The Relationship of the Quality Data Warehousing to Enhanced Perceived Net Profits and Decision Quality in the Enterprises

Chanin Amornbuth

Department of Management Studies, The University of Hertfordshire, UK

Abstract  Data warehousing is a result of business needs and technological advances. This study will investigate the extent to which DW quality relates to perceived net benefit and decision quality. This paper hopes to identify the strength of the relationships of the quality data warehouse to enhancing decision quality in the enterprise. The results showed that impact on decision quality eventually have some perceived net benefits from the use the warehouse. Moreover, it should add to the body knowledge on data warehouse.

Keywords  Data Warehousing, Quality Data, Net Benefit, Decision Quality

1. Introduction

Background

Information is pivotal in today’s business environment. Success is dependent largely on its decisive use. A lack of information is sure sign of failure. Gaining a competitive advantage as well as survival in many cases requires that companies accelerate their decision making process to respond quickly to change. One key to this accelerated decision making is having the right information, at the right time, and easily accessible.[1]

Problem Statement

Data warehousing is a result of business needs and technological advances. The business environment has become global, more competitive, complex, and volatile. In addition, Data warehousing has become an important technology for many organizations. Many organizations today depend on information from DW to make strategic decisions. A quality data warehouse can increases business profits by providing more reliable business information products that result in more usage, better decisions, and increased exploitation of business opportunities. Even though there are many success stories [2] a data warehouse project can be very expensive. A typical project cost over 1 million dollars in the first year alone [3]. There has been little academic research that systematically and rigorously investigates the quality of data warehousing. In addition, no empirical evidence is available to show the effects of data warehouse (DW) on decision quality[4].

Significance of the study

At the present time, forward thinking companies are beginning to use data warehousing to address enterprise level quality. In addition, data warehouse exists to serve the decision making needs of an enterprise. To achieve enterprise quality improvement, a quality data warehouse builds to serve the analytical needs of the enterprise. There is currently very little academic research on the use of quality data warehousing in the enterprises. This is the points to the significance of the study. Data warehouses are here to stay and they are beginning to play additional roles such as becoming part of firm’s infrastructure to support a customer focused strategy[5]. As a result, it should add to the body knowledge on data warehouse.

Conceptual or theoretical framework

The term, data quality, describes data that is fit for use by data consumers. In addition, data quality dimension defines as a set of data quality attributes that represent a single aspect or construct of data quality. Three approaches are used in the literature to study data quality: (1) intuitive, (2) theoretical, and (3) empirical approaches. The intuitive approach takes when the selection of data quality attributes for any particular study is based on the researchers’ experience or intuitive, implicit understanding about what attributes are important. Most data quality studies fall into this category. The cumulative result of these studies is a small set of data quality attributes that are commonly selected. The advantage of the intuitive approach is that each researcher can select the attributes most relevant to the particular goals of his or her
study [6].

A theoretical approach to data quality must also focus on how data may become deficient during the data process. Although theoretical approaches are often recommended, research offers very few examples. One such study uses an ontological approach in which attributes of data quality are derived on the basis of data deficiencies, which are defined as the inconsistencies between the view of a real world system that can be inferred from a representing information system and the view that can be obtained by directly observing the real world system [6]. The advantage of a theoretical approach is the potential of providing a comprehensive set of data quality attributes that are intrinsic to data product.

The empirical approach to data quality analyze information collected from data consumers to determine the characteristics they use to assess if data are fit for their intended use or application. The advantage of such an empirical approach is that it captures the voice of customers [6].

2. Review of the Literature

This part reviews previous research in the following order: (1) the data warehouse, (2) the quality of the data warehouse, (3) decision quality and (4) the measures of success in implementation research.

Inmon and Hackathorn[7], the father of DW, defined it as an integrated, time-variant, subject oriented, and nonvolatile collection of data in support of management’s decision making process and business intelligence. The data warehouse has resulted in many organizational benefits, including providing “a single version of truth” better data analysis and time savings for users, reductions in head count, facilitation of the development of new applications, better data, and support for customer-focused business strategies [8]. In addition, data warehouse implementation includes the conversion of data from numerous source systems into a common format. Since each data from the various departments is standardized, each department will produce results that are in line with all the other departments. As a result accurate data is available and accurate data is the basis for strong business decisions [9].

