SEMINAR

FÜR WIRTSCHAFTSINFORMATIK

UND INFORMATION SYSTEMS QUALITY

Prof. Dr. Ali Sunyaev

Hauptseminar Wirtschaftsinformatik

im Sommersemester 2014

Business intelligence and data quality:

Research on the causal link

vorgelegt von:

Schardein, Friederike

Content

Abbreviations

BA	Business analytics
BI	Business intelligence
CSF	Critical success factor
DQ	Data quality
ETL	Extraction-Transformation-Load

Figures

Tables

1. Inroduction and Motivation

Amounts of data are continuously growing. The International Data Corporation estimates that the total volume of data worldwide will increase from 2013 till 2020 by a factor of ten, that is from 4.4 trillion gigabytes to 44 trillion gigabytes.[1] Thus, it more than doubles every two years. The massive amounts of data available for organizations caused many to invest in technological systems to gain insights from internal and external data sources.[2] A business intelligence (BI) system is such a technical solution that offers integration and analysis of data to provide organizations with valuable information to support and improve their decision-making processes.[3] Especially proprietary data can offer competitive advantages since the value extracted from it cannot be copied by competitors.[4] Data quality (DQ) plays an important role for the value creation. The quality of data stored in BI systems affect the quality of the decision making.[5] In order to successfully implement a BI system in a company, a certain level of DQ might be needed. Thus, it is named as critical success factor (CSF) in theory and practical-oriented research.[6] However, DQ issues often become only visible when the BI system is introduced and the data from the source systems is loaded into the system.[7] These different perspectives raise the question about the causal link between BI and DQ. Moreover, research is divided concerning the causal link between the two. While some authors see DQ as CSF for BI,[8] a recent study of IŞIK ET AL (2013) revealed a negative correlation between the two indicating that an increase in DQ will not lead to better outcomes of BI.[9] Since it is not clear in which ways DQ effect the success of BI and how BI enhances DQ, this paper will contribute to this topic by examining the causal link and therefore the relevance between the two. Knowledge about the direction

---

[1] See for this and the next sentence Turner et al. (April 2014), p. 2.

[2] See Marchand, Peppard (2013), p. 104.

[3] See Hočevar, Jaklič (2010), pp. 89, 91 and Popovič et al. (2012), p. 729.

[4] See Breur (2009a), p. 21.

[5] See Wieder, Ossimitz, Chamoni (2012), p. 25.

[6] See Wixom, Watson (2001), p. 35 and Yeoh, Koronios (2010), p. 28.

[7] See Sammon, Finnegan (2000), p. 87 and Watson, Fuller, Ariyachandra (2004), p. 438

[8] See Wixom, Watson (2001), p. 35 and Yeoh, Koronios (2010), p. 28.

[9] See Işık, Jones, Sidorova (2013), p. 20.

of causal link between BI and DQ might offer benefits, for example by simplifying BI project implementation or gaining more value from BI projects. Therefore, the research question is:

What causal link exists between BI and DQ and vice versa?

By conducting a systematic literature review studies in the field of BI and DQ are identified and analyzed with regard to the question: What role does DQ play in the context of BI use and BI success and in which ways can the use and success of BI contribute to DQ?

Through the categorization of the studies by the direction of the indicated causal link it is possible to point out in which ways the different perspectives on the relationship between BI and DQ complement each other or can be combined. As a subgoal relevant context factors of the different studies are taken into account in order to explain possible inconsistencies between the study results. The results of this work will provide a better understanding for BI success and use of BI systems in practice with regard to DQ issues.

This paper is structured as follows. In chapter two the theoretical background is provided by defining the terms of BI and DQ. Then, in chapter three the methodology used for the literature research is described. The results of the literature research are presented in chapter four. First, a general overview of the primary studies is given which demonstrates how the studies are categorized. Then, the results are further described in the two subsections depending on the direction of the causal link. In chapter five the results presented in the previous chapter are discussed and implications are given. Last, the findings are summarized in chapter six and further research is suggested.

2. Theoretical background

This chapter serves as the foundation of this paper in order to give the relevant basis for the main part. The chapter is divided into two subchapters distinguishing between the terms BI and DQ.

2.1 Business intelligence

Business intelligence (BI) evolved out of decision-support systems (DSS) which were designed in the 1970s to support decision-making.[10] The decision-support domain has expanded over the years through the emergence of various applications in this field. The term BI was shaped by the analyst Howard Dressner of the Gartner Group in the early 1990s. Today there exist various definitions and views on the term of BI in the literature.[11] Some business perspectives focus on BI as data reporting and visualization, whereas technical views on BI emphasize the extraction, transformation and integration of data. In order to cover the different understandings of BI AZVINE ET AL. (2006) define the term in a broad manner. According to them BI deals with capturing and understanding as well as analysis and transformation of raw data into meaningful information in order to increase business performance.[12] WATSON (2009) regards BI as an umbrella term comprising "applications, technologies, and processes for gathering, storing, accessing, and analyzing data to help business users make better decisions".[13] Through BI information can be obtained within a short period of time while at the same time being presented in a simple and efficient way allowing the user to understand the meaning of the information.[14] With the support of BI important trends can be derived, customer behavior can be analyzed and decision-making is facilitated.[15]

BI systems explore, integrate, aggregate and analyze data from different information resources.[16] They are technical solutions that transform data into information and

---

[10] See for this and the next two sentences Watson, Wixom (2007), p. 96.

[11] See for this and the next sentenceAzvine et al. (2006), p. 1.

[12] See Azvine et al. (2006), p. 2.

[13] Watson (2009), p. 491.

[14] See Hočevar, Jaklič (2010), p. 89.

[15] See Hočevar, Jaklič (2010), p. 91.

