

# Stakeholder Perceptions of Data Quality in a Data Warehouse Environment

Anthony Giannoccaro\*

Graeme Shanks\*\*

Peta Darke\*

\*School of Information Management and Systems,  
Monash University, Melbourne, Australia.

\*\*Department of Information Systems,  
The University of Melbourne, Melbourne, Australia.  
g.shanks@dis.unimelb.edu.au

## Abstract

*The quality of the data provided is critical to the success of data warehousing initiatives. There is strong evidence that many organisations have significant data quality problems, and that these have substantial social and economic impacts. This paper describes a study which explores the perceptions of various stakeholders in the data warehousing environment concerning their data quality requirements. The study involved the development of a conceptual framework representing the relationships between stakeholder types and data quality dimensions, and the application of the framework in a case study of data quality in a data warehouse at a large transport company. The case study findings suggest that improved training in data warehouse usage is necessary to help data consumers better understand the data, that reward structures should be established to help ensure data quality compliance, and that the importance of data quality needs to be better communicated.*

## Keywords

Data quality, information quality, data warehouse

## INTRODUCTION

Many organisations are currently developing data warehouses in order to reduce the costs associated with the provision of data, to support a focus on complete business processes, and to achieve high estimated returns on investment (McFadden 1996). A key factor in the success of data warehousing initiatives is the quality of the data provided (Kimball 1996, Wang 1998). It is essential therefore that data quality is understood and that data quality assurance procedures are developed and implemented. Whilst many organisations are aware of the importance of data quality for their ability to compete successfully in the market place, research and industry surveys indicate that organisations are increasingly experiencing problems with data quality (Redman 1998, Wand and Wang 1996) and that these have substantial economic and social impacts (Strong *et al.* 1997). There has been a lack of methods and frameworks for measuring, evaluating, and improving data quality (Wand and Wang 1996), however, and little discussion of the management, economic or organisational aspects of data quality (Firth and Wang 1993).

A number of researchers have identified various data quality dimensions. However, these dimensions are often overlapping, vaguely defined and not soundly based in theory (Shanks and Darke 1998a). Some frameworks have been developed which organise important concepts for defining and understanding data quality (see for example Wand and Wang 1996, Shanks and Darke 1998b), and

support methodical approaches to improving data quality processes within organisations (see for example Kahn *et al.* 1997, Wang 1998). Different stakeholder groups have also been identified according to whether they are responsible for producing, maintaining or consuming data (Strong *et al.* 1997, Wang 1998). There has been little research, though, which has examined the relationship between various stakeholder groups and the data quality process, in particular stakeholders' differing data quality requirements and concerns.

This paper describes an empirical study of stakeholder perceptions of data quality in a data warehouse environment. A conceptual framework for understanding the relationships between different stakeholder types in the data warehousing environment and categories of data quality dimensions was developed from an analysis of the data quality and data warehouse literature. The framework was used to propose specific relationship instances between particular stakeholder types and specific categories of dimensions. It was then applied in a case study of a large transport company's data warehouse which examined the interactions between data quality dimensions and various stakeholder types. The case study was used to determine the empirical validity of the proposed relationships, and provided useful insights into the nature of the various stakeholders' data quality requirements and concerns.

Organisations are often aware of data quality problems. However, their improvement efforts generally focus narrowly on only the *accuracy* of data, and ignore the many other data quality attributes and dimensions that are important (Wang and Strong 1996). The data quality requirements of different stakeholder groups need to be considered (Shanks and Darke 1998b). This research focuses on understanding particular stakeholder groups' views and perceptions of data quality in terms of a broad range of dimensions. The case study findings suggest that improved training in data warehouse usage is necessary to help data consumers better understand the data in data warehouses, that reward structures should be established to help ensure data quality compliance, and that the importance of data quality needs to be better communicated.

The paper first defines data quality and reviews existing research in data quality. The next section describes the research approach adopted in this study. The conceptual framework for understanding relationships between stakeholder types and categories of data quality dimensions is discussed in the fourth section. The case study of data quality in a data warehouse is then described, and an analysis of the case study data is presented. Implications of the case study findings for data quality practice are discussed, and the paper concludes with some suggestions for future research.