Data quality is frequently discussed in the data warehousing literature as well; providing high quality data to decision makers is the fundamental reason for a building a warehouse[10]. Data quality is a major concern for many operational systems as well as data warehouses. Validation of accuracy, timeliness, completeness, and consistency remain major problems for many organizations even internal information systems where users are trained and managed by the organization. These problems are multiplied in information systems that are exposed to customers, vendors, and other partners. The result can be a disaster for a data warehouse that depends on such systems for its content. Mechanisms for protecting a data warehouse from poor quality data are crucial [11].

Decision quality (DQ) defines as the extent to which the decision outcome (a choice) is optimal, both in an objective and subject sense, the best choice available and the best choice for the decision maker. When individual preferences are consider, researchers often replace decision quality with the most preferred (highest ranked) choice without articulating the function by which the user does the ranking [4].

Data warehouse can be used effectively in the management of quality. This technology helps in the improvement of quality management, enables users to effectively access the data Optimizing Data Warehouse that they need, and at the same time issue further analysis on them. It helps in saving time users and also provides reliable and accurate analytical results. Helps management to predict accurately the quality of the product and thus provides a better support for managerial decision making [12]. Moreover, in every system, accuracy of a data is very important for any decision to be reached, for example; if a company wants to take decision on the marketing of their product to a specific area, they can only deliberate and produce a good marketing strategy if they have the right data. For knowledge discovery of the data, its privacy and security must be assured [13].

To develop the research model, data warehousing, and success literature were reviewed to identify factors that potentially affect data warehousing success. A review of literature produced a consolidated list of eight warehousing success measures: (1) ease of use, (2) speed information retrieval, (3) more information, (4) better-quality information, (5) improved productivity, (6) better decisions, improved business process, and increased competitive position. The first two measures are related to the quality of data warehouse itself and the next two are concerned with the quality of the system’s outputs. The following two benefits improved productivity and better decision are associated with how these outputs and benefits are associated with how these outputs can benefit individual users, while the last two measures involve how the output may produce firm-level benefits[14].

DeLone and McLean [15] developed a comprehensive IS success model which categorizes IS success into six major constructs: system quality, information quality, use, user satisfaction, individual impact, and organizational impact. The model makes two important contributions to understanding of IS success. First, it provides scheme for categorizing the multitude of IS success measures that have been used in the literature. Second, it suggests a model of temporal and causal interdependencies between the categories. This IS success model postulates the following interrelationship among the six constructs. Moreover, this paper was a very important contribution to IS success measurement research as it came up with a set of success measures that did not exist before. This provided researchers and practitioners a framework to measure the success of their IT investments and the value they could derive from it.
3. Methodology and Model

This part describes the research methods for the study. The population for this study consists of system users from organizations who utilize a data warehouse. To empirically test the relationship between the quality data warehouse and enhanced decision quality in enterprises, a survey instrument designed using scales with demonstrated reliability.

Figure 1 presents the research model for this study. The research model depicted in Figure 1 is based on a review of the existing literature on data warehousing led to identifying factors that characterize data warehousing success and related variables. This included exploring different measures of successful DW implementation. Seddon’s[16] six dimensions of system success selected as being the most appropriate for this study: system quality, data quality, system use, user satisfaction, perceived net benefits, and decision quality. Empirical studies[17] have found that these six dimensions relate to one another.

From research model for hypothesis presents the research model for this study. The model proposes that system quality and data quality singularly and jointly affect both system use and user satisfaction. Additionally, the amount of system use can affect the degree of user satisfaction positively or negatively-and the degree of user satisfaction also affects system use. System use and user satisfaction are direct antecedents of perceived net profit. System quality can affect the degree of decision quality-positive and the degree of data quality also affects decision quality. Lastly, this impact on decision quality should eventually have some perceived net benefits from the use the warehouse.