[16] See Olszak, Ziemba (2007), p. 136.

knowledge and therefore provide support for effective decision-making, strategic thinking and acting in organizations.[17] BI systems include information tools that assist users in acquiring the required information in an efficient and simple manner.[18]

A BI system typically consists of the following components:[19]

- *Operational and external databases* serving as data sources
- *Extraction-Transformation-Load tools* (ETL) which include data collection from different sources, error checking, transformation of data into a standardized form and saving them into a data warehouse
- A *data warehouse* which serves as the central database within a company to store and access data separately from operational systems
- *Analytical tools* which access the data and translate them into information.

Organizations deploy BI in order to achieve business success.[20] BI success refers to the value created for the organization through BI initiatives. Often it is difficult to quantify the economic benefit of BI since several factors might contribute to the overall profitability. A preferred measure of BI success is how much the BI system contributes to the company's performance.[21] Depending on the individual company, performance can be evaluated on different aspects, such as revenue, profitability, cost reduction, growth and efficiency. Other measures focus on intangible benefits such as improved quality of information and an increased sensing of opportunities and threats.[22] Different BI success measures are used across organizations and even across business units depending on the expected benefits from the BI investment.[23] Therefore, BI success can be broadly defined as the "positive benefits organizations achieve through their use of BI".[24]

---

[17] See Olszak, Ziemba (2007), p. 137.

[18] See Hočevar, Jaklič (2010), p. 92.

[19] See for this paragraph Hočevar, Jaklič (2010), p. 92 and Olszak, Ziemba (2007), pp. 138, 139.

[20] See for this and the next two sentences Mohanty (2008), p. 21.

[21] See fot this and the next sentence Howson (2008), p. 54.

[22] See Hannula, Pirttimäki (2003), p. 595.

[23] See Işık, Jones, Sidorova (2013), p. 14.

[24] Işık, Jones, Sidorova (2013), p. 14.

CSFs for BI success have been investigated by many researchers (such as OLZAK & ZIEMBA (2012), WIXOM & WATSON (2001), YEOH ET AL. (2010)). DQ has been identified as the most important technical aspect for successful BI.[25] It is necessary that a BI system contains data of high quality in order to create value for the organization.[26]

2.2 Data quality

Data quality (DQ) is a prerequisite for the quality of the decision-making.[27] Demands for better and more timely reporting also increased the importance of DQ.[28] The quality of data depends on the intended use of it as well as the data itself.[29] Therefore, "data has quality if it satisfies the requirements of its intended use".[30] This so called fitness-for-use approach emphasizes the relativity of DQ since the consumer's viewpoint as well as the context need to be considered in terms of DQ aspects.[31] This means the person who uses the data, that is the data consumer, must be satisfied with the quality provided.[32] Low DQ, for example through missing or inconsistent data, can have negative consequences such as inefficient business decisions, low performance, increased costs and customer dissatisfaction.[33]

Quality of data can be best analyzed through multiple attributes or dimensions.[34] WANG & STRONG (1996) developed a framework to gather those aspects of DQ that are significant to data consumers and divided them into different DQ dimensions. A DQ dimension is defined as a "set of DQ attributes that represent a single aspect or construct of DQ".[35] WANG & STRONG (1996) identified those dimension by collecting DQ

---

[25] See Howson (2008), p. 99.

[26] See Wixom, Watson (2001), p. 35.

[27] See Ballou, Tayi (1999), p. 73.

[28] See Breur (2009a), p. 20.

[29] See Olson (2003), p. 24.

[30] Olson (2003), p. 24.

[31] See Wang, Strong (1996), p. 6 and Tayi, Ballou (1998), p. 54.

[32] See Wang, Strong (1996), p. 6.

[33] See Daniel et al. (2008), p. 133 and Redman (1998), p. 80.

[34] See Tayi, Ballou (1998), p. 56.

[35] Wang, Strong (1996), p. 6.

attributes from data consumers. These dimensions were then grouped into the following four categories:[36]

- *Intrinsic DQ:* Data have quality in their own right, that is high-quality data should be inherently good. Dimensions of quality are accuracy and objectivity of data as well as their believability and their reputation.
- *Contextual DQ:* DQ needs to be appropriate to the context of the task. Dimensions in this category include the relevancy, timeliness and completeness of data as well as the appropriate amount of data and adding value.
- *Representational DQ:* Data need to be clearly represented in order to be interpretable and easy to understand. Concerning the format of data they need to be represented in a concise and consistent way.
- *Accessibility DQ:* Data needs to be accessible to the data consumer but also secure.

Intrinsic DQ deals with the nature of data, contextual DQ is associated to the task at hand whereas representational and accessibility DQ are related to the role and requirements of systems. Fig. 2-1 gives a graphical overview of the DQ dimensions.

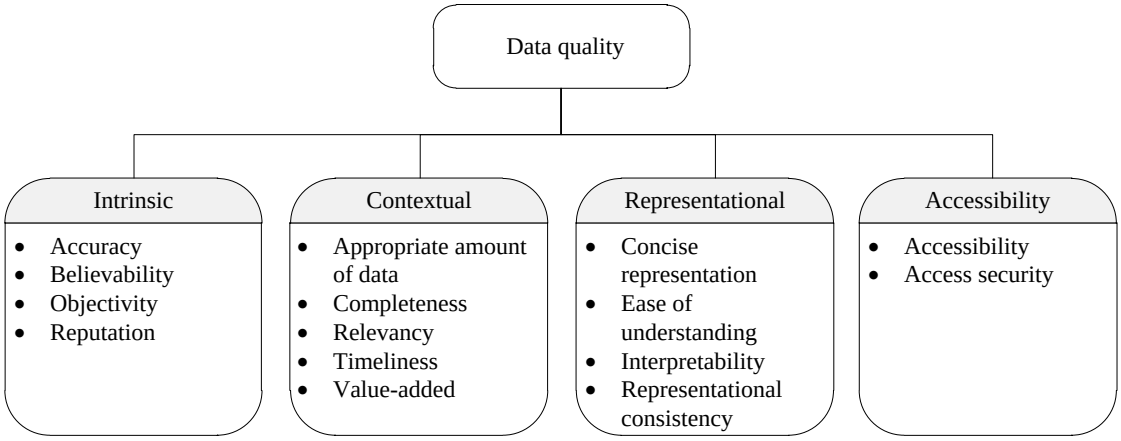

Fig. 2-1: DQ dimensions adapted from Wang & Strong (1996)[37]

KNIGHT (2011) recently identified through a literature review the common DQ dimensions among DQ researchers over the last decade.[38] According to this research the three most important dimensions are reliability, accuracy and timeliness.