## **DATA QUALITY IN THE DATA WAREHOUSE ENVIRONMENT**

Data warehouses are seen as a means of providing the data management infrastructure for executive information systems, decision support systems, and other management support systems (McFadden 1996, Gartner Group 1997). A data warehouse is a set of databases created to provide information to managers and decision makers through an integrated software and hardware environment that is optimised for retrieval rather than for update integrity and transaction throughput (Shanks *et al.* 1997). Effective business decision-making depends on good quality data, and poor data quality can be costly and sometimes disastrous (Wang and Strong 1996). A key factor in the success of data warehousing is the quality of the data provided (Kimball 1996, Wang 1998). Organisations therefore need to understand data quality, and establish procedures to assure the quality of data in data warehouses.

Despite their expenditure on information technology, many organisations still do not have the accurate, timely, and useful data they require for effective operations and decision-making. Data quality problems are becoming widespread in practice (Wang and Wang 1996, Redman 1998), and can have significant social and economic impacts (Strong *et al.* 1997). Before the issues involved in

managing data quality can be addressed, it is necessary to first understand what data quality means (Firth and Wang 1993).

In the quality literature, a widely-used concept is “fitness for purpose” (Wang and Strong 1996). In relation to data quality, this should encompass not only the intrinsic characteristics of the data itself but also assessments by users of the data (Strong *et al.* 1997). Data in a data warehouse must be useable and useful to data consumers, and support their work practices. Data quality may be defined at two levels: structure (metadata) and content (data) (Shanks and Darke 1998a). The quality of the structure of a data warehouse is defined as the quality of the conceptual model that is the basis for the design of the data warehouse. Quality metadata is important for all stakeholders in the data warehousing process so that they understand what the data warehouse contains and how to access data in the data warehouse. Quality data is important so that users of the data warehouse can use the data effectively in their decision making activities. This paper focuses on content (data) quality.

Attempts to define data quality have mainly concentrated on providing lists of desirable data quality dimensions (see Wand and Wang 1996 for an overview). Typically, dimensions such as accuracy, reliability, importance, consistency, precision, timeliness, understandability, conciseness, and usefulness are included. However, these are frequently overlapping, vaguely defined, ambiguous, and not based in theory (Shanks and Darke 1998a). There is no clear consensus of what constitutes the definitive set of data quality dimensions (Firth and Wang 1993), although data accuracy, currency and completeness are considered important (Redman 1992, Wand and Wang 1996).

More recently, some researchers have developed comprehensive frameworks that organise and structure important concepts in data quality. Four *categories* of data quality dimensions have been identified by Wang and Strong (1996). *Intrinsic* dimensions define the quality of data in its own right, *contextual* dimensions define data quality within the context of the task at hand, *accessibility* dimensions emphasise the role of information systems in providing data, and *representational* dimensions define data quality in terms of the presentation and delivery of data. These categories separate intrinsic data quality from other categories of data quality. Kahn *et al.* (1997) focus on delivering quality information in the sense of both product and service. Their framework categorises data quality dimensions as sound or useful (relating to *product* quality), or usable or effective (relating to *service* quality), and provides a means of determining the level of sophistication of data quality management within organisations. However, the data quality dimensions in these two frameworks are not based in theory.

A framework that provides formal definitions of several data quality dimensions (completeness, lack of ambiguity, meaningfulness and correctness) has been proposed by Wand and Wang (1996). The framework anchors data quality dimensions in ontological constructs relating to concepts of representation, interpretation, inference and the internal view of an information system, and uses an approach of mapping real-world systems to information systems to identify inconsistencies between them which are seen as data deficiencies. Only intrinsic data quality dimensions are included in the framework.

Shanks and Darke (1998b) proposed a useful framework for understanding data quality in a data warehouse environment that includes both intrinsic and extrinsic data quality dimensions and is based in semiotic theory, the study of the use of signs and symbols to convey knowledge. An important feature of their framework is the separation of data quality goals from the means to achieve them. Other components of the framework include stakeholders, improvement strategies, measures, weightings and ratings. Four types of stakeholder are identified: data producers, data custodians, data consumers, and data managers (Strong *et al.* 1997, Wang 1998). Data producers are those who create or collect data for the data warehouse. Data custodians are those who design, develop and operate the data warehouse. Data consumers are those who use the data in their work activities.

Data managers are those who are responsible for managing the entire data warehousing process. The framework supports the evaluation by individual stakeholders or stakeholder groups of data quality and the identification of data quality requirements in practice.

