



Decision making in the context of business intelligence and data quality

L. Marshall

Cape Peninsula University of Technology
Cape Town, South Africa

[R. de la Harpe](#)

Cape Peninsula University of Technology
Cape Town, South Africa
delaharper@cput.ac.za

Making decisions in a business intelligence (BI) environment can become extremely challenging and sometimes even impossible if the data on which the decisions are based are of poor quality. It is only possible to utilise data effectively when it is accurate, up-to-date, complete and available when needed. The BI decision makers and users are in the best position to determine the quality of the data available to them. It is important to ask the right questions of them; therefore the issues of information quality in the BI environment were established through a literature study. Information-related problems may cause supplier relationships to deteriorate, reduce internal productivity and the business' confidence in IT. Ultimately it can have implications for an organisation's ability to perform and remain competitive. The purpose of this article is aimed at identifying the underlying factors that prevent information from being easily and effectively utilised and understanding how these factors can influence the decision-making process, particularly within a BI environment. An exploratory investigation was conducted at a large retail organisation in South Africa to collect empirical data from BI users through unstructured interviews. Some of the main findings indicate specific causes that impact the decisions of BI users, including accuracy, inconsistency, understandability and availability of information. Key performance measures that are directly impacted by the quality of data on decision-making include waste, availability, sales and supplier fulfilment. The time spent on investigating and resolving data quality issues has a major impact on productivity. The importance of documentation was highlighted as an important issue that requires further investigation. The initial results indicate the value of research to investigate information quality in a BI environment.

Key words: Decision making, business intelligence, information quality, data quality

Received 9 February 2009; accepted 15 May 2009

Contents

1. [Introduction](#)
 - 1.1 [Research problem](#)
 - 1.2 [Research objectives and research methodology](#)
2. [Business intelligence \(BI\)](#)
3. [Information quality](#)
 - 3.1 [Information quality definition](#)
 - 3.2 [Information quality assessment - analysis](#)
 - 3.3 [Information quality assessment - findings](#)
4. [Practical analysis based on research findings](#)
 - 4.1 [BI department introduction](#)
 - 4.2 [Identified data quality issues](#)
 - 4.3 [Identified data quality impacts](#)
 - 4.4 [Identified data quality impacts \(from a business perspective\)](#)
 - 4.5 [Role of documentation – further analysis](#)
5. [Findings based on practical observations and investigations](#)
6. [Conclusion](#)
7. [References](#)

1 Introduction

Information that can be easily utilised must be available in order to make intelligent and effective business decisions based on facts. If this is not the case, decision-making processes can become extremely challenging and, in some cases, impossible. This, in turn, is likely to create serious implications for an organisation's well-being, particularly if the outcome of the required decision(s) relates to strategic objectives that create opportunities for increased competitive advantage, and thus influence an organisation's ability to prosper.

In this article, the research problem, objectives and methodology are discussed first, followed by a summary of the main findings from the literature review. The literature was reviewed to establish the information requirements for business intelligence (BI). A brief discussion on information quality and data roles follows. Three data quality frameworks were considered for this part and a consolidated view forms the proposed framework used for this study. Although data and information are two different concepts, they are used interchangeably for the purpose of this article (Pipino, Yang, Wang and Wang 2002). The findings of the empirical part of this study, based on the data collected from a large retail organisation in South Africa, are discussed next. The article is concluded with a discussion of the findings and the conclusions drawn from the interpretation.

1.1 Research problem

The research problem for this research can be summarised with the following problem statement: Decision making within a business intelligence (BI) environment is often considered challenging, owing to information that is difficult to utilise. It is not clear how the quality of information impacts on the decision-making process.

1.2 Research objectives and research methodology

Objectives of this research were aimed at: a) creating an awareness of underlying factors that prevent information from being easily utilised; and b) investigating how these factors influence the decision-making process, particularly within a BI environment.

The research explored key concepts and frameworks that had been identified in existing literature that focused on information quality and the assessment thereof. This was followed by a practical investigation that used a qualitative approach in order to understand the impacts and influences that information quality had on decision-making processes. Unstructured interviews were conducted with the main team members of the BI Department in the head office of a large retail organisation in South Africa. The collected data were analysed using the proposed framework as a theoretical lens. The findings were further interpreted to establish the factors that influenced the decision-making process. In this article, the results are presented as summaries of the main findings and conclusions that have been drawn from the interpretations.