---

[36] See for the following paragraph Wang, Strong (1996), pp. 19, 22.

[37] See Wang, Strong (1996), p. 20.

[38] See for this and the next sentence Knight (2011), p. 216.

3. Methodology of literature research

The literature research conducted in this seminar paper focuses on articles issued in internationally respected journals as well as conferences, in the research field of information systems. The selection of articles follows a five-step process (Fig. 3-1) which is derived from the literature search process by VOM BROCKE ET AL. (2009).[39]

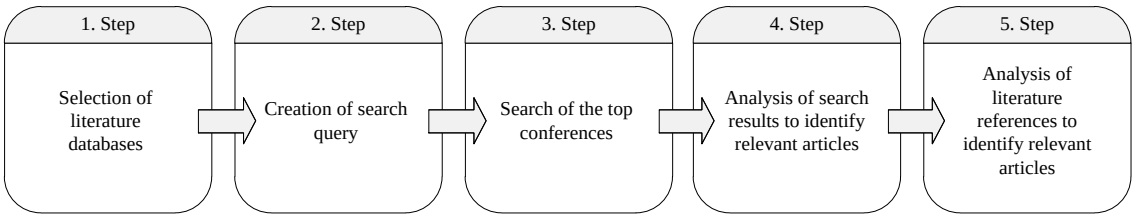

Fig. 3-1: Methodology of literature research

In the *first step* the literature databases which will be examined in the context of this work are selected. In order to cover the minimum requirement of including the *Senior Scholars‘ Basket of Journals*, the following three online literature databases are chosen: Ebsco (Academic Search Complete, Business Source Complete), Sciencedirect und ProQuest. In addition, conference proceedings are included through the electronic library of the Association of Information Systems (AISel).

The *second step* involves the creation of a search query which is executed on the prior chosen databases. The databases are browsed via keywords with a combination of the fields’ *title*, *abstract* and *keyword*. A research in advanced helped to identify suitable keyword combinations for the analysis. This results in the following search string:

```
TITLE OR ABSTRACT OR KEYWORD ("business intelligence") AND

TITLE OR ABSTRACT OR KEYWORD ("DQ" OR "quality of data")
```

The above mentioned syntax was each transformed to the specific syntaxes of the literature databases. In addition, the search was limited to articles that were published since 2001 and peer-reviewed. Thus, it is ensured that the articles focus on current research as well as have been reviewed by an expert. Through this search 42 articles were found.

In the *third step* the two top conferences in the field of information systems are searched in detail. In order to avoid missing relevant literature that was excluded by the search

[39] See vom Brocke et al. (2009).

query in step two, the respective BI tracks of the International Conference on Information Systems and the European Conference on Information Systems in 2013 are searched. Therefore, it is ensured that actual research is included into the work. This search revealed two more relevant articles.

The *fourth step* serves to evaluate the articles found in the second and third step. Through an assessment of the articles‘ titles and abstracts, a preselection of relevant articles is made. Altogether 14 articles are identified to be taken under further consideration. The excluded articles are not further considered. Based on the preselected articles, a full text analysis is performed to exclude further irrelevant articles, if necessary. Only full research papers are considered. After completing the full text analyzes 10 articles are identitfied as relevant.

In the *fifth step* an examination of the literature references from the articles identified in step four is made. Based on the title of the literature it is decided which articles are taken under further consideration. These articles are evaluated using the method described in the fourth step and, if identified as relevant, included to the pool of results. The forward search did not reveal any further articles. Through the backward search 6 additional articles were marked as relevant. In total 16 relevant articles are identified through the literature research, further reffered to as primary studies (Appendix Tab. 0-1).

4. Comparison of results

In this section the results of the relevant studies identified through the systematic literature review are presented. First, a general overview of the primary literature is given. Then, the other two subsections focus on specific results concerning the impact of DQ on BI and vice versa.

4.1 Overview of results

As shown in Tab. 4-1 the primary studies are classified into different categories concerning the direction of the causal link between DQ and BI, the used research method (contextual or empirical) and the perspective of BI they are covering (BI use or BI success). Studies falling into the category of BI use either address initiatives that facilitate the use of the BI system or how the use of the BI system can enhance DQ. Articles within the category of BI success focus on how DQ and BI initiatives influence the benefits of a BI implementation. The impact of DQ on BI is mainly linked to BI success while the impact of BI on DQ is more often associated with BI use. The majority of the primary studies, that is 12 out of 16, deal with the impact of DQ on BI. Also the studies are more often connected to BI success than to BI use. Concerning the research method the empirical approach is more commonly used among the primary studies.

The articles differ in their understanding of BI and DQ while some do not offer any definition (Tab. 4-2). A definition is here considered as explicitly describing the term of BI or DQ. The majority regards BI as an umbrella term, which is a collection of tools, technologies and processes for the creation of organizational knowledge and thus supporting decision-making of an organization. Some studies focus on BI as an analytical process to transform data into knowledge, while others see it as technical solutions to support decision making. Three studies do not offer a BI definition at all. DQ is only defined by a few studies, mostly through DQ dimensions. The fitness-for-use approach (see 2.2) is mentioned twice. For those studies that define DQ, Tab. 4-3 shows which DQ dimensions by WANG & STRONG (1996) (see 2.2) are covered. While accuracy, completeness and representational consistency are the dimensions mostly named, the dimensions objectivity, reputation, value-added and concise representation are not mentioned in any definition.

