

## Issues in User Perceptions of Data Quality and Satisfaction in Using a Data Warehouse - An Australian Experience

Amit Rudra  
School of Information Systems  
Curtin University of Technology, Perth  
AUSTRALIA  
RudraA@cbs.curtin.edu.au

Emilie Yeo  
School of Information Systems  
Curtin University of Technology, Perth  
AUSTRALIA  
emergper@singnet.com.sg

### Abstract

*Data warehousing has been touted to be the "solve-all" for the decision making managers. This is an exaggeration, as the problem of lack of data quality in data is one of the fundamental obstacles in the current data warehousing environment. The erroneous data can significantly affect an organisation, not to mention the frustration of the users of the data warehouse. In this paper, the researchers present some user perceptions of data quality and their satisfaction rating in using a data warehouse. These findings were accumulated from a survey conducted at a large organisation in Australia..*

### 1. Introduction

Data warehousing is one of the hottest topics of today in the information systems industry. And, judging from the number of papers presented at different conferences and other forums, data warehousing seems to be a discipline that is going to be around for a while (Mattison, 1996). Inmon (1994), the father of the term, defines it to be a subject-oriented, integrated, time-variant, non-volatile collection of data in support of management's decision making process to make timely, accurate decisions (Cobb, 1996; McClanahan, 1996). In comparison to an operational database, a data warehouse is managed data situated after and outside the operational systems (Gupta, 1997). Since a data warehouse contains historical data to supply a time-related view (e.g. for trend analysis), it is mainly intended for analytical applications such as decision support systems and executive information systems (Gallagher, 1995). Lately, its development has received a great boost due to the Internet

and the fact that information is now almost entirely captured electronically (Raden, 1996; Gupta, 1997; George, 1996). The data warehouse allows both simple and complex queries to the database, and provides the tools and procedures that let users manipulate the data to get only the pieces they need in the most useful form to support decision making (Lais, 1996). When modeled multidimensionally, it allows different combinations (commonly called data cubes) of results into reports for efficient query processing (Bussert, 1997; Chaudhuri and Dayal, 1997).

Recently, there has been a rise in expectations about what the data warehouse can do for the organisation (Marshall, 1996). However, a great problem in the way is the issue of data quality. One of main reasons for lack of data quality is the use or rather non-use of data (Orr, 1998). Orr (1998), in one of his rules of data quality, points out "if data is not used it cannot be correct for very long". It is therefore of interest to know not only how users access data in the data warehouse, but also, whether they find it useful and relevant at all. This research study aims to provide a real-world analysis of data warehousing from the user's point of view. It explores how some of the users of a large government organisation in Western Australia view or perceive system quality and information quality in a data warehousing environment. It also analyses their satisfaction with the system. It is expected that organisations considering implementing a data warehouse may benefit from the findings of this study. Moreover, quality issues and the organisation's role in achieving data quality may persuade the whole organisation to participate and take particular note of the users' views into account in the development of a data warehouse.

## 2. Research Study Objectives

Early research into data warehousing focused on the changing culture of corporate decision making which reflected the need and justification of data warehousing. However, the issues of user perceptions of information quality in data warehousing have not been fully explored. Current literature focuses on maintaining data quality in the area of general awareness and the usage of new tools to automate processes. For example, English (1996) and Strehlo (1996) both discuss data quality, but their focus was on tools to sustain the data warehouse. Some articles like Little and Gibson(1999) are on identifying factors that affect the implementation of data warehousing. Furthermore, English (1998a) considers the costs of low quality data and how it could affect the organisation. In other articles, he has formulated a self-assessment test on quality issues and advocates training as the solution to data quality problems (English, 1998b; English, 1998c). Nonetheless, most authors such as English (1998a), Strehlo (1996) and Cipriano (1995) have not fully covered the important point of how users view the quality aspect of data in a data warehouse.. A related study (Rudra and Yeo, 1999) looked at the key issues in achieving data quality and consistency in data warehousing among large organisations in Australia, but this was done from the perspective of the data administrators and not users of the data warehouse, therefore, our research aims to look into this problem. We found that even though the users have various views of data quality they all found accessing a data warehouse useful. In the next few sections we discuss these findings. However, we would like to first discuss some terms, like data quality, used in our research.