[top](#)

2 Business intelligence (BI)

Data are important organisational resources and individuals within the organisation use them for different purposes to support their organisational activities and decisions. Schlögl (2005), in his study to make more sense of the term information management, refers to the general purpose of information management to make the right information available at the right time and at the right place. Information not available to users adds little value to a decision-making process. However, when the required information can be obtained from different sources, integrated to provide a more complete view and presented in a way that promotes analysis, it can be used as a powerful tool to answer certain 'what', 'why', 'how' and 'what if' questions. The purpose of a BI implementation is to provide business with this functionality and a platform upon which information can easily be collected, analysed and converted into knowledge so that important decisions regarding operational, tactical or strategic actions or issues can be made (Cody, Kreulen, Krishna and Spangler 2002; Friedman and Strange 2004; Golfarelli, Rizzi and Cella 2004; Lonnqvist and Pirttimaki 2006; Strong, Lee and Wang 1997).

Golfarelli *et al.* (2004) express BI as a process through which data are 'converted into information and then into knowledge' via the use of various technologies, which include a data warehouse infrastructure, as well as analytical and reporting tools that provide users with the means to gather and analyse data in order to gain essential knowledge that will improve decision making.

However, BI is only as good as the underlying information it presents. Therefore, in order for BI to be useful

and valuable to a decision making process, and to the well-being of an organisation, information must reflect a high degree of quality (Friedman and Strange 2004; Herring 1992; Shankaranarayanan, Watts and Even 2006).

[top](#)

3 Information quality

For the purposes and focus of this article, the terms 'data quality' and 'information quality' are used interchangeably. Strong *et al.* (1997) do, however, provide a clear distinction between the two: data are considered to be raw, unprocessed facts, which are then organised, given context and transformed into information that can be utilised and analysed by a data consumer and converted into knowledge. This process of transformation is referred to as the 'data manufacturing system'.

A 'data consumer' refers to a person in a role within this system and, in the context of this article, can be seen as a decision maker. Two other roles include those of the data producer and data custodian (Parker, Stofberg, de la Harpe, Wills and Venter 2006; Strong *et al.* 1997):

- Data consumer: Individuals who use data
- Data producer: Individuals/sources who produce data
- Data custodian: Individuals who are generally responsible for the data and provide necessary resources to manage (process and store) the data.

3.1 Information quality definition

A widely and commonly accepted definition of information quality within existing research literature and industry is as follows: information that is 'fit for use' and satisfies the purpose for which it is intended (Lui and Chi 2002; Strong *et al.* 1997). In the context of BI, this means that information should reflect certain characteristics that the data consumer identifies as important in order to be regarded as useful to a decision making process.

This definition also suggests that quality should be assessed from a data consumer perspective and that there is more to information quality than mere correctness and accuracy (Lui and Chi 2002; Parker *et al.* 2006; Strong *et al.* 1997).

3.2 Information quality assessment - analysis

What underlying factors prevent information from being utilised easily?

Poor quality of production data that reside in organisational databases can create false perceptions that can impact a decision maker's ability to obtain insight into the business and make accurate and effective business decisions (Huang, Lee and Wang 1999; Redman 1995). Furthermore, it is far better to know that there are data quality issues than to be unaware of them – if decision makers know that there are quantifiable data quality issues, they will be more inclined to be cautious during a decision-making process (Snow 2007). Understanding the key issues or characteristics that render information invaluable is therefore important.

Strong *et al.* (1997), Lui and Chi (2002) and Helfert, Zellner and Sousa (2002) have each established a framework that provides a good starting point for this assessment through identification of certain quality characteristics. The following section discusses each framework separately and concludes with a consolidated view of their differences and similarities.

3.2.1 Four categories by Strong *et al.* (1997)

Strong *et al.* (1997) identify four categories of data quality dimensions, namely intrinsic data quality, accessibility data quality, contextual data quality and representation data quality, which are outlined below.

- *Intrinsic data quality* (accuracy, objectivity, believability and reputation): *When discrepancies across disparate sources of data exist, believability concerns are raised regarding credibility and accuracy of the underlying data. As time goes by and these concerns become common knowledge, a poor reputation of the data and data source develops, resulting in data not being used, since data consumers are unlikely to use data that they consider untrustworthy or which does not fulfil their needs.*

Data that are produced or derived as a result of human interpretation are often considered subjective and potentially biased, which can also create believability concerns.

- *Accessibility data quality* (accessibility and security issues): Human and technical aspects such as a lack of certain skills and expertise, as well as insufficient computing resources (lack of physical devices, computing power, network space and memory), can prevent access to information that is stored in central databases or shared repositories. Also, it often takes time to acquire these resources, which may result in required information not being available when needed

In addition to the above accessibility issues, security constraints can also cause information to become inaccessible as it generally takes time to obtain the necessary permission(s) in order to gain access. However, in several cases, these constraints are necessary or compulsory as they are often enforced by organisational policies or government acts regarding privacy and confidentiality of information.

