant theory are practices and a procedure of carrying out embryonic DM.*

Identifying the CSFs and a corresponding process model for embryonic DM integration are the main focus of the subsequent sections. Though related, each task is distinct and requires a different research approach. This includes specific problem identification and literature review. Therefore, in sections 4 and 5, the problem presented above is further elaborated in light of the pertaining research aim.

3. Research Approach

This research applied a design science research methodology (Hevner et al., 2004; Peffers et al., 2007) as a general framework (see Figure 3). Contrary to the research approaches that are used for exploring or confirming hypotheses, this study follows a design science approach because our primary goal is to develop a new artifact, i.e. a methodological framework for management of embryonic DM implementation. The design science research approach focuses on clarifying the objectives of a solution, i.e. research artifact (a construct, method, model or instantiation), and on building and carefully evaluating the utility of the artifacts (Hevner et al., 2004). To a lesser degree, design science research evaluates an artifact's reliability and validity (Hevner et al., 2004).

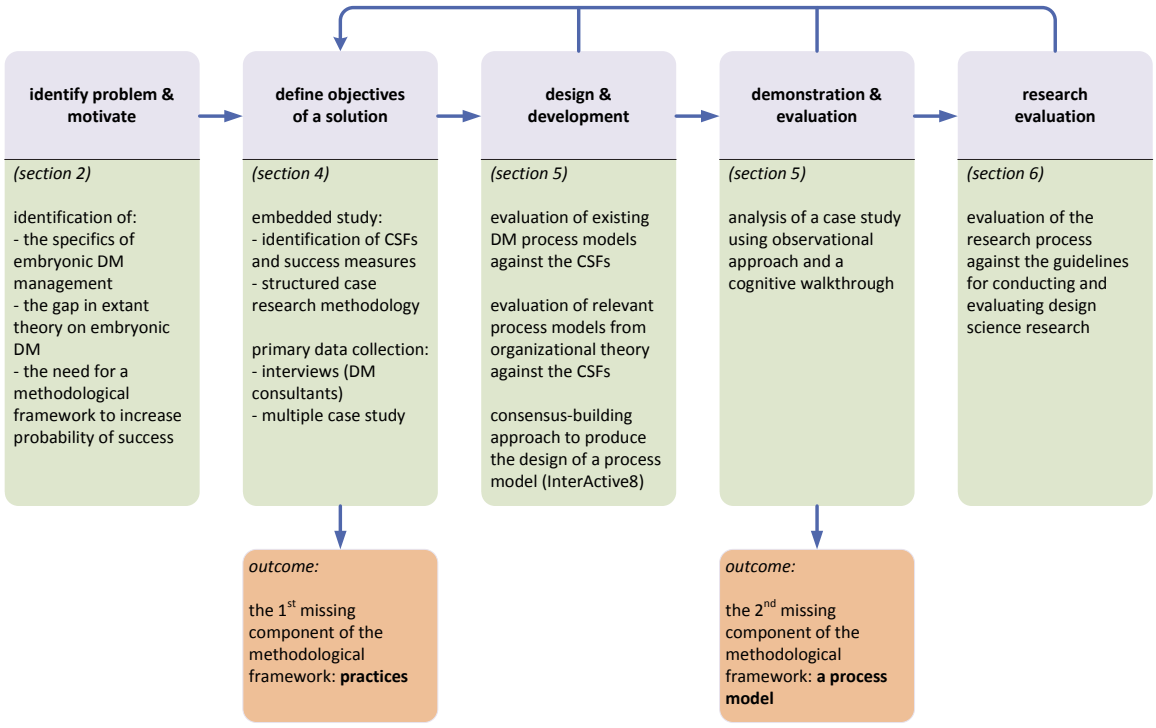

Figure 3: *The research process.*

Following (Peffers et al., 2007), the design science research process involves several stages and is iterative. Although it accommodates various entry points, for presentation purposes, in the nominal sequence the first stage is *identifying the problem and its importance*, which was presented in section 1 above. Next, *the objectives for a solution* should be defined. The methodological framework should respond to the issues that are critical for embryonic DM integration management. Hence, the objectives of our solution are defined by the CSFs of embryonic DM implementation. Given the absence of extant

theory on embryonic DM process, and contradicting findings on the CSFs of DM in general, we carried out an embedded study with a different research approach (see section 4) to identify the CSFs. Since the CSFs are arguably not an artifact, a different research methodology, i.e. the structured-case (Carroll & Swatman, 2000), was adopted. Its aim was to develop a conceptual framework consisting of the definition of embryonic DM success and the corresponding CSFs. Initially, we present the weaknesses of existing empirical research on the CSFs of DM. Subsequently, we propose a conceptual framework drawn from an extensive literature review. Next, primary data is collected in interviews with four experienced DM consultants in search for additional candidate CSFs. Finally, the conceptual framework is validated in a multiple case study carried out through 17 semi-structured interviews in eight organizations.

A process model for embryonic DM *is developed* in section 5. First, we present and evaluate existing process models for DM with respect to the CSFs. We then look at organizational and management literature for related process models. Three such models are selected based on their compliance with some or many of the CSFs of embryonic DM. Along with CRISP-DM, the most popular existing process model for DM, they served as the basis for the development of common elements for the proposed process model. The *solution is demonstrated and evaluated* through a case study based on three semi-structured interviews. Due to the complexity of the proposed methodological framework, both the definition of the design criteria and the development of the methodological framework required several iterations. This dissertation describes the final iteration of the research process.

To avoid potential weaknesses of qualitative research (Saunders, Lewis, & Thornhill, 2009; Yin, 2003), the investigation was performed by four researchers. Each research cycle was identical in terms of the division of roles among the researchers. First, a detailed *plan* for each phase was proposed by two authors and reviewed by all four authors in a meeting. Second, one author always carried out *data collection* and preliminary coding and analysis of data (Table 2 presents an overview of primary data collection). Third, the preliminary *coding and analysis* was cross-checked by another researcher. Finally, the findings were presented to the remaining pair of researchers for a systematic *reflection* through discussion. In doing so we followed the five points suggested by the structured-case framework: (1) review of the research process, (2) evaluation of the outcomes of analysis, (3) review of the structures of the structured-case, (4) looking beyond the data to build theory, and (5) changing the conceptual framework to incorporate the knowledge accumulated and theory built (Carroll & Swatman, 2000). The specifics of the research methodology applied in each phase of this investigation are explicated in the corresponding sections.

Table 2: *Primary data collection overview.*

Research phase	*Data collection method*	*Informant type*	*Number of interviews*	*Countries represented*
Defining best practice rules (CSFs)	Explorative interviews	DM consultants	4	USA, UK, Germany, Canada
	Multiple case study (8 organizations)	DM experts	9	Slovenia, Russia, Netherlands (2), Germany, Belgium Austria, Argentina
		business stakeholders	8	
Design of the process model	Case study	DM expert	1	Slovenia
		business stakeholders	2	
		TOTAL	**24**	

4. Defining Practice Rules: the CSFs of Embryonic DM

Our overall objective for this dissertation is the development of a methodological framework for embryonic DM introduction in organizations. As suggested by Peffers et al. (2007), a methodological framework is a system of principles, practices, and procedures and should guide practitioner decision making. The aim of this section is to identify those practices that improve the chances of success of embryonic DM integration, namely the CSFs.

Since extant DM methodologies have been criticized for primarily addressing the concerns of DM experts (Cao, 2010; Elder, 2007; Fayyad et al., 1996; Khabaza, 2007), the design of a new methodological framework should aim at also incorporating organizational concerns. (4.2) To achieve this goal we planned to put heavy emphasis on soliciting the input of business stakeholders in addition to that of DM experts in our primary data collection aimed at CSF identification.

To capture the complexity of the problem, we first atomize it conceptually by exploring the embryonic DM process requirements. The methodological framework should increase the likelihood of success of embryonic DM integration. Hence, below we first define success by suggesting relevant success measures. Next, we explore the factors that lead to success of embryonic DM management as the design requirements for the solution.

In the absence of extant theory on embryonic DM, a multi-method strategy was applied to increase the robustness of the results (Remus & Wiener, 2010). To ensure that the design requirements for the methodological framework are grounded in both theory and practice, data collection consisted of a combination of secondary and primary sources. Based on an extensive DM, BI, and IS implementation literature review, an initial conceptual framework was developed. Subsequent qualitative investigation consisted of semi-structured interviews with DM practitioners and was used to refine the theoretical framework, which was finally validated in a multiple case study.

The research process was iterative as suggested by structured-case (Carroll & Swatman, 2000) research methodology. New findings of each phase suggested additional exploration of the literature and analysis of primary data gathered previously. Although the investigation is presented as a straightforward, linear process with a framework, empirical study, analysis, and theorizing, the reality was a far more iterative process with alterations between theory, empirical data, and analysis – as suggested by Carroll and Swatman (2000). The essence of their structured-case methodological framework is that it forms an iterative, four-phased research cycle upon a formal theoretical framework, i.e. plan, investigate existing theory, collect data, and reflect on data/analysis. Below we

present only the results of the last iteration, namely the CSFs found relevant to the design of the methodology after the interviews with DM practitioners and case studies.

4.1 Theoretical framing

Immediate adoption of the CSFs of DM projects proposed by existing empirical studies (Hilbert, 2005; Nemati & Barko, 2003; Sim, 2003) has been judged inappropriate see their proposed CSFs in Table 3). Their resaerch presents two major weaknesses that call for an explorative study focused on embryonic DM:

(1) They do not distinguish between embryonic DM and subsequent, enterprise-wide DM implementation, which may be the reason for conflicting findings (more on this topic in section 4.3.2).

(2) Although they claim that there is a scarcity of prior research on DM management, they essentially build their theoretical framework from a literature review and validate it in a survey. Such a research strategy implies a supposition that the limited prior literature, mostly anecdotal reports, contains the "best set" of issues that contribute to successful DM implementation (Sim, 2003). The validity of this assumption should therefore be examined through an explorative investigation focused on the discovery of additional factors that may have been overlooked in earlier reports.

Table 3: *The CSFs of DM as proposed by extant theory.*

CSFs of DM		
(Hilbert, 2005)	(Nemati & Barko, 2003)	(Sim, 2003)
top management commitment	level of end-user DM expertise	interpretation & use of results
change management	project scope & length	data quality & management
a fixed budget for the project	resources availability	top management support
DM integration IT landscape	DM outsourcing strategy	output accuracy & reliability
high data quality	data quality & integration	clear business goal
	integration of technology	user consultation
	technological expertise	business environment

Given the absence of robust a priori theory, we sought to first develop a conceptual framework based on a literature review of DM, BI, and IS implementation. In addition, we explored different relevant organization science fields. The resulting framework was intended to help us remain cognizant of extant theory during our subsequent qualitative explorations of the embryonic stages in DM management. Our research sought to develop understanding by eliciting the salient beliefs of informants and typically assumed that the views of interest were explicitly known. Hence, our a priori theoretical framework served two key purposes. First, it helped us to ensure that we remained theoretically aware during our discussions. Second, it improved our ability to prompt informants concerning some potentially relevant issues that they either felt were of little importance or were unaware of at the time of our discussion.

Furthermore, CSF methodology is widely used for research and management, e.g. (Ang, Sum, & Yeo, 2002; Guynes & Vanecek, 1996; Lu, Huang, & Heng, 2006). It can be used for most organizational initiatives that need to be managed for 'success' (Bullen, 1995). Success is defined in terms of the DeLone and McLean IS Success Model, while candidate CSFs are presented according to the major dimensions of interest, namely organization, process, and technology. The rationale for the success measure and factor selection is described below.

4.1.1 'Success' measures

The DeLone and McLean IS Success Model (1992, 2003) was used to guide the identification of appropriate success measures. It proposes six dimensions of success: *systems quality, service quality, information quality, use/intention to use, user satisfaction,* and *net benefits*. When constructing a research model, researchers should treat IS success as a multi-faceted construct, choose several relevant success measures based on the research objectives and the phenomena under investigation, and consider possible relationships among the success dimensions (Delone & McLean, 1992). For new BI system implementations, Yeoh and Koronios (Yeoh & Koronios, 2010) identify *system quality, information quality, system use,* and *perceived net benefits* as the most appropriate success measures. Drawing on these definitions of success proposed for IS and BI systems implementations in general, we constructed a specific success model for embryonic DM success.

Systems quality and *service quality* constructs were not adopted as they were considered less appropriate for embryonic DM initiatives. On the one hand, *system quality* focuses on the system itself. It measures the information processing system in terms of flexibility, integration, response time, and reliability (Delone & McLean, 1992). On the other hand, *service quality* measures the support of the IS function to the IS users (Delone & McLean, 2003). We found it difficult to separate the system from the service in the case of embryonic DM. New information and knowledge is obtained through a process that is labor intensive, particularly in the data preprocessing and model building phases of the DM process (P. Chapman et al., 2000; Fayyad et al., 1996; Feelders, Danieils, & Holsheimer, 2000). Moreover, many different DM tools are likely to be tried before finding the one that is the most suitable for the given problem and the objective (Fayyad et al., 1996; E. A. King & Rathburn, 2010; E. A. King, 2005). Hence, it would be difficult to determine what system to evaluate. If, however, we were to treat embryonic DM as a service, we would run into an evaluation problem. The DM process is highly collaborative in all of the remaining phases, i.e. business understanding, data understanding, evaluation, and deployment, meaning that service quality depends on the contribution of both DM and domain experts (P. Chapman et al., 2000; Fayyad et al., 1996; Feelders et al., 2000). Our major data source in the validation phase would be the interviews with these stakeholders. We foresaw that, particularly in unsuccessful initiatives, each side would blame the other such that it would be impossible to reliably evaluate *service quality* (this foresight was confirmed in the case study research phase).

We therefore excluded both *systems quality* and *service quality* from the a priori theoretical framework.

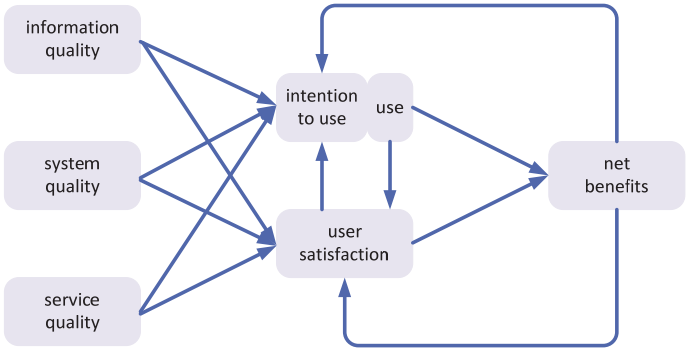

Figure 4: *DeLone & McLean IS Success Model.*

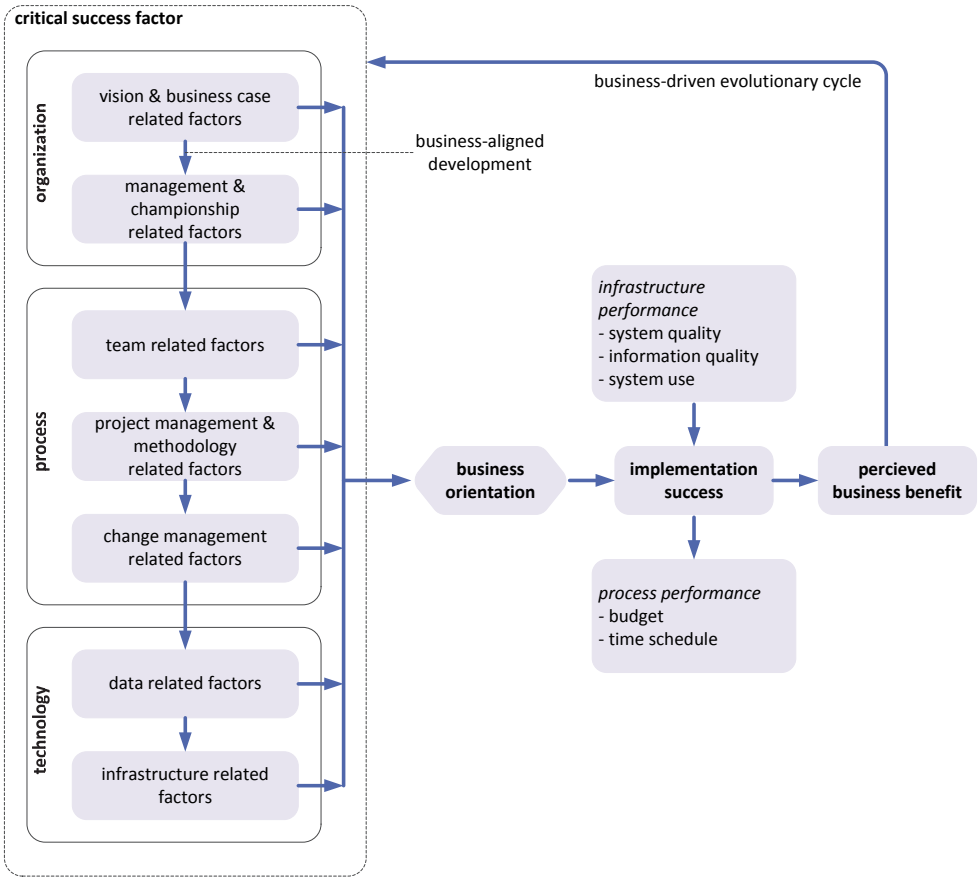

Figure 5: *Yeoh and Koronios's (2010) CSFs Framework for Implementation of BI Systems.*

Information quality refers to the understandability, usefulness, and relevance of DM as judged by business users (Delone & McLean, 1992). This factor has received considerable attention from the DM community, e.g. (Lavrač et al., 2004; Sim, 2003; Viaene & Van Den Bunder, 2011). The most commonly used term is *actionability* of results. Providing actionable information to decision makers is the fundamental reason for engaging in DM (Davenport & Harris, 2007). Cao (2010 p.755) defines actionable as

"knowledge that is business friendly, and can be taken over by business people for seamless decision making." Despite the prevalence of this term in the DM community, we use equivalent terminology suggested by DeLone and McLean (Delone & McLean, 1992) for its wider acceptance in the IS and business communities.

Information use is defined as the "recipient's consumption of the output of an information system" (Delone & McLean, 1992). Informed and effective use of an IS is a strong indication of its success (Delone & McLean, 2003). However, in many embryonic DM initiatives, the use of DM output, strictly speaking, may be delayed in time. Initial DM assessment may show that the organization is not ready for DM for technical or other organizational reasons (E. A. King, 2005; Lavrač et al., 2004). It may also happen that the DM project does not yield the answer to the question being pursued because there is no guarantee that one exists. However, this may still lead to learning (Hermiz, 1999). Moreover, DM often requires changes in the way of thinking (Kohavi et al., 2004; Weiss, 2009). This may be a slow process,(Cohen & Levinthal, 1990), taking months or sometimes years, but it is a necessary condition for obtaining DM results and using them (Davenport & Harris, 2007). In these cases, information *use* may not be applied as a measure of success. However, organizations in the midst of one of these scenarios may still provide valuable insights into our research question. We therefore adopt *use (intention to use)* from the DeLone & Mclean IS success model (Delone & McLean, 2003) and define it as the attitude of the stakeholders towards the use (current or future) of DM. This is an alternative way of judging "organizational learning" and was used as a success measure for organizational DM in (Nemati & Barko, 2003).

The *User satisfaction* construct measures "recipient response to the use of the output of an IS (Delone & McLean, 1992)." In a process sense, it must be preceded by the *use* of an IS (Delone & McLean, 2003). Yet, as suggested in the above paragraph, the use of new knowledge generated by DM may be delayed in time. We therefore exclude it from our theoretical framework. However, although the use of DM output is deferred, the use of DM is not. Business stakeholders do "use" DM indirectly through collaboration with DM experts. Their participation and contribution are critical for the co-production of actionable knowledge (Cao, 2010; Feelders et al., 2000). Hence, to some degree *user satisfaction* is implied by greater *intention to use* DM in the future. Such a relationship between these two constructs has already been suggested by DeLone and McLean (Delone & McLean, 2003).

Our literature review also identified two possible ways for quantitative and hence more objective measures of *net benefits*: *return on investment*, e.g. (Kohavi et al., 2004; Lavrač et al., 2004), and *efficiency increase*, e.g. (Davenport et al., 2010; Nemati & Barko, 2003). Either of the approaches is recommended although there are many cases where it is difficult to assign a quantitative (monetary) value to improvements generated by DM outputs, e.g. improved browsing experience or better customer satisfaction (Kohavi et al., 2004). An internet poll conducted in 2008 implies that ROI was estimated in approximately one half of DM projects (KDnuggets, 2008). It is therefore unclear whether these quantitative measures are operational for the purpose of our research;

however, we included them in the framework at this stage of the research process as an issue to be explored in the latter phases.

Top management support may be defined as the extent to which top management commitment to the use of DM is obtained through embryonic DM initiatives. This construct is not included in the DeLone and McLean IS Success Model as a success measure. Instead, most studies of IS success identify it as a success factor, e.g. BI (Yeoh & Koronios, 2010), data warehousing (Wixom & Watson, 2001), and Enterprise Resource Planning (ERP) (R. Bose & Luo, 2011). Moreover, it has also been found as a success factor of enterprise-wide DM (Davenport et al., 2010) because, beyond the embryonic stage, DM implementation must cross departmental boarders to deliver full benefits (Davenport et al., 2010; Moss & Atre, 2003). Enterprise-wide DM implementation implies many different forms of organizational resistance that can only be overcome by committed top executives (Davenport & Harris, 2007). As shown above, such executive sponsorship a priori is still uncommon. The embryonic DM process is required precisely to generate it. *Top management support* should therefore be included in our theoretical framework as a success measure. This is, to the best of our knowledge, unique in the IS implementation success literature.

Table 4 summarizes the two success models that were used to guide our identification of success measures for embryonic DM. Similar to the case of BI systems implementation, *information quality, use* (*intention to use)*, and *net benefits* constructs were judged appropriate given our research objective. In addition, due to the specific aim of the embryonic DM process to build executive commitment, we propose *top management support* as an explicit success measure. This is unique in relation to most other IS and BI implementations where, conversely, executive support is considered a success factor.

Table 4: *Summary of the general IS, BI, and embryonic DM success models.*

Success model / ***Success measure***	***General IS*** (Delone & McLean, 2003)	***General BI*** (Yeoh & Koronios, 2010)	***Embryonic DM***
systems quality	*	*	
service quality	*		
information quality	*	*	*
use/intention to use	*	*	*
user satisfaction	*		
net benefits	*	*	*
top management support			*

4.1.2 Candidate CSFs

Our effort to identify potentially relevant factors to early DM implementation success was guided by the Organization-Process-Technology framework suggested by (Wixom & Watson, 2001) and adopted by (Yeoh & Koronios, 2010) for data warehousing and BI systems implementations. It identifies three broad categories of organizational context that influence new technology implementation. The main contribution of this framework

is that it encourages the researcher to take into account the broader context of implementation. We therefore draw upon this work as the basis for exploring and categorizing the factors that we identify as potentially contributing to early DM implementation success.

In CSF taxonomy (J. J. Williams & Ramaprasad, 1996) factors have been classified in four levels: factors linked to success by a known causal mechanism, factors necessary and sufficient for success, factors necessary for success, and factors associated with success. This study considers the CSFs *necessary* for success such that the absence of one CSF would likely lead to the failure of the initiative.

The initial literature review resulted in 29 concepts as potential candidates for CSFs. These were used as the basis for constructing the interview guide for phase two of this research – interviews with DM consultants (section 4.2). Subsequent interview analysis combined with additional literature scrutiny served to reduce the number of constructs. This resulted in a framework consisting of nine potential CSFs that was used as the basis for the construction of the case study protocol and interview guide (section 4.3). Case study analysis led to the discovery of one new concept while suggesting further reduction of the framework. The concepts validated as the *necessary* CSFs of embryonic DM are presented in the final conceptual framework (section 4.3.4).

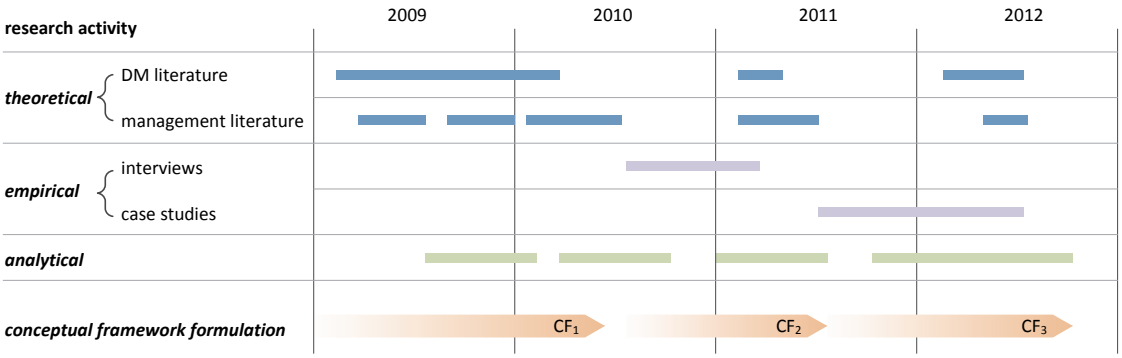

Figure 6: *Schematic overview of major theoretical, empirical, and analytical phases and the evolution of the Conceptual Frameworks (CF_i).*

While extant empirical research on the CSFs of DM and BI (Hilbert, 2005; Nemati & Barko, 2003; Sim, 2003; Yeoh & Koronios, 2010) is unanimous on some issues, it also presents contradictory findings. In the technology category, data issues, namely the quality and availability of data, have been found to be critical by all investigations. None of the reports, however, suggested DM tools, i.e. algorithms, as CSFs. This was surprising given that algorithms are the focus of most of the DM literature (Fayyad et al., 1996; Feelders et al., 2000). Three reports (Hilbert, 2005; Nemati & Barko, 2003; Yeoh & Koronios, 2010) conclude that the general level of IT integration and IT/DM expertise within an organization are CSFs. With respect to the DM process, Nemati and Barko (2003) found that expectations management, scope, and the availability of resources influence the success of DM projects. Sim (2003) proposes CSFs user participation and

DM alignment with the business strategy. However, the latter was not confirmed in (Nemati & Barko, 2003). In the organizational context, Hilbert (2005) suggests the importance of corporate culture, i.e. its openness to innovation; Sim (2003) proposes the criticality of the business environment, i.e. competitive pressure; and Nemati and Barko (2003) put forth the existence of outsourcing strategy as a CSF in the absence of sufficient internal DM expertise.