The research described in this paper draws on the data quality and data warehouse literature described in this section to propose a conceptual framework for understanding stakeholder data quality requirements which is used to define specific relationship instances between different stakeholder types in the data warehousing environment and specific categories of data quality dimensions. The framework was applied in a case study which examined the interactions between stakeholder groups and data quality dimensions in order to develop a better understanding of which dimensions are of most importance to which stakeholders.

## **RESEARCH APPROACH**

The research described in this paper consisted of two phases. The first phase involved development of the conceptual framework for representing relationships between stakeholder types and categories of data quality dimensions by synthesising concepts from the data quality and data warehouse literature. The research method adopted was conceptual study, which captures and articulates the researcher's views and is effective in developing new concepts and insights (Galliers 1991). The second phase of this research sought to examine the applicability of the framework concepts and proposed relationships between them to data quality requirements in a data warehouse environment in practice. The case study research method (Yin 1994) was used in this phase of the research. Case study research is appropriate where a contemporary phenomenon is to be studied in its real-life context, and is useful where "research and theory are at their early, formative stages" (Benbasat *et al.* 1987, p. 369).

The case study examined the data quality requirements of various stakeholders involved in implementing, maintaining and using a data warehouse in a large transport company. Data collection sources for the case study included semi-structured and unstructured interviews conducted with data warehouse users, implementation and management personnel, as well as project and system documentation and published information about the organisation. Qualitative data analysis methods (Miles and Huberman 1984, Myers 1998) were used to classify and analyse the case data. The unit of analysis in the case study was the stakeholder groups and their interaction with information systems project teams in using the data warehouse. The purpose of the case study was to investigate stakeholder groups' perceptions of data quality dimensions and to determine the empirical validity of the framework concepts and relationships.

## **A FRAMEWORK FOR UNDERSTANDING RELATIONSHIPS BETWEEN STAKEHOLDER GROUPS AND DATA QUALITY DIMENSIONS**

In order to meet the data quality requirements of all potential stakeholder groups in the data warehouse environment, it is essential to capture and understand the aspects of data quality that are important to each group and to identify ways of improving data quality. The framework for understanding stakeholders' data quality requirements proposed in this paper combines the *stakeholder*, *property*, *improvement strategy*, and *measure* concepts from Shanks and Darke's (1998b) framework for understanding data quality with the *category* and *data quality dimension* concepts from Wang and Strong's (1996) framework of data quality. *Stakeholder* identifies four stakeholder groups responsible for creating, maintaining, using, or managing data in the data warehouse environment: data producers, data custodians, data consumers, and data managers. The *property* and *data quality dimension* concepts are similar, although Shanks and Darke (1998b) do

not define *property*. Wang and Strong’s (1996) definition of a data quality dimension as “a set of data quality attributes that represent a single aspect or construct of data quality” is therefore adopted. *Improvement strategies* are the processes used to achieve data quality dimensions and a *measure* is a systematic way of measuring a data quality dimension. *Category* refers to Wang and Strong’s (1996) grouping of data quality dimensions into four data quality categories which capture the requirement for high quality data to be “intrinsically good, contextually appropriate for the task, clearly represented, and accessible to the data consumer” (Wang and Strong 1996, p. 22). The framework components and their interrelationships are shown below in Figure 1.