### 2.1. Data Quality

One of the critical success factors in data warehousing is the quality of data that is there in the data warehouse (Perkins, 1998; Chittenden, 1998). Decisions based on information derived from the data warehouse are only as good as the data quality of the data warehouse. It's therefore, important to provide a definition to this important term. Data quality refers to how relevant, precise, useful, in context, understandable and timely data is (Firth, 1997; Barry and Parasuraman, 1997; Miller, 1996). Perkins (1998) states very emphatically a data warehouse that contains "trusted, strategic information becomes a valuable enterprise resource for the decision makers at all organisational levels. If its users discover that it contains bad data, the data warehouse will be ignored and will fail". From the research of Delone and McLean (1992), the seven most important items that emerged were -- information accuracy, output timeliness,

reliability, completeness, relevance, precision, and accuracy. As information quality holds the same qualities as data quality, many authors have used the terms interchangeably (Firth, 1997; Barry and Parasuraman, 1997). To minimise integrity errors and improve information quality, there must be quality control (Clements, 1990).

It has been found that often, many end users, including managers are unaware of the quality of data they use in a data warehouse (Lambert, 1996). Data quality in the data warehouse is generally poor and there are many foreseeable setbacks (such as - economic failure, ineffective planning of business strategies). There have been some surveys of data quality in last few years. For example, Cipriano (1995) found that there is a correlation between quality management practices and quality information flows. Information flows and technologies also contribute toward obtaining high quality performances. In another research conducted by Forza (1995), quality information can be obtained with quality management practices, quality information systems and quality performance.

Thus, the issue of data quality in a data warehouse is of great importance. Its success depends on two important processes namely, data cleansing and data-quality improvement (English, 1996).

### 2.2. Data Inconsistency

One of the fundamental obstacles in the current data warehousing environment concerns the existence of inconsistent data. Data inconsistency occurs when there are different versions of the same data in the database (Awad and Gotterer, 1992). This can be caused by various stages of update or when a change has been made in one file and not in the remaining files. Inconsistencies in stored data are one of the most common sources of errors in a computer application (McFadden and Hoffer, 1999). They lead to inconsistent documents and reports as well as cause poor integrity of the system (Brathwaite, 1985). There is a negative relationship between data consistency and data redundancy. This means that at low levels of data redundancy, data consistency is high. On the other hand, there is a direct relationship between data consistency and data integrity. As Awad and Gotterer (1992:15) suggest, "data which is logically inconsistent lacks integrity because it cannot be depended upon". We can achieve data consistency by controlling or eliminating data redundancy. Coupled with good data administration this will promote a high level of data integrity.

### 3. Research Methodology

Since there has been little research (in what was being done) in regard to the issue of data quality in data warehousing, our aim was to find out how users of organisational data warehouse perceive data quality. Cipriano's (1995) survey research was about data quality in the United States. Therefore, the researchers thought it to be useful to conduct research on data quality in an Australian perspective. The survey (Appendix A) consisted of both close-ended and partially open-ended (where users could provide comments if needed) questions. The respondents were asked to rank, on a seven-point scale (whether they strongly agree to strongly disagree regarding system quality issues; never and always for information quality issues; adequate to inadequate for overall satisfaction issues), some of the issues concerning system quality, information quality and overall satisfaction. They were also asked whether they perceived the relationship between data consistency and data quality (yes or no). Quantitative analysis was used to analyse the data received (using SPSS for Windows).

#### 3.1. The Sample

A large government organisation was targeted (due to confidentiality reasons, the name and other details of the organisation are not revealed). The prospective 16 respondents were handed the questionnaire in person and the replies were received by mail. Respondents were assured of confidentiality and they included persons of various roles (from technical to managerial). The final replies received included 10 users or prospective users of the organisational data warehouse.