4. Research Design

This study will use a survey instrument composed of existing measures that previously developed based on the data warehousing and success literature. The data warehousing literature emphasizes that the purpose of data warehousing is to provide quality data that is accurate, timely, complete and consistent. This purpose reflects facets of success measure-information quality, and the measures of data warehousing success should be selected with this in mind. Data quality is the measure of information systems output, and it has been used to measure success in many studies[18]. SPSS for windows will be used to analyze the data. First, demographic summaries will be presented and then the means and standard deviations for all variables. After presenting a correlation matrix with all variables, the research hypotheses will be tested with regression analyses.

5. Research of Data Analyses

This part presents an analysis of the data collected in support of this research. A lot of individuals were invited to participate in a web-based survey, and 214 responses were received. Data were then prepared, examined, and screened for outliers and missing values. The hypotheses were then tested using a combination of simple regression analysis, and subgroup analysis. Support was found for the entire main effect hypothesis.

Construct Analysis were followed to conduct the construct analysis of the survey instrument. Each of the two parts of the survey containing target questions was evaluated separately. Upon completion of the construct reliability analysis, data were analyzed through two different statistical methods: descriptive statistics and regression analysis.

Reliability analysis

Prior to data analysis, the research instrument was assessed for its reliability. Cronbach’s coefficient $\alpha$ was computed for each variables to test for reliability. The result is the following: data quality=.985, system quality=.962, system usage=.924, user satisfaction=.716, perceived net benefit=.926, decision quality=.824. In general, noted that a reliability coefficient of.07 or higher is consider acceptable.
In this study indicates that Cronbach’s $\alpha$ values range from .716 to .985, so the reliability of this study is acceptable.

**Part I Descriptive Statistics for Variables**

The data information shows that the question 1 (Do you get the information you need in time?) and the Question 2 (the information is consistent.) were at 3.327. Next, the question 4 (the data from the DW is accurate.) was at 3.107. The remaining question 3 (the data are usually complete.) was at 2.995. The degree of opinion all variables were medium level. In term of perception of the system quality, the response time was at 3.444 which were high level above of all. The reliability of the system was at 3.327. The flexibility of the system was at 3.220. Finally, the integration of the system was at 3.215. The degree of opinion of reliability, flexibility, and integration were at medium level.

As data shows that planning was at 3.893. Communicating with others was at 3.888. Making decisions and controlling and guiding activities were at 3.664. Looking for trend historical reference was 3.220. Find problems and alternatives were 3.336. Budgeting was 3.103. Lastly, historical reference was 2.88. In conclusion, the degree of opinion of planning, communicating with others, controlling and guiding activities, and making decisions were at high level. The remaining variables were at medium level.

As can be seen from the data, the question 2 (overall, are you satisfied with the system?) and the question 4 (how efficient is the system used for your area of responsibility?) were at 3.551. The question 3 (how efficient is the system?) was at 3.444. The question 1 (how effective is the system?) was at 3.332. All three variables were at high level but one variable was at medium.

Moreover, task productivity was at 3.84. Then management control was at 3.589. Customer satisfaction was 3.403, the degree of opinion of all variables were at high level.

In term of the quality DW improved decision quality, making the institution successful overall was at 4.107. Enhancing the institution's reputation was at 3.888. Improving decision making was at 3.664 and lastly, distinguishing your institution from similar institution was at 3.551. The degree of opinion was at high level in all of variables.

**Table 1. Summary Descriptive Statistics for all Variables**

| Scale                | Mean   | SD.    | Degree of opinion |
|----------------------|--------|--------|-------------------|
| Data quality         | 3.1893 | 1.1208 | Medium            |
| System quality       | 3.3014 | 1.0454 | Medium            |
| System usage         | 3.4562 | .80495 | High              |
| User satisfaction    | 3.4696 | .58614 | High              |
| Perceived net benefit| 3.6132 | .59526 | High              |
| Decision quality     | 3.8026 | .63577 | High              |

As data shows data quality and system usage were at 3.1893 and 3.3014. This information could explain that participants had considerable experience in their data quality and system quality was in medium level. In term of system usage, user satisfaction, perceived net benefit, and decision quality, the respondents concluded that high data quality and user satisfaction can lead to net benefits perceived net profits and also provide decision quality.