- *Contextual data quality* (relevancy, timeliness, completeness and amount of data): Large data volumes can affect availability of information owing to the time that it takes to process. If information is not available when it is needed, it will not fulfil its purpose and not be considered useful to a decision-making process.

If data is incomplete or missing as a result of integration, operational, scheduling, and/or internal aggregation errors, it is unlikely that the resulting output will be of high relevance. This is also true if the existing information base is not sufficient to cater for new reporting or decision-making needs.

- *Representation data quality* (interpretability, ease of understanding and consistent representation): Challenges experienced in summarising, integrating and analysing inconsistently represented data make information inaccessible for use owing to the minimum amount of value that it will have for a consumer's decision-making process.

Information is also considered inaccessible if it is presented in a way that makes it too difficult or complex to understand or interpret. Therefore, it is important that information is represented in a way that makes it intuitive to understand and takes language, symbols and clear definitions into account (Pipino *et al.* 2002).

3.2.2 Data quality in the context of data evolution by Lui and Chi (2002)

Data flow through a life cycle of stages, each of which generally involves some form of transformation in order to satisfy its intended use (Lui and Chi 2002; Strong *et al.*, 1997). Reviewing the quality of data in the context of this life cycle is important as each stage of transformation or 'evolution' can introduce different types of quality issues that can affect the usefulness of data in different ways.

- *Collection quality*: The collection stage relates to processes that obtain and/or produce data and includes characteristics such as: a) bias and ambiguity involved during observation; b) poor accuracy of data; c) reliability of the data collector or producer; and d) completeness in terms of sufficiency for use.
- *Organisation quality*: Organisation quality relates to how data are stored. It is influenced by characteristics that pertain to collection quality, in addition to factors such as the lack of consistency of data or information across multiple data repositories (likely owing to a lack of automated processes that update equivalent data in multiple places), timeliness of data retrieval and ease of navigating the information.
- *Presentation quality*: This includes characteristics that relate to collection and organisation quality, as well as those which involve obtaining and/or producing information. Presentation quality primarily relates to consistent data semantics and format (the same data should be defined with the same meaning and format) and emphasises that data should be clear and easy to interpret and should reflect 'neutrality' (bias in terms of which data should be presented and which data should be hidden from a data consumer).
- *Application quality*: Quality of application or utilisation relates to all those characteristics that are noted under presentation quality, and thus organisation and collection quality, including those that impact effective use of information, for example:
 - Availability of information
 - Accessibility as a result of security and privacy agreements
 - Relevancy in terms of data volumes and/or whether the presented information is useful
 - The extent to which data or information can be easily analysed and manipulated for its intended purpose.

3.2.3 Semiotic assessment by Helfert *et al.* (2002)

Helfert *et al.* (2002) suggest an approach that satisfies specific requirements by relating relevant quality

characteristics, which are identified by the user, to various semiotic levels.

- *Syntax level*: Syntax level deals with issues that relate to how data are represented, formatted and transported between a source and destination system. It includes characteristics such as consistent representation, security and data accessibility.
- *Semantic level*: This level deals with the semantics of data such as content and meaning. Characteristics that are important at this level include aspects such as interpretability, data correctness or accuracy, ease of understanding, consistent values across disparate data sources and the objective nature of the data (and, therefore, includes believability and reliability).
- *Pragmatic level*: Pragmatic level relates to how information will be utilised and includes characteristics such as relevance, completeness and timeliness.

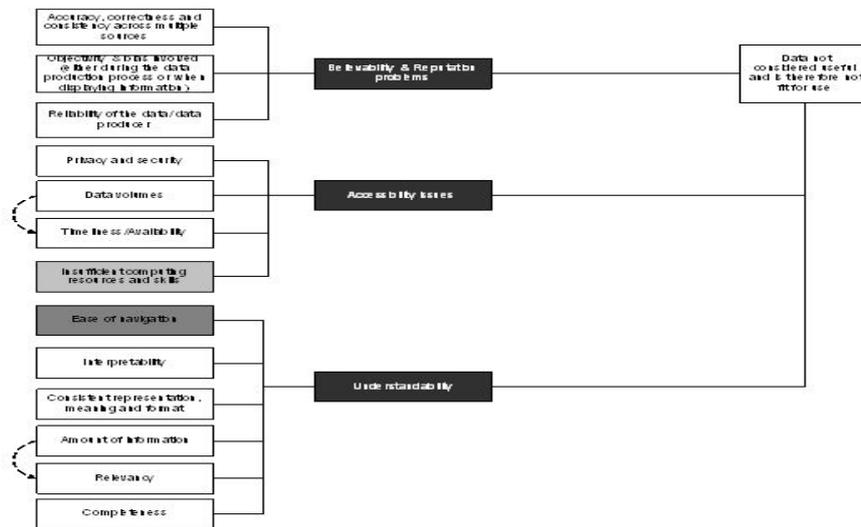
3.3 Information quality assessment - findings

Although the above frameworks differ in the approach used to categorise various quality characteristics, similarities exist in terms of the characteristics that are produced. These similarities and differences are summarised in the figures that follow.