The most intriguing contradictions refer to the CSFs top management support, business championship, and change management. In accordance with the findings in BI (Yeoh & Koronios, 2010), Hilbert (2005) and Sim's (2003) studies put forth change management and top management support as CSFs of DM projects. These conclusions, however, are challenged by Nemati and Barko (2003). In addition, their study found no supporting evidence for business championship, which was found as a CSF of BI implementations (Yeoh & Koronios, 2010). This incongruence might be explained by the fact that the above-mentioned studies do not distinguish between embryonic DM and subsequent, broader DM implementation. As suggested in (Davenport et al., 2010; Davenport & Harris, 2007), after separating these two different contexts, it becomes clear that in embryonic initiatives business championship is a success factor, but change management and top management support are not. These two do, however, become critical in enterprise-wide DM implementations.

Factor selection was a systematic, iterative, and initially expansive process. First, the literature review was enhanced through further reports of DM practitioners, e.g. (Berry & Linoff, 2004; Hermiz, 1999; E. A. King & Rathburn, 2010; E. A. King, 2005; Kohavi et al., 2004) and researchers, e.g. (Blumenstock, Hipp, Kempe, Languillon, & Wirth, 2006; Cao & Zhang, 2006; Lavrač et al., 2004; Weiss, 2009). This yielded a large number of concepts (29) within our initial conceptual framework (see Table 5). Consequently, we refined (expanded or reduced) the framework upon primary data analysis. To illustrate this process, we briefly refer to *external pressure* and *DM algorithms*, which were eventually eliminated from the *necessary* CSF list. *External pressure* was confirmed in the interviews with DM consultants as all but one informant insisted on its importance. Although case study data suggested that this factor is not *necessary* for success (section 4.3.2), we include it in the a priori conceptual framework CF_1 since it may be relevant to some practitioners. DM consultants and software vendors may draw on it when assessing the probability and timeframe of a project or software sale.

DM algorithms and related technical issues also formed part of our a priori theoretical framework. However, in the second phase of this research, the informants refrained from speaking about the algorithms even when prompted. Case studies presented additional evidence against the inclusion of *DM algorithms* among the *necessary* CSFs. For example, one company had been developing new algorithms for a very specific problem for two years. Although, at the time of the interviews, they were still not satisfied with the solution, the embryonic DM initiative had already generated executive commitment. Moreover, several informants from other organizations suggested that "*in terms of tools, everything is available. For business success we already have everything that is needed,*

and we really have to focus on solving the case in the sense that it is useful to the business user." Based on this evidence we concluded that, as implied in (Hermiz, 1999; Viaene & Van Den Bunder, 2011), the existence of appropriate *DM algorithms* may be a necessary condition to solve a specific problem. However, an embryonic DM initiative achieves its final goal of *top management support* through a series of different DM-related business problems. If available, DM tools cannot provide satisfactory solutions to one specific problem, other problems may be tackled and successfully solved. Hence, the availability of suitable *DM algorithms* cannot be a *necessary* CSF for embryonic DM initiatives.

Table 5: *The initial conceptual framework, result of an extensive literature review.*

Candidate CSFs		
competitive context	business demand/sponsorship	have analytics strategy
data quality	data oriented and fact based culture	user training
change management (processes, culture)	expect major results take time (iterative discovery process, incremental improvements, evolving optimization problem)	IT budget
build analytic capability (knowledge of the business, tech, data, communication skills, insightful, business savvy analysts, interpersonal and team skills)	understand high performance drivers of the organization	managing risks
symbiosis (business, subject matter experts, IT, analytics, end-users	technical readiness	the right team
absorption capacity (stakeholders understand the potential of DM)	start small/baby steps	structuring analytical resources
concrete goals and metrics	project planning	follow a process-methodology (CRISP-DM)
strategy alignment	business mission	prototyping (show ROI on small sample)
	speed to value (show value in less than 18 months)	wait with the purchase of software solution
	project assessment	downplay the importance of models, methods, algorithms and optimal model performance
		select well your first DM projects

Below we present candidate CSFs and the rationale for their classification within the Organization-Process-Technology framework. As explicated above, this is a refined version, and the result of an iterative and systematic factor selection process spanning all three phases of this investigation. Due to practical reasons of relevance and space, the a priori conceptual framework (CF_1) presented below has only seven concepts.

4.1.2.1 Organization

A ***business champion*** is someone within the organization who understands the potential of DM, has a business problem that may be solved with the use of DM, is actively involved in the DM process, and promotes DM internally. The literature provides abundant evidence that top management support is a CSF of IS initiatives, e.g. (R. Bose & Luo, 2011; Dong, Xu, & Zhu, 2009; Laicity, Khan, & Willcocks, 2009). In data warehousing, decision support systems, and BI, the distinction is made between top management support and the champion, who is likely to have closer ties to the daily actions and goals of the project team (Curley & Gremillion, 1983; Wixom & Watson, 2001; Yeoh & Koronios, 2010). However, this research is aimed at embryonic DM

initiatives where top management support is not available. Instead, it is the champion's primary objective. A sponsor tends to be a mid-level manager whose goal is to build some initial DM successes locally in order to attract executive support for enterprise-wide DM implementation (Davenport & Harris, 2007). The champion's top concern is therefore organizational and, because of this, classified in the *organization* category. However, it is also recognized that to obtain top management support, the champion must also provide effective support for the DM process, i.e. provide information, access to data, material resources, and political support (Davenport & Harris, 2007; Hilbert, 2005; Weiss, 2009). Sponsors do so through their people skills, and by pushing for more data and analysis, teaching others, focusing DM efforts where they make most difference, knowing the limits of DM, etc. (Davenport et al., 2010).

External pressure is defined as the existence of outside factors supporting embryonic DM implementation and the extent to which they are leveraged in its favor. Competitive pressure or competition intensity have long been recognized as a technology adoption driver in the IS literature, e.g. (R. Bose & Luo, 2011; Crook & Kumar, 1998; Zhu, Kraemer, & Xu, 2003). Analysis of the strategic rationale underlying competitive pressure as an IS adoption driver proposed that, by adopting IS, organizations might be able to alter the rules of competition, affect the industry structure, and leverage new ways to outperform rivals, thus challenging the competitive environment (Porter & Millar, 1985). DM has proved this proposition true. There are many leading organizations that compete on analytics such that, through good practices spill-overs, rivals are soon forced to follow suit (Barton & Court, 2012; Davenport et al., 2010; Davenport & Harris, 2007). The organizations with the highest propensity to apply DM tend to compete in mature industries with low profit margins (Sim, 2003). Another reported source of external pressure on DM adoption is regulation (Baesens, Mues, Martens, & Vanthienen, 2009; Cornolba & Giudici, 2004).

We chose a broader term, *external* (not *competitive*) pressure to make it applicable also to organizations that do not compete in the market. For instance, Moon (2002), Norris and Moon (2005), and Von Haldenwang (2004) have indicated various environmental factors that contribute to e-government adoption. There are many reports of DM use in governmental institutions, such as tax collecting agencies and law enforcement, e.g. (Chen et al., 2004; DeBarr & Eyler-Walker, 2006; PAW, 2012). Reported examples of external pressures on DM adoption include terrorist attacks (Cate, 2008) and budget cuts (Reilly, 2011). Similar to the competitive context, plausible external pressures are also spill-overs of good practices from similar organizations operating in other geographic areas. External pressure is an outside force that accelerates the embryonic DM process to achieve its goal: *top management support* for organization-wide DM implementation. It therefore operates at the organizational level, which suggests its classification in the *organization* category.

4.1.2.2 Process

The concepts *stakeholder participation, interdisciplinary learning*, and *focus on problem solving action* are categorized within the *process* category, which is in agreement with the

reports on the CSFs of BI and data warehousing implementations (Wixom & Watson, 2001; Yeoh & Koronios, 2010).

Stakeholder participation may be defined as the degree to which unity of effort is achieved among the contributors to the DM initiative. The stakeholders may include the champion, domain experts, DM experts, IT experts, and end users, (E. A. King, 2005; Sim, 2003). IS and BI literature suggest that user participation is associated with implementation success, particularly when the requirements for the system are initially unclear, e.g. (Hartwick & Barki, 1994; Wixom & Watson, 2001; Yeoh & Koronios, 2010). It is argued that when users participate in IS projects, they help maintain the focus on their requirements and needs. Moreover, they better understand what the IS will provide and are hence more likely to use it when delivered. In the case of DM, in addition to the users, other business people must participate beyond the provision of requirements or specifications. If a solution is to be actionable, they need to co-produce it by contributing their domain knowledge throughout the process (P. Chapman et al., 2000; Fayyad et al., 1996). For this reason, it is important to gain stakeholder commitment (compliance is insufficient) and trust through frequent interactions, explaining and clarifying the thinking behind the decisions or DM models, and managing participant expectations (Blumenstock et al., 2006; Viaene & Van Den Bunder, 2011; Weiss, 2009). Similar practices are suggested in the knowledge worker management literature, e.g. (Davenport, 2005; Kim & Mauborgne, 2003).

Interdisciplinary learning is defined as the effort to coordinate interdisciplinary collaboration between DM experts and business people. Several authors have suggested effective change management of processes, attitudes, and organizational culture as a factor of DM success, e.g. (Hilbert, 2005; Lavrač et al., 2004; Moss & Atre, 2003), similar to other IS initiatives, e.g. (Ehie & Madsen, 2005; S. Williams & Williams, 2006; Yeoh & Koronios, 2010). However, these initiatives tend to suppose larger-scale changes requiring top management involvement, which is not the case in localized embryonic DM implementations. Instead, the dynamics of the change process are more similar to those of interdisciplinary product or business development. DM requires the collaboration of a DM expert with a domain expert (Feelders et al., 2000) who may know very little (and often nothing at all) about DM (Davenport et al., 2010; Fayyad, 2007; Lavrač et al., 2004). Despite some successful attempts to automate parts of the DM process (Féraud, Boullé, Clérot, Fessant, & Lemaire, 2010; Fogelman Soulié, 2008), the consensus is that DM cannot be automated to the extent that a DM expert would not be needed (Berry & Linoff, 2004; Coppock, 2002). DM literature suggests that the cooperation between DM and domain experts requires the development of a common language (P. Chapman et al., 2000; Kohavi et al., 2004; Lavrač et al., 2004), management of the differences in time perception (Nemati & Barko, 2003; Weiss, 2009), and business savvy DM experts (Davenport & Harris, 2007; Eckerson, 2007; Sim, 2003). Similarly, management research has suggested that cross-functional problem solving is complex, iterative, and time consuming, e.g. (Argyris, 1977; Boland Jr. & Tenkasi, 1995; Cohen & Levinthal, 1990). Building on this body of knowledge, IS literature posits that the process of an individual's

knowledge becoming part of the solution to a complex problem, from being in the mind of one person to becoming a team's constructed knowledge, is an iterative, dynamically evolving process with no best structure (Markus, Majchrzak, & Gasser, 2002). It requires (1) externalization of tacit knowledge, (2) its internalization, and (3) the negotiation of meaning in order to arrive to a common understanding, which may only then become the basis for constructive interdisciplinary problem solving (Beers, Boshuizen, Kirschner, & Gijselaers, 2005; Markus, 2001). Both the problem definition and the solution are likely to evolve as stakeholders develop shared cognition and a common understanding of the problem (Du Chatenier, Verstegen, Biemans, Mulder, & Onno, 2009).

Focus on problem solving action is defined as the extent to which the stakeholders are able to adapt the DM process in light of emerging findings while keeping in mind the initiative's goal, the action that solves the identified business problem. The DM problem solving process requires an experimentation-based approach to project management. The stakeholders will normally have a vision and an aim; however, the path is often unclear because the requirements are often uncertain and changing, and because the DM algorithms to be used may be non-existent, new, or unknown to the team (Viaene & Van Den Bunder, 2011). Therefore, while project planning and management with timelines, milestones, and fixed deliverables are necessary to some degree, they cannot be binding as in an engineering, data warehousing, or BI project (Hermiz, 1999; E. A. King & Rathburn, 2010). In such contexts, IS and project management literature increasingly favors an adaptive, agile approach with less detailed planning and requirement specifications, and experimental and evolutionary design with significant on-going learning and change, e.g. (Gemino, Reich, & Sauer, 2008; Highsmith, 2009; Howel, Windahl, & Seidel, 2010). The name – *focus on business problem solving action* – was chosen because the terms *business problem* and *actionable solution* are frequently used in DM reports as a means to ensure DM's alignment with the business strategy and to provide focus to a DM initiative, e.g. (Berry & Linoff, 2004; Cao, 2010; P. Chapman et al., 2000).

4.1.2.3 Technology

Two data-related concepts, i.e. *data availability* and *data quality*, were identified as potential CSF candidates within the Technology category.

Data availability is defined as the extent to which data is accessible and ready for DM analysis. Data is a precondition for anything analytical; hence the DM process is greatly facilitated when an organization has integrated its data from different sources in a data warehouse, separated from transactional IS (Davenport et al., 2010). At the other end of the spectrum, data sets may be scattered over an organization, possibly in different, incompatible formats (Berry & Linoff, 2004; Lavrač et al., 2004). For this reason data acquisition may be one of the most problematic stages of the DM process (Weiss, 2009). Data tends to be more easily available in organizations that gather large quantities of data automatically without human intervention (Lavrač et al., 2004).

Data quality is the degree to which data is clean, i.e. free of noise and errors; complete, i.e. needed attributes are available and filled to an acceptable degree; and

sufficient, i.e. relevant to support analysis and for drawing conclusions (Davenport et al., 2010; Lavrač et al., 2004). In DM as in other IS and BI initiatives, data quality is often used as a success factor because one of the aims of the information processing system is to provide users with data that is accurate, complete, timely, relevant, consistent, and useful (Delone & McLean, 1992; Wixom & Watson, 2001; Yeoh & Koronios, 2010). The output of DM are insights from data analysis whose quality is conditioned by the quality of the data on which they are built (Eckerson, 2007). Therefore, data quality should be regarded as a success factor. Data sets should contain a sufficient number of records (lines) and attributes (variables) in order to facilitate relevant findings (Hermiz, 1999; Lavrač et al., 2004). Moreover, the records should be complete such that there are few missing values and free of erroneous, i.e. misleading, data (Eckerson, 2007). However, DM may also be performed with less than perfect data. Skillful DM experts have ways of dealing with missing or erroneous data (Davenport et al., 2010). Nevertheless, lower data quality will slow down progress towards DM results (Weiss, 2009).

4.1.2.4 A Priori Conceptual Framework Summarized

Figure 7 therefore presents the resulting research framework. Due to the cyclical nature of the embryonic DM integration process, success in one project builds the momentum for the next until *top management support* has reached the point where there is commitment to broader DM integration.

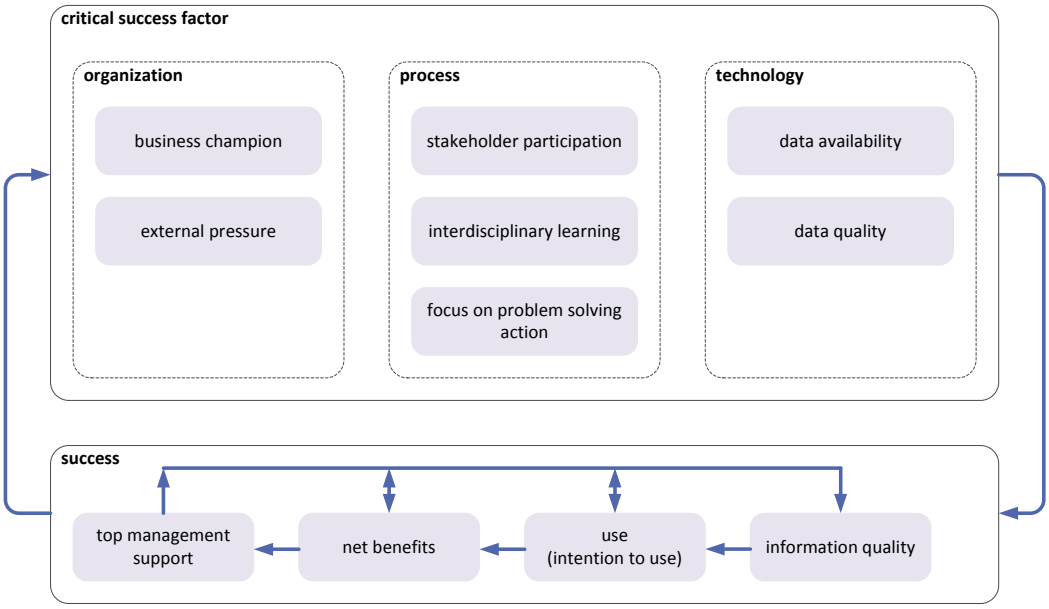

Figure 7: *Initial conceptual framework (CF_I).*

The sequence of the success measures is consistent with the levels of communication suggested by communication theory (Delone & McLean, 1992). First, information is produced. Depending on the problem and/or the problem-solving stage, the result of DM may be an insight, a model, or a model embedded in an IS. When this information is understandable and subsequently evaluated as useful and relevant (*information quality*), it induces its use or intention to use. In addition to the specific information produced, good

information quality potentially also motivates stakeholder *intention to use* DM in the future. Next, the use of information generates quantifiable *net benefits* (ROI/efficiency increase) and hence potentially *top management support.* Each of these measures may also influence the future performance of the preceding success measures by stimulating *stakeholder participation*. This is consistent with the feedback loops in the DeLone &McLean IS Success Model (Delone & McLean, 2003).

4.2 Interviews with Experienced DM Consultants

The purpose of the a priori theoretical framework (CF_1) is to sensitize us to the wide range of factors (in reality 29) potentially impacting early DM implementation. It is, however, largely based on literature that does not distinguish between embryonic DM management (without top management support) and subsequent DM implementation (counting on executive backing). Hence, the underlying assumption was that existing literature, and subsequently CF_1, also contains a "best set" of factors that contribute to successful embryonic DM implementation. We therefore engage in an explorative study specifically designed to test this assumption.

Data for this stage of the research was collected via semi-structured interviews conducted with experienced DM practitioners who were sufficiently familiar with embryonic organizational DM to adequately discuss the subject in an interview session. Semi-structured interviews were used to permit in-depth exploration of the research question with every study participant and to develop an understanding of the relevant issues as seen from the independent perspective of the DM practitioners (Blumberg, Cooper, & Schindler, 2008). An interview guide was purposefully constructed to permit comprehensive exploration of the factors impacting DM introduction in an organization. The use of semi-structured interviews permitted informants to freely express their views while also affording us the opportunity to raise issues suggested by our a priori framework (Saunders et al., 2009). The use of a semi-structured approach also permitted us to dynamically test our understanding of informant remarks throughout the course of each interview.

We considered the richness of the practitioners' DM experience, i.e. active participation in different types of DM problems, organizations, sectors, and countries. Together they bring to this research over 85 years of experience, much of it in embryonic DM initiatives. The interviewees were:

Udo Sglavo, SAS, Global Analytic Solutions Manager. He won the SAS CEO Award of Excellence in 2010. His particular interest is in helping organizations understand how to apply analytics to solve business problems.

Tom Khabaza, Ph.D., Institute of Data Miners. He has been a practitioner since 1992, much of the time as a Director of DM at SPSS. He co-authored Clementine, the SPSS's DM workbench, and the CRISP-DM methodology.

John F. Elder IV, Ph.D., Elder Research. He founded his DM consultancy in 1995. His company currently employs over 20 DM experts and works with numerous organizations

on different problems (text mining, stock selection, image recognition, cross-selling, drug efficacy, etc.).

Richard Boire, MBA, BoireFillerGroup. His experience in DM began in 1983 at Reader's Digest and further developed at American Express. In 1994 he formed his own consultancy and has since worked to solve database marketing problems in many different organizations.

According to Strauss and Corbin (1998), new informants should be identified until theoretical saturation is achieved. Because we insisted on the richness of the practitioners' experience and were pressed by our research schedule, the number of informants is low. In addition, we ran the risk of introducing bias into the research since, for the same reasons, all informants were external consultants. We planned on compensating for these weaknesses in the following stage (multiple case study) with a significant number of informants with different roles in the DM implementation process.

Table 6: *The concepts identified by DM consultants.*

Candidate CSFs		
data (integration, access, quality, quantity, management...), time to get data	concrete goals, clearly articulated business problem and baseline, metrics, measure improvement;	speed to value (show in 3-6-12-18 months), prototyping/pilot projects (assess ROI on small sample)
analysts' skills (business, communication, insight, interpersonal, team, educational, sales, intuition, experience, manage political issues, humility)	start with analytical strategy - what is the model going to be used for, the process should be carried out with deployment in mind	wait before buying a software solution
domain expert & IT support, user buy-in, trust, business/SMEs understand potential of analytics, confidence in technology, the right team, symbiosis	change management (processes, attitudes, way of thinking, culture)	project assessment, assess benefits potential early
business sponsorship, business champion, project owner; DM project must help them achieve personal goals	seamless model integration, embedding DM; monitoring of the model	well defined process/follow an analytical methodology, i.e. CRISP-DM
decision centricity	competitive context, how can analytics help, alignment with business strategy, understand high performance drivers of the organization	project planning, explicit deliverables
		technical readiness
		define terminology, well defined collaboration principles
		fixed budget for a project

Data collection comprised face-to-face interviews that were conducted with each practitioner. Interviews lasted on average just under 50 minutes. They were recorded and transcribed by the researchers to yield a total of approximately 40 single-spaced pages of text (Bole, 2013). After the interview, further clarifications (if necessary) were made by follow-up phone calls and email communications. The data were analyzed by content analysis technique, a constant comparison technique (Miles & Huberman, 1994), to identify major themes. In other words, the qualitative data were examined thematically and emergent themes were ranked by their frequency and later categorized. As a result, the theory developed reflected the actions, problems, and issues practitioners face. Finally, the results were reflected upon in a discussion with the remaining authors,

additional clarifications from the literature, and a review of the findings by the interviewees as recommended in (Carroll & Swatman, 2000).

The interviews led to the identification of a wide range of factors that might impact embryonic DM implementation (see Table 6). Many of these were, however, cited quite infrequently and were therefore judged to be of minor importance. As expected, the results show a biased, sales-oriented view. Given our interest in offering generalizable findings, we therefore judged that, beyond confirming the seven dimensions of the existing framework, the evidence was insufficient to introduce changes. CF_2 thus remains identical to CF_1. To avoid repetition, a selection of the statements supporting each potential CSF is presented in Table 7.

Table 7: *Summary of potential CSFs and supporting findings.*

***Potential CSF* / Informant support**
Business champion
You need somebody that is going to be engaged. If not, I would walk away from it. We have had cases when the champion has left the company and the project just slowly stalls.
It is the person with the highest interest in the success of a project, who puts his head down and says: I'll make sure that this is a success. Otherwise these kinds of projects never end. (…)You need to have a strong sponsor inside the company saying to the management: This is the impact we made.
The sponsor is the one who knows how the company works and how to set up things nicely.
External pressure
DM may help fulfill a legal requirement, such as compliance with Basel II in the case of a bank. A telecom operator may be rapidly losing users and corresponding revenues and profits to a competitor. These would be examples of a high level of "pain". Conversely, a dominant retail chain in a high growth market may be losing market share to new entrants, but its sales volume and profit continue to grow nevertheless. This company is not likely to try DM because they might not perceive the need to do anything differently (abr.).
Stakeholder participation
When you are deploying something, if it threatens anybody, they might be working hard to cause it to fail. That hurdle is always harder than it looks. You have to build allies by helping people look good, by helping them achieve their career goals, otherwise they are going to be blocking you more than helping you.
Even if you successfully implement it and the users are not using it, then it is a dead end.
An organization can command: Do this! But people will just turn it off if they do not see how it helps them meet their personal goals. Most people are not on commission and therefore not as easily motivated as salesmen. But it increases a company's success if they are able to convince the employees that it will help them do their job better.
Everything in these projects is trust related. That is the secret to success. If you run into problems, which will always occur, you need to have a trusted relationship so that you can work together to overcome these problems (abr.).
It is about the level of trust, that they trust you, if there is a problem, that you can work it out.