Tab. 4-4 gives an overview of the addressed topics by the primary studies which are presented in detail in the following two subchapters.

Causality	Article	Type of research			BI perspective	
		Conceptual research	Exmprical research		BI use	BI success
			Qualitative	Quantitative		
Impact of DQ on BI	Dawson & Van Belle (2013)		x			x
	Işık et al. (2013)			x		x
	Malladi & Krishnan (2013)			x	x	
	Missi et al. (2004)	x				x
	Pula et al. (2003)		x			x
	Tamm et al. (2013)		x			x
	Hwang & Xu (2008)			x		x
	Hawking & Sellitto (2010)		x			x
	Olszak & Ziemba (2012)		x			x
	Wixom & Watson (2001)			x		x
	Yeoh & Koronios (2010)		x			x
	Yeoh et al. (2012)	x				x
Impact of BI on DQ	Breur (2009a)	x				x
	Breur (2009b)	x			x	
	Wieder et al. (2012)			x		x
	Daniel (2008)	x			x	
	Yeoh et al. (2012)	x			x	

Tab. 4-1: Overview of causal directions, used research method and BI perspective in primary studies

Article	BI definition			DQ definition	
	Umbrella term	Analytical Process	Technical solution	DQ dimensions	Fitness-for-use
Breur (2009a)				x	x
Breur (2009b)					
Daniel (2008)			x	x	
Dawson & Van Belle (2013)	x				
Hawking & Sellitto (2010)					
Hwang & Xu (2008)			(x)		
Işık et al. (2013)	x			x	
Malladi & Krishnan (2013)		(x)	(x)		
Missi et al. (2004)		x			
Olszak & Ziemba (2012)	x				
Pula et al. (2003)					
Tamm et al. (2013)			x		
Wieder et al. (2012)		x		x	
Wixom & Watson (2001)			(x)	x	
Yeoh & Koronios (2010)	x		x		
Yeoh et al. (2012)	x				x

x = term explicitly defined; (x) = parts of the term defined, e.g. data warehouse

Tab. 4-2: Overview of definitions covered by the articles

	Intrinsic DQ				Contextual DQ					Representational DQ				Accessibility DQ	
Article	Believability	Accuracy	Objectivity	Reputation	Value-added	Relevancy	Timeliness	Completeness	Appropriate amount of data	Interpretability	Ease of understanding	Representational consistency	Concise representation	Accessibility	Access security
Breur (2009a)		x				x	x	x						x	x
Daniel (2008)	x							x				x			
Işık et al. (2013)		x				x	x			x	x	x			
Wieder et al. (2012)		x				x		x	x			x			x
Wixom & Watson (2001)		x						x				x			

Tab. 4-3: DQ dimensions addressed by primary articles

Causality	BI perspective	Topic	Respective articles
Impact of DQ on BI	BI use	Need for strong data-related infrastructure and data management practices	Malladi & Krishnan (2013)
	BI success	DQ as CSF for BI success	Dawson & Van Belle (2013); Hawking & Sellitto (2010); Hwang & Xu (2008); Olszak & Ziemba (2012); Wixom & Watson (2001); Yeoh & Koronios (2010)
		Lack of DQ inhibits BA success	Tamm et al. (2013)
		Better DQ does not lead to greater BI success	Işık et al. (2013)
		Detection and prevention of errors before and during BI use	Missi et al. (2004); Pula et al. (2003); Yeoh et al. (2012)
Impact of BI on DQ	BI use	Revelation of DQ issues through report usage	Daniel (2008); Yeoh et al. (2012)
		Quality enhancement of data entry through user-friendly systems	Breur (2009b)
	BI success	Use of modeling paradigm to detect DQ issues	Breur (2009a)
		Positive influence of BI management on DQ	Wieder et al. (2012)

Tab. 4-4: Adressed topics by primary studies

4.2 Impact of DQ on BI

4.2.1 BI use

MALLADI & KRISHNAN (2013) examine the effects of data infrastructure sophistication and data management on the extent of BI usage in business activities.[40] Since DQ is crucial to gain reliable information from BI systems, a strong data-related infrastructure with a focus on data collection and cleansing promotes BI use.[41] On the contrary, the authors identify a negative impact on BI use if companies have not established strong data management practices. Those companies might be more likely to deal with DQ issues and thus face greater challenges in relying on their data due to mistrust. These

[40] See Malladi, Krishnan (2013), p. 7.

[41] See for this and the next two sentences Malladi, Krishnan (2013), pp. 7, 13.

findings suggest the need for technology and process readiness in order to provide appropriate data that can be used for BI.[42]

4.2.2 BI success

Several studies investigate CSFs for data warehouse and BI success. On the basis of the study by WIXOM AND WATSON (2001) which identified that a high level of DQ positively influences the level of perceived warehousing success,[43] several subsequent studies confirm this result with regard to BI success. The studies use different methods for data collection including online surveys, Delphi method, case study and industrial presentations. DQ is important for BI success since it influences the quality of information which in turn results in positive organizational outcomes,[44] such as increased quality of management reports and thus improved decision outcomes.[45] These results are connected to specific sectors (DAWSON & VAN BELLE (2013): financial services sector; HAWKING & SELLITTO (2010): ERP systems environment; OLSZAK & ZIEMBA (2012): small and medium enterprises), as well as across industries (HWANG & XU (2008); YEOH & KORONIOS (2010)). The need of readily available relevant high-quality data is also pointed out by TAMM ET AL. (2013) with a focus on business analytics (BA).[46] BA is defined as using data in order to have a better base for decision-making which is enabled through BI.[47] Interviews with senior managers involved in BA activities confirm that a lack of DQ inhibits BA success.[48] One interviewee points out that a good analytical model is of no value if the data itself has poor quality and lack of confidence.[49]

In contrast to these findings above IŞIK ET AL. (2013) which made a survery with business managers do not find support for the hypothesis that better DQ in an

---

[42] See Malladi, Krishnan (2013), p. 16.