Framework	Description of approach used
Four categories by Strong <i>et al.</i> (1997)	No underlying theory or practical explanation for organising these characteristics into the four categories. The defined dimensions are based on a qualitative investigation of three companies that have introduced data quality projects.
Data quality in the context of data evolution by Lui <i>et al.</i> (2002)	The identified quality characteristics are mapped to each of the data life cycle stages as it is suggested that each stage introduces additional issues.
Semiotic assessment by Helfert <i>et al.</i> (2002)	The characteristics, of which the majority are consistent with those that are highlighted by Strong <i>et al.</i> (1997), are related to various semiotic levels. Helfert <i>et al.</i> (2002) also indicate that the priorities and importance of these characteristics should be defined by the data consumers in terms of their needs

The three frameworks discussed in the previous sub-sections were combined into the framework that was used by this study (Helfert *et al.* 2002; Lui and Chi 2002; Strong *et al.* 1997). All the unshaded blocks in Figure 1 represent the characteristics common to all the referred frameworks. The shaded block (ease of navigation) only applies to the framework of Lui and Chi (2002). The lighter shaded block (insufficient computing resources and skills) only applies to the framework of Strong *et al.* (1997). The arrows indicate the influence that a characteristic can have on another characteristic, for example data volume on timeliness or availability and amount of information on relevancy. The blocks representing the characteristics are grouped into three group characteristics, namely, believability and reputation problems, accessibility issues and understandability. All three contribute towards data considered as not useful and therefore not fit for purpose (refer to the definition adopted for this study for data quality). The proposed framework is presented in Figure 1 below.

Figure 1 Proposed framework incorporating similarities in terms of the identified quality characteristics



[top](#)

4 Practical analysis based on research findings

How do underlying factors that prevent the utilisation of information influence a decision-making process, particularly within a BI environment? To practically examine and understand how the quality of information influences a decision-making process, a qualitative investigation was conducted within the BI department of a well-known and respected retail organisation that offers a selected range of clothing, food, beauty, digital, homeware and financial products and services.

The information gathering technique that was used in this investigation involved both formal and informal discussions with individuals from various areas within the BI department and business, using the similarities identified in the previous section as a foundation to guide the interview process and obtain an understanding of associated impacts.

4.1 BI department introduction

The BI department comprised five dedicated teams, which represented business analysis, meta data, datawarehouse (DWH), online application processing (OLAP) and information delivery. These teams worked together to provide their customers (the business) with information that could be easily utilised to make effective business decisions.

4.1.1 Business analysis team

According to Respondents 1 and 8, the team of business analysts and business process analysts comprised representatives from each area of the business, namely foods, clothing/home/beauty/digital, stores, customer and finance.

The primary focus of this team was to provide support to business users and to interpret and translate reporting and decision-making needs of the business into a business requirements specification (BRS) that could be used by systems analysts, across design, reporting and OLAP teams, to create system specifications.

4.1.2 Meta data team

The meta data team comprised two functional areas, namely DWH design and data quality.

The DWH design group provided a link between business analysts, DWH developers and information delivery system analysts. Their focus was on analysis, design and documentation of logical data structures and meta data-related information [tables and field characteristics, field definitions and meanings, specific extract, transform or transfer and load (ETL) requirements and source system details], which fulfil requirements that were specified by business analysts and support information delivery needs.

Experience gained from observing the number of data quality-related production issues that had (and continued to have) an impact on operational activities and the broader business had created a sense of recognition for the importance of data quality within the BI department. This could be seen through the establishment of a dedicated data quality group within the meta data team. The role of the data quality group

was to manage, resolve and prevent production issues timeously in order to maintain and improve the quality of data that was used for decision making and analysis (Respondent 5).

4.1.3 Informix DWH team

This DWH team comprised a group of Informix developers that were responsible for managing and supporting the DWH infrastructure in order to satisfy the business and data quality requirements; improving DWH performance in order to meet defined service level agreements (SLAs); liaising with database administrators in terms of space, backups and security; creating and implementing change requests; and providing technical support to the meta data team (Respondents 6 and 9).

