AN INVESTIGATION INTO DATA QUALITY ROOT CAUSE ANALYSIS

(Research-in-progress)

Philip Woodall
University of Cambridge
phil.woodall@eng.cam.ac.uk

Andy Koronios
University of South Australia
andy.koronios@unisa.edu.au

Jing Gao
University of South Australia
jing.gao@unisa.edu.au

Ajith Kumar Parlikad
University of Cambridge
ajith.parlikad@eng.cam.ac.uk

Elaine George
University of South Australia
Elaine.George@unisa.edu.au

Abstract: Data and information quality is a well-established research topic and gradually appears on the decision-makers' top concern lists. Many studies have been conducted on how to investigate the generic data/information quality issues and factors by providing a high-level abstract framework or model. As a result, a number of data and information quality methodologies and frameworks have been developed and made available to organisations. Across all examined methodologies and data quality improvement frameworks, this study found that although data quality root cause analysis (RCA) is regarded as an essential data quality improvement method, there is limited guidelines on how to conduct RCA to investigate data quality problems.

Key Words: Data quality, root cause analysis, root causes, data quality improvement, data quality methodology

INTRODUCTION

To begin, it is noted that data and information are often used synonymously particularly when addressing quality issues. In practice, managers differentiate information from data intuitively, and describe information as data that has been processed. Unless specified otherwise, this paper will use data interchangeably with information, as well as use Data Quality (DQ) interchangeably with Information Quality (IQ).

It is generally agreed now that a lot of intellectual property is locked in enterprise data repositories. Enterprise data and information is increasingly recognized as one of the most valuable, if not the most valuable and proprietary resources that enterprises possess. It is also true however that the value of this asset needs to be unlocked for the enterprise to use it to make better decisions and to gain competitive advantage. Data and information development and management is essential for improving or developing new contexts to support the management process and make, strategic decisions([29]; [40]; [45]). For example, managers derive information from data to enable them to make operational decisions related to produc-

tion, ordering and scheduling. Accountants use data to prepare financial statements and documents including financial reports. The importance of data has been increasing Information is an ever increasing important business resource that supports organisational decisions, and, therefore, management of the quality of this information is critical for organisational success.

As presented in existing literature and case studies, poor data quality (DQ) is often discussed as problems in organizations [19]. Regardless of the actual problem context, the typical problem solving exercise often involves Root Cause Analyses (RCA). Indeed, in industries such as manufacturing and aviation, RCA is a critical and compulsory activity for any problem solving exercises.

Although RCA had its genesis in engineering environments, many such techniques and methods have been used in a number of different contexts. Table 1 below lists the more frequently used RCA techniques and illustrates some of the limitations of each.

Method/ Tool	**Type**	**Defines Problem**	**Defines all causal relationships**	**Provides a causal path to root causes**	**Delineates evidence**	**Explains how solutions prevent recurrence**	**Easy to follow report**
Events and causal factors	Method	Yes	Limited	No	No	No	No
Change analysis	Tool	Yes	No	No	No	No	No
Barrier analysis	Tool	Yes	No	No	No	No	No
Tree diagrams	Method	Yes	No	No	No	No	No
Why-why chart	Method	Yes	No	Yes	No	No	No
Pareto	Tool	Yes	No	No	No	No	No
Story telling	Method	Limited	No	No	No	No	No
Fault tree	Method	Yes	Yes	Yes	No	Yes	No
Failure modes and effects analysis (FMEA)	Tool	Yes	No	Limited	No	Limited	No

Table 1 Comparison of selected RCA methods and Tools (Modified table from [18])

Within the data quality improvement context a number of additional RCA techniques have been used. These include Scatter diagram and Stratification/ Is/Is Not Analysis. [5] describes RCA within the DQ improvement context as a means of investigating and categorizing the root causes of IQ problems. It seems that the plethora of data quality improvement frameworks are deficient in their inclusion of RCA, yet only when we are able to determine why IQ problems have occurred, are we able to specify workable corrective measures that prevent future problems and result in sustained data quality improvement [5].

Although it is acknowledged that many DQ frameworks and methodologies are abstract in nature with little guidance on how these frameworks could be operationalized into prescriptions for data quality improvement and thus little guidance on performing RCA one of the most critical activities in preventing data quality errors from recurring. This research endeavours to address these shortcoming by investigating how RCA in data quality contexts is currently performed and how might this be improved.