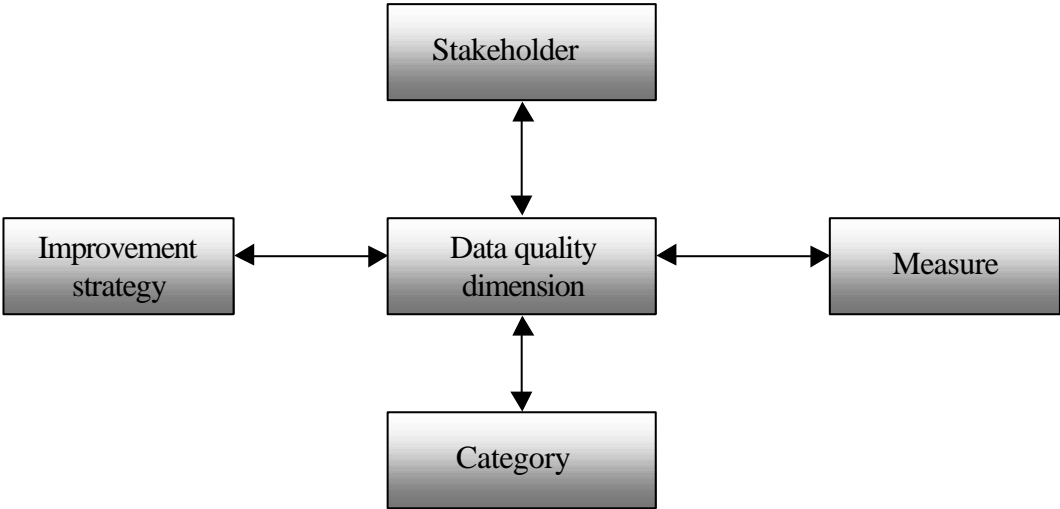


Figure 1: The framework for understanding relationships between data quality dimensions and stakeholder groups in a data warehouse environment

Different stakeholders may have different perspectives on data quality. The framework relates stakeholders to data quality dimensions to support identification of differing data quality requirements. Dimensions are related to measures to enable evaluation of data quality by stakeholders, and improvement strategies provide for identification of activities that will improve data quality in terms of specific dimensions. The categories provide a means of organising the many data quality dimensions that have been identified in the literature into a classification scheme that captures the essential data quality characteristics that are important to data consumers.

The framework provides a set of concepts that can be used as a basis for empirical studies of stakeholders’ data quality requirements and quality improvement processes in practice. Three components of the framework, *stakeholder*, *category* and *data quality dimension*, were used to investigate the nature of the various stakeholders’ data quality requirements in a data warehouse environment in the case study reported in this paper. Drawing on the available data quality and data warehouse literature, specific relationship instances between particular stakeholders types in the framework and particular categories of dimensions in the framework were proposed. These were then empirically evaluated in the case study that examined the various stakeholders’ perceptions of the data quality dimensions that are important to them. The data quality dimensions identified by Wang and Strong (1996) were used in the study. The relationship instances that were proposed are shown in Figure 2 below.

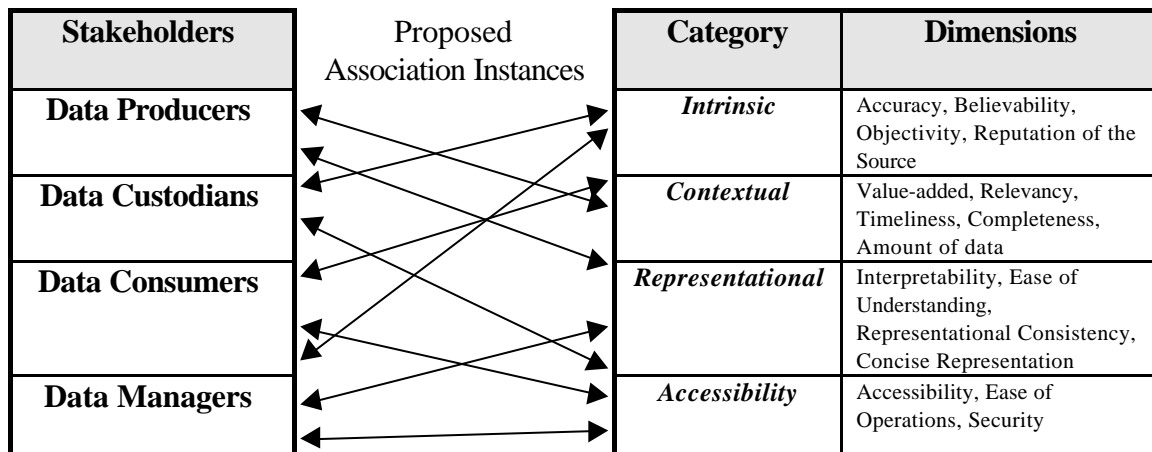


Figure 2: Proposed relationship instances between stakeholder types and data quality categories and dimensions.

It was considered likely that data producers would perceive contextual and representational data quality categories to be important, in particular “amount of data” and “ease of understanding”, as they are responsible for generating and processing data in particular situations. It was considered that data custodians would be concerned with the intrinsic and accessibility categories, in particular the “accuracy” and “security” dimensions. They are responsible for providing and managing the resources for storing, processing and accessing the data. To ensure usage of the data warehouse, data must be accurate, complete and accessible. Data consumers were considered likely to perceive the accessibility and contextual categories as most important, in particular the “accessibility”, “timeliness” and “completeness” dimensions, as they are reliant on the data warehouse for provision of timely, complete and accessible data to support their decision making and other tasks. Incomplete data is a major source of data consumers’ complaints about inadequacies of data (Strong *et al.* 1997). Data managers were considered likely to perceive the intrinsic, representational and accessibility categories as important, in particular the “accuracy”, “interpretability” and “security” dimensions. They are responsible for managing the operation of the data warehouse, including the accuracy, security, and representation of the data.