**Table 2. Coefficients between System quality and System usage, User satisfaction, Decision quality**

| Variables                              | B     | SE   | B   | p       |
|----------------------------------------|-------|------|-----|---------|
| System quality and System usage        | 1.882 | .144 | .477| .000    |
| System quality and User satisfaction   | 2.001 | .081 | .445| .000    |
| System quality and Decision quality    | 2.179 | .085 | .492| .000    |

In addition, the coefficients between system quality and system usage were at the level of statistical significance at 0.01. The system quality increased by 1, the system usage would increase by 0.477. Moreover, the coefficients between system quality and user satisfaction were at the level of statistical significance at 0.01. The system quality increased by 1, the user satisfaction would increase by 0.445. The coefficients between system quality and decision quality were at the level of statistical significance at 0.01 and would increase by 0.492.

**Table 3. Coefficients between Data quality and System usage, User satisfaction, Decision quality**

| Variables                              | B     | SE  | B  | p       |
|----------------------------------------|-------|-----|----|---------|
| Data quality and System usage          | 1.741 | .110| .538| .000    |
| Data quality and User satisfaction     | 2.172 | .076| .407| .000    |
| Data quality and Decision quality      | 2.425 | .085| .431| .000    |

As can be seen from the data, the coefficients between satisfaction and decision quality, and the coefficients between data quality and decision quality were at the level of statistical significance at 0.01. The coefficients between data quality and system usage were at .538. The coefficients between data quality and user satisfaction were at .407. The coefficients between data quality and decision quality were at .432. All three variables are related positively.

**Table 4. Coefficients between System usage, User satisfaction, Decision quality and Perceived net benefit**

| Variables                             | B     | SE  | B  | p       |
|---------------------------------------|-------|-----|----|---------|
| System usage and Perceived net benefit| 1.973 | .138| .475| .000    |
| User satisfaction and Perceived net benefit | .352 | .093| .940| .000    |
| Decision quality and Perceived net benefit | .981 | .167| .692| .000    |

As data shows the coefficients between system usage, user satisfaction, decision quality and perceived net benefit were at the level of statistical significance at 0.01. The coefficients between system usage and perceived net benefit were at .475. The coefficients between user satisfaction and perceived net benefit were at .940. Finally, the coefficients between
decision quality and perceived net benefit were at .692.

Table 5. Coefficients between user satisfaction and system usage

| Variables                              | B   | SE  | B   | p   |
|---------------------------------------|-----|-----|-----|-----|
| User satisfaction and System usage    | .238| .245| .927| .000|
| System usage and User satisfaction    | 1.770| .131| .492| .000|

Moreover, the coefficients between user satisfaction and system usage were at the level of statistical significance at 0.01. The coefficients between user satisfaction and system usage were at .927. The coefficients between system usage and user satisfaction were at .492. A P of 5% or less is generally accepted point at which to reject the null hypothesis.

Part II Hypotheses Testing-main Effect

Research Model and Regression Analysis Results

As hypothesized, the model proposes that system quality and data quality singularly and jointly both system use and user satisfaction. Two paths had positive effects with coefficient of .477 and .407 respectively. System quality was associated with system usage, user satisfaction and decision quality. Three paths had positive effects with coefficients of .477, .455 and .492 respectively. User satisfaction was associated with system usage. Two paths had positive effects with coefficients of .927 and .492 respectively. Data quality was associated with system usage, user satisfaction and decision quality. Three paths had positive effects with coefficients of .538, .407 and.432 respectively. Perceived net benefits was associated with system usage, user satisfaction and decision quality. Three paths had positive effects with coefficients of .475, .940 and .692 respectively.