Table 7: *Summary of potential CSFs and supporting findings (Continued).*

Interdisciplinary learning
The changes – it may not even be a question of change, except maybe a change in mind-set. (…) It is not exactly knowledge. It is a way of thinking. There is a difference, because you could do a Ph.D. in statistics or DM, and still not have the right way of thinking. It is not expertise. It is knowing what could be done, how it relates, how it is relevant.
More and more we are finding that companies want to see you having done almost exactly what they want you to do for someone else already. So that you go through the painful learning process on somebody else's nickel.
We try to impart learning that they would otherwise not have. That way we get them engaged.
Learning is not merely domain but also organization specific (abr.).
Focus on problem solving action
I always want to think that change is a by-product of the real process, which is fitting solutions to problems. (…) It is a problem of targeting, defection or whatever. They just need to work out how to do it a little bit better.
[The business problem] should be in parallel or even around [the DM process]. It is before and after and in the middle.
[For the DM expert there is always a] temptation to do the thing that they know how to do and not the thing that needs doing. That is the hardest part, looking at the problem from the client's viewpoint.
Data availability
You have the budget and time, but not the authority to get the data (…) The huge stopping point is getting the data. (…) A lot of the projects will be delayed for months and then ultimately cancelled when nobody was able to draw the data out.
They can have ugly data, on all kinds of different legacy systems, as long as you can get it. You do not need fancy systems to be able to do something.
You have to figure out if they have data available, and if not, the first thing to advise them is to start collecting it.
Data quality
The biggest risk is the data. If you do not have good data…
Data quality is just one issue …

Some issues discussed in depth by multiple informants reveal previously unreported insights that we consider valuable for the stakeholders in early DM initiatives. Salient considerations on *relationship management*, *interdisciplinary learning*, *focus on problem solving action*, and *data* are presented below.

4.2.1 Relationship management

An obvious, though frequently forgotten issue, is that the first DM initiative tends to also be the champion's first. This implies that their comprehension of DM is likely to be vague. Their expectations about the benefits and about the process may be unrealistic or erroneous, which is potentially frustrating for both the DM expert and the champion: *"It astonishes me, how often the client says: What should we do next? This is the first for them and they are hesitating. Frequently they will also say: Do you have the data? … and*

we say: No, our value is the service – the analytics. We can advise you on how to get the data… I think they would like to get a complete solution with a bow tie around it."

To make things more complex, the DM expert, who might be expected to provide advice on the process and benefits, is unable to do so beyond providing a very general framework. This is due to several factors. The first is that the DM process is explorative in nature. Moreover, the DM expert does not know the specific business model of the organization, its culture, the people who will be involved, and the data that is available: *"We do a lot of work for Nestle. What you need to understand is how marketing works for the business, how they are selling products. It is not enough to know some marketing principles. It is more the understanding of the mechanics of Nestle, how do they sell chocolate or baby food, how the different components work together."*

A DM expert must possess consulting skills in order to build relationships, *"in the concepts of a particular problem and the understanding of the problem and knowledge and building up a picture and a solution. And they* [consultants] *are building that picture all the time, from the first day they meet their prospect."* This indication prompted an overview of management consulting literature. Most of the informants' considerations in this section were confirmed and additional ones were found in (Schein, 1999).

The relationship is yet to be built which implies uncertainty. Knowing that, for both the champion and the DM expert, the particular DM initiative is their first, both need to be patient with each other, expecting from the other many questions and erroneous suppositions. Informants recommend patience because relationship management may be time-consuming and sometimes frustrating, which at times leads to wariness. However, with time, an appreciation of the other person and the other domain will develop, and importantly result in mutual trust. *"Everything in these projects is trust related. That is the secret to success. If you run into problems, which will always occur, you need to have a trusted relationship so that you can work together to overcome these problems."* And, "*It is about the level of trust, that they trust you, if there is a problem, that you can work it out."* The informants tend to speak of stakeholder support and commitment in terms of trust, which reflects stakeholder awareness and confidence in DM, and their understanding of how DM is relevant to their daily tasks and business in general.

An important part of relationship building is that a DM expert helps *"other people* [business champion in particular] *look good. You have to help them achieve their career goals by your success; otherwise they are going to be more blocking you than helping you."* Another practitioner corroborates this idea, *"you want to make sure that the business champion is looking good. They need to get something out of it. It might be a promotion or whatever they are looking for in their career."*

Interestingly, humility is an important characteristic of a DM expert. *"We focus on hiring people who are humble. Arrogant consultants are not liked. They have to know they don't have all the answers. They may have solved a few problems, but here is someone who has worked in the field for 20 years, maybe someone older than them. You need to listen carefully. You try to ask the questions that will help you formulate your response. Then you go back and write something. It shows you have considered the*

problem, maybe even consulted with other people. It shows respect."

Another related lesson proposed by various practitioners is to avoid the temptation to immediately show analytics results when meeting larger groups of stakeholders. It is a common misconception that the analytics results and related benefits will immediately win everyone over. This rarely happens, particularly in early analytics initiatives. Team members entering into such a meeting tend to be cautious, and sometimes defensive, due to their uncertainty about analytics. For them it is a new and unknown way of doing things, and they are still trying to find out exactly how it will change their work life. Their true concerns rarely surface in larger meetings. Instead, the cumulative defensiveness of stakeholders easily erupts in emotionally-charged attacks on analytics.

It is therefore much more effective to review the results with stakeholders individually. In such a setting, real fears are more likely to be openly discussed. With their insecurities addressed, stakeholders start to contribute their knowledge and slowly adopt the analytics solution as "their own". This leads to commitment. When all or most stakeholders reach this point of identification with the analytics solution, a larger meeting with all stakeholders may be constructive. Experience has demonstrated time and again that the individual approach is the only effective one and in the end much faster than trying to win everyone over at once.

4.2.2 Interdisciplinary Learning & Focus On Problem Solving Action

Taking DM on board does imply changes *"in particular if the whole art of becoming analytic is new to them, because most of the time their business processes will need to be adapted"*. However, a *business development* mind-set is more appropriate than a *change management* approach, because changes are incremental and evolutionary. "DM is not about radical change. It is very different from that. An organization has a straightforward business problem which analytics can help with. It is not really much of change content. It's a problem of targeting, or a problem of defection, or a problem of resource allocation, or whatever. They just need to work out how to do it a little bit better." Besides, *"if you start with the word change, that* makes DM sound very expensive and disruptive. People will resist it. (…) *It may not even be a question of change, except maybe a change in mind-set.* (…) I*t is not about knowledge. It is a way of thinking. There is a difference, because you could do a Ph.D. in statistics or DM, and still not have the right way of thinking. It is not expertise. It is knowing what could be done, how it relates, how it is relevant."* It is important to note that neither business people nor DM experts can competently answer these questions except through collaboration, i.e. *interdisciplinary learning*.

One way to facilitate interdisciplinary learning is by provoking frequent interactions among the stakeholders. The practitioners use different techniques, e.g. *"rapid prototyping"* and *"project milestones"*, or they simply insist on breaking longer projects into shorter ones of no more than three month's duration. The objective is always the same: to manage expectations via frequent interactions rooted in the problem and emerging solutions.

When the potential champion is hesitant about DM, informants also recommend an introductory project, which they might call *"day's assessment", "data discovery", "proof of concept",* or *"assessment workshop"*. Essentially it is a mini consulting project, with two high level aims: to reduce the risk of the organization by limiting its cost, and to facilitate *"knowledge transfer"*. The DM experts learn about the organization (its business, people, and data) while business people may learn more about the DM expert, DM process, and the specific benefits DM could offer them. Moreover, after this exercise showing potential benefits, further project planning and estimation of future gains and costs becomes more manageable.

All of the informants agree on the need to focus on the problem and a solution throughout the process, because for the DM expert there is always a "*temptation to do the thing that they know how to do and not the thing that needs doing. That is the hardest part, looking at the problem from the client's viewpoint.*" In different terms, another practitioner summarizes a similar idea, a complaint of many DM experts, "*We have CRISP-DM, which tells us how data mining is done, to get good DM results. But there is another half to the whole process, which is helping the organization take this on board and become a DM organization, rather than just consult them to put the proof of concept in the corner. We need something to help us help the organization, a bit like CRISP-DM, but the other half of the equation – the organizational half.*" The lack of such a methodology appears to be the root cause of why *"a third of our projects are successful all the way through the research phase* [development of a DM model] *and don't get implemented for some reason."*

4.2.3 Data considerations

The practitioners stressed that in the first DM projects, the inability to obtain data often delays or kills a project. Unless an organization has already built their data warehouse, IT will be asked to supply the data, which is often *"the huge stopping point… a lot of the projects will be delayed for months and then ultimately cancelled when nobody was able to draw the data out."* The IT department often finds excuses: *"We have a lot of work, this is extra work for us…"* The practitioners *"try to avoid work for them"* by offering to pull data out themselves or try to convince IT to give them the data in whichever form it comes out the easiest. Still, the IT personnel have difficulty understanding *"how important it is to get the data and they can just ignore you and kill you off".* At this stage the only thing that helps is if the sponsor has enough power to say to them, *"this is your top priority. Get it as soon as possible!"*

An important finding of the DM readiness assessment might be *"to advise the client to start collecting data."* If an organization starts by collecting data counting on a DM expert's advice, it will take a big step forward. In a few weeks or months, it might already accumulate enough data to meet the quality requirements and enable DM analysis.

4.3 Conceptual Framework Validation: A Multiple Case Study

Finally, we test the conceptual framework CF_2 in a multiple case study. Case study methodology was chosen because it has a distinct advantage when 'how' and 'why' questions are asked about a contemporary set of events over which the investigator has little or no control (Yin, 2003). Thus, it provides better explanations and understanding of the examined phenomenon, which would be lost in quantitative designs (Benbasat, Goldstein, & Mead, 1987).

Considering the several known potential weaknesses of the case study method (Benbasat et al., 1987), a case study protocol was designed, with careful documentation of all procedures relating to the data collection and analysis phases of the study. Data collection entailed semi-structured interviews with 17 stakeholders of early DM projects and a number of other sources, including feedback sessions, meeting minutes, presentations, internal communications, web sites, and field-notes, to facilitate data triangulation. The interviews were semi-structured, each completed by the same researcher within approximately one hour. All interviews followed the same structure and format (as pre-specified by the case study protocol) and commenced with an open discussion on perceived success/failure factors. Subsequently, the individual constructs of the research framework were introduced and the informants' opinions on the relevance and importance of these issues were sought. This approach enabled the researchers to obtain new ideas to enhance the framework while simultaneously validating existing a priori constructs. Reliability was enhanced through the use of a detailed case protocol and a case database containing all relevant data collected (Benbasat et al., 1987; Yin, 2003).

All 17 interviews were recorded and transcribed, amounting to a total of approximately 180 single-spaced pages of text (Bole, 2013). Data analysis was an on-going process that started immediately after the first interview, with sorting information into the framework dimensions through coding. Information not possible to classify was given special codes and marked for potential future extension of the framework. The codes were refined as the analysis evolved. Initially, any direct or implied existence of the constructs of CF_2 was coded. Later, the analysis of the information already coded under each of the constructs was refined to distinguish between citations that indicated mere existence of the constructs and those that specified their criticality. These implicit and explicit counts were then reflected as a basis for making our judgments. A cross case analysis approach was used to gain better understanding and increase the generalizability of the findings (Saunders et al., 2009). The research cycle was completed by formal reflection on the findings in meetings of the authors.

A multiple-case design is recommended for explorative studies (Benbasat et al., 1987; Yin, 2003). In contrast to sampling logic, a case study is an investigation following replication logic that leads to analytic generalization (Yin, 2003). Thus, multiple cases in this research should be regarded as multiple experiments and not multiple respondents in a survey (Yin, 2003), i.e. relevance rather than representativeness is prioritized in case selection. We selected eight organizations (Amsterdam Police, Domel, GZA Hospitals,

IMH, ING, MercadoLibre, Merck KGaA, and Telekom Austria) that had recently introduced DM (the process of DM introduction was the unit of analysis). In none of the cases was the DM proponent a top executive.

To support the generalizability of the findings, the organizations were selected such that they differ in various dimensions. Two are not-for-profit organizations (one non-governmental). The commercial institutions operate in different industries (e-commerce, electronics, financial, pharmaceuticals, steel, and telecommunications). They vary in size: 1,000-10,000 employees (4), 10,000-50,000 employees (3), and over 100,000 employees (1). In their embryonic DM projects, they addressed an array of problems including a recommendation system, fraud detection, text mining, database marketing, quality control, process mining, and product lifetime prediction. The organizations are based in Argentina, Austria, Belgium, Germany, the Netherlands (2), Russia, and Slovenia. Table 8 elaborates on the antecedents leading to the first DM initiatives for each organization.

Table 8: *The antecedents leading to the first DM projects.*

	Organization / description of antecedents leading to DM initiative
A	The company runs into a production quality problem. R&D encounters difficulties in discovering its root causes. They had recently seen a presentation by a DM services company and asked them for help with data analysis.
B	This organization's top priority was to become intelligence led. Their data warehouse was full of unstructured data (text) that they wanted to leverage. They contracted a DM consultant.
C	The company was engaged in basic research to design a prediction model for the longevity of one of their products. In the literature they observed similar models built with DM. They started working with an external DM expert.
D	A person responsible for process optimization had some intuitions about their processes but no tools to investigate them. Based on industry reports that recommend use of DM she decided to try a DM approach with an external consultant.
E	The organization had been piling client data in transactional databases. They observed that DM was being used within the industry to optimize targeting costs. They hired a DM expert.
F	The organization's data warehouse had been implemented. Their Head of BI felt it was time to start leveraging their data with DM. They employed a DM expert to work on different projects.
G	The organization had been buying results of DM analysis from a DM company, sector specialist. They wanted to have more influence on the model development process. They hired an external DM consultant to help them with client segmentation.
H	The first predictive model for churn was provided by the mother company. The users did not understand the model and felt it was becoming outdated. They employed a DM expert to build a new one.

The informants had to have first-hand experience of the embryonic DM projects. In addition, to enable the view of the CSFs from different perspectives, the interviewees span all significant roles participating in the DM process. Nine interviewees were DM

experts (5 internal and 4 external consultants, 3 of them academics). This balances well with the four practitioners interviewed in the previous stage of this research (non-academic, external consultants). The remaining 8 interviewees were from the business side: end users (2), domain experts (2), IT experts (2), and business champions (2).

4.3.1 'Success' Measures

In section 4.1.1, our literature review yielded four success measures of embryonic DM initiatives, i.e. *information quality, use/intention to use, net benefits,* and *top management support*. In this section we describe their validation through the case study.

The case analysis indicates that in embryonic DM initiatives success tends to be judged based on stakeholder perception of efficiency gains. ROI was not estimated in any of the eight organizations we studied. When inquired about the ROI, informants replied that in the early phases they did not perform any financial calculations. Only in two cases, where DM was used for database marketing, did the organization evaluate and compare the efficiency of predictions with and without DM. In all the remaining cases, the informants essentially stated, *"I believe what we are doing will be used a lot and will be beneficial. But we have not calculated ROI."* Similarly, another informant commented, *"We didn't do it yet. We are planning to do that in the future, but not at this moment, because we are not that far in developing this system. (...) But we are aware of the advantages of what we are doing."* The DM proponents tend to see in the technology an investment into the organization's future in the sense of business or product development. This observation led us to re-specify the success measure *net benefits* before proceeding to the evaluation of success. ROI was abandoned. Moreover, since a quantifiable measure of success could not be relied on in the majority of the cases, we propose to use *(perceived) net benefits* instead. It may be defined as the degree to which stakeholders are confident that the DM initiative will contribute towards the intended improvement of organizational performance.

The evaluation of success measures (Table 9) shows that it is possible to obtain key decision-maker support even if the perceived efficiency increase resulting from the first projects is merely acceptable (C) or even poor (E). Hence, a lack of immediate and considerable business improvements does not necessarily determine an initiative's effectiveness in achieving *top management support*. For example, a project may fail to reach its goals, but the organization learns from this failure and is eager to start another project based on the newly acquired knowledge. This implies that the organization has understood how DM can help it improve. Conversely, a project may successfully reach its goals but results in an aversion to DM by some key stakeholders. Similar observations are also reported in (Hermiz, 1999; Weiss, 2009). Therefore, when the final aim is broader DM implementation, *top management support* is a good proxy for the overall success of early DM projects. This rationale is also supported by the majority of informants (8 explicitly and 6 implicitly).

Table 9: *Success in early DM projects.*

Organization / *Success measures*	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Information quality	?	o → +	-	+	o → +	+	?	+
Intention to use	o	+	+	+	- → +	+	-	+
(Perceived) net benefits	-	+	o	+	- → +	+	-	+
Top management support	-	+	o	+	+	+	-	+
Overall success	U	S	MS	S	U → S	S	U	S

Note: + = Good, o = Acceptable, - = Poor, ? = Inconclusive
→ = change in time
S = Successful, MS = Moderately Successful, U = Unsuccessful

Based on this evidence, we evaluated those initiatives that led to unquestioned support by top management as successful, those where executive backing was not secured as unsuccessful, and those where top management showed increased interest in DM but required more time and evidence to grant full sponsorship as moderately successful. The analysis of triangulated results for all eight organizations yields four instances in which undisputed success emerged (B, D, F, H). One initiative was judged as successful although performance in the different dimensions of success was mixed (E). One case was moderately successful (C) and two unsuccessful (A, G).

In both of the unsuccessful initiatives, the break down resulted from the inability to arrive at a common understanding of the problem. DM experts claim that after showing the results from the initial data analysis, the domain experts refrained from volunteering additional insights to enable further iterations. Domain experts, on the other hand, claim that nothing worth pursuing was presented to them. Due to contradictory evidence, it was not possible to evaluate information quality. In both cases projects were suspended; however, in organization A the DM promoter demonstrated increased interest in DM and hoped that in the near future they would have the opportunity to try DM again. The initiative has also raised his awareness of the data quality and access issues that the organization must resolve. For this reason intention to use was evaluated as acceptable.

The stakeholders of the moderately successful case (C) show a clear inclination to continue to use DM in the future. The statements of the informants, domain experts in particular, show pre-existing awareness of the fact that the initiative would be interdisciplinary, iterative, and uncertain. This is the most salient difference with respect to the unsuccessful cases presented above. Initial DM results were not actionable. However, the initiative has helped them understand that the problem originated in the low quality of their data. As a result, they learned what would have to be improved in the data collection process to enable broader DM implementation - they state several possible applications. Their top management has been intrigued by the DM results. However, to be motivated to back broader DM implementation, they wish to see further tangible benefits.

Among the successful instances, we single out the one with low scores in some success dimensions (E). This anomaly later served as particularly useful in cross case analysis. In

this case, the DM expert initially had difficulties in securing trust. The domain experts and end users remained skeptical of both technology and the DM expert for an extended period of time. This was reflected in the average quality of insights, their low use and a lack of perceived business performance improvements. However, the organization did not succumb to the repetitive temptation to abandon DM. Thus trust was established, albeit slowly, through repetitive iterations. Subsequently, *information quality*, *intention to use*, and *(perceived) net benefits* also improved and eventually executive buy-in was achieved.

The DM implementation process among some of the unquestionably successful cases (B, D, F, and H) shows glitches in interdisciplinary collaboration and subsequently in the initial quality of information provided by DM analysis. However, these problems were overcome due to different favorable conditions. For example, the IT/database expert in organization B was a wise, mature person with well-developed interpersonal skills. He had substantial informal authority among his colleagues and extremely good knowledge of the organization. Although he was not the *business champion*, he soon became interested in DM and turned out to be an excellent guide to the external DM expert. The stakeholders in D and F were accustomed to interdisciplinary collaboration and to data driven decision making. In addition to these circumstances, the DM expert hired by organization H had both business and analytic backgrounds. These conditions supported the DM implementation process to yield actionable insights early (D, F, and H) while helping overcome the initial shortcomings in *information quality* in organization B. This set the stage for good performance on the remaining success measures.

4.3.2 'Success' Factors

Following the evaluation of success, we next demonstrate how overall success compared against the management of the CSFs. A cross-case analysis (Table 10) was completed as recommended by (Miles & Huberman, 1994) and subsequently used as the point of departure in analyzing patterns in the data. (For easier overview and analysis we group the organizations based on overall success.)

Most of the informants emphasize the key role of a *business champion.* From a total of 71 citations, 14 express its criticality. We mark with a "+" those instances where a business champion was directly involved in the initiative by closely overseeing and managing the team and the process. In G, C, and D the champion's involvement included provision of most of the domain expertise. In F the champion was the head of the BI department and in H the head of product marketing and customer relationship management (CRM). They both closely followed the initiatives to the point of active presence in critical meetings with domain experts, end users, and IT. Marked with an "o" are those cases where the business champion supported the initiative from a distance, i.e. without direct involvement in its management. In B and E the champion contracted a DM consultant and beyond that point completely entrusted the initiative to their database expert and the DM expert, respectively. A (marked with a "-") was the only instance in which the DM project was not backed by anyone in the organization. Instead, DM was promoted by a close but external collaborator as an extension of a government subsidized

R&D project.

Table 10: *Evaluation matrix: management of the CSFs.*

Overall success	U		MS	S	S			
Organization / *CSF*	*A*	*G*	*C*	*E*	*B*	*D*	*F*	*H*
business champion	-	+	+	o	o	+	+	+
external pressure	+	-	+	+	-	+	+	+
stakeholder participation	-	o	+	- → +	o → +	+	+	+
interdisciplinary learning	-	+	+	- → +	o → +	+	+	+
focus on problem solving action	+	o	o → +	o → +	o → +	+	+	+
data availability	+	+	o	o → +	+	+	+	+
data quality	o	o	- → o	+	- → o	+	+	+
process facilitation			**	*	*	**	**	**

Note: + = Good, o = Acceptable, - = Poor, → = change in time
* = personal function, ** = shared responsibility
S = Successful, MS = Moderately Successful, U = Unsuccessful,
S = Successful, initially not successful (see Table 9)

External pressure was the least cited issue with 11 general citations among which two express its importance. Regulatory pressure was present in instance D. Organizations A and C were pressed by the increasing product quality offering of their competitors. On the other hand, E, F, and H compete in the service sector where many of their competitors had already implemented DM, primarily to improve their understanding of their customers, i.e. CRM and marketing. Instances G and B present no evidence of external forces influencing DM adoption. Among these, B was judged successful in early DM management. This evidence suggests that *external pressure* is not a factor *necessary* for success.

The criticality of *stakeholder participation* management is reflected in its second score in the number of citations (62 referring to its importance and 165 in total). In A stakeholder participation was minimal due to a lack of a business champion. G and B show acceptable levels of stakeholder participation. In G there appeared to be internal commitment of the business champion and domain experts throughout the project; however, the final results were discarded. Presumably they did not support the champion's hidden agenda, which was never disclosed to the DM consultant, except implicitly through requests to change some assumptions (in the domain of DM), which were judged unreasonable by the DM consultant. In E and B *stakeholder participation* was not managed initially. There was a significant level of mistrust towards DM and DM experts. This difficulty slowed down the process, but was eventually overcome. In the remaining cases, *stakeholder participation* management motivated commitment to the DM initiative from its beginning.

The cases reveal some previously unreported (in DM literature) approaches in *stakeholder participation* management. Several informants recommend that to gain trust and acceptance, DM experts must not make other stakeholders feel inferior, which is

consistent with the recommendations in management consulting literature on how to build helping relationships, e.g. (Schein, 1999). Another often recommended strategy is to understand the points of view and to gain the trust of individual team members separately before attempting to achieve the commitment of the whole multi-disciplinary team. Attempting to gain the support of everyone at once, i.e. in the same meeting, tends to be too complex. The research on cross-functional collaboration management is supportive of this finding, e.g. (Holland, Gaston, & Gomes, 2000; Kirkman, Rosen, Gibson, Tesluk, & McPherson, 2002; Thamhain, 2004).

Management of *interdisciplinary learning* was by far the most cited factor across all sites and informants. Nearly a third of 222 citations explicitly state its importance. A common denominator of the instances that perform well in this dimension is that the group of stakeholders was accustomed to fact based decision making (D, F, H) or interdisciplinary innovation (C, D). In accordance with related research, e.g. (Beers et al., 2005; Fosfuri & Tribó, 2008; Holland et al., 2000), the stakeholders in such organizational cultures were attentive to the potential dangers of interdisciplinary collaboration. Conversely, in more traditional environments (A, G, E, B), the pitfalls of interdisciplinary collaboration came as a surprise. Among these, instance G was a special case because the DM consultant had previously done a great deal of work in the same industry. Organizations E and B show clear improvements in the awareness and management of the issue during the DM initiative.

Focus on problem solving action was also among the most cited issues with 132 citations, of which 37 explicitly refer to its importance. The informants from all the sites confirm that it is important to focus on actionability, on a clear connection between DM and an unambiguous business improvement. The DM initiative in organization C was applied to basic research. The problem definition therefore crystalized with the passing of time. In cases A, D, F, and H, the understanding of the problem to be solved was common among all stakeholders from the beginning to the end of the initiative. Conversely, in instances G, E and B, discrepancies in problem interpretation became evident with time. In case G (and partly in E) it was due to hidden agendas. A more common source of misunderstanding was the impatience in the problem definition stage (E, B), which resulted in merely apparent agreement on the problem definition, i.e. the stakeholders erroneously believed they had achieved common understanding of a problem. In all the cases the error was corrected, once discovered. The only negative consequence was a delay in the time to results. Informants recommend more interactions and clear communication to avoid the slip.