4.1.4 OLAP team

The OLAP team was responsible for developing, managing and maintaining BI's cube infrastructure, which provided a powerful base for the rapid analysis of large volumes of data. There were 14 cubes in production (Respondent 3).

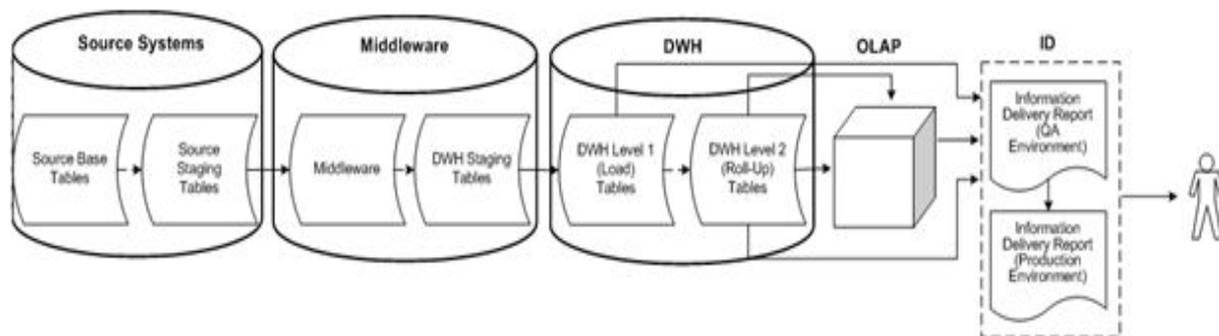
4.1.5 Cognos reporting/information delivery team

This concerns management, systems analysis, presentation and support of information by using modern enterprise business intelligence technology, keeping the business up to date and better informed (Respondent 2).

4.2 Identified data quality issues

The data flow of the retail organisation is depicted by in Figure 2 below. A number of links existed within the ETL flow of data from source, via middleware through to the BI environment, as well as internally within BI in order to reach the business user who had a reporting or decision-making need. Having a view of this data flow is important as each link or entity had its own processes and dependencies that created an additional point of possible failure, which could pose a greater threat to the quality of data.

Figure 2 Data flow between source and Information delivery report



Following discussions with various individuals within respective BI teams, a number of data quality issues and their associated impacts were identified, all of which affected utilisation of information. These issues, their potential causes and impacts are summarised in Table 2.

List of identified data quality issues	Description	Cause	MD	DWH	BA	ID
Information accuracy						
Variations between DWH and operational data sources	Business users who had access to DWH and operational data raised concerns regarding accuracy of the data reflected in DWH when discrepancies existed between the two.	A majority of these discrepancies were caused by incomplete documentation and technical issues such as	X	X		

		ETL, aggregation and batch failures or space issues that occurred between A to E in Figure 2 above.				
Discrepancies of same data across different tables within DWH	Alignment exercises showed a history of discrepancies between certain 'level 1' and 'level 2' tables - DWH consisted of two levels of tables, namely those that were referred to as 'level 1' tables and were directly loaded from source (operational system); and those, which were internally aggregated into 'level 2' tables.	<p>1) Lack of process dependencies and late delivery of source data: Time-based rollup processes that began before source data had loaded often resulted in data being excluded at higher levels.</p> <p>2) Transaction updates: Program logic that only executed if the transaction date was within a certain date range. This is risky as late transactions were often received from the source.</p> <p>3) Resend exercises that were not carried through to higher levels: A data refresh, which was usually necessary in the event of an extensive data issue, required a reprocessing of data at higher levels.</p>	X	X	X	X
Under- or overstated data values	Users had also logged incidents relating to data values that either appeared under- or overstated.	<p>1) Duplication at source or within DWH.</p> <p>2) Data that had been</p>	X		X	X

		incorrectly mapped or calculated. 3) A failure that occurred in the required ETL processes between the source and DWH or internally to DWH, for example, between the load and aggregation levels.				
Logic used in calculations to derive business measures was sometimes inaccurate or inconsistently applied	For example, a key measure called availability was calculated by dividing FD_NUM_AVAIL_DAYS by FD_NUM_CATLG_DAYS. A majority of the DWH programs reflected the correct logic, however, in one of the programs FD_NUM_CATLG_ADJ_AVAIL_DAYS was used instead of FD_NUM_CATLG_DAYS.	1) Program errors. 2) Incomplete BRS and/or meta data documentation in terms of field definitions or ETL specifications. 3) Pressurised environment of fast delivery. 4) Lack of thorough testing. 5) Inexperienced resources.	X		X	X
Inconsistencies						
Irregular or inconsistent data formats	Examples of inconsistencies within the BI/DWH included: 1) DWH stored a field called LOCATION_NAME as Varchar 20 where the source defined it as Varchar 40. Result: Truncation and potential duplicates. 2) DWH usually stored date fields as dd/mm/yyyy; however there had been instances where mm/dd/yyyy was used. Result: Data appeared missing.	1) Lack of consistent and up-to-date meta data. 2) Inexperienced developers and systems analysts. 3) Poor analysis. 4) Technical constraints.	X	X		
Understandability						
Understandability	Incorrectly defined or understood requirements, which resulted in a design or data fix that did not satisfy the business	Owing to the urgency of certain design	X	X	X	