*Limitation of the survey:* As the survey was conducted in a limited way, e.g. in one organisation with not many users of the data warehouse, the researchers would like to point out that the conclusions drawn for the survey is typical of the surveyed organisation. There may or may not be much resemblance with users of another organisation. Nevertheless, it still offers us a number of interesting observations.

### 4. Findings

The 10 responses received (response rate of 62%) were analysed. There were a significantly large number of users (60%) with limited or no usage (or are anticipating usage in future) of the data warehouse. According to various users, the data warehouse is generally used for statistical research

into client data, evaluating performance, analysing business processes, and, ad-hoc queries for parliamentary questions, finances and auditors' requests. The data also helps the users to make informed decisions. The data retrieved from the warehouse will eventually be presented to senior management and other strategically oriented sections of the ministry in the form of reports, for example annual and quarterly reports. However, at present there are more users (60%) who find the system difficult to use as they are unfamiliar with this new technology. Various reasons that have been attributed to this are:

1. Not using the data warehouse enough.
2. Difficult to navigate around the warehouse.
3. Inadequate training.

One particular user found that the data warehouse was overwhelming, as training was not user friendly to her as a beginner.

There has been mixed reactions to whether the data warehouse provides easy retrieval of data. This is shown by the high standard deviation of 1.54. The warehouse's cubes have been criticised as unreliable. According to the respondents, the cubes have not been tested and they do not know of the actual intended purpose as originally specified. Two respondents commented that it is necessary to have a good understanding of the system to ensure that the data is what you want. Another user suggested that there needs to be considerable amount of work done to the data warehouse in order to achieve business needs. Responding to the question of whether the data warehouse is a failure, 9 of the respondents gave rankings of 6 and 7 (Disagree and Strongly Disagree respectively) to show their support for it. This is a remarkable finding as even though they find the system difficult to use they do not consider it to be a failure. Using the binomial test this factor was found to be significant. Moreover, some users mentioned that although the data warehouse has limited use due to the semi-implemented state of some systems, it would change in the next few years when the data warehouse becomes invaluable.

In the following sections the findings are presented in detail.

### 4.1. Perceptions of System Quality

Here the respondents were asked to rank, on a seven-point scale (strongly agree to strongly disagree), how they perceived using the data warehousing system itself. Table 1 shows the findings.

**Table 1. Ranking of user perceptions of system quality**

Factors	Mean Score	Std. Dev.
I find that data warehousing system is a failure	6.29	1.11
The data warehousing system does not meet my needs	4.63	1.85
It is easy to learn how to use it	4.44	1.01
The data warehousing system is easy to use	4.11	1.27
It is easy to retrieve data from the system	3.89	1.54
Using the system is frustrating	3.63	1.77
Using the system requires a lot of mental work	3.13	1.55

As discussed earlier, the users definitely do not see the data warehousing system itself as a failure. This is shown by the high mean of 6.29. Upon performing a significance test (binomial) this value was significant. This is an important finding, i.e. even though the users are divided on all other factors of usage of the data warehousing system (for example, they do not necessarily consider that the system meets their needs or is easy to use or learn to use it), they still find that it serves them well and is not a failure. The users commented that with additional training and documentation, they foresee that the mental work involved in using the warehouse would significantly decrease. 55.5% of the respondents polled that the data warehouse was not easy to learn and another 66.6% voted that the data warehouse was not easy to use. A further 62.5% polled that using the system was frustrating. This could be due to a lack of a proper interface designed for the users.

### 4.2. Perceptions of Information Quality

In this question the respondents were asked to rank on a seven-point scale (never to always) how they perceived the quality of information provided by the data warehousing system. Table 2 shows the findings. As can be observed from the mean scores and standard deviation, the users are well divided in their opinion here.