Most informants cite *data availability* as a relevant factor. It is, however, the second least cited issue (28 total citations, 4 citing its importance). This is partly due to the fact that prior to their first DM initiative, several organizations had already built a data warehouse (B, F, H) or routinely used their transactional data for analytical purposes (A, G, D), i.e. they had resolved the data access issue. Subsequently, the informants from these organizations consider data availability as self-evident. In none of the remaining instances was data impossible to obtain, although accessing it slowed down the process of

DM implementation either due to the fact that it had to be obtained from the transactional systems (E) or several different databases (C). At site C the existing data also proved inadequately collected and too scarce for DM analysis; hence it had to be made available through additional testing and measurements.

Data quality is consistently cited across informants and sites. The issue was mentioned 69 times by the informants (14 of these specifically citing its importance). In the cases where data quality was initially poor (C and B), the issue was approached opportunistically. The DM initiative helped the stakeholders understand the necessity to revise and improve the process of data gathering. For example, one organization that applied DM for text analysis spent a lot of time detecting errors. *"We started discovering many things which were not so good with the definition, training of* [employees]*, a lot of faulty case labels... so we discovered a lot of additional domain knowledge, enriched domain knowledge."* This effect was present at most sites, as well as among those that had already built a data warehouse or habitually used their data for other kinds of analysis (D, F, H). These, however, clearly demonstrate fewer difficulties related to data quality.

4.3.3 Cross-Case Analysis

Early cross-case analysis was inconclusive. On the one hand, instances C, D, F, and H demonstrate that above average score in the CSFs leads to the success of a DM initiative. The unsuccessful case A conversely confirms that bad performance on the CSFs leads to an unsuccessful outcome. Instance G appears to prove the contrary, as the CSFs were reasonably managed; however, its failure is explained by the presumed hidden agenda of the business champion. The disconcerting case was E and to a lesser extent B. While eventually successful, they initially show poor or average management of *stakeholder participation*, *interdisciplinary learning,* and *focus on problem solving action*. The deficit in the management of these process related factors was particularly acute in the early stages of these initiatives. Importantly, the two instances challenge the the assumption that the business champion "is actively involved in the DM process." Surprisingly, in both instances at some point the management of the process factors improved considerably although the champion remained distant throughout the initiative.

The explanation was found in case E where embryonic DM projects performed poorly or average on *information quality, use (intention to use),* and *(perceived) net benefits* yet managed to secure *top management support* and hence final success. In this initiative there is a turning point that coincides with the involvement of a new person (an additional DM expert). Detailed analysis revealed that her contribution was primarily that of managing the process CSFs. The informants from site E, hence, speak of her contribution as that of facilitation. We therefore adopt the term *process facilitation* to denominate this role. Moreover, it was confirmed that a facilitator role is recommended in the literature on management of interdisciplinary collaboration, although other terms such as moderator, coordinator, or improvement leader may also be found, e.g. (Harkness, Kettinger, & Segars, 1996; Markus, 2001; Senge, 2006). Following this lead, we analyzed the remaining cases. In instance B we discovered that a facilitator also made a difference. The

data expert who was entrusted the initiative by the champion possessed the qualities of the process facilitators and, like his counterpart in E, managed the process CSFs without any formal power. In the two cases the role was not appointed. It emerged to cover a deficit recognized by the persons who naturally, as a consequence of their personal characteristics and desire to help the initiative, evolved in the role.

We also found evidence of this role in organizations C, D, F, and H. In these cases, however, the role was shared by all stakeholders as a consequence of culture. These organizations were accustomed to fact-based decision-making, and to interdisciplinary and innovative initiatives, which was not the case in the more conservative environments of B and E. This insight was suggested by one informant from F when she contrasted the experiences in the current organization, which is very innovative and interdisciplinary, with the previous organization, which was much more conservative. Moreover, the insight was also confirmed in D. The role of a coordinator of an interdisciplinary team had been made formal prior to the DM initiative. The goal of the team had been to analyze internal processes and improve them. The first DM initiative changed little of the functioning of the team (one additional professional, i.e. the DM expert, and a different way of analyzing data, i.e. DM). However, based on her experience, the coordinator anticipates greater resistance when she sets out to improve other processes with new teams unaccustomed to interdisciplinary process dynamics and to making recurrent improvements based on analytical results.

Case data indicates that the *process facilitation* role may be taken up by any stakeholder. The DM expert helped facilitate the process in E and the IT expert in B. The most salient characteristic of the process facilitators in our study is that they are naturally attentive to the people issues involved in management of *stakeholder participation* and *interdisciplinary learning*. This is consistent with the suggestion of one DM practitioner that DM initiatives often bring together people from different departments within an organization that would otherwise never communicate. Process facilitators also exhibit a tendency for business development and a mind focused on actionability of proposed solutions, i.e. management of *focus on problem solving action.* In addition, the facilitators are inclined to internally propagate DM, laterally and towards the top of the organization.

4.3.4 Revised Conceptual Framework

The findings summarized above are reflected in the final conceptual framework (CF_3) as shown in Figure 8. Following our case study analysis, conceptual framework CF_2 is re-specified. Success dimension *net benefits* was modified to *(perceived) net benefits* because there was no evidence of ROI calculations to be found. Moreover, only in two cases was an efficiency increase quantified, while in the remaining instances the stakeholders relied on their perceptions to judge business improvements.

External pressure is downplayed because it was not identified across three of the sites studied. Since two of them were found to be successful in their embryonic-stage DM initiative, we conclude that *external pressure* is not a necessary CSF. It may, however, be considered as an important facilitating factor in the decision to adopt DM.

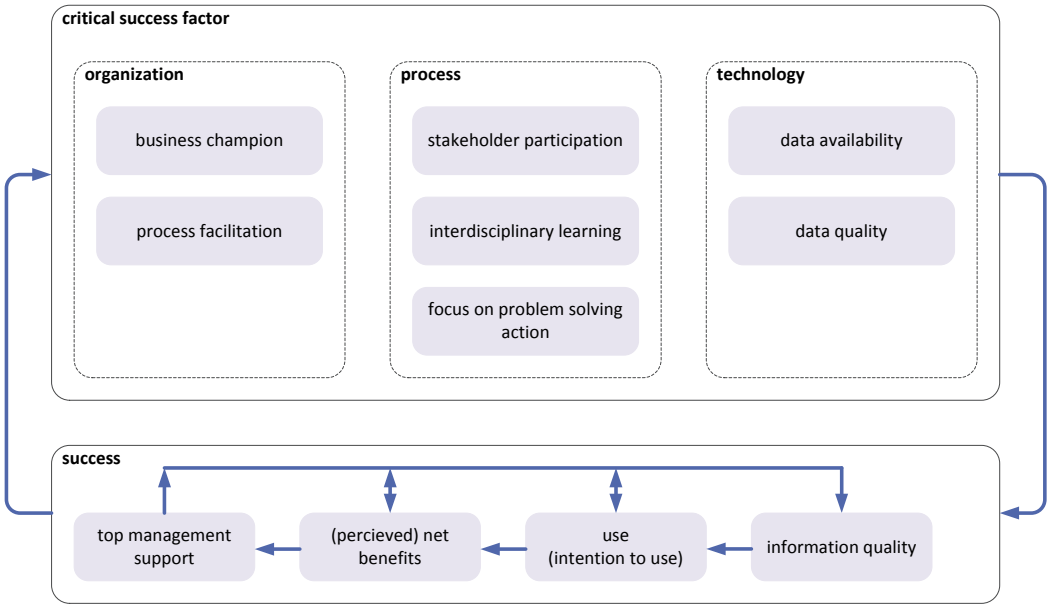

Figure 8: *Re-specified conceptual framework (CF_3).*

The above contraction in the *organization* category of the framework is offset by an expansion: *Process facilitation* is a role in a DM initiative which helps manage the DM process and also supports the internal promotion of DM. Case study data shows that: (1) a process facilitator brings to a DM initiative an important improvement to the DM process, particularly in its multi-disciplinary collaboration aspect; (2) the role may be shared or taken up by any stakeholder, including business champion; and (3) the appointment of a process facilitator was not intentionally managed as part of any DM initiative (a facilitator either emerged or had existed prior to DM introduction). Therefore, it is important that a DM champion considers this role, particularly in conservative environments not prone to interdisciplinary product or business development and fact-based decision making.

The surprising discovery of the role of the *process facilitation* factor suggests a modification in the definition of the CSF *business champion*. The existing definition suggests that a champion "is actively involved in the DM process." The evidence proves this statement false. In three cases the DM pioneer was a rather distant figure. We therefore refine the definition in the following terms: A *business champion* is someone within the organization who understands the potential of DM, has a business problem that may be solved with the use of DM, and is actively *or through delegation* involved in the DM process and in the internal promotion of DM.

4.3.4.1 Interdisciplinary Collaboration Elaborated

Management of interdisciplinary collaboration merited the most attention by our informants. We elaborate on the issue since previous DM theory has not addressed the root causes of its complexity. Interdisciplinary collaboration research shows that the source of difficulty in multi-disciplinary team collaboration is in mental models or cognition, e.g. (Beers et al., 2005; Du Chatenier et al., 2009; Senge, 2006). Different mental models (cognitive distance) are at the root of the differences in the way we see, comprehend, express, approach, and solve problems. Mental models are a set of basic

assumptions about the world and how it works (Senge, 2006). People need them as a way of dealing with information overload that is far beyond their processing capacity (Argyris, 1977; Simon, 1996). A person's assumptions depend on that person's cultural, educational, and professional background (Boland Jr. & Tenkasi, 1995; Du Chatenier et al., 2009). The more diverse these are, the more difficult communication and collaboration are, resulting in break-downs in interdisciplinary activities, i.e. conflicts and project failures (Boland Jr. & Tenkasi, 1995; Tidd, Bessant, & Pavitt, 2005).

When properly managed, however, different mental models may be a source of creativity in problem solving (Ritter & Gemünden, 2004; Senge, 2006). Constructive interdisciplinary collaboration calls for the development of shared ways of thinking, or shared cognition (Beers et al., 2005; Markus, 2001; Spector & Kim, 2012). The more developed the shared cognition, the more effective and efficient joint efforts become as communication improves and mutual confidence and trust grow (Du Chatenier et al., 2009; Renzl, 2008). This process requires time and methods of coordination that enable stakeholders from different disciplines to resolve their disagreements and achieve unity of effort (Boland Jr. & Tenkasi, 1995; Harkness et al., 1996; Lichtenstein, Bendall, & Adam, 2008; Markus et al., 2002). Interestingly, research implies that methods of coordination are necessary even in multi-functional teams that have been collaborating for a long time within the same organization (Harkness et al., 1996; Senge, 2006). Therefore, DM implementation, be it with internal or external DM experts, requires a method of coordination. This need is even more pronounced in embryonic DM initiatives, as was highlighted in our interviews with DM practitioners (section 4.1).

According to interdisciplinary collaboration theory, the process of an individual's knowledge becoming part of the solution to a complex problem, from being in the mind of one person to becoming a team's constructed knowledge, requires (1) the externalization of tacit knowledge (mental models), (2) its internalization by the remaining team members, and (3) the negotiation of meaning in order to arrive to a common understanding (see Figure 9). Only when this level of shared thinking is reached can it become the basis for constructive interdisciplinary problem solving (Beers et al., 2005).

When the concept of mental models is understood, it becomes evident why the DM problem definition process is uncertain and requires iterations. It also suggests why the requests for a clearly defined business problem of many DM reports, e.g. (Hermiz, 1999; Kohavi et al., 2004; Lavrač et al., 2004; Weiss, 2009), are difficult to satisfy without iterations, particularly when introducing DM in an organization. Moreover, there is evidence in all instances of our case study that similar difficulties also emerge at lower levels, in the details. For example, a commonly reported difficulty for a DM expert is to go off track by (often implicitly and always unintentionally) making domain-related assumptions that had not been addressed explicitly by the domain expert (also unintentionally because they could not know a priori which pieces of information would be relevant). Analogously, domain experts were reported to go astray due to the same causes.

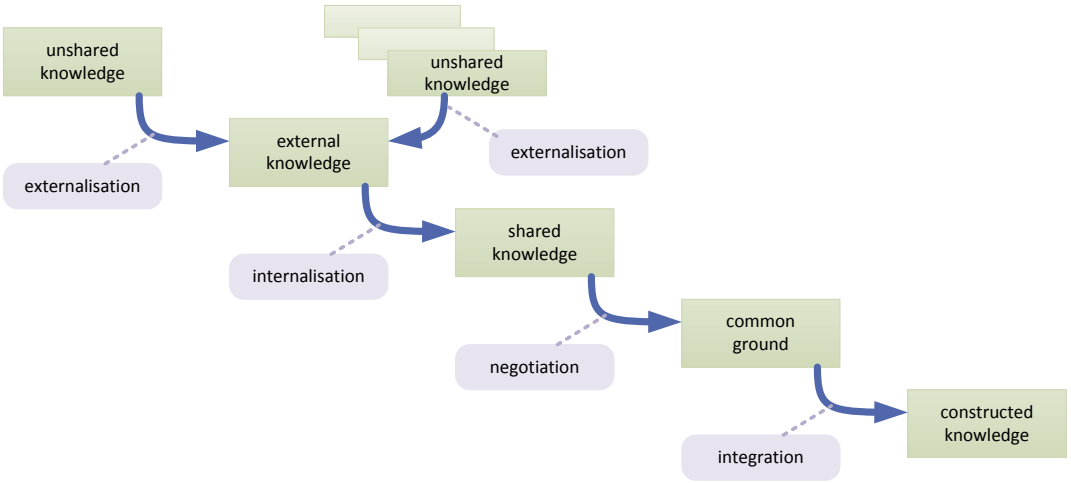

Figure 9: *From unshared knowledge to constructed knowledge; adopted from (Beers et al., 2005).*

In dealing with interdisciplinary collaboration, beyond the common advice (frequently given in DM literature) of making DM digestible by refraining from technical language and the use of visuals, a salient lesson learned by the DM experts we interviewed in the case study was the strategy of remaining silent as soon as the first end user or domain expert catches the idea. *"If one has understood it, then he or she can perfectly explain it to his colleagues. Better than me because they are speaking a different language! I am not speaking their language. They have different language they use, quite a different language!"*

While recommended, additional communication cannot always help. The unfortunate circumstance is that often these unconsciously made suppositions can only be discovered by the other party once they materialize in analysis results, a model, or action. Hence, to maintain the desired focus and avoid delays, it is beneficial to hold frequent meetings where outputs or plans of action of one or the other side are reviewed, as recommended by interdisciplinary innovation research and the work on Emerging Knowledge Processes, e.g. (Fagerberg, 2005; Harkness et al., 1996; Markus et al., 2002). Hence, "one shot modeling" (the introduction of DM in terms of the client providing the data, describing the problem in one meeting and expecting the DM expert to comes back with a solution in a few weeks) is not likely to yield satisfactory results for the client. The informants, who are external DM consultants, reaffirm this claim by drawing on all of their experiences. They suggest that client satisfaction tends to be conditioned on their a priori disposition to engage in a longer (some months) problem-solving collaboration with a DM expert. Unfortunately, the overwhelming majority of clients still expect and demand a one shot modeling DM process.

Erroneous expectations of a linear process in the other domain may point to one of the greatest difficulties of the embryonic DM process. As suggested by Davenport and Harris (2007), the embryonic-stage DM integration process is exploratory and iterative from the organizational point of view. Business people understand this; however, as shown in the previous paragraph, they erroneously expect the DM process to be linear. Yet, the DM process is also explorative and iterative (P. Chapman et al., 2000; Fayyad et al., 1996).

Moreover, our case study data shows that DM experts often also fall into a similar trap of expecting a linear process on the business side. They mistakenly expect the business people to provide a *clearly defined business problem*. This is also suggested in various reports (Hermiz, 1999; Lavrač et al., 2004; Weiss, 2009). Similarly, when DM experts present well-performing DM models (in terms of classification accuracy, ROC analysis) to business people they are often surprised when they do not get a positive response. This behavior shows that DM experts also erroneously expect a linear business process of DM integration in an organization. Unfortunately, business people are usually unable to identify a DM problem, much less clearly define it, without the support of DM experts. This is due to their deficient knowledge of DM. The involvement of both DM experts and business people is therefore required at every step of the process.

These erroneous and conflicting expectations point to an important gap. Currently the interface between the organizational and analytical process models has not been defined. This gap may be bridged through another process model. It should integrate the interplay between the two explorative and iterative processes to help merge the DM and the organizational perspectives on the same embryonic DM process. Such a method of coordination should facilitate the development of shared cognition and stakeholder participation to yield an increased unity of effort in embryonic DM initiatives.

4.3.4.2 Saliency of the Process CSFs

The case study informants speak significantly more (nearly 3 times as many citations) about the process CSFs than about the remaining CSFs. This might be explained by the nature of the CSFs. Organization and technology CSFs cannot be managed in the same way as process CSFs. The former provide fewer opportunities to be commented on because they are either present or not. As such, they might be better viewed as enablers of embryonic DM initiatives. The *business champion* is either present or there is no DM. External pressure, if present, can be leveraged to develop a sense of urgency, but the DM proponent cannot influence it. *Data availability* can only be managed if an organization is in possession of data and the champion has sufficient authority. Similarly, *data quality* can only be managed if the champion has enough power, which is not a likely scenario. *Process facilitation* is a special case. Its existence and criticality has been established in this investigation; hence in the early interviews, prior to its identification, the issue had not been discussed. The organization and technology CSFs therefore enable and facilitate the DM initiative, but cannot be managed in the same sense as the process CSFs.

The CSFs are systemic in in the sense that they are highly interrelated and interconnected, particularly the process related factors. Above, we have shown how the organization and technology CSFs support the start of a DM initiative and ensure the management of the DM process. Cross case analysis suggests that among the CSFs the best predictor of overall success in an embryonic DM initiative is *stakeholder participation* (see Table 6). The involvement of key constituents is particularly important in a DM process because the results must be co-produced. Better still, the domain experts and end users should own the problem. They are therefore the ones who produce the results, with the necessary help of a DM expert. Schein (1999) insists that only this

approach favors the creation of an atmosphere of mutual acceptance and trust necessary to build a successful problem-solving relationship. In our case study *stakeholder participation* is the second most cited issue, the first being *interdisciplinary learning.* The two issues are closely interconnected. *Interdisciplinary learning* enables *stakeholder participation* while the latter reinforces the former because the more committed the stakeholders are, the more knowledge they volunteer to the initiative, further promoting *interdisciplinary learning*. This reinforcing relationship is reflected in our data; 93 statements mention both issues simultaneously.

Moreover, in accordance with previous theory, this research shows similar interrelationships between all three process CSFs; 63 statements refer simultaneously to both *interdisciplinary learning* and a *focus on problem on solving action.* Forty-two were coded as simultaneously mentioning *stakeholder participation* and *focus on problem solving action* while 29 statements concur in all three process CSFs. This interdependence is consistent with IS research. In interdisciplinary initiatives it is necessary to develop shared cognition as the foundation to a common definition of the problem. This enables effective new knowledge construction, e.g. (Beers et al., 2005; Markus, 2001). In turn, the business relevance of the proposed solutions influences the ability to recognize the value of a technology, thus increasing the stakeholders' absorptive capacity, i.e. the ability for interdisciplinary learning (Cohen & Levinthal, 1990; Zahra & George, 2002). IS success literature (Delone & McLean, 1992, 2003) also implies these relationships with the reinforcing association between "net benefits" and "use" or "intention to use" success dimensions. The three process CSFs should therefore be carefully managed through *process facilitation*.

This research (section 4.3.3) also revealed that the intensity of *process facilitation* and its form is contingent on organizational culture. Organizational propensity to fact-based decision making and interdisciplinary and innovative initiatives (Davenport et al., 2010; Detert, Schroeder, & Mauriel, 2000; E. A. King, 2005) should be assessed early in embryonic DM initiatives. Our case study showed that in the organizations that display such a culture, the process was facilitated by the culture itself. In instances C, D, F, and H *process facilitation* was shared by all stakeholders. This was denoted by "**" in Table 10. Conversely, cases E and B (marked with "*") evolved in conservative environments and hence required explicit *process facilitation*, which was painstakingly carried out by one stakeholder. Moreover, we found that organizational culture assessment may be localized. For the success of an embryonic DM initiative, it is sufficient if it is limited to the stakeholders directly involved in the process. These findings agree with reports that studied organizational issues from the contingency perspective and identified internal environment as a contingent factor, e.g. (Khazanchi, 2005; Scott, 2003).

4.4 CSF Summary

Practice rules and guidelines that promote embryonic DM success have been expressed as seven CSFs. The CSFs are: a *business champion, process facilitation, stakeholder participation, interdisciplinary learning, a focus on problem solving action, data*

availability, and *data quality.* The CSFs were defined as follows:

A *business champion* is someone within the organization who understands the potential of DM, has a DM-related business problem, and is actively or through delegation involved in the DM process and in the internal promotion of DM;

Process facilitation is a role that helps manage the DM process and supports the internal promotion of DM. It does so especially by managing DM's multi-disciplinary collaboration aspect. The role may be taken up by any stakeholder and is particularly critical in conservative organizational cultures;

Stakeholder participation is the degree to which unity of effort is achieved among contributors to the DM initiative;

Interdisciplinary learning provides guidelines for the coordination of interdisciplinary collaboration between DM experts and business people;

Focus on problem solving action defines the ability to adapt the DM process in light of emerging findings while keeping in mind the initiative's goal, the action that solves the identified business problem;

Data availability refers to the accessibility of data and its readiness for DM analysis; and

Data quality expresses the degree to which data is clean, complete, and sufficient. In addition to the CSFs, the process model should also integrate the two simultaneously running iterative processes, i.e. the DM and the organizational DM integration processes.

In the interviews and case studies described above, we also asked the informants whether they followed a methodology in their embryonic initiatives. One informant suggested that they followed C-K design theory (Hatchuel & Weil, 2009). A careful review of the methodology showed that it does not meet the design criteria defined by the CSFs of embryonic DM. Some DM experts either mentioned CRISP-DM or a similar method of their own. Most, however, invented the process as they went along. Importantly, none of the business stakeholders could name or describe a methodology that was followed. This confirmed the necessity for an empirically derived methodology useful for both DM experts and business people.

Having defined the embryonic DM practice rules we may proceed to the design of the last component of the methodological framework, i.e. the process model.

5. Development of a Process Model: InterActive8

The completion of a methodological framework of embryonic-stage DM implementation requires the design of a process model. Peffers et al. (2007) suggests that such a procedure is necessary to provide a generally valid process for carrying out a methodology. The above-defined guidelines for successful embryonic DM implementation suggest the following design requirements for the process model:

(1) It should incorporate both DM and organizational perspectives;

(2) It should integrate two concurrent, iterative and explorative processes, i.e. the organizational process and the DM process;

(3) The model should increase the likelihood of success of embryonic DM initiatives by facilitating the management of the process related CSFs: stakeholder participation, interdisciplinary learning, and focus on problem solving action.

5.1 Evaluation of Existing Process Models Against the CSFs

Several "real-life" process models for DM (see Figure 10) have been proposed in the past, e.g. (Berry & Linoff, 2004; Cao & Zhang, 2006; P. Chapman et al., 2000; Fayyad et al., 1996). Among these reasonably similar models, the CRoss Industry Standard Process for DM (CRISP-DM) stands out as one that is based on empirical research. It is also the best known and the most frequently used (KDnuggets, 2007b). Moreover, CRISP-DM includes the most comprehensive guide for the DM process. We therefore justify the use of CRISP-DM as the basis for evaluation of all existing process models.

CRISP-DM was developed in the late 1990s (P. Chapman et al., 2000). The motivation for the project stemmed from the fact that DM was new in the organizational context, forcing DM practitioners to invent the process as they went along (Khabaza, 2007). The design of CRISP-DM was based on a series of workshops in Europe and the US with members of a special interest group (over 200 DM practitioners) and on the rich DM experience of the consortium members. The methodology was tested in a series of DM projects implemented by the industrial partners of the consortium (P. Chapman et al., 2000) and later widely adopted by the DM practitioner community (KDnuggets, 2007b). CRISP-DM emphasizes that DM is not a linear or waterfall style process but rather an iterative one. The top-level diagram shows six phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment (see Figure 10, b). The user may also drill down into the detail of each phase's tasks and outputs.