Table 6. the Matrix of relationship of all Variables

|                    | System quality | System usage | Data quality | User satisfaction | Perceived net benefit | Decision quality |
|--------------------|----------------|--------------|--------------|-------------------|-----------------------|------------------|
| System quality     | Pearson Correlation | Sig. (2-tailed) | .619** .000 | .961** .000 | .794** .000 | .866** .000 | .809** .000 |
| System usage       | Pearson Correlation | Sig. (2-tailed) | .642** .000 | .470** .000 | .642** .000 | .470** .000 |
| Data quality       | Pearson Correlation | Sig. (2-tailed) | .778** .000 | .844** .000 | .844** .000 | .762** .000 |
| User satisfaction  | Pearson Correlation | Sig. (2-tailed) | .926** .000 | .739** .000 | .926** .000 | .739** .000 |
| Perceived net benefit | Pearson Correlation | Sig. (2-tailed) | .809** .000 | .470** .000 | .739** .000 | .739** .000 |

From the above information, the marks ** show the relationship between all variables had statistical significance at 0.01. All studied variables had the level of statistical significance at 0.01. Pearson’s r is close to 1 this mean that there is a strong relationship between two variable. This means that changes in one variable are strongly correlated with changes in the second variable. In this study person’s r is 0.985. This number is very close to 1. For this reason, we can conclude that there is a strong relationship of all variables.
### 6. Summary of Hypothesis Testing Results

In terms of hypothesis testing, we found that the results of the simple regression analysis are accepted as tenable and all $H_0$ are rejected. We could conclude that all variables were related to the others. In this study, it concludes that a data warehouse with good data quality and system quality improves the way data is provided to decision-support applications and decision makers. This supports the data warehousing literature that emphasizes that data warehouses must contain high-quality data, flexibly respond to users' requests for data, and integrate data in the ways that are required by users, all in order to create value for the organization.

### 7. Conclusions and Limitation

Data warehousing technology predominantly aims to structure the data in a summarized way which supports improved access to and use of the data in an efficient and effective manner. Despite compelling evidence that lack of attention to quality data warehousing problems leads to substantial human and economic losses because of inaccurate and inadequate data in their operational databases. Moreover, Hwang [2] showed how poor data quality results in the operational impacts (e.g., poorer decision making, more difficulties in building data warehouses, increased organizational mistrust, and strategic impacts (difficulties in setting and executing strategies, diversion of management attention). In addition, in the midst of this upheaval, technology presents the concept of data warehousing as one way to handle information overload. Data warehouses help managers and decision makers turn raw data into valuable information. Finally, data quality refers to the quality of the data that are available from the data warehouse. In addition, data quality is frequently discussed in the data warehousing literature as well; providing high quality data to decision makers is the fundamental reason for a building a warehouse.

### 8. Implication and Recommendation

The data warehousing exists to serve the decision making needs of the enterprise. Data warehousing is continuously gaining importance as organizations are realizing the benefits of decision-oriented data base. The success of data warehousing initiatives depends on the quality of data stored in it. Moreover, if your organizations want to continue to reap the benefits of reliable business intelligence, forecasting and analysis, ongoing data quality management and monitoring has to be high on the agenda. With this in mind the data warehouse offers an ideal platform for which to initiate an enterprise data quality program that will benefit the entire organization. Organizations must establish standards and guidelines for all personnel to follow to ensure that data quality is addressing the entire lifecycle of a system. Good quality data ensures user’s trust in data warehouse system making it more usable and optimizes the business benefits gained. Therefore, a careful and thorough maintenance of high quality is a necessity. Information should be checked and assured on a regular basis in order to ensure that the quality requirement is met and that information continuous to meet the intended need. Finally, the greatest potential benefits of data warehousing are when the warehouse is used in the redesign of business processes and to support strategic business objectives. Improved decision making usually results from the better information available from a data warehouse.

In addition, for a data warehouse to succeed, it must provide data, information, and knowledge that allow a user to solve their business problems. More important than building the system right (i.e. Critical Success Factor for data warehousing implementation), is building the right system (i.e. providing a solution to the business problem). To solve a user’s business problem, the IT analyst must understand the problem. To understand the business problem, the analyst must engage the user, elicit their problem, and agree together on the solution [19]. Additional investigation was conducted on data warehouse end users and their data warehouse usage, the value of the data warehouse in supporting managerial and...
knowledge workers’ tasks. This study confirmed that the system was critical in serving a variety of tasks from routine work to complex planning and decision making. Users had a high regard for the role of the data warehouse in enhancing their work productivity. In general, the study indicates that the IS success model introduced by DeLone and McLean [18] can become a good framework in understanding the success of data warehousing. The study indicated that the system helped the company in a variety of ways. Decision making and other organizational operations became more fact based, thereby improving the quality and effectiveness of decision-making processes and outcomes. Fact-based operation facilitated distributed decision making by more people. The distribution of decision making, in turn, improved organizational efficiency by reducing procedural overhead and quickly accommodating fluctuations in the marketplace.