**Table 2. Ranking of user perceptions of information quality**

Factors	Mean Score	Std. Dev.
Do you think the output is useful	4.86	1.46
Is the information clear	4.33	1.22
Does the information content meet your needs	3.86	1.68
Do you get the information you need in time	3.83	2.14
Does the system provide up-to-date information	3.67	2.35
Is the system accurate	3.63	1.60
Does the system provide the precise information you need	3.57	1.72
Does the system provide sufficient information	3.43	1.40
Are you satisfied with the accuracy of the system	3.38	1.69

The two factors that score above the rest are the usefulness of the information provided by the data warehouse and the clarity of the information provided. This has relevance with what was found in the previous section (section 4.1) viz. that most of the respondents perceived the system to be useful and not a failure. However, upon significance test (binomial), no factor was found to be significantly high enough.

There have been mixed reactions to users being satisfied with the accuracy of the system (standard deviation 1.69). One user noted the lack of complete source data in the system. The issue of cubes being untested has led to unsatisfied respondents. User 9 indicated that data integrity on operational systems is unsatisfactory and that quality systems need to be developed. User 10's perception was that because business areas do not update the information in the data warehouse, the retrieved information would be inaccurate. Also, user 8 commented that some users might not know if the data was accurate and some guidance from the data warehousing team may be needed to ensure confidence. Many users have stated that additional help and documentation is needed to help users get the sort of data they are looking for.

Under "clarity of information", due to the lack of familiarity with the systems, some users find the abbreviations of the field names bothersome. In the opinion of the researchers, for the data warehouse to provide up-to-date information, it is subject to the time the source data was last refreshed. In the instance of providing sufficient information, users feels that more relevant cubes need to be developed as the current data in the warehouse does not provide the information needed for performance indicators. Moreover, user 9 stated that the data warehouse does not meet the needs of the ministry.

As for effectiveness of the data warehouse, users feel that they are limited by their own effectiveness. The need for further training is again exemplified. Additional training according to the users could be, for example, in areas of how best to apply the data warehouse to one's needs.

### 4.3. Perceptions of Overall Satisfaction

In this question the respondents were asked to rank on a seven-point scale (adequate to inadequate) how they perceived the overall satisfaction with the warehousing system. Table 3 shows the findings.

**Table 3. Ranking of user perceptions of satisfaction with the system**

Factors	Mean Score	Std. Dev.
How effective is the data warehouse	3.43	1.99
How adequately do you feel the data warehouse has met the information processing needs	2.40	2.07
Overall are you satisfied with the data warehouse	2.00	1.81
How efficient is the data warehouse	1.80	0.84

When asked whether the data warehouse has met the users' information processing needs, 80% ranked it as sufficiently adequate. However, in the case of data warehouse effectiveness, there were mixed reactions viz. 2 and 6 with highs of 28.6% respectively. A wide spread of opinions on this question were noted from the graph as this is further supported by its standard deviation of 1.99. The most probable reason for this is that there are two categories of users, one group are those who are frequent users of the warehouse and the other novices of the data warehouse.

In terms of efficiency, 80% of the respondents agree that the data warehouse has good performance levels in delivering the queries to the user in the shortest amount of response time. In addition, 70.5% of users polled that they were satisfied with the data warehouse in general but additional improvements could be made to further enhance the performance and usage of the data warehouse. Although there is trend towards the satisfied side of the voting scale, it is necessary to pay attention to the 14.3% of users who polled 6 out of 7, voting that they were unsatisfied with the data warehouse. Again, some reasons are better documentation of information in the warehouse and how the users can use the warehouse to its full potential in their work.

### 4.4. Perceptions of Data Consistency and Data Quality

In this section the respondents were asked either a yes or a no to find out their views regarding data consistency and data quality in a data warehouse. The frequency distribution here indicates that most of the users (60%) do not know the importance of the link between data consistency and data quality. However, they have noted the fact that some data in the warehouse is inconsistent. For example, User 1 commented that the different departments have a number of different databases with very different data structures. User 3 acknowledges the existence of some systems, which have poorer data. He feels that time and training of users will rectify the existence of poor data. However, time and training of users are not the only two factors involved. There needs to be commitment from all parties involved as well as ownership of data along with various data quality checks to ensure that poor data gets rectified into becoming useful and rich data which users can utilise.