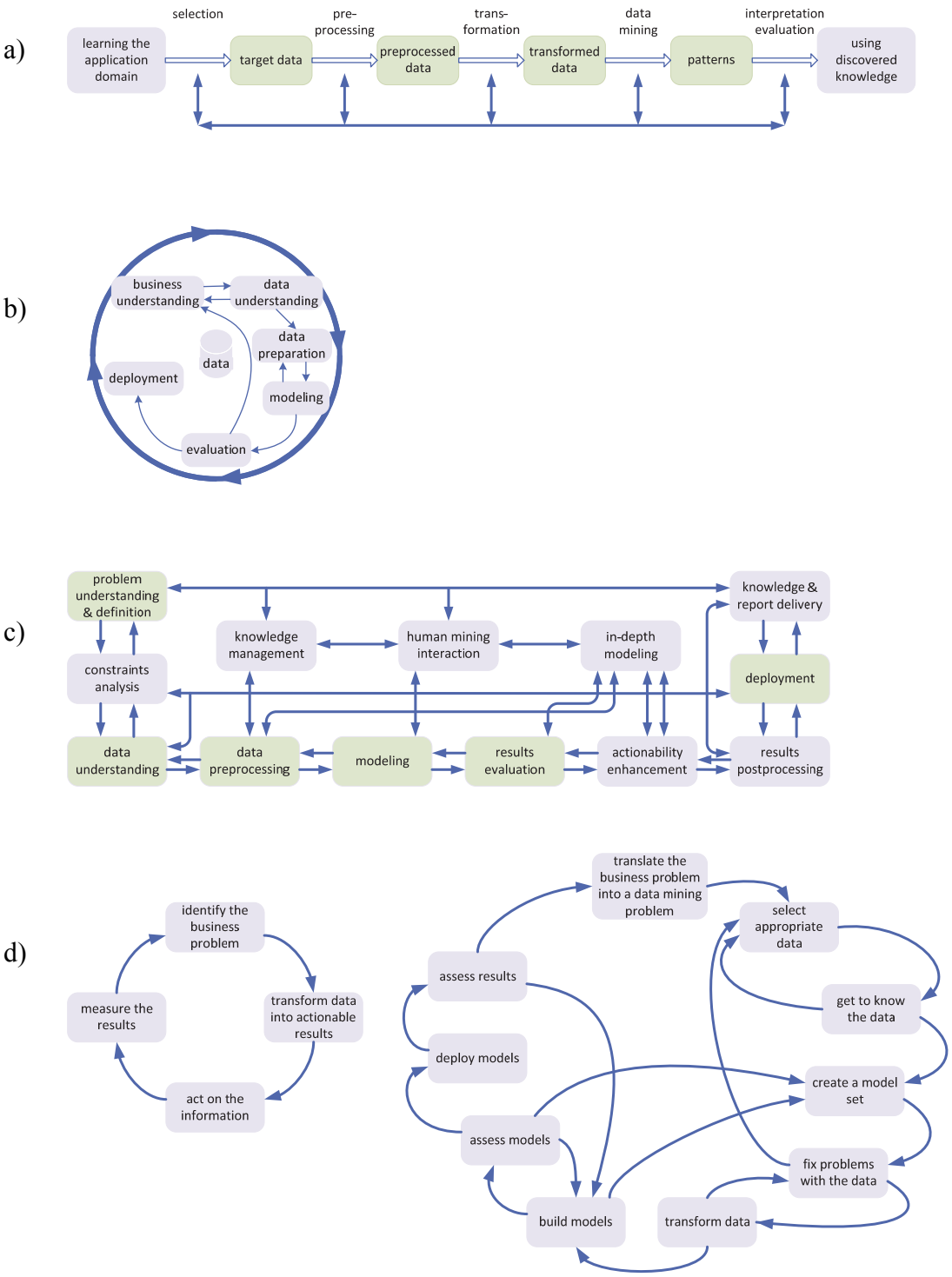

Figure 10: *a) KDD Process (Fayyad et al., 1996). b) CRISP-DM (P. Chapman et al., 2000). c) Domain-Driven In-Depth Pattern Discovery (Cao & Zhang, 2006). d) the Virtuous Cycle of DM (left) and the corresponding methodology for the DM process (right) (Berry & Linoff, 2004).*

In 2006 the CRISP-DM consortium attempted to motivate a CRISP-DM 2 project (Khabaza, 2007) due to gaps that had surfaced in the methodology. Essentially, CRISP-DM has been found to address the process from the point of view of DM experts but not that of business people (Khabaza, 2007). Reportedly, the DM experts themselves had observed that there is another important perspective to the whole process: how an

organization integrates DM to become a DM organization. They suggest that an integrating methodology is needed to help an organization in the process of introducing DM into the organizational business processes. Specifically, the CRISP-DM special interest group members propose that a new methodology should induce frequent interactions between DM and organizational processes and structures, i.e. lead to *interdisciplinary learning.* CRISP-DM does, however, imply the importance of *interdisciplinary learning* in that it suggests a compilation of a glossary of business and DM terminology relevant to the project (P. Chapman et al., 2000).

Another important area of improvement identified in (Khabaza, 2007) is *focus on problem solving action.* The document recognizes that CRISP-DM 2 should assist the stakeholders in the effort to align the DM effort and business needs by inciting early planning for the deployment of DM results. This should also facilitate the mapping from business goals to DM goals. These issues are recognized as particularly difficult for individuals and organizations new to DM, as in the case of embryonic DM. The CRISP-DM 2 initiative was extremely well received by the DM community (Khabaza, 2007). This highlights the lack of such a methodology as an important and relevant problem for DM research and practice. Unfortunately, the CRISP-DM 2 project has not been carried out.

In a careful scrutiny of CRISP-DM and other cited process models, we discovered further strengths and weaknesses beyond the suggestions identified in (Khabaza, 2007). The data related CSFs (*data quality* and *data availability*) are covered in detail, albeit from a technical (DM expert's) perspective. People issues, on the other hand, are taken for granted. The importance of identifying the stakeholders, particularly the internal sponsor (*business champion*), is recognized as an activity in the first, Business understanding, phase only in CRISP-DM. No further advice is given for building the relationship between the stakeholders despite the fact that most of our informants in the CSF study insisted on its criticality. Moreover, the CSFs *process facilitation,* and *stakeholder participation* are neither explicitly mentioned nor implied although they were found to be fundamental success factors to embryonic DM. Importantly, no guidance is provided to the *business champion* on how to manage the integration of DM into the organization. This is one of the manifestations of the greatest weakness of all four process models. Their target audience is DM experts, not business users. As they do not incorporate the business view on the process, they cannot be regarded as holistic process models of embryonic DM integration in an organization. Nevertheless, CRISP-DM remains valid as a very good guide to DM experts in their effort to obtain good DM results in an organization.

5.2 Relevant Process Models from Organizational Theory

To propose appropriate elements to an integrating process model for embryonic DM, we looked at extant theory and current thought. We sought to build upon what researchers said in key prior literature about how processes similar to embryonic-stage DM should be carried out. Our aim here was to develop a process model that would serve as a

commonly accepted process for embryonic DM implementation based on the requirements determined in section 4, i.e. the CSFs. Instead of focusing on the differences among various models, we sought to use a consensus-building approach to produce the design. Consensus building was important to ensure that the embryonic DM process model was based on well-accepted elements.

In addition to process models specific to DM, a number of other methods were reviewed. Based on their overlap with the design criteria for embryonic DM, we chose three to contribute ideas for the final process elements. We draw from practice-based models of organizational problem solving: the PDSA cycle (Deming, 2000a), the process consultation model for group problem solving process (Schein, 1999), and the Action Design Research model (Sein, Henfridsson, Purao, Rossi, & Lindgren, 2011). The rationale for selecting these models is explicated below and summarized in Table 11.

PDSA Cycle (Figure 11)**.** Deming's PDSA Cycle (Deming, 2000a, 2000b) was chosen because of its profound impact on the way organizations do business (Senge, 2006) and because it is rooted in statistics, a domain analogous to DM. At the core of the PDSA cycle is the integration of statistics into organizational processes (Deming, 2000b). This is important for the integration of DM because DM and statistics have common aims in that both are concerned with discovering structure in data (Breiman, 2001; Hand, 1999) to help organizations learn and improve (Deming, 2000b; Khabaza, 2011). The overall aim of both statistics and DM is therefore to either describe or to predict phenomena based on previous data. However, DM arguably requires further specialization as compared to statistics (Breiman, 2001; Fayyad et al., 1996). This is likely to increase the cognitive distance between DM experts and business people. Subsequently, the management of DM is more complex in the sense that it requires careful attention to interdisciplinary interactions between the stakeholders.

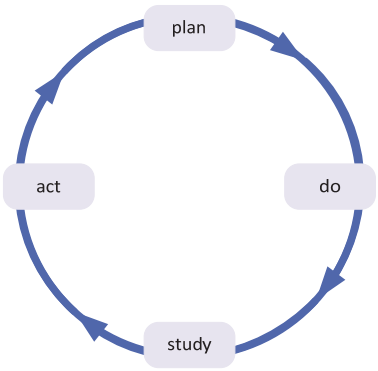

Figure 11: *The PDSA cycle (Deming, 2000a).*

The PDSA cycle is an adaptation of the scientific method (Cleghorn & Headrick, 1996), which is an explicit representation of how people solve problems and learn in everyday life (Argyris, 1982). It is also known as the PDCA, Shewhart, or Deming cycle (Deming, 2000b; Shewhart, 1939). It is a methodology used in most sectors and types of organizations for continuous improvement and total quality management initiatives, e.g.

(Bessant & Caffyn, 1997; R. L. Chapman & Corso, 2005; Cleghorn & Headrick, 1996; Frakes & Fox, 1996; Owlia & Aspinwall, 1997).

PDSA Cycle complies with three CSFs of embryonic DM. Deming provides a detailed description of the personal characteristics of a leader, i.e. *business champion,* and also suggests the required competences. The management philosophy he proposes is focused on obtaining employee commitment to the improvement process and complies very well with *stakeholder participation*. Moreover, throughout the cycle its aim is organizational improvement, and hence the steps are concerned either with a problem or a solution, namely the *problem solving action.*

Table 11: *Overview of the main arguments for the selection of process models similar to embryonic DM integration.*

Embryonic DM CSFs[1]	***PDSA Cycle*** (Deming, 2000a)	***MSGPS*** (Schein, 1999)	***ADR model*** (Sein et al., 2011)
business champion	*A leader* should, among other things, have some sense of theory (understand technology) and be able to explain his plan of action and predict the results to those in power.	*Primary client* The individual who ultimately owns the problem, and typically pays the consulting bill.	
process facilitation		It is the consultant's principal role in all phases of the process model.	
stakeholder participation	Suggests detailed guidelines to achieve full benefits from stakeholder participation. Recommends participation of a statistician at all steps of the process "for economy, speed and protection from faulty conclusions" (Deming, 2000b).	Suggests detailed guidelines to achieve full benefits from stakeholder participation. - classification of roles - explains intrapsychic processes and cultural rules of interaction - argues for early involvement of stakeholders to increase commitment and avoid miscommunication	Implies the importance of the contribution of business stakeholders (particularly end-users) and researchers in the design and development of an artifact within the BIE (Building, Intervention, Evaluation) stage.
inter-disciplinary learning		Suggests the Dialogue[2] (managed conversation) to become aware of some of the tacit assumptions. Dialogue assumes that every person comes with different assumptions such that mutual understanding is in most cases an illusion. It facilitates creation of shared cognition and a "common" thinking process.	Principle 3: Reciprocal Shaping - inseparable influences mutually exerted by the two domains. Principle 4: Mutually Influential Roles - mutual learning among researchers (knowledge of theory and technology) and business stakeholders (practical hypothesis and knowledge of organizational practices).
focus on problem solving action	The focus of the PDSA cycle is improvement, i.e. solution of a problem. The first step (Plan) focuses on deciding what action is required to solve the problem. The third (Study) and fourth (Act) steps focus on the evaluation of the proposed improvement and its wider implementation.	At each stage of the problem solving process the discussion may reveal new features that lead to reformulation of the problem. Such recycling prevents working on the wrong problem and thus wasting resources.	Principle 5: Authentic and Concurrent Evaluation – evaluation is not a separate stage that follows artifact building. Decisions about designing and shaping are interwoven with ongoing evaluation.

[1] CSFs data availability and data quality are specific to the DM domain and are therefore not considered by any of the three process models.

[2] This methodology should not be confused with the general understanding of the world "dialogue" as a mere two-way conversation between two or more people.

The Model of the Stages of Group Problem Solving (MSGPS). We chose the process consultation model for group problem solving (Schein, 1999) because it complies with all CSFs but *data availability* and *data quality,* which are specific for DM. The primary aim of MSGPS is to establish a trusting relationship based on a business problem. Hence it builds commitment (*stakeholder participation)* while *focusing on problem solving action.* Successful embryonic DM has been found to require a trusting relationship between the DM champion, a DM expert, and other stakeholders (see section 4.2.1). MSGPS provides a detailed categorization of stakeholder roles, with particular emphasis on the *business champion*. It provides abundant advice for building trust and commitment from the point of view of any stakeholder.

Moreover, DM experts are often expected to advise business stakeholders on the DM problem solving process. The ultimate goal of process consultation is therefore to establish a helping relationship (Schein, 1999). It is based on the assumption that one can only help a human system to help itself, because a consultant can never know enough about the particular situation and culture of an organization to make specific recommendations. This assumption is always true in embryonic DM. A DM expert is always a newcomer to the organization, regardless of the particular contractual relationship between the DM expert and the organization. On the other hand, business stakeholders tend to be new to DM. In this manner MSGPS is centered around *process facilitation*, which has been found particularly critical in traditional organizational cultures.

MSGPS also considers the complexity of *interdisciplinary learning*, which characterizes embryonic DM processes. The model proposes a form of managed conversation called the Dialogue. It assumes that every person comes with different assumptions and that mutual understanding a priori, in most cases, is an illusion. Dialogue helps people learn by making them more conscious of their own tacit assumptions. In addition, it helps them recognize that others may be operating from different assumptions. Hence MSGPS facilitates creation of shared cognition and a "common" thinking process.

The model is singular in that it highlights that the problem solving process is composed of two different cycles of activity. One occurs before the decision to act has been taken and one after (see Figure 12). The first cycle involves: problem formulation, producing proposals for action, and forecasting consequences of solutions or testing proposed solutions and evaluating them conceptually before committing to action. This cycle ends when the group has made a formal decision on what to do. The second cycle follows and is formed by three stages: action planning, action steps, and evaluation of the outcomes of the action. It often leads back to the first cycle with problem redefinition. An important skill of a process consultant (*process facilitator*) is to help the stakeholders appreciate the importance of the first cycle. This insistence on problem redefinition in order to solve the right problem resembles the CSF *focus on problem solving action* and presents process consultation as an Emerging Knowledge Process.

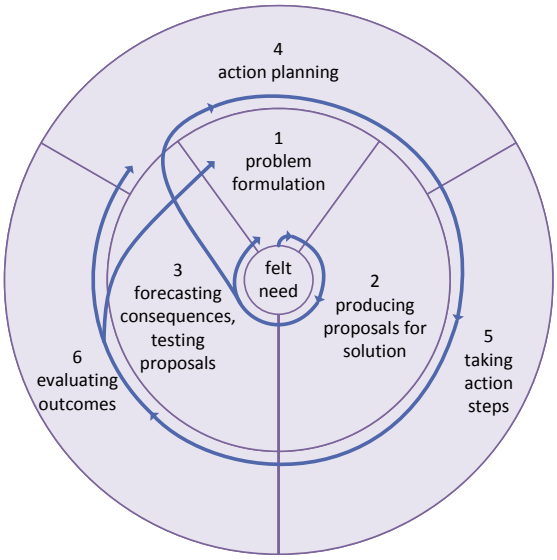

Figure 12: *The Model of the Stages of Group Problem Solving (Schein, 1999).*

Action Design Research (ADR) Model. ADR is similar to DM in that it is concerned with the development of an artifact while paying due attention to organizational context in its shaping. In DM the artifact is a DM result, i.e. a model and new findings (P. Chapman et al., 2000). However, as we showed earlier (sections 4.2.1, 4.2.2), this model is necessarily shaped by the organizational context during its development and use through the collaboration of DM experts and business stakeholders. Although the aim of embryonic DM is not to produce theory, we draw from ADR methodology proposed by Sein et al. (2011) precisely because it captures these characteristics. They achieve it by integrating two complementary research approaches: action research and design science. The latter is more focused on the technical view of the IT artifact that is being designed (Sein et al., 2011). Action research, on the other hand, is focused on practical relevance. It studies organizational phenomena while actively trying to change them. As such, it is strongly oriented toward collaboration involving both researchers and business stakeholders (Baskerville & Myers, 2004).

Particularly interesting for the design of the process model for embryonic DM is action design research's second phase, *Building, Intervention and Evaluation* (BIE - see Figure 13). BIE is concerned with the design and development of a solution to a specific organizational problem (other phases focus on the development of theory). Like DM, BIE is carried out as an iterative process. The problem and the solution are continuously developed and repeatedly tested through organizational intervention as they are subjected to participating stakeholders' assumptions, expectations, and knowledge, which require *stakeholder participation*. The principles that shape BIE (reciprocal shaping, mutually influential roles, authentic and concurrent evaluation) are akin to those defined by the DM process related CSFs *interdisciplinary learning* and *focus on problem solving action*.

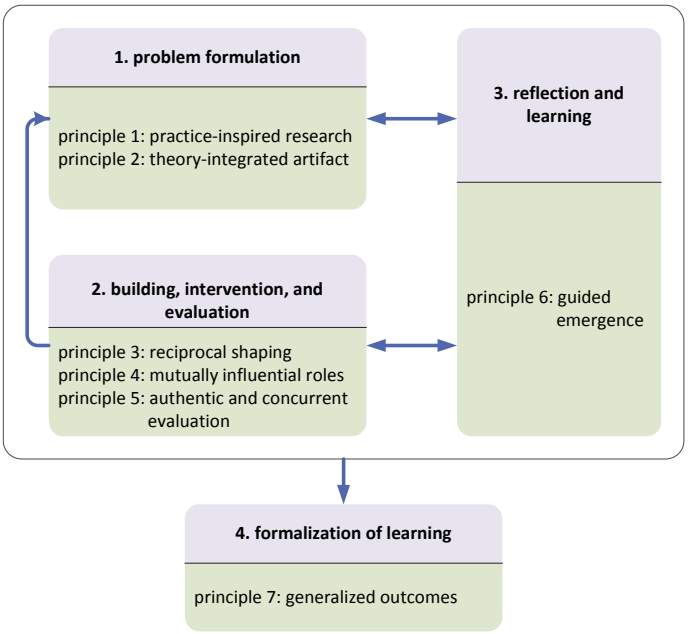

Figure 13: *The Action Design Research Method: Stages and Principles (Sein et al., 2011).*

5.3 InterActive8: the Process Model

Table 12 shows the process elements as stated or implied within the selected process models. These elements are substantially similar. Table 12 also presents our synthesis: the components of InterActive8 process. InterActive8 is our proposal for the embryonic DM process model. It was designed based on the CSFs of embryonic-stage DM and the process models presented in sections 5.1 and 5.2. The result of our synthesis is a process model consisting of five activities justified and described below and graphically in Figure 14. Below, we first present some general remarks about the process model and follow with the discussion of each of its stages.

The name, InterActive8, is a play of words – *activate interdisciplinary interactions* – which were found critical to the success of embryonic DM. It also refers to the shape of the process model. Instead of a cycle, as in the case of PDSA, MSGPS, ADR and CRISP-DM, it takes the shape of a Figure 8[3]. The upper half is predominantly the organization's domain and the lower half is the domain of DM. Stakeholders move through the stages together to jointly solve business problems.

Common problem (re)formulation is a singular activity within the methodology in response to the *focus on problem solving action* CSF. It assures business relevance, i.e. *actionability*, of a DM initiative and hence the continued support of key decision-makers. When the initiative goes through the bottom half of InterActive8 (solving the DM problem), it generates a new challenge for the organization. Business people need to comprehend the findings of DM analysis, reflect on their business through the perspective of the findings, propose improvements, plan them, and carry them out. Conversely, going

[3] The shape was suggested by Jure Žabkar.

through the upper half of InterActive8 generates new data, i.e. a new DM problem.

Table 12: *Design of InterActive8: The synthesis elements and an overview of the elements from the three selected practice-based process models of organizational problem solving.*

Common process elements	***PDSA cycle*** (Deming, 2000a)	***MSGPS*** (Schein, 1999)	***ADR model*** (Sein et al., 2011)	***CRISP-DM*** (P. Chapman et al., 2000)
Common Problem (Re) Formulation	**Plan** a change or a test, aimed at improvement: -generate suggestions -choose the one to test -predict possible results -choose the most promising option	**Deciding what to do**, the conceptual cycle: -problem formulation -propose solutions -forecast consequences, test proposals -choose validation method(s)	**Problem formulation**: -structure the problem -identify possible solutions -guide design	
Plan Action		**Action planning** -can be treated as a new problem; problem formulation, idea production and testing -short-circuiting leads to blame being attributed to deficient proposal -assign clear responsibilities -attention: proposal communication		**Plan deployment** -develop alternative plans for deployment -how to propagate results to its users -how to deploy the result within organization's IS -identify possible pitfalls of deployment -plan monitoring and maintenance
Take Action	**Do** – Carry out the change or test (preferably on small scale)	**Taking action steps**	**Intervene** in the organization	
Repeat[4]: Common Problem (Re) Formulation				**Business Understanding:** -determine business objectives -assess situation
Plan DM Analysis		Prior to Evaluation the team should agree on: (1) the criteria of evaluation, (2) the timetable, and (3) who is responsible for reporting evaluation results		**Business understanding:** -determine DM goals -produce project plan **Data understanding**
Analyze Action	**Study** the results. What did we learn? What went wrong?	**Evaluating outcomes** -be psychologically prepared to go back to problem reformulation (not merely suggesting new solution alternatives)	**Build** the artifact	**Data preparation:** select, clean, construct, and integrate data **Modeling:** select modeling technique, generate test design, build and assess model
Repeat: Common Problem (Re) Formulation	**Act –** Adopt the change, or abandon it, or run through the cycle again	**Problem (re)formulation**, possible during/after any of the above four stages	**Evaluate** the artifact **Assess** need for additional cycles, repeat	**Evaluation** in light of business objectives -determine next steps -may trigger new, more focused business questions

[4] This element differs from the element *problem (re)formulation* in the first line. The corresponding phase in CRISP-DM is necessarily succeeded by DM activity (Plan analysis), while the stages of the three organizational problem solving models necessarily lead to stages that refer to action planing or taking. InterActive8 allows either of the two options to follow the problem (re)formulation stage.

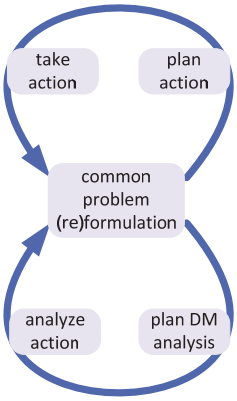

Figure 14: *InterActive8: completing the process model for embryonic DM.*

InterActive8 is intuitive for business people because it follows the natural problem solving process and for DM experts because it integrates the principal steps of the DM process. Table 12 suggests the mapping of individual phases of CRISP-DM to InterActive8:

- *business understanding* takes place within common problem (re)formulation and plan DM analysis,
- *data understanding* is an integral part of DM analysis planning,
- *data preparation* and *modeling* belong to action analysis,
- business *evaluation* suggests a return to problem formulation, and
- *deployment* is part of action planning.

Similarly, our proposed process model consists of the same activities (problem formulation, action planning, execution, and evaluation) as the human cognitive process (Argyris, 1982). Moreover, the names of the stages were primarily adopted from the PDSA cycle, which is widely used in business for continuous improvement initiatives (R. L. Chapman & Corso, 2005). We may therefore expect that InterActive8 will be readily adopted by both business people and DM experts.

A major objective of InterActive8 is to stimulate frequent interactions between business people and DM experts. In our study of the CSFs of embryonic DM, recurrent interactions were found to generate *stakeholder participation* by enabling *interdisciplinary learning*. Therefore, InterActive8 recommends active participation of all stakeholders in the problem formulation stage. In addition, by design, it ensures that this stage is the most frequently visited in the process from a business problem to a DM based solution (see Figure 15). Moreover, InterActive8 advocates the involvement of all stakeholders in both loops of the 8, i.e. the analytical and the organizational. The relative intensity of involvement, however, differs (see Figure 16). It should be equal in the problem (re)formulation stage; slightly less so in the two planning stages; and dominated by the DM experts in the action analysis stage and by the business people in the action taking stage.

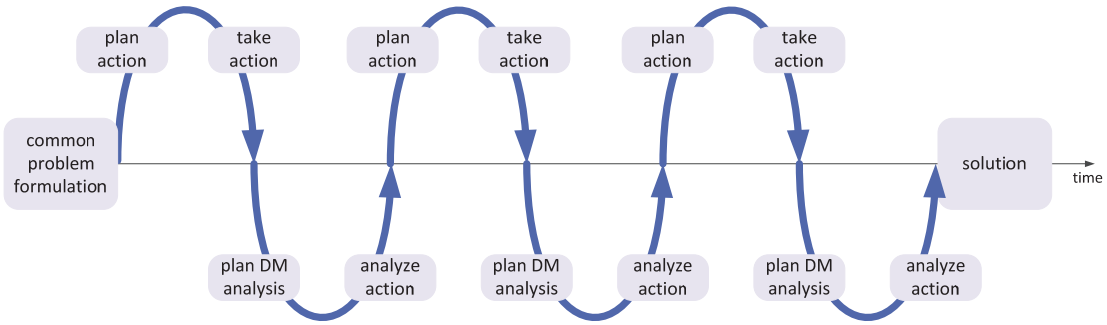

Figure 15: *Time-decomposed InterActive8. Common problem (re)formulation is the most frequent activity, which takes place each time the arrow crosses the time-line.*

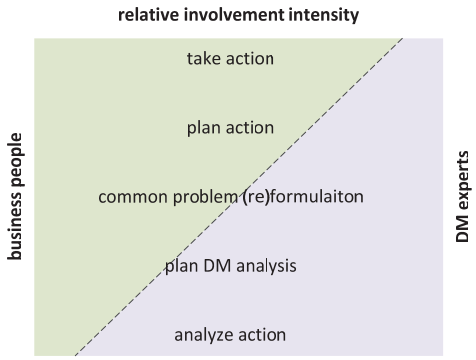

Figure 16: *The relative intensity of involvement of DM experts and business people at different stages of InterActive8.*

InterActive8 facilitates learning through action and reflection. When all stakeholders participate throughout the process, as suggested by Figure 16, the result is another important property of InterActive8. Adult learning and hence organizational learning is a result of acting and reflecting (Argyris, 1982; Senge, 2006). The design of InterActive8 provides the opportunity for both. The DM experts may reflect on their analysis and models by observing how they are used in practice (the upper half of the 8). Conversely, business people are induced to reflect on their business actions while participating in the analytical activities (the bottom half).