	requirement.	requirements and data fixes, thorough testing, impact analysis and requirement analysis were often neglected, which resulted in greater issues after the requested changes had been implemented into the production environment.				
Information availability						
Information not available when needed	Information required immediately or before a specific time of day was not always available for use when needed.	<p>1) Delay or failure of ETL processes between source, middleware and DWH.</p> <p>2) Access issues (not everyone had been granted access rights to DWH data).</p> <p>3) Insufficient history as a result of inaccurate or incomplete conversion specifications.</p>	X	X	X	X
Existing information did not fulfil new requirements	Current infrastructure did not always fulfil new information requirements, which resulted in new interfaces that should be developed.	The business environment was changing all the time, requiring new decisions in reaction to certain internal or external events. There were also instances where the business requirement had not been properly considered, defined incorrectly or were largely	X		X	

		incomplete.				
--	--	-------------	--	--	--	--

4.3 Identified data quality impacts (from a BI perspective)

Individuals within the BI department (ranging from technical specialists, systems analysts and managers to business representatives) identified a number of impacts that specifically related to the issue categories that are described above, namely information accuracy, consistency, understandability and availability. Their feedback was analysed, consolidated and summarised in Table 3 below.

Issue group	Impact(s)	Impact description
Information accuracy	1) Believability and trust issues. 2) Business confidence in BI, which resulted in data not being used. 3) Confusion. 4) Ineffective decisions were being made. 5) Delays in decision-making process.	Users tended to believe the data that resided in the operational (source) system as it represented the place of data origin. When users doubted the trustworthiness of the data reflected in the BI reports and DWH, they would likely seek the truth at source, which would not only affect business' confidence in BI, but also made a decision-making process challenging.
Inconsistencies	Data duplication and perception that data were missing.	In terms of the first example (refer to Table 2 above), duplication might occur as a result of truncation if the data exceeded 20 characters. This scenario could have a major impact on availability of information as Respondent 10 (2008) indicated that a duplication of this nature could cause the OLAP cubes to fail. In addition, inconsistent date formats, as per the second example, could create issues for ad hoc users if they had become accustomed to a certain format – when ad hoc users tried to access their data using data in the irregular format, no values were returned, which created a perception that data were missing.
Understandability	Ineffective use of time and money.	According to Respondent 9 (2008), the manager of the DWH team, this was one the most common issues that impacted delivery within the BI department, and consumed a majority of resources in terms of staff and time or rework.
Information availability	Information not considered useable or useful and BI's reputation in terms of delivery.	Information that was not available when the data consumer required it, could not be considered valuable or useful. Although some of the issues were triggered by other systems, it was the BI department's reputation that was at stake.

In addition to the impacts listed above, it was also noted that the time spent on investigating and resolving data quality issues within and across the BI department, had a major impact on productivity as well as cost. The majority of the respondents within the business analysis, DWH and meta data teams estimated that they collectively spent more than 80% of their daily activities on investigating and resolving data quality issues, leaving only a small amount of time to focus on other important responsibilities such as performance tuning, enhancements, business support and requirement specifications (Respondents 1, 3, 5, 6, 8 and 9).

Also, owing to the number of data quality-related issues that were logged on a daily basis, more time was spent on reactively resolving issues, as opposed to focusing on more efficient processes or mechanisms that would prevent or at least reduce the impact of these data issues.

4.4 Identified data quality impacts (from a business perspective)

Following consultation with forecast analysts and BI data consumers of the business, it was evident that waste, availability, sales and supplier fulfilments were key measures that were directly impacted by quality of data on decision making.