An interesting point brought up by user 6 was that it could be better to develop people who specialise in using the data warehouse and use them regularly since, unless one uses the warehouse regularly, the skills are not developed. Consultation could then be given to those people who require information. This in the opinion of the researchers would irradiate the whole purpose of the data warehouse, as information can be retrieve at any time using ad-hoc queries to the database.

## 5. Conclusion

This research found that a significant portion of the users in the organisation found the data warehouse to be useful and not a failure. While they were mixed about their perceptions regarding quality of information provided by the warehouse, they did think that it provided useful information. However, they were not sure if the data quality provided was good due to inconsistencies in data structures across various systems. They were overall satisfied with the system, but were not sure if it was effective. However, they perceive that it met their needs to certain extent. The users are however, not sure about the link between data consistency and data quality.

Data quality in data warehousing is an important field that is still emerging in the industry and in coming years it would assume even greater importance. The researchers believe that this research has highlighted some important issues regarding the quality of data in a data warehouse especially, with regard to users of the system and that it would benefit the practitioners, viz. the data administrators, to keep in mind some of the factors when planning to set up a data warehouse. To the academics it is hoped that the research has sparked some ideas for future research in this field of study.

## 6. References

- Awad, E.M. and Gotterer, M.H., 1992, *Database Management*, Boyd & Fraser Publishing Company, Danvers, Massachusetts.
- Barry, L.L. and Parasuraman, A., 1997, *Listening to the Customer - The Concept of a Service-Quality Information System*, Sloan Management Review, Spring 1997, Vol. 38(3), pp. 65-76.
- Bussert, J., 1997, *Data Management 101*, MIDRANGE Systems, Cardinal Business Media Inc, Jan 17, 1997, Vol. 10 n1, pp. 44(1).
- Brathwaite, K. S., 1985, *Data Administration: Selected Topics of Data Control*, John Wiley & Sons Inc, New York.
- Chaudhuri, S. and Dayal, U., 1997, *An Overview of Data Warehousing and OLAP Technology*, SIGMOD Record, Vol.26, n1, pp. 65-74.
- Chittenden, J., 1998, *The Data Warehousing Research Project Report*, Available [Online] at: <http://www.state.tn.us/finance/oir/prd/data.htm>
- Cipriano, F., 1995, *The Impact of Information Systems on Quality Performance: An Empirical Study*, International Journal of Operations and Production Management, Vol. 15(6), pp. 69-83.
- Clements, R.B., 1990, *Creating and Assuring Quality*, ASQC Quality Press.
- Cobb, I., 1996, *Data Warehousing: The Latest Fad?*, Management-Accounting-London, June 1996, Vol. 74(6), pp. 44-46.
- Delone, W.H., McLean, E. R., 1992, *Information Systems Success: The Quest for the Dependent Variable*, Information Systems Research, 3:1, pp.60-95.
- English, L.P., 1996, *Help for Data Quality*, Information Week, Available [Online] at: <http://techweb.cmp.com/iw/600/00oldat.htm>
- English, L., 1998a, *Data Quality: Meeting Customer Needs*, Pitney Bowes, Available [Online] at: <http://www.pitneysoft.com/products/DQEng001.htm>
- English, L., 1998b, *DQ Point 6: Institute Training on Data Quality*, DM Review Magazine, Available [Online] at: [http://www.dmreview.com/issues/1998/may/articles/may98\\_72.htm](http://www.dmreview.com/issues/1998/may/articles/may98_72.htm)