5.3.1 The Stages of InterActive8

Next, we provide a detailed description of each of the five activities of InterActive8. Table 13 at the end of this section highlights the activities comprised within each stage, its outputs, and the CSFs addressed by the activity.

The four process models, which were used as the basis to build InterActive8, include or imply an initial trigger to the process. The motivation that activates the process might be an idea for improvement (Deming, 2000a), felt need (Schein, 1999), perceived or anticipated problem (Sein et al., 2011), or a DM-related business problem (P. Chapman et al., 2000). Deming's "idea for improvement" could be associated with the business champion's "recognition of DM's potential". It is implied that the trigger comes from

within the organization. This is in accordance with our definition of *business champion,* who "has a DM-related business problem". In CRISP-DM the first explicit stage is Business Understanding in which a DM expert is to "understand, from a *business perspective,* what the client [organization] really wants to accomplish" (P. Chapman et al., 2000). This implies the assumption that business people are, on their own, able to adequately define a DM-related business problem. However, in embryonic DM this assumption rarely holds, as suggested by *interdisciplinary learning* and *focus on problem solving action* CSFs. Therefore, following the remaining three models, we first propose an activity to determine what the problem is and what can/should be done about it.

Common Problem (Re)Formulation is central to InterActive8, as suggested in Table 12 and Figure 14. Following the CSF *stakeholder participation*, a DM initiative should return to Problem (Re)Formulation often to stimulate team-member interactions and to decide what to do next. This is in accordance with Schein (1999), who suggests that problem (re)formulation may be necessary during or after any stage. Moreover, we draw on his argument that action planning should be treated as a new problem that requires new problem formulation; hence, planning for DM analysis also requires its own problem formulation stage. Problem (re)formulation also incorporates part of CRISP-DM's Business Understanding phase, namely determining business objectives and a thorough assessment of the situation.

Returning to the problem (re)formulation phase is consistent with the CSF *focus on problem solving action.* It guarantees alignment between the business problem and DM initiative. Moreover, problem (re)formulation is necessary because business people, who are the principal implementers of business action, need to transfer the lessons learned to DM experts. Such knowledge transfer is also required after completing the DM analysis loop of InterActive8, although in this case it is DM experts explaining lessons learned from DM analysis to business people. In this way problem (re)formulation meetings facilitate the emergence of shared cognition, which is fundamental to effective interdisciplinary problem solving (Beers et al., 2005). Hence, this step is also in accordance with the CSF *interdisciplinary learning.* The speed of new knowledge transfer increases with the tenure of the interdisciplinary team (Holland et al., 2000) involved in a DM initiative. Similarly, knowledge transfer is facilitated by the intensity of involvement of DM experts in the business loop and of business people in the analytical loop of InterActive8.

> ***Common Problem (Re)Formulation.*** Define the problem such that it incorporates both business and DM perspectives. A common problem is different from a business problem or DM problem. It is the answer to the question: How can DM help solve the business problem? This activity is conceptual and, in addition to common problem formulation, involves identification of possible solutions and forecasting the consequences to test the proposals. Moreover, in this stage a validation method should be specified, i.e. business success criteria should be agreed on. The Dialogue (Schein, 1999) is the recommended form of structured conversation involved in this activity and is carried out in meetings of

preferably all the different stakeholders or their representatives.

The final outcome of this stage is the decision of what to do next. Unless the decision is to stop, the primary problem ownership is transferred from DM experts to business people or vice versa. The team may decide to address a business problem by moving into the business action loop of InterActive8. Alternatively, a decision to analyze an action might lead the initiative into the analytical loop. It is also possible to decide to tackle both business and analytical activities concurrently. This activity is central to the embryonic DM process. Stakeholders should expect to come back to it often and reformulate the problem in light of new findings and improved shared cognition. In this manner this stage also involves joint evaluation of the stages previously carried out, whether analytical or organizational.

InterActive8 may also be applied to provide structure and focus to problem formulation. DM experts and business people can only foresee obstacles and opportunities in the part of the process that each side understands. This is of little use to the common initiative unless they are able to make these issues explicit and understood by the remaining team members (Beers et al., 2005). This problem may be overcome by explicitly simulating the InterActive8 process as a team. In conversing about the journey through the upper, business loop of the 8, DM experts may gain an appreciation of the business. Similarly, discussing the analytical part of the 8 may help business people develop an understanding of DM. As a result, problem formulating conversations will gain in structure and focus.

Process facilitation may play an important role in the Problem (Re)Formulation stage. A process facilitator should look for premature shortcuts in reasoning and problem misdiagnosis. For this reason it helps when they are not under the time pressure that is so prevalent among managers (Schein, 1999). They may also help steer the process away from interpersonal and interdepartmental conflicts and frustrations by helping the team focus on the task rather than relationships (Holland et al., 2000).

A decision to implement a proposal within an organization leads to the deployment of a DM solution (P. Chapman et al., 2000). Although in CRISP-DM this phase is denominated *Deployment*, the model in reality provides no guidance beyond the deployment planning phase. This is because DM experts are unlikely to get involved in deployment execution (P. Chapman et al., 2000). However, the non-participation of DM experts may lead to failure. MSGPS suggests that communication of the proposal and the deployment plan is a frequent source of breakdowns (Schein, 1999). When implementation is handed from the proposal generating group to a different group of people, they may neither understand it clearly nor be sufficiently committed to the proposal or solution. This is reminiscent of the difficulties identified in our case studies and discussed in section 4.3.4.1. Unless aided by DM experts in implementing DM

results, business people may go astray in DM results implementation or test by unconsciously making erroneous DM-related assumptions.

To avoid this pitfall, a high degree of overlap between the solution generating and implementing groups is desirable (Schein, 1999). Ideally, they would be the same problem-solving group. When this is not possible, a good way to enable the implementation group to get completely on board is to bring its representatives into the problem-solving process at the earliest possible stage. Alternatively, the two groups may, at least, review thoroughly and completely all the previous steps that led to the solution proposal.

> ***Plan Action.*** This stage may be treated as a new problem requiring its own problem formulation, solution ideation, and proposal testing. Short-circuiting or avoiding these phases may lead to inadequate proposal implementation. Subsequently, the users may erroneously conclude that the proposal was deficient, instead of blaming insufficient action planning. A key role of the *process facilitator* may be to slow the group down and encourage careful planning before leaping into action. *Process facilitation* may also consist of helping the group comprehend how difficult it is to communicate a complex action proposal to an implementer. Communication breakdowns may be avoided if this is understood early enough in the embryonic DM process. DM experts play an important role in helping communicate the proposal to the implementers, i.e. business people. They should also provide advice on data collection issues to insure *data quality* and *availability*, propose how a model might be deployed within an organization's IS, identify possible pitfalls, and propose a plan for DM model monitoring and maintenance.
>
> An action plan should clearly allocate responsibilities to individual team-members for specific actions. This ensures that actions are taken and also provides grounds for responsible implementers to raise questions about the implementation that had not been considered before.

Action planning is followed by the implementation of the plan in all four process models, although in CRISP-DM this activity is only implied given that "in many cases it is the customer, not the data analyst, who carries out the deployment steps" (P. Chapman et al., 2000). However, Deming (2000b), in accordance with both ADR and MSGPS methodologies, recommends that a statistician participates in all phases of the PDSA cycle. This should avoid faulty conclusions, and save time and other resources. We may therefore conclude that the participation of a DM expert in the implementation phase is also desirable, although the involvement may be less intensive than in the common problem formulation and action planning stages.

Beyond the above advice, existing process models provide strikingly little guidance for the action implementation phase. They imply that successful implementation is

conditioned on whether the previous phases are conscientiously carried out and the commitment of implementers and end users is secured. In addition, PDSA suggests that changes or tests are first implemented on a small scale, which is in accordance with rapid prototyping suggested by our informants in section 4.2.2.

> ***Take Action.*** Implement the action plan. For economy and speed DM expert assistance is desirable, although the involvement need not be operational. If possible, changes and improvements should first be tested on a small scale. The result is new data/information concerning the problem and the proposed solution that is to be studied in the next stage, i.e. common problem (re)formulation.

When the team resolves to proceed with DM analysis in the Common Problem (Re)Formulation stage, the result is a new DM problem. CRISP-DM recommends careful DM project planning in order to avoid expanding "a great deal of effort producing the right answers to the wrong questions" (P. Chapman et al., 2000). Planning for the DM analysis step of InterActive8 encompasses CRISP-DM's stages Business Understanding, i.e. *determine DM goals* and *produce project plan*, and Data Understanding, i.e. *collect, describe, explore*, and *verify quality* of data. The practitioners in our study of CSFs attributed high importance to the management of *data availability*. Therefore, throughout the analytical loop, data access issues must be carefully overseen and managed by either the business champion or the process facilitator.

> ***Plan DM Analysis.*** Prepare a plan for DM analysis including initial exploration of the data. The plan is based on the business objectives, business success criteria, common problem definition, and a careful assessment of the situation. Assessing the situation requires an elaboration of the resources available to the initiative and a list of requirements, assumptions, and constraints. Furthermore, risks should be identified and contingency plans suggested. Finally, it is advisable to elaborate a glossary of business and DM terminology relevant to the project. Based on business understanding and situation assessment, DM goals should be defined in technical terms, including DM success criteria. Data Understanding involves an assessment of *data availability* and *data quality.*
>
> Planning for DM analysis requires relatively high involvement from the business side in order to validate that DM experts correctly understand the issues involved. The role of *process facilitation* may be to slow the group down and encourage careful planning before leaping into DM analysis. Moreover, a business champion or process facilitator must manage the *data availability* issue throughout this and the following phase.

Next, the plan is to be executed. CRISP-DM is primarily used to inform this stage since Analyzing Action within the DM context is most clearly a technical activity,

principally carried out by DM experts.

Table 13: *Summary of the stages of InterActive8: activities, outputs, and the CSFs addressed by each of the stages.*

Stage	*Activities*	*Outputs[5] & consequences*	*CSFs addressed*
common problem (re)formulation	initiate embryonic DM process[6] evaluate results from previous stages; decide what to do next define problem in common terms propose possible solutions forecast consequences specify validation method define success criteria	DM & business alignment DM and/or implementation problem decision on what to do next: - stop, or - go to the analytical loop, or - go to the business loop, or - do both concurrently.	business champion focus on problem solving action interdisciplinary learning stakeholder participation process facilitation
plan action	formulate implementation problem ideate possible solutions test proposals produce project plan	allocation of responsibilities emergence of new questions about the implementation proposal decision on what to do next: - return to *common problem (re)formulation* - go to *take action*	focus on problem solving action stakeholder participation interdisciplinary learning data quality data availability process facilitation business champion
take action	implement the action plan	new data/information concerning the problem and the proposed solution	stakeholder participation interdisciplinary learning data quality
plan DM analysis	formulate DM problem determine DM goals and criteria for evaluation assess data availability and quality ideate possible solutions test proposals produce project plan	allocation of responsibilities data assessment emergence of new questions about the proposal for DM analysis decision what to do next: - return to *common problem (re)formulation* - go to *analyze action*	focus on problem solving action stakeholder participation interdisciplinary learning data quality data availability process facilitation business champion
analyze action	implement the DM analysis plan - data preparation - modeling	clean, merged and reformatted dataset with derived attributes DM model(s) model assessment (in DM terms)	stakeholder participation interdisciplinary learning data quality

Analyze Action. Carry out the Data preparation and Modeling stages of CRISP-DM. Decide on the data to be used for analysis based on the DM goals and technical constraints. Then the data should be cleaned to construct additional, new attributes, records, or transformed values of existing attributes. Finally, data should be integrated from multiple tables or records and formatted to fulfill the requirements of the modeling tool. In building a model, a modeling technique is to be selected first. Then, a test design should be generated, and the selected modeling tool is to be run on the prepared dataset to create one or more models. Finally, the DM expert assesses the model by interpreting it according to his/her domain knowledge to prepare the ground for its evaluation together with domain experts in the subsequent, Problem (Re)Formulation stage. The DM expert

[5] Outputs that are implied in the previous (activities) column are not listed

should stay firmly aware of the limitations of his/her domain knowledge. If the model is to be evaluated by a larger, multidisciplinary group, it might make sense to present it first to each individual, as suggested in section 4.2.1 on relationship management.

5.4 Demonstration of InterActive8: A Case Study

IS artifacts should be rigorously evaluated via well-executed methods to demonstrate their utility, quality, and efficacy, i.e. how well they work (Hevner et al., 2004). The evaluation method may be observational, analytical, experimental, testing, or descriptive (Hevner et al., 2004; Hevner, 2007). Design Science research is considered to be relevant to IS and IT practitioners if the resultant artifact addresses the problems they face and maximizes opportunities from the interaction of people, organizations, and IT (Hevner et al., 2004). Likewise, in this research, it was vital to determine the practical relevance of (developing) the InterActive8 process model. We therefore evaluate InterActive8 by analyzing a case study (observational method) to inform the process model design and evaluate the example instance in a real world context. The case study afforded us the possibility to study InterActive8 in depth within the organizational environment (Benbasat et al., 1987; Hevner et al., 2004; Yin, 2003).

In data collection we followed the principles introduced for the multiple case study described earlier (section 4.3). To avoid potential weaknesses of the case study method (Benbasat et al., 1987) and enhance case study reliability (Yin, 2003), we purposefully designed a new case study protocol (including documented procedures of data collection and analysis). Data collection entailed semi-structured interviews with DM practitioners and several other sources, i.e. meeting minutes, presentations, feedback sessions, etc. to facilitate data triangulation. All data was archived in a case study database (Yin, 2003).

In total three interviews were carried out, one with a DM expert, an academic, and an external consultant, and two with a domain expert and a process facilitator. All interviews were recorded and transcribed, totaling approximately 30 single-spaced pages of text. The interviews were completed by the same researcher within one to two hours and followed a two-phased structure as pre-specified in the case study protocol. The first phase took place prior to exposing the informants to InterActive8. The embryonic DM process was openly discussed. This enabled us to gather case data that were not influenced by our proposed model. The information was sorted with respect to the CSF framework dimensions through coding. Having completed all interviews, summary reports of the first phase of interviews (open discussion of the DM integration process prior to exposing the informant to InterActive8) were prepared for later comparison with the second phase, the walk through review of the model. The comparison served as the basis for judging the utility of the process model.

In the second phase the interview proceeded in the form of a cognitive walkthrough

[6] This activity is carried out only in the first iteration.

(Preece, Rogers, & Sharp, 2007). The aim of the walkthroughs was to obtain practical insights on the collaborative and procedural aspects of the InterActive8 process model and its relevance to embryonic DM implementation. A walkthrough is a step-by-step review and discussion with practitioner(s) about the activities that make up a process to reveal errors that are likely to hinder the effectiveness and efficiency of the process (or method) in realizing its intended plan (Kolfschoten & de Vreede, 2007). Walkthroughs generally involve one or more evaluators (or experts) performing a stepwise review of a scenario (or representation of the design of an artifact) so as to note possible problems (Preece et al., 2007). Several variations of walkthroughs are commonly used in software development to find errors in software code and functionality, verify software requirements, validate software against predefined standards, reduce risks of discontinuity, and generally improve software quality (Paul, 2006). Similarly, in collaboration engineering, walkthroughs are used as one of the methods for evaluating and validating the design of a collaboration process (Kolfschoten & de Vreede, 2007). Since embryonic DM is fundamentally an interdisciplinary collaboration process, a walkthrough was judged appropriate for its evaluation.

The interviewees were expected to comment on the relevance of InterActive8. Moreover, they were asked to review the requirements and activities described in the model with the focus of identifying faults and ambiguities, and giving practical insights into eliminating them. Finally, they were encouraged to verify (based on their experience with an embryonic DM initiative) the relevance of the defined requirements and activities in achieving the general aim of the research. This was carried out through a stepwise discussion of the inputs to the walkthrough and yielded both an overall and a detailed assessment of the InterActive8 process model. The research cycle was completed by formal reflection on the findings in meetings with co-researchers.

The company UCS d.o.o. was selected for the case study. With no prior experience in DM, they decided in 2009 to attempt to integrate DM into one of their products. UCS offers footwear manufacturers and retailers sophisticated solutions to provide the best fitting footwear to their customers. One of their products is a system to recommend shoe sizes to online shoppers. Their customer list includes Botisto, Humanic, Decathlon, Sport Lentsch, and W. L. Gore & Associates. One of the greatest risks for on-line shoe retailers is that their customers cannot try shoes on and therefore may not be satisfied with the shoes once delivered. UCS had been working on a recommendation system aimed at reducing this risk and the associated costs.

In 2009 no recommendation systems for buying footwear on-line were being used by retailers. To seize this opportunity, UCS started collecting data by scanning shoes and people's feet. Then they asked them to try the shoes on to record whether in reality they fit or not. When the company began to wonder what to do with this data, somebody suggested they try DM. This need triggered the DM integration process at UCS. At that time nobody had first-hand experience with DM. Based on the hype around DM, the CEO was willing to try this "new" technology, but with minimal commitment. He was not involved operationally, and government subsidies were sought to fund the project.

Subsequently, they contacted a DM expert.

Below we show that the embryonic DM integration process at UCS followed the InterActive8 process model. Each phase is described, including its effects on the following phases. Issues related to Common Problem (Re)Formulation were referred to the most frequently by our informants. This is reflected in the lengthier description of this phase. As suggested above, in reality UCS did not follow the InterActive8 process model because it had not been hitherto available. In the summary of the case that follows, we use the language of InterActive8 to interpret the embryonic DM process actually used by the team at UCS. This enabled us to determine how well the designed process model fits with the embryonic DM process actually carried out at UCS.

Common Problem (Re)Formulation. At the time of the first meeting the issue of formulating a common problem had not been addressed because the participants had no awareness that such a concept existed. UCS was under the impression that they would quickly brief the DM expert on the problem and provide their data. They expected the results within approximately a week. The DM expert's concerns, on the other hand, principally revolved around data quality, data availability, and the algorithms to be used. Each side was only concerned with their respective issues, unaware of those of the other side. In retrospect they understand this as a problem although at the time they were not aware of it. The first meeting ended quickly with the only available option that fit into UCS's understanding of DM at the time – a one-shot modeling, i.e. going through the analytical loop of InterActive8. Interviewee testimonies show that practically no shared cognition was developed as grounds for DM analysis. This behavior is consistent with our findings on interdisciplinary collaboration discussed in section 4.3.4.1.

The absence of common problem understanding and shared cognition nearly undermined the DM initiative at the time of the second meeting. Given UCS's high expectations, the meeting was a disappointment for them. They were hoping for 95 % accuracy. Instead, after the first iteration through the analytical loop of InterActive8, they achieved only around 60 % accuracy of prediction. Besides, the business people and the DM expert could not agree on the accuracy measure. Each side had difficulties accepting the other side's definition of accuracy. In retrospect they realize that they were both right and both wrong, but they were not aware of it. However, given the government grant, the competitive context, and the DM expert's insistence on DM being an iterative endeavor, they decided to continue.

Our informants recalled several examples of the difficulties related to interdisciplinary collaboration. One was the frustration of the DM expert with the need for a recommendation system. In his personal experience, if shoes were size 44, they fit him. He automatically assumed it was that way for everyone. It took time and many questions (seemingly silly to UCS) for him to develop an appreciation of the fact that for a lot of people selecting shoe size is not that simple. Finally, he was able to correct this assumption. Another example was when UCS staff could not understand the DM expert's concerns over data. For a long time they insisted that only numerical data from scans of

feet and shoes could be used as input. Again, it took several problem (re)formulation meetings before they opened up to the possibility of exploring other alternatives with more emphasis on behavioral data. In addition, they slowly accepted the fact that DM requires an explorative, iterative, and interactive process.

The dynamics and the difficulties set in the first two meetings continued in subsequent sessions although with time communication became more fluent. Looking back on the process and having seen InterActive8, they realize that they were implicitly (re)formulating the common problem over and over again as each side learned more about the other domain. In each meeting it seemed that a new difficulty would appear in one form or another. Communication was challenging and misunderstandings common, which made it difficult to achieve the necessary level of mutual trust. At first the DM expert principally communicated with an IT expert. However, it soon became clear that UCS's Technical Director was a better choice. He understood the business implications in addition to the technical issues.

A thorough examination of the UCS case showed that in accordance with our findings from the CSF study (section 4.3.3), the Technical Director assumed the *process facilitation* role. He had a natural inclination towards people issues and quickly started to mitigate the pitfalls of interdisciplinary collaboration. For example, the DM expert, when calling to set up the next meeting, often announced that the results of his analytical modeling were very good. Initially, the Technical Director immediately communicated this enthusiasm to the CEO. However, time and again the DM results that seemed to be good as judged from the technical point of view of the DM expert proved disappointing after scrutiny by the domain experts. Thus, he learned to downplay the DM expert's enthusiasm to avoid disillusionment. He also resolved to delay giving the news to the CEO for the same reason. The protagonism of the Technical Director as a process facilitator was also very helpful because of his ability to manage the CSFs *focus on problem solving action* and *stakeholder participation*.

The merging of two perspectives of the problem into a common one was a painstaking process. It took a great deal of time and energy to overcome the mistrust and conflicts generated due to the lack of awareness of the existence of different mental models. Common problem definition never became an easy exercise. When exposed to the Dialogue (a methodology to manage interdisciplinary conversations), they agreed that its use would likely have helped them in the Common Problem (Re)Formulation meetings. In addition, our informants also agreed that InterActive8 could be helpful in providing structure and focus to the Common Problem (Re)Formulation meetings. They judged that following the steps of InterActive8 together as a team would expose everyone to the assumptions and mental models of the others and learn from each other.

Plan Action. In UCS action planning was for a long time carried out by UCS staff without counting on the assistance of the DM expert. However, with time they learned that not counting on the DM expert's advice at this stage may be costly. Such an example was UCS's effort to collect data about shoes, feet, and the fit or misfit between the two.

The idea was that this data would serve as the basis for developing the recommendation system. UCS used special 3D scanners they had developed to measure shoes and feet. The experiment involved several dozens of people. First they had their feet measured. Subsequently, they were asked to try on different pairs of shoes and answer the question: "Do you like these shoes?"

Having collected the data, UCS gave it to the DM expert. Initial analysis showed strange inconsistencies in the data. This led the DM expert to investigate the data collection process. He immediately noted a deficiency both in the ambiguity of the question and the fact that people's eyes were not covered. As a result, a negative answer to the question "Do you like these shoes?" may be based on bad fit, on the persons dislike of the appearance of the shoes, or both. Thus, the costly experiment needed to be repeated to obtain reliable data. UCS could have avoided this cost if they had consulted a DM expert in the first place. One informant from UCS also attested that they encounter similar difficulties with their clients, who, without counting on UCS's advice, design experiments to test the recommendation system. Having repeated similar mistakes many times, UCS eventually resolved to count on the DM expert's presence in most meetings relevant for the design of the recommendation system, even when they are not directly related to DM.

The interviewees recalled several other similar examples of this behavior; however, they eventually started insisting on the DM expert's presence in action or experiment planning. From that point on several tests were designed and supervised by the DM expert, although implemented by the business people. Alternatively, the DM expert would at least review a proposed action plan and provide his feedback on which data to collect and how. The Technical Director summarized: *"when you start talking about the* [implementation] *details, you suddenly realize that things are not as straightforward* [as initially proposed]*, and that there are many details that need to be considered."* This is in accordance with the recommendation suggested by Schein (1999) and adopted in InterActive8 that implementers are likely to raise questions about the implementation that had not been considered in the initial proposal. Upon considering these details, UCS often needed to go back and (re)formulate the problem and action proposal.

The process facilitator played an important role in the coordination of the action planning process. On the one hand, his knowledge of the organization and persuasion skills were instrumental in finding the right people and in gaining their participation. On the other hand, he helped avoid many communication pitfalls. In informal conversations with the implementers, he gathered their opinions with respect to the proposal. He then communicated the dilemmas expressed by the business people to the DM expert, such that he could prepare convincing arguments ahead of his meetings with business people.

Take Action. In commenting action taking, the participants agreed that the most critical issue is having adequately carried out the previous steps, i.e. adequately defined the common problem and planned the action with the assistance of a DM expert. One domain expert observed: *"Today, as I look back, I am amazed how naïve we were in committing those ridiculous mistakes, or that we didn't even think of some things. Of course now I*

have been into this [DM] *for quite some time. I know that those things were very wrong and we would not repeat those types of mistakes."* This observation corroborates the need for a methodological framework for embryonic DM integration. The informants had no further comments regarding proposal implementation that would be specific to embryonic DM.

Plan DM Analysis. The domain expert informants from UCS feel that this step was initially entirely omitted. As a consequence, the models built by the DM expert lacked relevance to the real business context. For example, early in the DM initiative the DM expert carried out analysis in some way that did not correspond to the reality of the processes analyzed. Often this resulted in models that by technical standards performed very well, e.g. above 90 % prediction accuracy. When the team met to jointly evaluate the model, they realized that in reality the results were much worse. This was disappointing. The Technical Director suggested that in DM analysis planning *"teamwork is key. Before exploring a certain direction, a DM expert should sit down with us and verify whether his idea makes practical sense. Proceeding in this way many failed attempts would have been avoided or redirected towards a more realistic scenario."* Therefore, joint planning of DM analysis, conversing about the domain, and the analytical approach also helps avoid costs while building shared cognition.