## THE CASE STUDY

The case study concerns data quality issues in a data warehouse development project within a large transport company with over 10000 employees. Over the last three to five years there have been significant reductions in staff numbers due in part to increased use of information technology and increased contracting out of services. The company has more recently adopted a customer service orientation and is moving towards a more business driven approach to service delivery.

A strategic aim of the company was to create an organisational structure which facilitated change and lead to fulfillment of the company’s goals through more effective leadership, decision making, communications and performance monitoring. In order to support this aim, the company introduced outsourcing of some of its information technology functions and services, developed corporate standards for the use of information technology, and established accountability on the part of each business area for their use of information technology.

Recently a data warehouse project was initiated within the company in order to provide high quality information to support critical business decisions, in particular for the sales and marketing departments. The data warehouse was implemented incrementally over a period of eighteen months. Within 2 years there were over 1000 users of the data warehouse.

The data warehouse project team consisted of a project manager (data manager), a data warehouse system supervisor responsible for data in the data warehouse (data custodian), several data production analysts responsible for sourcing and transforming data (data producers), and business solutions managers responsible for interacting with business units using the data warehouse (data consumers).

Data quality was an important consideration within the data warehouse project team, and a number of data quality principles were developed:

1. data quality problems should be fixed in the source systems where possible;
2. data in the data warehouse is assigned a business area responsible for its quality;
3. data in the data warehouse is “published” after being checked by the responsible business area.

### **Stakeholder Views on Data Quality**

Representatives from each of the four stakeholder groups defined in the Wang and Strong (1996) framework were interviewed about their perceptions of data quality. Additional information was obtained from reports, memos, and notes from meetings.

#### *Data Producers*

Data sourced from the company’s information systems was carefully analysed prior to uploading to the data warehouse. Consistency of representation of the data and referential integrity were seen as very important. It was found that the consistency of the data had a significant effect on the perceived data quality. One data producer noted:

“If our data consumers find that our data has consistency problems, then there is a real danger that they will lose confidence in the data warehouse”.

The accuracy and believability of data were also seen as important:

“All our data is system generated. However, we (data producers) must be vigilant to ensure that the accuracy and believability of the data remains high”.

Data producers were generally concerned that there were no incentive programs in place at the company to improve the quality of data at its source. It was suggested that a reward structure for data quality compliance would increase the believability and reputation of the source data and therefore increase the perception of data quality in the data warehouse.

Information requirements of data consumers were carefully collected to ensure that data loaded into the data warehouse was relevant. Data producers therefore had actual associations with the *intrinsic* and *representational* data quality categories and a weaker association with the *contextual* data quality category.

#### *Data Custodian*

The design of the data warehouse was of concern to the data custodian, being a trade-off between the volume of stored data and achieving the desired level of granularity for drill-down queries. Another area of concern was ensuring that the accuracy and reputation of the data were considered high. Edit checks and database integrity constraints were employed to ensure the consistency of the data remains high.

Timeliness and relevancy were considered important data quality dimensions.

“the data needs to be contextually sound (timely and relevant) to satisfy the requirements of numerous departments within the organisation. For example, our airport operations department requires high quality data for time critical aircraft scheduling analysis”.

The data custodian therefore had actual associations with the *intrinsic* and *contextual* data quality categories.

#### *Data Consumers*

Data consumers were typically business analysts who required detailed and accurate data in order to make appropriate and sound business decisions. In general, data consumers believed that their perception of data quality was relative to the tasks they needed to perform.

“During any given time, the same data I receive from the data warehouse may be needed for multiple closely-linked tasks that require a different set of quality dimensions (timeliness, relevancy, consistency, and security) ... invariably we are chasing a continually moving target”.

Difficulty in accessing data was perceived as a significant concern. Data accessibility includes both the need for easy access to required data and the ease with which data can be manipulated (aggregated and formatted).

“... the data has to have intuitive appeal ... there must be a degree of ease with which I can manipulate the data to suit the business task that I am working with at any one time”.