Recommendations for the Future Research

Data quality research faces new challenges that arise from ever changing business environments, regulatory requirements, increasing varieties of data forms/media. Data quality issues have been classified and described in a way that should data warehouse practitioners, implementers, and tool providers find and resolve data quality problem issues as they move forward with each phase of data warehousing. In addition, the ability to integrate big data technologies, analytics technologies, back office system, and traditional data warehouse has the potential to fundamentally change the economics of the data warehousing for the better.

Limitation of Study

First this study will examine data warehouse development and implementation with in different organizations. As such, the ability to generalize the results of the study may be somewhat limited. Second, this study will examine measure the quality of the data. Previous experience with the decision making process impact a participant view of what quality measure is most important to their own decision making process. One person may believe accuracy is the most important, while another may see completeness as the most important criteria of data quality.

REFERENCES

[1] Poe, V., editor. Building a data warehouse for decision support. 1996, New Jersey: Prentice Hall.
[2] Hwang, M.I. and H. Xu, A Survey of data warehousing success issues. Business Intelligence Journal, 2005. 10(4): p. 7.
[3] Blackwood, P., Data warehousing solution focuses on quality improvement. Data Warehousing Technology, 2008. 41(2): p. 6.
[4] Park, Y.-T., An Empirical investigation of the effects of data warehousing on decision performance. Information & Management, 2006. 43(1): p. 51-61.
[5] Radding, A., Data warehousing in enterprise resource planning systems. Journal of Information System, 2000.
[6] Wang, R., Strong, D. Beyond accuracy : what data quality means to data consumers. Journal of Management Information Systems, 1996. 12(4): p. 5-43.
[7] Immon, W., Data warehouse—a perspective of data over time. Database Management, 1992: p. 370-390.
[8] Ariyachandra, T. and H. Watson, Which data warehouse architecture is most successful? Business Intelligence Journal, 2006. 11(1).
[9] Sharma, S.J., R, Enhancing business intelligence using data warehousing: a multi case analysis. Journal Of Advance Research in Computer Science and Management, 2013. 1(7): p. 160-167.
[10] Wixom, H., Watson,J., An epirical investigation of the factors affecting data warehousing success. MIS Quarterly,2001. 1(1): p. 17-41.
[11] Tebourksi, W., W.B.A. Karâa, and H.B. Ghezala, Semi-automatic data warehouse design methodologies: automatic data warehouse design methodologies: automatic data warehouse design methodologies: a survey. 2013.
[12] Yalman, M., Tunga, A., Examing university students' perceptions of computer and WWW (World Wide Web) with respect to certain variables: a case from Turkey. International Journal of Science, Innovation and New Technology, 2013. 1(6).
[13] Timothy, M., Duggal,K. , Review on security levels of data warehouse. Journal of Advance Research in Computer Science and Software Engineering, 2013. 3(11).
[14] Arnott, D., Success factors for data warehouse and business intelligence systems. ACIS 2008 Proceedings, 2008: p. 16.
[15] DeLone, W.H. and E.R. McLean, Information systems success: the quest for the dependent variable.Information Systems Research, 1992. 3(1): p. 60-95.
[16] Seddon, B., Kiew, Y., A partial test and development of DeLone and McLean’s model of IS success. Australian Journal of Information Systems, 1996. 4(1): p. 90-109.
[17] Roldanl, M. Analysis of the information systems success dimensions interdependence: an adaptation of the DeLone & McLean’s model in the Spanish EIS field. in BIT World 2000.
[18] Delone, W.H. and E.R. McLean, The DeLone and McLean model of information systems success: a ten-year update. Journal of Management Information Systems, 2003. 19(4): p. 9-30.
[19] Shin, B., An exploratory investigation of system success factors in data warehousing. Journal of Association for Information System, 2003. 4: p. 141-170.