However, joint planning of the analysis cannot substitute exploration in modeling. *"Now we know that in DM you cannot know which method would yield good results. Thus, we had to try many methods only to discard them as inadequate for our problem."* In addition to more communication, full iterations through both loops of InterActive8 were therefore also necessary for interdisciplinary learning. They tended to lead to a discovery of a new problem, often in the form of an assumption not known either by the DM expert or UCS. *"Thus we gradually progressed towards a common or the same understanding of the problem."*

The necessity to return to problem (re)formulation was characteristic to all the phases of both the business and the analytical loops of InterActive8. Our informants suggest that *"it is impossible to foresee everything. On both sides* [business and DM] *it is important to realize that you know very little about the other domain and that it is necessary to learn a lot and listen to one another."* Such iterations gradually led to a significantly different approach to solve the recommendation system problem than the one proposed at the outset. Therefore, *"it is very useful to expect potential failure of any given approach rooted in mutual misunderstandings, however, it is better to consider such a failure as a necessary step towards a better result."*

Analyze Action. As in the case of action implementation, this stage did not involve characteristics specific to embryonic DM. It was nearly exclusively carried out by the DM expert, who followed a customized version of CRISP-DM steps of Data preparation and Modeling. In the absence of the complexity of the collaboration with the business people from UCS, the execution of these steps was easier. The real test of the effectiveness of

this stage, however, was the evaluation performed by UCS in the subsequent, Common Problem (Re)Formulation stage. The only recurrent difficulty was the previously mentioned lack of attention to expectation management. The DM expert frequently announced good results before exposing them to the domain expert's evaluation.

The above-described demonstration shows that the embryonic DM initiative at UCS followed the sequence of steps to DM integration as described by InterActive8. Moreover, based on the comments of our informants, we may conclude that had they been exposed to our model, many difficulties of the process could have been avoided. In particular, they highlighted the pitfalls inherent in the problem formulation phase and to a great extent also the pitfalls intrinsic in the two planning steps. Moreover, we may also conclude that InterActive8 is easy to understand and follow.

The case study showed that planning for both action and DM analysis often required the team to return to the Common Problem (Re)Formulation phase. This may be captured in the model diagram by adding two arrows leading from Plan Action and Plan DM Analysis to Common Problem (Re)Formulation as shown in Figure 17. The same effect is achieved if the take action stage is carried out simply by returning to common problem (re)formulation. To conserve the simplicity of the process model, we chose the second option, which leaves InterActive8 as initially proposed in Figure 14.

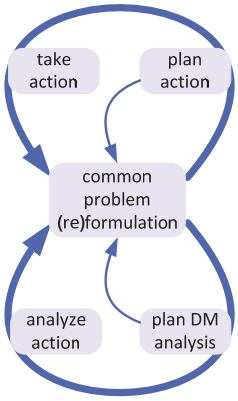

Figure 17: *Case study findings suggest two additional recursive arrows from Plan Action and Plan DM Analysis to Common Problem (Re)Formulation.*

5.5 Evaluation of the Proposed Process Model

The preceding CSF identification and subsequent design and demonstration of InterActive8 indicate that it has met the requirements of being theoretically grounded and useful to practitioners. Moreover, both the consensus design method and its demonstration also show that InterActive8 meets the three requirements of the solution specified at the beginning of section 5: it incorporates both DM and organizational perspectives, it integrates two concurrent, iterative and explorative processes, and it

increases the likelihood of embryonic DM initiative success.

First, InterActive8 incorporates both DM and organizational perspectives. InterActive8 is grounded in both organizational problem solving and DM theory. By applying a consensus method to design the InterActive8 process model, both perspectives are inherent in the outcome of the design process. For example, CRISP-DM (P. Chapman et al., 2000), Deming's PDSA cycle (2000a), Shein's (1999) model of the stages of group problem solving, and the Action Design Science model (Sein et al., 2011) may be roughly mapped to the InterActive8 process. Importantly, InterActive8 is also consistent with the organizational embryonic DM process defined by Davenport and Harris (2007). As follows from the interdisciplinary nature of embryonic DM, InterActive8 is embedded within each of the three steps of the organizational process, i.e. identification of a DM problem, implementation of a localized project, and documentation and propagation of the DM benefits. InterActive8, therefore, does not substitute either of the processes (organizational and analytical) but is consistent and compatible with both of them.

Second, InterActive8 integrates two concurrent iterative and explorative processes, namely the DM process and the organizational process. The organizational process may be visualized in the left diagram of Figure 18. The right of Figure 18 shows the DM process in its most popular representation, i.e. CRISP-DM (P. Chapman et al., 2000). The integration is achieved principally by situating at the center of the process model Common Problem (Re)Formulation, which must be carried out together by business people and DM expert(s). The generation of shared cognition among the stakeholders is effectively the integration of the two iterative processes. To some degree the two planning phases also contribute to the generation of common understanding of the problem. The integration of the two processes therefore also depends on the extent in which all stakeholders take part in action planning and DM analysis planning.

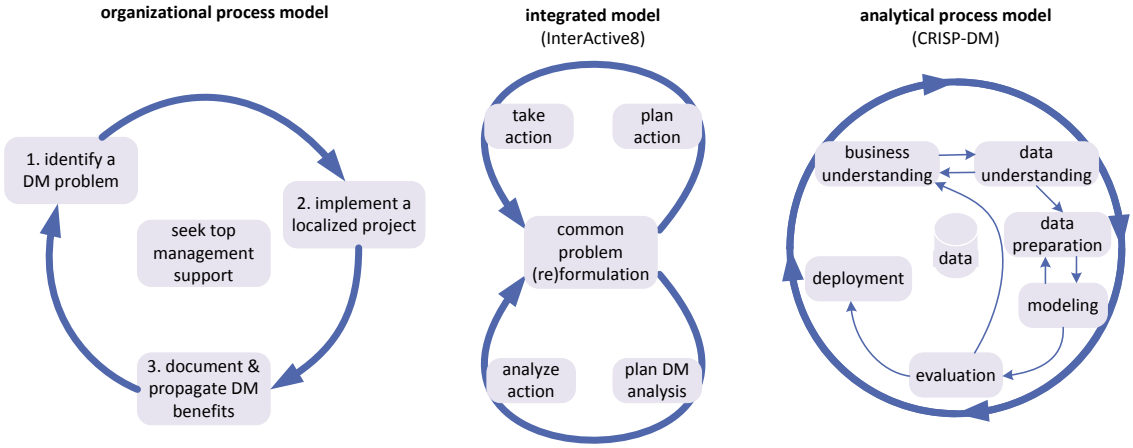

Figure 18: *InterActive8 (center) integrates an organizational embryonic DM process (left) and an analytical DM process (right), two concurrent, iterative and explorative processes.*

Third, InterActive8 is an improvement over existing process models. Designed to facilitate the management of process related CSFs, InterActive8 increases the likelihood

of successful embryonic DM implementation, i.e. achievement of top management support. The model is centered around the problem to be solved and the business action which is to solve the problem. This helps the users manage the CSF *focus on problem solving action*, which is the most direct way of influencing the success measure *information quality*, i.e. actionability of DM solutions. InterActive8 also insists on interdisciplinary interactions in all of the phases of the process, but particularly in the problem formulation phase. Together with the recommended use of the Dialogue, a form of managed interdisciplinary conversation, this helps manage the CSFs *interdisciplinary learning* and *stakeholder participation.* The increased stakeholder commitment to the DM initiative thus improves the success measures *use (intention to use)* and *(perceived) net benefits*. As suggested in Table 13, InterActive8 also helps to focus on the remaining CSFs. Since the CSFs and the success measures work as a highly interrelated and interconnected system (see Figure 8 and section 4.3.4.2), DM practitioners may expect an increase in the probability of success and the speed to success from InterActive8. This was also demonstrated in the UCS case study. By following the proposed process model, embryonic DM initiatives are more likely to obtain *top management support* for wider DM implementation and do it faster.

6. Research Evaluation

Hevner et al. (2004) proposes seven guidelines for conducting and evaluating design science research in IS. We use them to evaluate and discuss the methodological framework and the research that led to it.

The first guideline, ***Design as an Artifact***, is to ensure that design-science research produces a viable artifact in a form of a construct, model, method or an instantiation. A methodological framework for embryonic DM integration fits into their definition of an artifact. It is an artifact at the intersection of people, organization, and IT. Moreover, it defines the ideas, practices, and technical capabilities through which the implementation of DM can be effectively and efficiently accomplished.

The second guideline is to ensure ***Problem Relevance***. The objective of the research should be to develop technology-based solutions to unsolved and important business problems. This is to ensure that the research contributions are clear, verifiable, new, and interesting. We have shown in section 1 that embryonic DM requires a specific managerial approach that is in some aspects different to BI implementation or other IS initiatives. Moreover, we show that the embryonic stages of DM integration process are under-researched, although the embryonic phase is necessary in a great majority of organizations (Davenport & Harris, 2007).

The third guideline, ***Design Evaluation***, is to ensure an artifact's utility, quality, and efficacy. They must be demonstrated through rigorous application of established evaluation methods. Our methodological framework evaluation is grounded in evidence collected through a case study method, which is a recommended observational evaluation method for in-depth studies of an artifact within an organizational context (Hevner et al., 2004).

Research Contributions is the fourth guideline aimed at ensuring clear and verifiable contributions of the research to theory and practice. Given the importance of this topic, we discuss them in a separate subsection below. We also included a subsection on the relevance of this research to practitioners.

The fifth guideline, ***Research Rigor***, is to ensure that the construction of an IS artifact is justified using prior theory and that the evaluation is carried out with appropriate research methods. The design of the methodological framework for embryonic DM management described in the previous sections includes a careful justification of each step using prior theory and evidence from explorative interviews and case studies. The design of the artifact is informed by existing theoretical frameworks, e.g. the Structured-case research method (Carroll & Swatman, 2000), the Design Science Research Model (Peffers et al., 2007), the DeLone & McLean IS Success Model (Delone & McLean,

1992, 2003), data warehousing and BI success models (Wixom & Watson, 2001; Yeoh & Koronios, 2010), etc.

Design as a Search is the sixth guideline. It suggests that design science research is inherently an iterative search for an effective solution to a problem. This study used ongoing comparisons between emerging solutions, interview findings, case study evidence, and continuous literature scrutiny to develop a reliable, valid, and useful methodological framework. In this way, both the identification of the CSFs and the design of InterActive8 required several iterations over an extended period of time.

The final, seventh guideline, ***Communication of Research***, suggests that the results of IS design-science are to be communicated effectively to technology-oriented and management-oriented audiences. Throughout the case studies participant feedback indicated that our findings are accessible to practitioners and can produce outputs that are useful for embryonic DM initiatives. Moreover, the findings have been communicated to relevant academic and business audiences in technology-oriented research conferences (Bole, Jaklič, Žabkar, & Papa, 2011; Bole & Papa, 2012), Predictive Analytics London 2011 practitioner conference, and in selected academic journals (Bole & Papa, 2011; Korošec, Bole, & Papa, 2013).

7. Discussion

A multi method initiative was undertaken to design a methodological framework for successful management of embryonic DM initiatives. Its aim is to guide practitioner decision making and also facilitate comparative research studies and the development of cumulative knowledge. The methodological framework is composed of three components: the definition of embryonic DM, best practice guidelines, and a process model. In this section, we discuss the contributions of this investigation and its relevance to DM practitioners.

7.1 Contributions to Theory

The introduction of DM in an organization is an important issue for both researchers and practitioners; however, no studies have empirically assessed the practices in embryonic-stage DM projects. Although some guidelines exist, none have been subjected to rigorous empirical testing. The methodological framework developed in this study may guide DM practitioners and may also facilitate variable selection in future research. Hence, this dissertation presents the first rigorous enquiry that develops an understanding of the factors that affect the execution of embryonic DM initiatives and the related process model. The reported success framework also provides clear guidance on how to practically measure the effectiveness and efficiency of embryonic-stage DM initiatives.

Our findings help advance DM theory by assisting DM researcher in obtaining and carrying out applicative research. As DM is fundamentally an applicative science, the development of new techniques and algorithms greatly depends on scientists' exposure to real-world problems. To date researchers could rely on existing methodologies, e.g. CRISP-DM, KDD Process, however as demonstrated earlier in this work these methods present several weaknesses. The proposed methodological framework improves and complements the existing methodologies. Therefore, it should aid DM researchers to obtain and implement applicative research projects more successfully thus advancing the DM theory towards new and better techniques and algorithms.

The InterActive8 process model is a unique contribution in that it describes the embryonic DM process by integrating two concurrent, iterative and explorative processes. Existing models, which were used to inform the design of InterActive8, tend to address only the organizational process or only the DM process. Similarly, the model is unique in the DM literature in that it attempts to incorporate both organizational and DM perspectives into one model. These two characteristics make InterActive8, to the best of our knowledge, unique also within the IS and organizational problem solving fields.

Moreover, this study extends current theory on DM integration management CSFs. It

suggests that prior research on the CSFs of DM (Hilbert, 2005; Nemati & Barko, 2003; Sim, 2003) presents conflicting findings because it does not distinguish between embryonic DM initiatives and organization-wide DM implementation. Our study presents the differentiating aspects of the embryonic DM initiatives. They are local initiatives, led by a mid-level manager, and their principal aim is to build executive commitment for enterprise-wide DM implementation. Consequently, the proposed formulation of the CSF theoretical framework to guide practitioner adoption is distinctive in various dimensions.

First, one of the four constructs that define success, *top management support,* is not to be found in extant IS (Delone & McLean, 1992, 2003) and BI (Wixom & Watson, 2001; Yeoh & Koronios, 2010) success literature. Second, contrary to what is implied for BI and DM in general (Hilbert, 2005; Sim, 2003; Yeoh & Koronios, 2010), *a well-established business case, top management support*, and c*hange management* cannot be CSFs in embryonic DM. Third, our research confirms the criticality of a *business champion* although it has been downplayed by all previous reports on the CSFs of DM (Hilbert, 2005; Nemati & Barko, 2003; Sim, 2003). Fourth, we found *interdisciplinary learning* to be the most frequent source of problems related to embryonic DM integration, yet it has not been explicitly considered by the existing DM literature. Fifth, in contrast with previous research on DM CSFs, this investigation was designed to search for additional concepts, beyond the constructs found in extant theory. The result is the CSF *process facilitation*, which is, to the best of our knowledge, unique in DM theory.

7.2 Practitioner Relevance

This research also suggests a number of important implications for DM pioneers and experts (internal and consultants), as well as for organizations that develop and sell DM software (Table 14: *Implications of this research for DM practice.*). Our findings allow organizations to identify and focus their scarce resources in the CSF areas. The analysis of the CSFs indicates that non-technical factors, including organizational and people-related issues, are far more influential and important than technological and data-related factors. The InterActive8 process model may be applied to educate the DM process stakeholders with respect to the embryonic DM process. In this manner the stakeholders may be off to a head-start in the painstaking process of interdisciplinary learning that embryonic DM requires. Moreover it may provide structure and focus to interdisciplinary meetings.

InterActive8 may also be used as a tool to manage stakeholder expectations as it is very explicit about the need to regularly (re)formulate the problem. This may help avoid frustration and build commitment faster. In accordance with research on process consultation (Schein, 1999), this research highlights the importance of building a helping relationship during and based on the problem-solving process. DM proponents that understand the necessity for an initial investment in this relationship can expect to achieve better results, i.e. stakeholder participation, efficiency increases, and top management support.

This study suggests that the level of *process facilitation* needed in a particular

organization depends on its culture. We found that innovative and interdisciplinary environments that foster fact-based decision making may facilitate the DM process. Therefore, the embryonic DM effort involves an adequate assessment of the degree to which the management of *stakeholder participation, interdisciplinary learning,* and *focus on problem solving action,* i.e. process facilitation, is required.

Table 14: *Implications of this research for DM practice.*

***Stakeholder* / Implications of the research**
DM pioneers – business managers, domain experts & end users
·Highlights the factors of importance in the management of a DM initiative, particularly the process factors in conservative organizational cultures ·Suggests that in traditional cultures the DM process may be managed by the *business champion* directly or indirectly through a process facilitator ·Suggests the skills required to manage the process – may be used for assessment of the candidate for a process facilitator ·Recommends patience and frequent interactions between the DM expert and domain experts as a way to stay focused, build trust, and avoid unintended erroneous assumptions ·Highlights the need to establish a method of coordination for efficient and unified DM effort (explorative process carried out by an interdisciplinary team) and proposes the InterActive8 process model as an alternative
DM experts & consultants
·Provides advice on building a trusting relationship with the remaining stakeholders of a DM initiative: patience (it is everybody's first to some degree), humility (avoid the trap of giving the impression of superiority), problem ownership (the problem and the solution belong to the business users), building of the relationship in the context of the business problem ·Suggests practical ways to facilitate interdisciplinary collaboration (remain silent as soon as the first domain expert has understood the idea, interdisciplinary learning also extends to knowing the client organization), business alignment (frequent interactions with domain experts), and data access issues (avoid work to the IT) ·Proposes ways to get the first project with a hesitant client (an assessment mini-project)
DM consultants & software vendors
·Improves the ability to identify customers and prospects more likely to start (existence of *external pressures*) and successfully complete DM implementation (facilitating culture, engaged *business champion*, or process facilitator)

DM consultants and software vendors can draw on the CSFs as a guide in their client prospecting efforts. *External pressure* on the organization is a good indicator of the likelihood of DM adoption. Moreover, an a priori assessment of all CSFs may be a good indication of the likelihood of success. Given the context-specific nature of how organizations understand the DM implementation process, frequent interactions and open communication between the consultants and their customers will be critical to ensure adequate management of potential shortcomings.

The findings of this research may also apply in DM integration in scientific research. Recently, Shmueli and Koppius (2011) have shown that predictive analytics, i.e. DM, are

necessary to build empirical models that predict well. Yet DM is still rare in theory building and testing in IS and in most social sciences. Shmueli and Koppius suggest that the under-representation may be due to the "unfamiliarity of most IS researchers with predictive analytics". This context implies that embryonic DM integration in research will follow similar processes and requirements as those suggested by our study.

8. Conclusions

A multi-method initiative was undertaken to design a methodological framework for the management of embryonic DM integration in organizations. The research involved the identification of the requirements and procedures (CSFs) and design of a process model (InterActive8) involved in the embryonic stages of DM implementation management.

First, the guidelines that lead to success were identified. In the absence of extant theory, a sensitizing framework was developed based on a literature review to guide exploratory interviews with experienced DM practitioners. This inquiry intended to re-focus the a priori theoretical framework on embryonic DM integration. The potential CSFs were then tested and further explored via a multiple case study. The results of this case study suggest that the process factors, i.e. *stakeholder participation, interdisciplinary learning,* and a *focus on problem solving action*, are the most salient issues impacting the success of embryonic stages in DM implementation. Their management must be carried out by the *business champion* or, when detached from the initiative, in collaboration with a process facilitator. *Interdisciplinary learning* and *process facilitation* CSFs are the unique contributions of this research that have not been previously identified in the DM literature. Exploratory interviews and case study results also served to enrich our understanding of the factors presented by prior research, the process CSFs in particular. These results may be regarded as a first step in the development of theory related to embryonic DM management to enable an appropriate foundation for the technology's later successful implementation.

Second, InterActive8, a process model for embryonic DM implementation, was designed. The design requirements were derived from the results of the preceding CSF study. In an extensive literature review, several reasonably similar process models for DM process on the one hand, and for organizational problem solving on the other, were identified and scrutinized based on the CSFs of embryonic DM integration. A consensus method was then applied in the design of InterActive8, which ensured its consistency with the selected DM and organizational problem solving models. The application of InterActive8 was subsequently demonstrated and evaluated in a case study. To the best of our knowledge, InterActive8 is a unique model in its characteristics of integrating two concurrent, iterative and explorative processes, i.e. the DM process and the organizational process of embryonic DM integration.

8.1 Research Limitations

Prior studies have examined the CSFs of DM in general. However, embryonic DM projects in established organizations where the support of key decision-makers has yet to

be secured present a greater complexity than other DM projects. No theory exists in relation to this context. Thus, it was not possible to rely on extant theory. The study draws heavily on referent and analogous domains to establish the initial set of candidate success factors and measures. It is recognized that these domains may have been overly influential (due to differences in context).

Interviews with DM practitioners and case studies of embryonic DM initiatives were conducted to empirically modify and further build the framework. The approaches selected and the data collected did not allow for rigorous testing of the relative influence of the success factors on the success measures – an inherent weakness of most success factor studies (Nandhakumar, 1996). This limitation presents an opportunity for future research.

Among the organizations subjected to our research, there were many different factors influencing the research process, and hence it was inevitable that the case studies were not completely controlled as well-structured processes. In addition, the CSFs were validated as generally applicable and useful for explaining how decisions in embryonic DM projects influence and are influenced by the context. However, users of the CSFs should be cautioned that the seven CSFs may not be sufficient. In the eight case studies used for their validation, there were other more or less important factors; however, they vary from one organization to the next. Therefore, they were not included among the CSFs necessary for success. Moreover, a user, e.g. a DM proponent, should adapt the knowledge to fit the specific problem situation and context. Interpretative research generally seeks to develop practical knowledge that can be used in designing, implementing, and managing IS initiatives (Gregor, 2006; Klein & Myers, 1999). This knowledge is abstract in the sense that it is not a recipe for designing and implementing a specific IS initiative.

It is also important to recognize that to obtain and validate both dependent and independent variables, our research relied on a single individual. As a result, the reported framework reflects an aggregation of individual perspectives rather than a shared mind set, i.e. the findings may have been influenced by the implicit or explicit biases of our respondents. Although the number of informants and the variety of their roles in the DM process suggests that the impact of biases is minimal, such analysis is not without limitations and certainly does not preclude all possible forms of bias. Therefore, to further eliminate this bias, our study might be extended by relying on other techniques such as focus groups (Morgan, 1997) or variants of the Delphi method (Okoli & Pawlowski, 2004).

8.2 Further Research

Given the lack of prior attention to embryonic DM implementation, numerous opportunities exist for future work. The concept of *process facilitation* is not addressed in the DM literature. It emerged from this research; however, its study was limited to the development of only a superficial view of the *process facilitation* role and the facilitator. Additional study is necessary to fully grasp this role. Moreover, the form of *process*

facilitation was found to depend on organizational culture. Additional work specifically designed to examine the embryonic DM process from the contingency perspective might reveal other relevant contingencies.

Our data suggests that the relationship between stakeholder participation and embryonic DM success is a very complex issue and depends on trust, commitment, stakeholder expectation management, etc. This is in accordance with the findings of other IS researchers who have expressed the complexity of user participation and its relation to success in terms of hands-on activities performed, psychological involvement and attitude, e.g. (Hartwick & Barki, 1994; Lin & Shao, 2000). Analysis of the influence of stakeholder participation on embryonic DM success at this level of detail was, however, beyond the scope of this study. To fully understand this mechanism, additional study is necessary.

Another possible direction for future research is the development of a dynamic system model for CSFs to describe and understand the internal causalities among the success factors and their relative influence on the success measures. The model should be operationalized for the purpose of a quantitative survey and subsequent statistical testing of model completeness and facility. These data might also facilitate testing the predictive power of the success factors.

Finally, it is anticipated that the framework reported in this study could also be readily adapted and tested within other analytics domains, i.e. forecasting/extrapolation, modeling, experimental design, simulation, and optimization.

9. Acknowledgements

I would like to thank my mentors and co-authors, Gregor Papa, Jurij Jaklič, Jure Žabkar, and Peter Korošec, who, each in a different way, helped me to complete this dissertation.

In moments of research crisis I also had the invaluable help of Stijn Viaene and Ray Hackney. They helped me to comprehend things at a higher level and move on with the research process.

This research would not have been possible without the generous support of the interviewees. Most remain anonymous for privacy reasons, but I am allowed to explicitly thank John Elder, Udo Sglavo, Tom Khabaza, and Richard Boire.

I am grateful to the EU (Operation part financed by the European Union, European Social Fund), the agency Spirit Slovenia, and the company Pikarp d.o.o. for economic support.

Thanks to my late friend Jose Luis Martos, who suggested to me over ten years ago that I should consider doing a Ph.D. I have felt his support from Heaven throughout the past five years.

Thanks to my parents. May this dissertation reward their efforts to pass on to me their virtues.

Thanks to my wife, Tea, for her support in times of exhaustion and crisis, and to my children, Terezija, Benedikt, and Veronika, for their inspiration.

I would like to thank God for giving me all of the above.

10. References

Ang, J. S. ., Sum, C.-C., & Yeo, L.-N. (2002). A multiple-case design methodology for studying MRP success and CSFs. *Information & Management*, *39*(4), 271–281.

Argyris, C. (1977). Organizational learning and management information systems. *Accounting, Organizations and Society*, *2*(2), 113–123.

Argyris, C. (1982). *Reasoning, Learning, and Action: Individual and Organizational*. San Francisco: Jossey-Bass Inc Pub.

Baesens, B., Mues, C., Martens, D., & Vanthienen, J. (2009). 50 years of data mining and OR: upcoming trends and challenges. *Journal of the Operational Research Society*, *60*, S16–S23.