- *Waste, availability and sales:* To get products into the stores and onto the shelves for sale to consumers, stock has to be ordered from suppliers. These orders were initially created by an order management system, which calculated recommended order quantities based on balance on hand (BOH), sales, stock on hand (SOH) and forecast data (Respondents 4 and 7). If this data reflected a poor degree of quality, and the quality of the data was unknown, too much stock could have been ordered, which was likely to result in waste (perishable products, in particular). Likewise, too little stock could have been ordered, resulting in lost sales and opportunity costs (if consumers continued to find that a certain product was never available, they would likely have taken their business elsewhere)
- Supplier fulfilment refers to the extent to which a supplier satisfies the order agreement with the retail organisation in terms of delivery to the distribution centre or warehouse. This measure is important as it determines: a) the rate at which a supplier will be penalised (for under delivery); b) rebates (for continuous delivery within the agreed tolerance); c) whether or not the contract with the supplier will be renewed or suspended; and d) what the supplier is paid. If the data, which specifies what the supplier delivered are inaccurate or inaccessible, then the retail organisation could penalise the supplier unnecessarily or pay him or her short, which could affect their relationship negatively (Respondents 4, 7 and 8). Doney and Cannon (1997) and Lee, Ha and Kim (2000) emphasize the importance of building good relationships with suppliers as they can influence competitiveness of the entire supply chain, as well as reputation of the brand.

4.5 Role of documentation – further analysis

A key factor that kept emerging from the practical investigation was the importance of documentation. Several activities required certain specifications upon which requirements were captured and data were obtained from the respective source systems. Respondent 5 indicated that he had concerns about the state of this documentation and highlighted that [interface](#) documentation, namely source specification documents ([SSDs](#)) and application integration specifications ([AISs](#)) were largely inconsistent across the various source and destination systems with respect to content and format.

This concern prompted further analysis, which was pursued by means of 68 interface audits. The results of these audits showed that 79% of information included in these documents were outdated and contained a number of discrepancies, including missing fields, incorrect sequence of fields, incorrect definitions of data types, field miss-mappings, incorrect formatting rules and incorrect ETL logic.

Respondents 5 and 9 agreed that a majority of these discrepancies could have leaked easily into the production environment and created a number of data quality issues, if not picked up during integration testing or this audit. In addition, outdated or incomplete BRS and meta data ETL requirements could also lead to data or information that was largely inconsistent and inaccurate (inconsistent if the logic that was used to derive a measure in one table differed to the logic that was used to derive the same measure in another table, and inaccurate if requirements were incomplete or incorrectly defined).

[top](#)

5 Findings based on practical observations and investigations

For information to be considered useful and to add value to a decision making process, it should meet certain quality requirements.

Accuracy, consistency, understandability and availability are key factors that affect the quality of information and create barriers during the utilisation process (Table 3). This can have severe implications for an organisation when outcomes of certain decisions affect profits, expenses, reputation, partner or supplier relationships and customer loyalty. In addition, these barriers can also influence the bottom line indirectly, as they tend to decrease productivity, occupy resource capacity and take a large amount of time and effort to investigate and resolve underlying causes.

Evidence also shows that documentation plays an important role in decreasing the number of defects or issues that begin to manifest after new interfaces are created and/or changed. It is therefore necessary to ensure that documentation is created during the development of the BI solution and kept up to date during the operational stage when the BI solution is used. When this is done, experiences related to quality problems can be recorded and shared.

Thorough integration and user acceptance testing is another key aspect that can improve the quality of information upfront in order to ensure that the required information satisfies the intended need. As a result, more time can be spent on establishing more proactive measures that detect quality issues before they are discovered by data consumers. This will also increase the business' confidence in BI and resolve ever-present trust and believability issues. Testing must be an integral part of the development of the BI solution, and a technical and business focused test plan to sufficiently provide for data quality testing is necessary.

The identified key factors, need for documentation and proper testing are important points derived from the findings that, once addressed, should result in better quality data for the BI solution. Only then will it possible for organisations to utilise BI effectively and gain competitive advantage from its information.

6 Conclusion

[top](#)

A decision-making process is an integral part of several key business activities, including those which determine future strategies and goals. Information upon which these decisions are based is, therefore, an essential asset that can influence an organisation's well-being.

BI is a powerful tool that aids decision-making processes by providing a means by which information can easily and quickly be analysed and converted into knowledge. However, as evidence and research have shown, information does not always reflect a high degree of quality or satisfy the intended need, which creates challenges during the utilisation process and delays in decision making. Furthermore, consequences of ineffective decisions and operational inefficiencies, which are created as a result of poor quality information, affect an organisation's bottom line.

Organisations, their leaders and data consumers should be aware of these issues and understand their associated impacts. Information should be checked and assessed on a regular basis in order to ensure that quality requirements are met and that information continues to meet the intended need. Various data quality frameworks and findings that are presented in this article provide a good starting point for this assessment.