[43] See Wixom, Watson (2001), p. 35.

[44] See Hwang, Xu (2008), p. 54.

[45] See Yeoh, Koronios (2010), p. 28.

[46] See Tamm, Seddon, Shanks (2013), p. 4.

[47] See Tamm, Seddon, Shanks (2013), p. 2.

[48] See Tamm, Seddon, Shanks (2013), pp. 9, 13.

[49] See Tamm, Seddon, Shanks (2013), p. 14.

organization leads to a greater BI success.[50] They suggest that the DQ provided by BI systems is perceived as being sufficient and further improvements may negatively affect other BI capabilities, such as flexibility.[51] However, the authors emphasize the danger of underestimating the need to continually ensure quality of data within the BI system. Hence, organizations might not recognize that their decisions are based on incorrect data.

DQ initiatives before or during the BI implementation are proposed by several authors to ensure BI success. The assessment of DQ at the beginning of the BI implementation allows discovering inconsistencies in data at an early stage and to initiate actions for improvement.[52] As there are different sources of DQ issues, for example source systems or ETL implementation, it is important to identify them to be able to fix the problems.[53] Also, a DQ program promotes the prevention of errors before data are integrated.[54] Thus, the decision input is improved and it is ensured that the data fulfill the requirements of the BI system.[55] As a result, business value and success is achieved earlier through the BI investment.[56]

Not only during BI implementation but also when the BI system is already in use, comprehensive DQ management should be done.[57] This includes monitoring and solving of DQ issues as well as educating BI users about DQ and the BI environment to ensure BI success.[58]

4.3 Impact of BI on DQ

4.3.1 BI use

The influence of BI use on DQ is addressed by DANIEL ET AL. (2008) and YEOH ET AL. (2012) through the use of reports from BI systems. Reports disclose the quality of the

---

[50] See Işık, Jones, Sidorova (2013), pp. 18, 20.

[51] See for this and the next two sentences Işık, Jones, Sidorova (2013), p. 21.

[52] See Pula, Stone, Foss (2003), pp. 336, 337.

[53] See Yeoh, Wang, Verbitskiy (2012), p. 2.

[54] See Missi, Alshawi, Fitzgerald (2004), p. 6.

[55] See Missi, Alshawi, Fitzgerald (2004), p. 10.

[56] See Pula, Stone, Foss (2003), p. 338.

[57] See Yeoh, Wang, Verbitskiy (2012), p. 2.

[58] See Yeoh, Wang, Verbitskiy (2012), p. 10.

data on which the generated results are based on.[59] Therefore, information needs to be presented in an understandable manner so that users can identify quality problems. Otherwise, if information presented in the report is misunderstood by the user this might lead to misuse of information.[60] This in turn can result in the user's lack of confidence towards the data presented.[61] A possible solution to prevent this problem might be the provisioning of additional information to users to improve understanding of the report as well as offering users the opportunity to give feedback about the report.[62] Also, interactive user involvement through which users can correct the information by including or excluding data to the report allows report updates on the fly.[63]

BREUER (2009b) discovered that a user-friendly system enhances the quality of data entry.[64] Supporting users with interfaces that facilitate faster data entry will result in more accurate data. Error rates can be improved through the provision of intelligent text input fields and smart drop-down menus.

4.3.2 BI success

Despite offering the least expensive solution, preventing errors in data before the data warehouse is set up is not always possible due to time constraints.[65] Governance of data warehouse projects is seen as success factor to create the most value from corporate data.[66] In order to meet the demands for more value and faster delivery of data, BREUR (2009a) suggests an approach for modeling the data warehouse. The so called Data Vault modeling paradigm is used to detect DQ issues when transforming data to a data mart, [67] which provide data in a usable format for analysis.[68] Since the Data Vault contains data as is, it allows the data warehouse team to identify discrepancies between

---

[59] See for this and the next sentence Daniel et al. (2008), p. 133.

[60] See Yeoh, Wang, Verbitskiy (2012), p. 6.

[61] See Yeoh, Wang, Verbitskiy (2012), p. 4.

[62] See Yeoh, Wang, Verbitskiy (2012), p. 6.

[63] See Daniel et al. (2008), pp. 134, 141.

[64] See for this and the next two sentences Breur (2009b), p. 120.

[65] See Breur (2009a), p. 26.

[66] See for this and the next sentence Breur (2009a), p. 27.

[67] See Breur (2009a), p. 27.

[68] See Solomon (2005), p. 27.

actual data and required data for business use.[69] This enables discovering root causes of errors and therefore supports the improvement of business processes causing these errors which in turn improve DQ. In addition, since data is loaded basically unchanged to the data store, business owners maintain control of the data they are using. [70] Also, auditing and tracing of data are improved for users which naturally enforce better DQ. Thus, Data Vault modeling is seen as an effective paradigm to ensure success of the data warehouse.[71]

Through a survey WIEDER ET AL. (2013) analyze the positive impact of BI management quality on DQ. BI management includes the setting of implementation standards as well as provisioning of required resources and scalable solutions.[72] The results confirm that the quality of managing BI is an antecedent of DQ and thus improves the quality of the decision-making.[73]

5. Discussion and implications

In the following the causal link between DQ and BI which is derived from the results from section 4 is discussed. The findings in section 4 revealed both similar and diverse viewpoints on the relationship between DQ and BI. It is discussed to which extent these viewpoints complement each other. Then, contradictory viewpoints and possible explanations are presented.

5.1.1 Complementary findings

The manifold studies on CSFs of BI confirm that high DQ remains an important driver for BI success throughout the years and is relevant across different sectors and industries. A lack of DQ inhibits BI success since it negatively affects the quality of the organizational decision-making. DQ management is suggested as concrete action to detect DQ issues and react to it in order to improve the level of DQ. Companies that fail to establish DQ management will struggle with reliability of data resulting in user's mistrust of the BI system and thus inhibit BI use. Next to DQ management, a strong

---

[69] See for this and the next sentence Breur (2009a), p. 27.