Barton, D., & Court, D. (2012). Making Advanced Analytics Work for You. *Harvard Business Review*. Retrieved October 20, 2012, from http://hbr.org/2012/10/making-advanced-analytics-work-for-you/ar/4

Baskerville, R., & Myers, M. D. (2004). Special Issue on Action Resarch in Inofrmation Systems: Making IS Research Relevant to Practice - Foreword. *MIS Quarterly*, *28*(3), 329–335.

Beers, P. J., Boshuizen, H. P. A., Kirschner, P. A., & Gijselaers, W. H. (2005). Computer support for knowledge construction in collaborative learning environments. *Computers in Human Behavior*, *21*(4), 623–643.

Benbasat, I., Goldstein, D. K., & Mead, M. (1987). The Case Research Strategy in Studies of Information Systems. *MIS Quarterly*, *11*(3), 369–386.

Berry, M. J., & Linoff, G. S. (2004). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. Wiley & Sons.

Bessant, J., & Caffyn, S. (1997). High-involvement innovation through continuous improvement. *International Journal of Technology Management*, *14*(1), 7–28.

Blumberg, B., Cooper, D. R., & Schindler, P. S. (2008). *Business Research Methods*. Berkshire: McGraw-Hill.

Blumenstock, A., Hipp, J., Kempe, S., Languillon, C., & Wirth, R. (2006). Interactivity Closes the Gap, Lessons Learned in an Automotive Industry Application. In *Proceedings of the KDD Workshop on Data Mining for Business Applications*. Philadelphia.

Boland Jr., R. J., & Tenkasi, R. V. (1995). Perspective Making and Perspective Taking in Communities of Knowing. *Organization Science*, *6*(4), 350–372.

Bole, U. (2013). Transcripts of semi-structured interviews with data mining practitioners. Ljubljana.

Bole, U., Jaklič, J., Žabkar, J., & Papa, G. (2011). Identification of important factors to success of organizational data mining. In *EPIA 2011, 15th Portuguese Conference on Artificial Intelligence* (pp. 535–549).

Bole, U., & Papa, G. (2011). Who are the Likeliest Customers: Direct Mail Optimization with Data Mining. *Contemporary Engineering Sciences*, *4*(6), 259–268.

Bole, U., & Papa, G. (2012). Optimization in Organizations: Things We Tend to Forget. In *Fifth International Conference on Bioinspired Optimization Methods and their Applications* (pp. 257–268). Bohinj.

Bose, I. (2001). Business data mining - a machine learning perspective. *Information & Management*, *39*(3), 211–225.

Bose, R., & Luo, X. (2011). Integrative framework for assessing firms' potential to undertake Green IT initiatives via virtualization - A theoretical perspective. *Journal of Strategic Information Systems*, *20*(1), 38–54.

Breiman, L. (2001). Statistical Modeling: The Two Cultures. *Statistical Science*, *16*(3), 199–231.

Bullen, C. (1995). Reexamining Productivity CSFs: The Knowledge Worker Challenge. *Information Systems Management*, *12*(3), 13–18.

Cao, L. (2008). *Domain Driven Data Mining (D3M)*. *2008 IEEE International Conference on Data Mining Workshops* (pp. 74–76). Washington, DC, USA: IEEE.

Cao, L. (2009). Introduction to Domain Driven Data Mining. In L. Cao, P. S. Yu, C. Zhang, & H. Zhang (Eds.), *Data Mining for Business Applications* (pp. 3–10). Boston, MA: Springer US.

Cao, L. (2010). Domain-Driven Data Mining: Challenges and Prospects. *IEEE Transactions on Knowledge and Data Engineering*, *22*(6), 755–769.

Cao, L., & Zhang, C. (2006). Domain-Driven Data Mining. *International Journal of Data Warehousing and Mining*, *2*(4), 49–65.

Cao, L., & Zhang, C. (2007). The Evolution fo KDD: Towards Domain-Driven Data Mining. *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)*, *21*(4), 677 – 692.

Carroll, J. M., & Swatman, P. A. (2000). Structured-case: a methodological framework for building theory in information systems research. *European Journal of Information Systems*, *9*(4), 235–242.

Cate, F. H. (2008). Government Data Mining: The Need for a Legal Framework. *Harvard Civil Rights-Civil Liberties Law Review*, *43*(2), 435–489.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R.

(2000). CRISP-DM 1.0, Step-by-step data mining guide.

Chapman, R. L., & Corso, M. (2005). From continuous improvement to collaborative innovation: the next challenge in supply chain management. *Production Planning & Control*, *16*(4), 339–344.

Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., & Chau, M. C. L. (2004). Crime data mining: a general framework and some examples. *Computer*, *37*(4), 50–56.

Cleghorn, G. D., & Headrick, L. A. (1996). The PDSA cycle at the core of learning in health professions education. *The Joint Commission journal on quality improvement*, *22*(3), 206–12.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, *35*(1), 128–152.

Coppock, D. S. (2002). Data Mining Automation. *Information Management Online*.

Cornolba, C., & Giudici, P. (2004). Statistical models for operational risk management. *Physica A*, *338*(1-2), 166–172.

Crook, C. W., & Kumar, R. L. (1998). Electronic data interchange: a multi-industry investigation using grounded theory. *Information & Management*, *34*(2), 75–89.

Curley, K. F., & Gremillion, L. L. (1983). The role of the champion in DSS implementation. *Information & Management*, *6*(4), 203–209.

Davenport, T. H. (2005). *Thinking for a Living*. Boston: HBS Press.

Davenport, T. H., & Harris, J. G. (2007). *Competing on Analytics: The New Science of Winning*. HBS Press.

Davenport, T. H., Harris, J. G., & Morison, R. (2010). *Analytics at Work: Smarter Decisions, Better Results*. HBS Press.

DeBarr, D., & Eyler-Walker, Z. (2006). Closing the gap: automated screening of tax returns to identify egregious tax shelters. *ACM SIGKDD Explorations Newsletter*, *8*(1), 11–16.

Delone, W. H., & McLean, E. R. (1992). Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research*, *3*(1), 60–95.

Delone, W. H., & McLean, E. R. (2003). The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of Management Information Systems*, *19*(4), 9–30.

Deming, W. E. (2000a). *The New Economics for Industry, Government, Education* (second., p. 266). Cambridge: MIT Press.

Deming, W. E. (2000b). *Out of the Crisis* (p. 507). Cambridge: MIT Press.

Detert, J. R., Schroeder, R. G., & Mauriel, J. J. (2000). A Frameowrk for Linking Culture and Improvement Initiatives in Organizations. *Academy of Management Review*, *25*(4), 850–863.

Dong, S., Xu, S. X., & Zhu, K. X. (2009). Information technology in supply chains: the value of IT-enabled resources under competition. *Information Systems Research*, *20*(1), 18–32.

Du Chatenier, E., Verstegen, J. A. A. M., Biemans, H. J. A., Mulder, M., & Onno, O. (2009). The Challenges of Collaborative Knowledge Creation in Open Innovation Teams. *Human Resource Development Review*, *8*(3), 350–381.

Eckerson, W. W. (2007). *PREDICTIVE ANALYTICS Extending the Value of Your Data Warehousing Investment*. *TDWI Best Practices Report*.

Ehie, I. C., & Madsen, M. (2005). Identifying criticial issues in enterprise resource planning (ERP) implementation. *Computers in Industry*, *56*(6), 545–557.

Elder, J. (2007). Successes, Failures and Learning from Them. *KDD 2007 conference*. The 13th International Conference on Knowledge Discovery and Data Mining. Retrieved March 07, 2011, from http://videolectures.net/kdd07_elder_sfle/

Fagerberg, J. (2005). Innovation: A Guide to the Literature. In J. Fagerberg, D. Mowery, & R. Nelson (Eds.), *The Oxford Handbook of Innovation*. New York: Oxford University Press.

Fayyad, U. (2007). The Data Miner's Story - Getting to Know the Grand Challenges. *KDD 2007 conference*. Retrieved June 17, 2010, from http://videolectures.net/kdd07_fayyad_dms/

Fayyad, U., Piatetsky-shapiro, G., & Smyth, P. (1996). The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communications of the ACM*, *39*(11), 27–34.

Feelders, A., Danieils, H., & Holsheimer, M. (2000). Methodological and practical aspects of data mining. *Information & Management*, *37*(5), 271–281.

Féraud, R., Boullé, M., Clérot, F., Fessant, F., & Lemaire, V. (2010). The Orange Customer Analysis Platform. In *Advances in Data Mining, Applications and Theoretical Aspects* (pp. 584–594). Springer Berlin Heidelberg.

Fogelman Soulié, F. (2008). Industrial Data Mining, Challenges and Perspectives. *ECML PKDD 2008 conference*. Retrieved June 28, 2010, from http://videolectures.net/ecmlpkdd08_soulie_idmc/

Fosfuri, A., & Tribó, J. (2008). Exploring the antecedents of potential absorptive capacity and its impact on innovation performance. *Omega*, *36*(2), 173–187.

Frakes, W. B., & Fox, C. J. (1996). Quality improvement using a software reuse failure modes model. *IEEE Transactions on Software Engineering*, *22*(4), 274–279.

Gartner. (2011). Gartner Identifies the Top 10 Strategic Technologies for 2012. Retrieved July 19, 2012, from http://www.gartner.com/it/page.jsp?id=1826214

Gemino, A., Reich, B. H., & Sauer, C. (2008). A Temporal Model of Information Technology Project Performance. *Journal of Management Information Systems*, *24*(3),

9–44.

Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly*, *30*(3), 611–642.

Guynes, C. S., & Vanecek, M. T. (1996). Critical success factors in data management. *Information & Management*, *30*(4), 201–209.

Hand, D. J. (1999). Statistics and data mining. *ACM SIGKDD Explorations Newsletter*, *1*(1), 16–19.

Harkness, W. L., Kettinger, W. J., & Segars, A. H. (1996). Sustaining process improvement and innovation in the information services function: lessons learned at the Bose corporation. *MIS Quarterly*, *20*(3), 349–368.

Hartwick, J., & Barki, H. (1994). Explaining the Role of User Participation in Information Systems Use. *Management Science*, *40*(4), 440–465.

Hatchuel, A., & Weil, B. (2009). C-K design theory: an advanced formulation. *Research in engineering design*, *19*(4), 181–192.

Hermiz, K. (1999). Critical Success Factors for Data Mining Projects. *Information Management Magazine*. Retrieved April 05, 2010, from http://www.information-management.com/issues/19990201/164-1.html

Hevner, A. (2007). A three cycle view of design science research. *Scandinavian Journal of Information Systems*, *19*(2), 39–64.

Hevner, A., March, S., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *Management Information Systems Quarterly*, *28*(1), 75–106.

Highsmith, J. (2009). *Agile project management: creating innovative products*. Boston: Addison-Wesley.

Hilbert, A. (2005). Critical Success Factors for Data Mining Projects. In D. Baier, R. Decker, & L. Schmidt-Thieme (Eds.), *Data Analysis and Decision Support* (pp. 231–240). Springer-Verlag.

Holland, S., Gaston, K., & Gomes, J. (2000). Critical success factors for cross-functional teamwork in new product development. *International Journal of Management Reviews*, *2*(3), 231–259.

Howel, D., Windahl, C., & Seidel, R. (2010). A Project Contingency Framework Based on Uncertaingy and Its Consequences. *International Journal of Project Management*, *28*(3), 256–264.

KDnuggets. (2007a). Poll: CRISP-DM the Most Popular DM Methodology. *KDnuggets*. Retrieved March 22, 2011, from http://www.kdnuggets.com/news/2007/n17/1i.html

KDnuggets. (2007b). Poll: Data Mining Methodology. KDnuggets. Retrieved June 16, 2010, from http://www.kdnuggets.com/polls/2007/data_mining_methodology.htm

KDnuggets. (2008). Poll: Data Mining ROI. *KDnuggets*. Retrieved July 26, 2010, from

http://www.kdnuggets.com/polls/2008/roi-data-mining.htm

Khabaza, T. (2007). *CRISP-DM Overview and Update*.

Khabaza, T. (2011). Let's Debunk the Myths about Data Mining. *IMIS Journal*, *21*(4), 30–31.

Khazanchi, D. (2005). Information Technology (IT) Appropriateness: The Contingency Theory of "Fit" and IT Implementation in Small and Medium Enterprises. *The Journal of Computer Information Systems*, *45*(3), 88–95.

Kim, W. C., & Mauborgne, R. (2003). Fair Process: Managing in the Knowledge Economy. *Harvard Business Review*, *81*(1), 127–136.

King, E. A. (2005). How to Buy Data Mining, A Framework for Avoiding Costly Project Pitfalls in Predictive Analytics. *Information Management Magazine*. Retrieved March 08, 2008, from http://www.information-management.com/issues/20051001/1038094-1.html

King, E. A., & Rathburn, T. A. (2010). 101 Ways to Sabotage Your Predictive Analytics Project, Here Are the Top Five. *BeyeNETWORK*. Retrieved July 21, 2010, from http://www.b-eye-network.com/view/13685

King, S. F., & Burgess, T. F. (2008). Understanding success and failure in customer relationship management. *Industrial Marketing Managemnet*, *37*(4), 421–431.

Kirkman, B., Rosen, B., Gibson, C., Tesluk, P., & McPherson, S. (2002). Five challenges to virtual team success: Lessons from Sabre, Inc. *Academy of Management Executive*, *16*(3), 67–79.

Kiron, D., & Shockley, R. (2011). Creating Business Value with Analytics. *MIT Sloan Management Review*, *53*(1), 56–63.

Klein, H. K., & Myers, M. D. (1999). A Set of principles for Conducting and Evaluating Interpretative Field Studies in Information Systems. *MIS Quarterly*, *23*(1), 67–93.

Kohavi, R., Mason, L., Parekh, R., & Zheng, Z. (2004). Lessons and Challenges from Mining Retail E-Commerce Data. *Machine Learning*, *57*(1/2), 83–113.

Kolfschoten, G. L., & de Vreede, G. J. (2007). The collaboration engineering approach for designing collaboration processes. In J. M. Haake, S. F. Ochoa, & A. Cechich (Eds.), *Groupware: Design, Implementation, and Use* (Vol. 4715, pp. 95–110). Berlin, Heidelberg: Springer.

Korošec, P., Bole, U., & Papa, G. (2013). A multi-objective approach to the application of real-world production scheduling. *Expert Systems with Applications*, *40*(15), 5839–5853.

Laicity, M. C., Khan, S. A., & Willcocks, L. P. (2009). A review of the IT outsourcing literature: insights for practice. *Journal of Strategic Information Systems*, *18*(3), 130–146.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big Data, Analytics and the Path From Insigths to Value. *MIT Sloan Management Review*, *52*(2), 21–32.

Lavrač, N., Motoda, H., Fawcett, T., Holte, R., Langley, P., & Adriaans, P. (2004). Introduction: Lessons Learned from Data Mining Applications and Collaborative Problem Solving. In *Machine Learning* (Vol. 57, pp. 13–34).

Lichtenstein, S., Bendall, D., & Adam, S. (2008). Marketing Research and Customer Analytics: Interfunctional Knowledge Integration. *International Journal of Technology Marketing*, *3*(1), 81–96.

Lin, W. T., & Shao, B. B. M. (2000). The relationship between user participation and system success: a simultaneous contingency approach. *Information & Management2*, *37*(6), 283–295.

Lu, X.-H., Huang, L.-H., & Heng, M. S. H. (2006). Critical success factors of inter-organizational information systems—A case study of Cisco and Xiao Tong in China. *Information & Management*, *43*(3), 395–408.

Luftman, J., & Ben-Zvi, T. (2010). Key Issues for IT Executives 2010. *MIS Quarterly Executive*, *9*(4), 263–273.

Markus, M. L. (2001). Toward a Theory of Knowledge Reuse: Types of Knowledge Reuse Situations and Factors in Reuse Success. *Journal of Management Information Systems*, *18*(1), 57–93.

Markus, M. L., Majchrzak, A., & Gasser, L. (2002). A Design Theory for Systems That Support Emergent Knowledge Processes. *MIS Quarterly*, *26*(3), 179–212.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. Sage.

Moon, M. J. (2002). The Evolution of e-Government among Municipalities: Rethoric or Reality? *Public Administration Review*, *62*(4), 424–433.

Morgan, D. L. (1997). *Focus groups as qualitative research*. Sage.

Moss, L. T., & Atre, S. (2003). *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications*. Addison Wesley.

Nandhakumar, J. (1996). Design for success?: critical success factors in executive information systems. *European Journal of Information Systems*, *5*, 62–73.

Nemati, H. R., & Barko, C. D. (2003). Key Factors for Achieving Organizational Data-mining Success. *Industrial Management & Data Systems*, *103*(4), 282–292.

Norris, D. F., & Moon, M. J. (2005). Advancing e-Government at the Grassroots: Tortoise or Hare? *Public Administration Review2*, *65*(1), 64–75.

Okoli, C., & Pawlowski, S. (2004). The Delphi method as a research tool: an example, design considerations and applications. *Information & Management*, *42*(1), 15–29.

Owlia, M. S., & Aspinwall, E. M. (1997). TQM in higher education - a review. *International Journal of Quality & Reliability Management*, *14*(5), 527–543.

Paul, J. (2006). Structured Walkthroughs and Formal Technical Reviews. Retrieved September 08, 2009, from http://www.jodypaul.com/SWE/WT/walkthroughs.html

PAW. (2012). Predictive Analytics World for Government. Retrieved October 04, 2012, from http://www.predictiveanalyticsworld.com/gov/2012/

Pechenizkiy, M., Puuronen, S., & Tsymbal, A. (2008). Does Relevance Matter to Data Mining Research? In T. Y. Lin, Y. Xie, A. Wasilewska, & C.-J. Liau (Eds.), *Data Mining: Foundations and Practice* (Vol. 118, pp. 251–275). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-540-78488-3

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), 45–77.

Porter, M. . E., & Millar, V. E. (1985). How infromation gives you competitive advantage. *Harvard Business Reviewa*, *63*(4), 149–160.

Preece, J. J., Rogers, Y., & Sharp, H. (2007). *Interaction Design: Beyond Human-Computer Interaction* (2nd ed.). Chichester: John Wiley & Sons.

Reilly, S. (2011). Software that predicts the future: Does it really help? *Federal Times*. Retrieved October 04, 2012, from http://www.federaltimes.com/article/20110918/AGENCY03/109180301/1001

Remus, U., & Wiener, M. (2010). A Multi-Method, Holistic Strategy for Researching Critical Success Factors in IT projects. *Information Systems Journal*, *20*(1), 25–52.

Renzl, B. (2008). Trust in management and knowledge sharing: The mediating effects of fear and knowledge documentation. *Omega*, *36*(2), 206–220.

Rexer, K. (2011). *4th Annual Data Miner Survey - 2010 Survey Summary Report*.

Rexer, K. (2012). *5th Annual Data Miner Survey - 2011 Survey Summary Report*.

Ritter, T., & Gemünden, H. G. (2004). The Impact of a Company's Business Strategy on Its Technological Competence, Network Competence and Innovation Success. *Journal of Business Research*, *57*(5), 548–556.

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research Methods for Business Students*. Financial Times/Prentice Hall.

Schein, E. H. (1999). *Process Consultation Revisited: Building the Helping Relationship*. Addison-Wesley.

Scott, W. R. (2003). *Organizations: Rational, Natural, and Open Systems*. *Phoenix Usa* (Vol. 5, p. 430). Prentice Hall.

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action Design Research. *MIS Quarterly*, *35*(1), 37–56.

Senge, P. (2006). *The Fifth Discipline: The Art and Practice of the Learning Organization*. Doubleday.

Shewhart, W. A. (1939). *Statistical Method from the Viewpoint of Quality Control*. Washington D.C.: Graduate School, Department of Agriculture.

Shmueli, G., & Koppius, O. R. (2011). Predictive Analytics in Information Systems Research. *MIS Quarterly*, *35*(3), 553–572.

Sim, J. (2003). *Critical Success Factors in Data Mining Projects*. Retrieved from http://www.library.unt.edu/theses/open/20032/sim_jaesung/dissertation.pdf

Simon, H. A. (1996). *The Sciences of the Artificial*. MIT Press.

Spector, J. M., & Kim, C. (2012). A Model-based Approach for Assessment and Motivation. *Computer Science and Information Systems*, *9*(2), 893–915.

Spiegler, I. (2003). Technology and knowledge: bridging a "generating" gap. *Information & Management*, *40*(6), 533–539.

Strauss, A., & Corbin, J. (1998). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (Vol. 2, p. 379). Sage.

Thamhain, H. (2004). Team Leadership Effectiveness in Technology-Based Project Environments. *Project Management Journal*, *35*(4), 35–46.

Tidd, J., Bessant, J., & Pavitt, K. (2005). *Managing Innovation: Integrating Technological, Market and Organizational Change*. Wiley.

Viaene, S., & Van Den Bunder, A. (2011). The Secrets to Managing Business Analytics Projects. *MIT Sloan Management Review*, *53*(1), 64–70.

Von Haldenwang, C. (2004). Electronic Government (e-Government) and Development. *European Journal of Development Research*, *16*(2), 417–432.

Wang, H., & Wang, S. (2008). A knowledge management approach to data mining process for business intelligence. *Industrial Management & Data Systems*, *108*(5), 622–634.

Weiss, G. M. (2009). Data Mining in the Real World: Experiences, Challenges, and Recommendations. In *Proceedings of the 2009 International Conference on Data Mining* (pp. 124–130). CSREA Press.

Williams, J. J., & Ramaprasad, A. (1996). A Taxonomy of Critical Success Factors. *European Journal of Information Systems1*, *5*(4), 250–250.

Williams, S., & Williams, N. (2006). *The Profit Impact of Business Intelligence*. Morgan Kaufmann.

Wixom, B. H., & Watson, H. J. (2001). An Empirical Investigation of the Factors Affecting Data Warehousing Success. *MIS Quarterly*, *25*(1), 17–41.

Wu, X., Yu, P. S., Piatetsky-Shapiro, G., Cercone, N., Lin, T. Y., Kotagiri, R., & Wah, B. W. (2003). Data Mining: How Research Meets Practical Development? *Knowledge*

and Information Systems, *5*(2), 248–261. doi:10.1007/s10115-003-0101-1

Yeoh, W., & Koronios, A. (2010). Critical Success Factors for Business Intelligence Systems. *Journal of Computer Information Systems*, *50*(3), 23–32.

Yin, R. (2003). *Case Study Research: Design and Methods*. Sage Publications.

Zahra, S. A., & George, G. (2002). Absorptive Capacity: A Review, Reconceptualization, and Extension. *The Academy of Management Review*, *27*(2), 185–203.

Zhu, K., Kraemer, K., & Xu, S. (2003). Electronic business adoption by European firms: a cross-country assessment of the facilitators and inhibitors. *European Journal of Information Systems*, *12*(4), 251–268.

Index of Figures

Index of Tables

Appendix: Bibliography

Journal papers

Bole, U., & Papa, G. (2011). Who are the Likeliest Customers: Direct Mail Optimization with Data Mining. *Contemporary Engineering Sciences*, *4*(6), 259–268.

Korošec, P., Bole, U., & Papa, G. (2013). A multi-objective approach to the application of real-world production scheduling. *Expert Systems with Applications*, *40*(15), 5839–5853.

Conference papers

Bole, U., Papa, G., & Žabkar, J. (2010). Organizational point of view of applying predictive analytics in organizations. In *Proceedings of the XIIth International Electrotechnical and Computer Science Conference ERK 2010* (pp. 82–85). Ljubljana: IEEE Region 8.

Bole, U., Jaklič, J., Žabkar, J., & Papa, G. (2011). Identification of important factors to success of organizational data mining. In *EPIA 2011, 15th Portuguese Conference on Artificial Intelligence* (pp. 535–549).

Bole, U., & Papa, G. (2012). Optimization in Organizations: Things We Tend to Forget. In *Fifth International Conference on Bioinspired Optimization Methods and their Applications* (pp. 257–268). Bohinj.

Technical reports

Bole, U. (2010). Interview with a data mining practitioner: Udo Sglavo. *JSI Technical report, 10782*. Ljubljana.

Bole, U. (2010). Interview with a data mining practitioner: Tom Khabaza. *JSI Technical report, 10780*. Ljubljana.

Bole, U. (2010). Interview with a data mining practitioner: Richard Boire. *JSI Technical report, 10781*. Ljubljana.

Bole, U. (2010). Interview with a data mining practitioner: John Elder. *JSI Technical report, 10779*. Ljubljana.

Bole, U., Jaklič, J., Papa, G., & Žabkar, J. (2011). Critical success factors of data mining in organizations. *JSI Technical report, 10893*. Ljubljana.

Bole, U. (2013). Transcripts of semi-structured interviews with data mining practitioners. *JSI Technical report, 11355.* Ljubljana.

Bole, U., & Žabkar, J. (2011). InterActive8 methodological framework for management of data mining integration in organizations. *JSI Technical report, 10894.* Ljubljana.

Appendix: Biography

In 1996, Uroš Bole obtained his B.Sc. degree in Applied Mathematics and Economics at Brown University, USA. After spending a year working in an engineering company in Slovenia, he enrolled in the full-time MBA program at the IESE Business School at the University of Navarra, Spain. After completing his studies in 2000, he worked in several managerial positions in Spain and Slovenia. As part of his current position, he co-founded a company to provide consulting services in data mining.

In 2008, he started working towards his Ph.D. degree at Jožef Stefan International Postgraduate School in Slovenia. Since then he has been investigating the challenges of the integration of data mining and analytics in organizations. His research interests include interdisciplinary collaboration, systemic thinking, organizational learning, and process consultation.