The representational data quality category was seen as very important:

“the quality of the decisions we make are largely associated with the quality of the data we have to run with .... If the data we get isn't entirely consistent, easy to understand and concise, then the recommendations that we deliver to the business units are useless”.

Data consumers could request training about using the data warehouse at any time. The training concerned content and usage of the data warehouse. There were no standard programs offered periodically throughout the organisation.

Data producers therefore had actual associations with the *contextual*, *representational* and *accessibility* data quality categories.

#### *Data Manager*

The data manager had a broad view of data quality. Relevance was seen as the most important dimension of data quality and after that accuracy and well categorised data that minimises subjectivity on the part of those who collect the data. The data manager noted:

“If data categorisation is not addressed in a meaningful way then the use made of data aggregation mechanisms lose their appeal”.

Accessibility issues including ease of access to relevant data, ease of manipulation of that data and adequate training. Completeness was also seen as important.

“All the necessary values should be present to ensure the data consumer's requirements are met. Completeness cannot be overlooked in our role”.

The perception of high data quality by data consumers was seen as very important. If data consumers do not consider data to be of high quality, then they will use the data warehouse less, if at all.



Due to his broad perspective on data quality, the data manager had actual associations with all four data quality categories: the *intrinsic*, *contextual*, *representational* and *accessibility* data quality categories.

## CASE STUDY ANALYSIS

The case study analysis consists of two main parts: a discussion of similarities and differences between proposed and actual associations between stakeholders and data quality categories, and a discussion of important implications that emerge from the case study for practitioners.

### Proposed and Actual Associations between Stakeholders and Data Quality Categories

It was proposed that data producers would have associations with the *contextual* and *representational* data quality categories. The case study evidence confirmed these two associations and suggested a further association with the *intrinsic* data quality category. Clearly, dimensions within the contextual and representational categories are of central interest to data producers. They are concerned with the amount of data to include in the data warehouse and its consistency and ease of understanding. Data producers in this case study were also very aware of the need for accuracy and believability of the data. This was surprising, as these dimensions are normally the concern of data consumers. It demonstrates a strong commitment to data quality among the data producers. This commitment is also demonstrated in their call for training programs to improve the quality of data in source systems.

It was proposed that data custodians would have associations with the *intrinsic* and *accessibility* data quality categories. The case study evidence confirmed the association with the *intrinsic* data quality category but there was no evidence to support the association with the *accessibility* data quality category. A further association was suggested with the *contextual* data quality category. Accuracy and reputation of data are of central concern to data custodians. These intrinsic dimensions of data quality will always be “core business” for data custodians. It was surprising that accessibility and security were apparently of little concern to data custodians. One explanation is that the data warehouse has been operating for some time now and these aspects have been largely addressed. The concern of data custodians with the contextual dimensions of timeliness and relevancy demonstrates a good understanding of wider data quality issues on their behalf. It is an attempt to meet and exceed customer expectations of data quality (Kahn et al. 1997).

It was proposed that data consumers would have associations with the *contextual* and *accessibility* data quality categories. The case study evidence confirmed these two associations and suggested a further association with the *representational* data quality category. Contextual and accessibility data quality categories are highly relevant to the need to support the tasks undertaken by data consumers, in particular the ability to easily manipulate and aggregate data. However, it was clear that data quality was seen as a relative concept, dependent on the nature of the task at hand and the decision making style of the particular data consumer. Data consumers were also concerned with the consistency, ease of understanding and conciseness of data. These representational data quality dimensions were important in justifying and supporting decisions made using the data.

It was proposed that data managers would have associations with the *intrinsic*, *representational* and *accessibility* data quality categories. The case study evidence confirmed these three associations and suggested a further association with the *contextual* data quality category. The data manager in this case study was clearly very aware of a broad range of data quality issues. He was also strongly committed to developing the perception of high data quality amongst data consumers to ensure the data warehouse was widely used. In general, data managers need to understand the data

quality requirements of all other stakeholders and manage expectations and perceptions of data quality.

A summary of the actual associations is shown below in Figure 3.