[70] See for this and the next sentence Breur (2009a), pp. 27, 28.

[71] See Breur (2009a), p. 28.

[72] See Wieder, Ossimitz, Chamoni (2012), p. 13.

[73] See Wieder, Ossimitz, Chamoni (2012), p. 25.

data-related infrastructure as well as high BI management quality enhance the quality of data and thus contribute to both use and success of BI systems. However, since DQ is hardly defined by the studies it cannot be stated if there are certain dimensions of DQ that are more important to be achieved than others. Accuracy, completeness and consistency of data are the dimensions mostly mentioned by the six studies that define DQ. This result is in contrast to KNIGHT (2011) who identified reliability, accuracy and timeliness as most important dimensions of DQ.

Still, the results imply that prior to the introduction of a BI system in an organizations, DQ cleansing processes must be done. Once the system is in use additional errors in data are detected, for example through report use, and thus the quality of data can be further improved. It is necessary to actively use data in order to be able to detect errors or determine if actual DQ still suits a changing business environment.[74] Providing user feedback therefore is an important mechanism to address such inconsistencies.[75]
Towards maintaining DQ, new data entering the BI system should fulfill the quality requirements. This can be enhanced by providing user-friendly systems that facilitate data entry.

5.1.2 Diverse and contradictory findings

Most striking are probably the different results of WIXOM & WATSON (2001) and IŞIK ET AL. (2013) concerning the impact of DQ on BI success. While WIXOM & WATSON (2001) find support that a high level of DQ positively influences warehousing success, IŞIK ET AL. (2013) detect no relationship between better DQ and increased BI success. A possible explanation for these conflicting results could be that WIXOM & WATSON (2001) limit their quantitative study to the success of data warehousing as a subset of BI, while IŞIK ET AL. (2013) focus on BI as a whole. In addition, the authors differ in their definition of DQ. While WIXOM & WATSON (2001) concentrate on accurate, complete and consistent data in a warehouse,[76] IŞIK ET AL. (2013) see more dimensions as crucial for BI success, including timeliness, relevancy and comprehensiveness of data.[77]

---

[74] See Orr (1998), p. 68.

[75] See Orr (1998), p. 67.

[76] See Wixom, Watson (2001), p. 19.

[77] See Işık, Jones, Sidorova (2013), pp. 14, 15.

It must also be considered that a period of about 12 years lies between the two studies. For both studies the research is based on a survey with managers using data warehouses or BI systems. While for WIXOM & WATSON (2001) the measure of DQ in the survey equals their definition of DQ (that is accuracy, completeness, consistency of data) it remains unclear if and how IŞIK ET AL. (2013) provided a common understanding of DQ for the participants of their survey. The same applies to the benefits of the data warehouse or BI system. WIXOM & WATSON (2001) measure perceived benefits of the data warehouse by data suppliers on the basis of three criteria: change of job through data warehouse, reduced time and effort for decision-making support for end-users.[78] IŞIK ET AL. (2013) define BI success in a broad manner by referring to positive benefits organizations achieve through BI use.[79]

IŞIK ET AL. (2013) explain their results by assuming that today's organizations are already aware of the importance of DQ and thus have achieved a reasonable level of quality. However, DQ underlying the BI systems should not be underestimated. As the case study by PULA ET AL. (2003) shows, transforming, cleaning and matching of data from source systems uncovered lots of DQ issues although the employees had previously assumed the data was of high quality. [80] Failing to recognize poor DQ can result in wrong or bad decisions without knowing it. An alternate explanation for these contrary findings is that better DQ does improve BI success however only to a certain extent. Once a certain threshold of DQ is crossed, it might have a negative influence on other factors of BI success. Strict DQ rules can hinder flexibility of BI systems such as decision-making in exceptional or urgent situations.[81] Therefore, IŞIK ET AL. (2013) conclude that DQ might be necessary but not sufficient for BI success.[82]

While some studies demand for DQ assessment at the beginning of the BI project to achieve early business values, BREUR (2009a) indicates that it is not always possible due to time constraints.[83] Regarding the definitions for BI and DQ these two viewpoints

---

[78] See Wixom, Watson (2001), p. 31.

[79] See Işık, Jones, Sidorova (2013), p. 14.

[80] See Pula, Stone, Foss (2003), pp. 329, 333.

[81] See Işık, Jones, Sidorova (2013), pp. 15, 21.

[82] See Işık, Jones, Sidorova (2013), p. 21.

[83] See Breur (2009a), p. 26.

are difficult to compare. BREUR (2009a) does neither give a definition for BI nor for data warehouse and data warehouse governance, but he defines DQ. YEOH ET AL. (2012) only mention the fitness for use approach while the others do not give any definition for DQ. PULA ET AL. (2003) also do not offer a definition for BI terms. The BI definitions by MISSI ET AL. (2004) and YEOH ET AL. (2012) focus on different aspects (analytical process vs. umbrella term). However, for a BI project to be successful it might be advisable to first try to resolve as many DQ issues as possible prior to implementation. Then, if time pressure does not allow this procedure, an adoption of the proposed Data Vault modeling paradigm might be beneficial.

5.1.3 Implications

DQ is considered as CSF for BI success since it positively influences the values derived from a BI system. However, not all DQ issues can be resolved before the BI system is in use. Still, one should try to identify and eliminate data inconsistencies before and during BI implementation in order to avoid rework and achieve early benefits of the BI system. With DQ management, DQ problems can be monitored and fixed. Some DQ issues will only occur when actually using the BI system which emphasizes the importance of the other direction of the causal link, which is the impact of BI on DQ. When working with the BI system users usually detect additional DQ problems. Therefore, feedback mechanism for users of the BI system should be established. Sustaining a level of DQ can be achieved by BI management through setting of implementation standards. It should be also ensured that new data entering the BI systems is of high quality, for example through the provisioning of user-friendly systems.