In addition, top business and IT management commitment to information quality is also essential. Standard quality requirements should be established to enforce adherence and awareness, particularly when it comes to testing and signing off requirement specifications and interface documentation.

Although data quality improvement techniques, testing approaches and document data management and control mechanisms are important to data quality, a detailed analysis of each was not part of the scope of this research. It is, therefore, suggested that further research should focus on these topics in relation to improving and maintaining data quality.

The nature of this study was exploratory and the results already indicate that more detail and comprehensive research is required. The findings confirm the problem statement and indicate that decision making within a BI environment is indeed challenging and information difficult to utilise. The proposed combined framework represents the underlying factors that prevent information from being easily utilised, based on the literature reviewed. The findings derived from the empirical data also offer an answer to the research question, 'how do underlying factors that prevent utilisation of information influence a decision-making process, particularly in a BI environment?'

Although the findings already indicate the problems experienced by the BI team of one of the departments of the retail organisation, more research is required to establish the extent of these experiences. Further research is also required to determine if the same results apply not only to the rest of the organisation, but also whether these are specific to the retail industry. A better understanding of the quality of the information on which decisions are based is required to fine-tune further research. The proposed framework also needs to be tested in practice to determine its usefulness by BI users. The proposed combined information quality framework (Figure 1) and data flow diagram (Figure 2) were found to be a useful starting point and provided insights in the problems experienced by BI users. It may also be necessary to include the impact of BI technologies, tools, etc. in further studies to establish how these influence the users' perceptions of information quality.

[top](#)

7 References

Cody, W.F., Kreulen, J.T., Krishna, V. and Spangler, W.S. 2002. The integration of business intelligence and knowledge management. *IBM Systems Journal* 41(4):697-713.

- Doney, P.M. and Cannon, J.P. 1997. An examination of the nature of trust in buyer-seller relationships. *Journal of Marketing* 61(2):35-51.
- Friedman, T. and Strange, K.H. 2004. Architecture: the foundation of business intelligence. Gartner, Inc. [Online]. Available WWW: http://www.gartner.com/DisplayDocument?id=430431&ref=g_sitelink. (Accessed May 2009)
- Golfarelli, M., Rizzi, S. and Cella, I. 2004. Beyond data warehousing: what's next in business intelligence? Proceedings of the 7th International Workshop on Data Warehousing and OLAP (DOLAP 2004), Washington DC.
- Helfert, M., Zellner, G. and Sousa, C. 2002. Data quality problems and proactive data quality management in data warehouse systems. Proceedings of the Business Information Technology World Conference, 2–5 June 2002.
- Herring, J.P. 1992. The role of intelligence in formulating strategy. *The Journal of Business Strategy* 13 (5):54–60.
- Huang, K., Lee, Y. and Wang, R. 1999. *Quality information and knowledge*. Upper Saddle River, NJ: Prentice Hall PTR.
- Lee, E.K., Ha, S. and Kim, S.K. 2000. Management of innovation and technology. An effective supplier development methodology for enhancing supply chain performance, 2: 815-820.
- Lonnqvist, A. and Pirttimaki, V. 2006. The measurement of business intelligence. *Information Systems Management Journal* 23(1):32-40.
- Lui, L. and Chi, L.N. 2002. Evolutional data quality: a theory-specific view. Proceedings of the Seventh International Conference on Information Quality (ICQ-02), 298-298.
- Parker, M., Stofberg, C., De la Harpe, R., Wills, G. and Venter, I. 2006. Data quality: how the flow of data influences data quality in a small to medium medical practice. Proceedings of community informatics for developing countries: Understanding and organising for a participatory future information society, 31 August-2 September 2006, Cape Town, South Africa.
- Pipino, L.L., Yang, W., Wang, L. and Wang, R. 2002. Data quality assessment. *Communications of the ACM* 45(4): 211-218.
- Redman, T.C. 1995. Improve data quality for competitive advantage. *Sloan Management Review* 36(2):99-107.
- [Schlögl](#), C. 2005. Information and knowledge management: dimensions and approaches. *Information Research* 10(4).
- Shankaranarayanan, G., Watts, S. and Even, A. 2006. The role of process metadata and data quality perceptions in decision making: and empirical framework and investigation. *Journal of Information Technology Management* 17(1):50-67.
- Snow, A. 2007. A holistic approach to measuring information quality. *The Information and Data Quality Newsletter* 3(1):1-4.
- Strong, D.M., Lee, Y.W. and Wang, R.Y. 1997. Data quality in context. *Communication of the ACM* 40 (5):104–108.



ISSN 1560-683X

Published by [InterWord Communications](#) for Department of Information and Knowledge Management,
University of Johannesburg