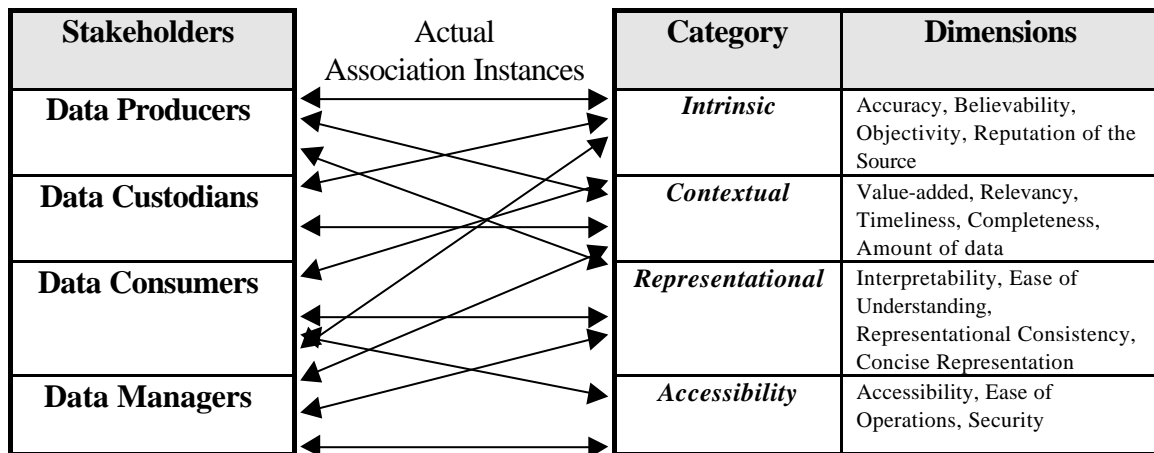


Figure 3: Actual relationship instances between stakeholder types and data quality categories and dimensions.

### Implications for Practice

Several important implications for practitioners emerge from the case study findings. These are:

#### *The Importance of Data Quality Needs to be Widely Understood*

Clearly, data quality issues are critical to the success of a data warehousing project. In this case study the data manager had a broad, clear understanding of data quality issues and had succeeded to some extent in communicating this knowledge to other stakeholders. The framework proposed in this paper provides a useful vehicle for planning the establishment of an awareness of data quality issues among relevant stakeholders. Knowing which data quality dimensions are important for particular stakeholders will help focus data quality initiatives within organisations.

Perceptions of data quality among data consumers is important to increased demand for data from the data warehouse. All stakeholders in the data warehousing process need to understand how they can improve the quality of data and therefore increase the level of perceived data quality. Data consumers need to be regularly surveyed for their views on the quality of data they are using. Increased usage of the data warehouse will also lead to feedback about data quality problems that need to be addressed (Orr 1996).

#### *Training in the Content and Usage of the Data Warehouse is Important*

Training in data quality for all stakeholders in the data warehousing process is important for improving data quality. On one level, training should be provided to develop a broad understanding of the scope of data quality as defined by the framework used in this paper. More focused training should be provided to particular stakeholders in the areas of data quality most relevant to them. In particular data consumers should be trained in the content of the data warehouse to increase data warehouse usage. Data providers and data custodians need to be aware of potential data quality problems and strategies for addressing those problems.

Training programs should be pro-actively organised rather than being demand-driven. In the case study, training programs were provided on request from data consumers. While the active data

consumers were generally aware of data warehouse contents and data quality issues, a more proactive training program would encourage expanded use of the data warehouse.

### *Reward Structures for Data Quality Compliance Should be Established*

Data producers commented that many data quality problems, in particular the intrinsic data quality problems, could be best fixed at the source of the data. In many cases this was in the databases and files of legacy application systems within the company and due to problems such as interpretation and coding errors at initial data entry or subsequent data transformation. Reward structures for data quality compliance for those people who perform initial data entry should be established. The resulting increase in intrinsic data quality will flow through to the data warehouse.

Data producers also noted that the source of some data quality problems was extremely difficult to find. Documenting information flows and process maps for an organisation's information systems would help data producers in this matter.

## **CONCLUSION**

Data quality is an important factor in the success of data warehousing projects. This paper has proposed a framework for understanding stakeholder data quality requirements that is used to define associations between stakeholder types and specific categories of data quality dimensions. A number of associations were proposed and then examined empirically in a case study. The case study led to a deeper understanding of stakeholder data quality requirements and suggested some changes to the proposed associations and some important implications for practitioners. The framework provides a useful basis for planning data quality initiatives and data quality training within organisations.

Areas for future research include:

- multiple case studies within different industry sectors to further develop the framework and to obtain a deeper understanding of stakeholder perceptions of data quality;
- improvements to the theoretical basis for explaining the associations between stakeholder types and specific categories of data quality dimensions;
- a survey of stakeholder perceptions of data quality requirements to extend the empirical basis of the framework.