Since a certain extent of DQ might inhibit other CSFs of BI and therefore harm the overall BI success, specification of a threshold of required DQ in an organization might be necessary.

6. Conclusion

This work shows that a causal link between BI and DQ exists in both directions. DQ plays an important role for a successful BI project and thus has to be assured. However, not all DQ issues can be resolved before the implementation of the BI system. Some DQ problems are only surfaced once the system is implemented and others even later when the system is already in use. Therefore, both directions of the causal link are important for a successful BI implementation.

It was a subgoal to compare the different context factors used by the studies in order to understand conflicting results. However, it turned out that only 5 of 16 studies explicitly give a definition for DQ. For BI terms most of the studies offered a definition. Still, for examining the causal link between BI and DQ, it is necessary to have a common understanding for both terms and also communicate these to the respective participants when doing empirical studies. DQ dimensions have to be fully understood to be able to address and resolve DQ issues.[84] Further research could also deal with identifying which of the DQ dimensions are most important for organizations and therefore should be top priority. Still, it is not clear if better DQ always leads to greater BI success. Thus, it should be examined at which level DQ does not contribute any longer to BI success. If a threshold of DQ exists it must be communicated to business to prevent negative consequences on other CSFs of BI.

Another interesting research area is about the impact of sensemaking on the usefulness of BI data. This topic is studied by HOUGHTON & MACKRELL (2012) but since the research is not completed at this time, it is not included into the work. Their first results indicate that sensemaking can hinder DQ of the BI system if users do not fully understand how work is undertaken and thus develop their own sense of work.[85] By further examining this topic, useful insights about the data users and their influence on DQ might be detected.

---

[84] See Tayi, Ballou (1998), p. 56.

[85] See Houghton, Mackrell (2012), p. 6.

References

Azvine et al. (2006)
B. Azvine, Z. Cui, D.D Nauck, B. Majeed: Real time business intelligence for the adaptive enterprise. In: IEEE Computer Society (Editor): Proceedings of the 8th IEEE International Conference on E-Commerce Technology and the 3rd IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services (CEC/EEE'06), June 26-29, San Francisco, California. Los Alamitos, California 2006, p. 29

Ballou, Tayi (1999)
Donald P. Ballou, Giri K. Tayi: Enhancing data quality in data warehouse environments. In: Communications of the ACM. Nr. 1, Vol. 42, 1999, pp. 73-78

Breur (2009a)
Tom Breur: Data quality is everyone's business - Designing quality into your data warehouse - Part 1. In: Journal of Direct, Data and Digital Marketing Practice. Nr. 1, Vol. 11, 2009, pp. 20-29

Breur (2009b)
Tom Breur: Data quality is everyone's business - Managing information quality - Part 2. In: Journal of Direct, Data and Digital Marketing Practice. Nr. 2, Vol. 11, 2009, pp. 114-123

Daniel et al. (2008)
Florian Daniel, Fabio Casati, Themis Palpanas, Oleksiy Chayka: Managing data quality in business intelligence applications. In: H. V. Jagadish (Editor): Proceedings of the VLDB Endowment, August 24-30, Auckland, New Zealand. 2008, pp. 133-143

Hannula, Pirttimäki (2003)
Mika Hannula, Virpi Pirttimäki: Business intelligence empirical study on the top 50 Finnish companies. In: Journal of American Academy of Business. Nr. 2, Vol. 2, 2003, pp. 593-599

Hočevar, Jaklič (2010)
Borut Hočevar, Jurij Jaklič: Assessing benefits of business intelligence systems - A case study. In: Management: Journal of Contemporary Management Issues. Nr. 1, Vol. 15, 2010, pp. 87-119

Houghton, Mackrell (2012)
Luke Houghton, Dale Mackrell: The impact of individual, collective and structural Sensemaking on the usefulness of business intelligence data. In: Hakikur Rahman, Anabela Mesquita, Isabel Ramos, Barbara Pernici (Editor): Proceedings of the 7th Mediterranean Conference on Information Systems, MCIS 2012, September 8-10, Guimarães, Portugal. Berlin, New York 2012, pp. 1-10

Howson (2008)
Cindi Howson: Successful business intelligence: Secrets to making BI a killer app. New York 2008

Hwang, Xu (2008)
Mark I. Hwang, Hongjiang Xu: A structural model of data warehousing success. In: The Journal of Computer Information Systems. Nr. 1, Vol. 49, 2008, pp. 48-56

Işık, Jones, Sidorova (2013)
Öykü Işık, Mary C. Jones, Anna Sidorova: Business intelligence success: The roles of BI capabilities and decision environments. In: Information & Management. Nr. 1, Vol. 50, 2013, pp. 13-23

Knight (2011)
Shirlee-ann Knight: The combined conceptual life-cycle model of information quality: part 1, an investigative framework. In: International journal of information quality. Nr. 3, Vol. 2, 2011, pp. 205-230

Malladi, Krishnan (2013)
Suresh Malladi, M. Krishnan: Determinants of usage variations of business intelligence & analytics in organizations - An empirical analysis. In: Richard Baskerville, Michael Chau (Editor): Proceedings of the 43th International Conference on Information Systems, ICIS 2013, December 15-18, Milan, Italy. 2013, pp. 1-22

Marchand, Peppard (2013)
Donald A. Marchand, Joe Peppard: Why IT fumbles analytics. In: Harvard Business Review. Nr. 1, Vol. 91, 2013, pp. 104-112

Missi, Alshawi, Fitzgerald (2004)
Farouk Missi, Sarmand Alshawi, Guy Fitzgerald: Towards a framework for realizing healthcare management benefits through the integration of patient's information. In: Timo Leino, Timo Saarinen, Stefan Klein (Editor): Proceedings of the 13th European Conference on Information Systems, ECIS 2004, June 14-16, Turku, Finland. 2004, pp. 1-14