- English, L., 1998c, *The High Costs of Low Quality Data*, DM Review Magazine, Available [Online] at: [http://www.dmreview.com/issues/1998/jan/articles/jan98\\_38.htm](http://www.dmreview.com/issues/1998/jan/articles/jan98_38.htm)
- Firth, C., 1997, *When Do Data Quality Problems Occur?*, Available [Online] at: <http://sunflower.singnet.com.sg/~cfirth/dq1.htm>
- Forza, C., 1995, *Quality Information Systems and Quality Management: A Reference Model and Associated Measures for Empirical Research*, Industrial Management and Data Systems, Vol. 95(2), pp. 6-14.
- Gallagher, S., 1995, *How to Make Use of Raw Data*, Government Computer News, Cahners Publishing Associates LP, April 3, 1995, Vol. 14 n7, pp. 23(3), Available [Online Database - Computer Database ASAP].
- George, T., 1996, *The Web meets the Data Warehouse*, Insurance Systems Bulletin, Nov 1996, Vol. 12(5), pp. 4.
- Gupta, V.R., 1997, *Introduction to Data Warehousing*, System Services corporation, Chicago, Illinois.
- Inmon, W., 1994, *Using the Data Warehouse*, John Wiley and Sons Ltd.
- Kelly, S., 1997, *Data Warehousing in Action*, John Wiley and Sons Ltd.
- Lais, S., 1996, *See Your Data in a New Light; Data Warehouses Give You Insight into Information*, Government Computer News, Cahners Publishing Associates LP, Nov 18, 1996, Vol. 15 n29, pp. 43(2), Available [Online Database - Computer Database ASAP].
- Lambert, B., 1997, *Break Old Habits to Define Data Warehouse Requirements*, Available [Online] at: <http://www.data-warehouse.com/resource/articles/lamber11.htm>
- Little, R. and Gibson, M., 1999, *Identification of Factors Affecting the Implementation of Data Warehousing*, HICSS-32 (Hawaii International Conference on System Sciences), Jan 1999, Maui, Hawaii.
- Marshall, M., 1996, *Data Warehouse Expectations Rise*, Communications Week, CMP Publications Inc, Feb 12, 1996, n 596, pp. 4-7, Available [Online Database - Computer Database ASAP].
- Mattison, R., 1996, *Data Warehousing - Strategies, Technologies and Techniques*, McGraw Hill.
- McClanahan, D., 1996, *Making Sense Of Enterprise Data*, Data Based Advisor, Data Based Solutions Inc, Nov 1996, Vol. 14 n11, pp. 76(4), Available [Online Database - Computer Database ASAP].
- McFadden, F.R. and Hoffer, J.A., 1999, *Data Base Management*, 5<sup>th</sup> Edition, The Benjamin / Cummings Publishing Company, Inc.
- Miller, H., 1996, *The Multiple Dimensions of Information Quality*, Information Systems Management, Spring 1996, Vol. 13(2), pp. 79-82.
- Orr, K., 1998, *Six Rules of Data Quality*, Data Warehouse Report, July 28 and August 4, 1998. Available [Online] at: <http://datawarehouse.dci.com/articles/1998/07/28orr.htm>
- Perkins, A., 1998, *Critical Success Factors for Data Warehousing Engineering*, Visible Systems Corporation.

Available [Online] at: <http://www.visible.com/about/whitep.html>

Porter, J.D. and Rome, J.J., 1994, *The Data Warehouse: 2 Years Later...Lessons Learned*, Arizona State University, 1994 CAUSE Annual Conference held in Orlando, FL.

Raden, N., 1996, *Warehouses and the Web*, Information Week, May 13, 1996, 579, pp. 80-86.

Rudra, A. and Yeo, E., 1999, *Key Issues in Achieving Data Quality and Consistency in Data Warehousing among Large*

*Organisations in Australia*, HICSS-32 (Hawaii International Conference on System Sciences), Jan 1999, Maui, Hawaii.

Strehlo, K., 1996, *Data Warehousing: Avoid Planned Obsolescence*, Datamation, Available [Online] at: <http://www.datamation.com/PlugIn/issues/1996/jan15/01bsw200.html>

Warmouth, M.T. and Yen, D., 1992, *A Detailed Analysis of Executive Information Systems*, International Journal of Information Management, Vol. 12, pp. 192-208.