## **REFERENCES**

- Benbasat, I., Goldstein, D.K. and Mead, M. (1987) The Case Research Strategy in Studies of Information Systems, *MIS Quarterly*, 11, 369-386.
- Firth, C.M. and Wang, R.Y. (1993). Closing the Data Quality Gap: Using ISO 9000 to Study Data Quality. TDQM Working Paper. Data Quality Research: A Framework, Survey, and Analysis. (TDQM-93-03) Total Data Quality Management (TDQM) Research Program, MIT Sloan School of Management, April 1993.
- Galliers, R.D. (1991) "Choosing Information Systems Research Approaches" in H-E Nissen, H.K. Klein, and R.A. Hirschheim (eds.) *Information Systems Research: Contemporary Approaches Emergent Traditions*, Proc. IFIP TC8/WG8.2 Working Conference, Dec. 1990, North-Holland, 327-346.
- Gartner Group (1997), *Four Myths About Data Warehouses*, Gartner Group Research Note.
- Kahn, B.K., Strong, D.M., and Wang, R.Y. (1997). A Model for Delivering Quality Information as Product and Service. *Proceedings of 1997 Conference on Information Quality*, (October), Massachusetts Institute of Technology, pp. 80-94.

- Kimball, R. (1996) *The Data Warehousing Toolkit*, John Wiley, New York.
- McFadden, F. (1996) Data Warehouse for EIS: Some Issues and Impacts, *Proc. 29th Annual Hawaii International Conference on System Sciences*, Hawaii, (January).
- Miles, M.B. and Huberman, A.M. (1984) *Qualitative Data Analysis: A Sourcebook of New Methods*, Sage Publications, Newbury Park.
- Myers, M.D. (ed.) (1998) *Qualitative Research in Information Systems*, URL  
<http://www.auckland.ac.nz/msis/isworld/>
- Orr, K. (1996) Data Quality and Systems Theory, *Proc. of the 1996 Conference on Information Quality*, Boston, Massachusetts, (October).
- Redman, T.C., (1992) *Data Quality: Management and Technology*. Bantam Books. New York
- Redman, T. (1998) The Impact of Poor Data Quality on the Typical Enterprise. *Communications of the ACM*, 41, 79-82.
- Shanks, G., O'Donnell, P. and Arnott, D. (1997) Data Warehousing: A Preliminary Field Study, *Proc. 8th Australasian Conference on Information Systems*, University of South Australia, Adelaide, (September).
- Shanks, G. and Darke, P. (1998a) Understanding Data Quality in a Data Warehouse, *The Australian Computer Journal*, 30, 122-128.
- Shanks, G. and Darke, P. (1998b) Understanding Data Quality in a Data Warehouse: A Semiotic Approach. *Proc. of the 1998 Conference on Information Quality*, Boston, Massachusetts, (October).
- Strong, D.M., Lee, Y.W. and Wang, R.Y. (1997) Data Quality in Context, *Communications of the ACM*, 40:5, 103-110
- Wand, Y. and Wang, R. (1996) Anchoring Data Quality Dimensions in Ontological Foundations, *Communications of the ACM*, 39, 86-95
- Wang, R.Y., Strong, D., and Guarascio, L. M. (1994) TDQM Working Paper. Data Quality Research: Beyond Accuracy: What data Quality Means to Data Consumers. (TDQM-94-10) Total Data Quality Management (TDQM) Research Program, MIT Sloan School of Management October 1994.
- Wang, Y. and Strong, D.M. (1996) Beyond Accuracy: What Data Quality Means to Data Consumers, *Journal of Management Information Systems*, 12, 5-34.
- Wang, R. (1998) A Product Perspective on Total Data Quality Management, *Communications of the ACM*, 41:2, 58-65
- Yin, R.K. (1994) *Case Study Research: Design and Methods*, (2<sup>nd</sup> ed.), Sage Publications

## **COPYRIGHT**

Anthony Giannoccaro, Graeme Shanks, Peta Darke ©1999. The authors assign to ACIS and educational and non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to ACIS to publish this document in full in the Conference Papers and Proceedings. Those documents may be published on the World Wide Web, CD-ROM, in printed form, and on mirror sites on the World Wide Web. Any other usage is prohibited without the express permission of the authors.