Mohanty (2008)
Soumendra Mohanty: Measuring the value of intelligence in business intelligence. In: DM Review. Nr. 12, Vol. 18, 2008, pp. 20-23

Olson (2003)
Jack Olson: Data quality: The accuracy dimension. San Francisco, CA 2003

Olszak, Ziemba (2007)
Celina M. Olszak, Ewa Ziemba: Approach to building and implementing business intelligence systems. In: Interdisciplinary Journal of Information, Knowledge & Management. Vol. 2, 2007, pp. 135-148

Orr (1998)
Ken Orr: Data quality and systems theory. In: Communications of the ACM. Nr. 2, Vol. 41, 1998, pp. 66-71

Popovič et al. (2012)
Aleš Popovič, Ray Hackney, Pedro S. Coelho, Jurij Jaklič: Towards business intelligence systems success: Effects of maturity and culture on analytical decision making. In: Decision Support Systems. Nr. 1, Vol. 54, 2012, pp. 729-739

Pula, Stone, Foss (2003)
Edith N. Pula, Merlin Stone, Bryan Foss: Customer data management in practice: an insurance case study. In: The Journal of Database Marketing. Nr. 4, Vol. 10, 2003, pp. 327-341

Redman (1998)
Thomas C. Redman: The impact of poor data quality on the typical enterprise. In: Communications of the ACM. Nr. 2, Vol. 41, 1998, pp. 79-82

Sammon, Finnegan (2000)
David Sammon, Pat Finnegan: The ten commandments of data warehousing. In: SIGMIS Database. Nr. 4, Vol. 31, 2000, pp. 82-91

Solomon (2005)
Martin D. Solomon: Ensuring a successful data warehouse initiative. In: Information Systems Management. Nr. 1, Vol. 22, 2005, pp. 26-36

Tamm, Seddon, Shanks (2013)
Toomas Tamm, Peter Seddon, Graeme Shanks: Pathways to alue from business analytics. In: Richard Baskerville, Michael Chau (Editor): Proceedings of the 43th International Conference on Information Systems, ICIS 2013, December 15-18, Milan, Italy. 2013, pp. 1-16

Tayi, Ballou (1998)
Giri K. Tayi, Donald P. Ballou: Examining data quality. In: Communications of the ACM. Nr. 2, Vol. 41, 1998, pp. 54-57

Turner et al. (2014)
Vernon Turner, David Reinsel, John Gantz, Stephen Minton: The digital universe of opportunities: Rich data and the increasing value of the internet of things. April 2014

vom Brocke et al. (2009)
Jan vom Brocke, Alexander Simons, Bjoern Niehaves, Kai Riemer, Ralf Plattfaut, Anne Cleven: Reconstructing the giant: On the importance of rigour in documenting the literature search process. In: Susan Newell, Edgar A. Whitley, Nancy Pouloudi, Jonathan Wareham, Lars Mathiassen (Editor): 17th European Conference on Information Systems, ECIS 2009, June 8-10, Verona, Italy. Verona 2009, pp. 2206-2217

Wang, Strong (1996)
Richard Y. Wang, Diane M. Strong: Beyond accuracy: What data quality means to data consumers. In: Journal of Management Information Systems. Nr. 4, Vol. 12, 1996, pp. 5-33

Watson (2009)
Hugh J. Watson: Tutorial: Business intelligence - Past, present, and future. In: Communications of the Association for Information Systems. Vol. 25, 2009, pp. 487-510

Watson, Fuller, Ariyachandra (2004)
Hugh J. Watson, Celia Fuller, Thilini Ariyachandra: Data warehouse governance: best practices at Blue Cross and Blue Shield of North Carolina. In: Decision Support Systems. Nr. 3, Vol. 38, 2004, pp. 435-450

Watson, Wixom (2007)
H.J Watson, Barbara H. Wixom: The Current state of business intelligence. In: Computer. Nr. 9, Vol. 40, 2007, pp. 96-99

Wieder, Ossimitz, Chamoni (2012)
Bernhard Wieder, Maria-Luise Ossimitz, Peter Chamoni: The impact of business intelligence tools on performance: A user satisfaction paradox? In: International Journal of Economic Sciences & Applied Research. Nr. 3, Vol. 5, 2012, pp. 7-32

Wixom, Watson (2001)
Barbara H. Wixom, Hugh J. Watson: An empirical investigation of the factors affecting data warehousing success. In: Management Information Systems Quarterly. Nr. 1, Vol. 25, 2001, pp. 17-32

Yeoh, Koronios (2010)
William Yeoh, Andy Koronios: Critical success factors for business intelligence systems. In: Journal of Computer Information Systems. Nr. 3, Vol. 50, 2010, pp. 23-32

Yeoh, Wang, Verbitskiy (2012)
William Yeoh, Te-Wei Wang, Yuriy Verbitskiy: Describing data quality problem through a metadata framework. In: Association for Information Systems (Editor): Proceedings of the 18th Americas Conference on Information Systems, AMCIS 2012, August 9-11, Seattle, Washington. Red Hook, NY 2012, pp. 1-10

Appendix

No.	Article	Identified through step 4	Identified through step 5	Respective reference
1	Breur (2009a)	x		
2	Breur (2009b)	x		
3	Daniel (2008)		x	16
4	Dawson & Van Belle (2013)	x		
5	Hawking & Sellitto (2010)		x	4
6	Hwang & Xu (2008)		x	13
7	Işık et al. (2013)	x		
8	Malladi & Krishnan (2013)	x		
9	Missi et al. (2004)	x		
10	Olszak & Ziemba (2012)		x	4
11	Pula et al. (2003)	x		
12	Tamm et al. (2013)	x		
13	Wieder et al. (2012)	x		
14	Wixom & Watson (2001)		x	4,7, 8,12, 13
15	Yeoh & Koronios (2010)		x	4
16	Yeoh et al. (2012)	x		

Tab. 0-1: List of primary studies