Research Question and Design

With the above consideration, a two-stage study has been developed in order to answer the following research question:

To what extent is RCA, as is recommended and described in DQ improvement methodologies, sufficient to extract the root causes of DQ problems?

The following sub-questions will be investigated in order to answer the research question

1. To what extent is RCA emphasised in DQ improvement methodologies?;
2. Are there sufficient guidelines for performing DQ RCA?; and
3. Does DQ RCA differ from generic RCA in real world practices?

This paper will focus in reporting on the findings of the first stage of this project.

The first stage of this study will examine a number of the most representative generic data quality frameworks and improvement methodologies in order to determine the role and importance of RCA. By a close investigation of these frameworks and methodologies, we also try to extract the DQ RCA guidelines which are currently available to data quality practitioners.

The second stage of this study will involve actual case studies. A number of organisations who have implemented DQ initiatives will be chosen with a particular focus on how they conducted data quality root cause analysis and the results.

This study is intended to increase understanding on how data quality root cause analysis is conducted in practice and highlight the problems and challenges in performing root cause analysis for data quality improvement. This study may also deliver a common root causes for some common data quality problems shared by organisations world-wide. More importantly, it may help the academic society enhance frameworks and methodologies for better adoption.

Within the data quality context, [5] describes RCA as a means to "Investigate and categorize the root causes of IQ problems. It is only when the reasons for each data quality error that exist have been identified that it will be possible to determine why IQ problems occurred and thus recommend appropriate corrective measures that prevent future data quality errors. This study attempts to obtain an insightful understanding on how RCA is conducted in DQ initiatives.

DATA QUALITY FRAMEWORKS AND IMPROVEMENT METHODOLOGIES

As the field of data quality developed a large number of data quality frameworks and improvement methodologies have emerged (e.g., [47]; [50]; [48]; [44]; [39]; [24]; [20]; [7]; [12]; [35]; [22]; [15]). These frameworks have attempted to organize and structure important issues in data quality from a number of different perspectives. Some of these are generic frameworks whilst others are specific to particular domains, for example health or asset management.

An analysis of the most well known data quality frameworks and improvement methodologies between 1990 and 2011 was identified through a literature review (the results are listed in Table 1). These frameworks and methodologies are constructed to provide a comprehensive coverage of IQ problems, related activities, and context-driven IQ dimensions [46]. Additionally, these IQ frameworks can also be used as knowledge resources to provide guidelines on how to ensure information quality for various environments ([37]; [46]).

Author (Year of Publication)	**Area**
[2]	Management Information Systems
[43]	Newspapers
[28]	Corporate Communications
[49]	DQ Research
[41]	Data Bases
[32]	Data Quality Methodology
[38]	Data Quality Methodology
[33]	Information Systems
[47]	Information Systems
[10]	Information Management
[13]	Corporate Communications
[3]	Data Warehouses
[23]	Information Systems
[21]	Knowledge Management
[6]	Data Bases
[11]	Data Quality Methodology
[42]	Data Quality Methodology
[24]	Information Systems
[27]	Information Systems
[31]	Corporate House holding
[34]	Information Systems
[4]	Data Quality Improvement
[16]	Data Quality Methodology
[14]	Data Quality Assessment
[8]	Portal Data Quality
[17]	Data Quality Metrics
[26]	Sensor Data
[9]	A quality framework to evaluate E-Government service delivery
[36]	Data Quality in Data Warehousing
[25]	Enterprise Knowledge Management
[30]	Data Quality Methodology
[48]	Data Quality Methodology
[1]	Information Quality Framework for e-Learning Systems

Table 2: Information Quality frameworks

Some of the above frameworks and methodologies are generic and others focus on a special area or industry. Nevertheless, these DQ frameworks will help researchers and practitioners to obtain an in-depth understanding of various data quality issues.

IMPROVEMENT METHODOLOGIES THAT INCLUDE RCA

From the above list, we have identified the DQ methodologies that recommend conducting RCA, and these are shown in Table 3. Interestingly, many of these are from leading experts in the DQ area who have produced methodologies that have gained wide acceptance. RCA is therefore not a peripheral topic in DQ improvement proposed by a one-off methodology. Rather, it has been independently agreed by experts as a necessary step to DQ improvement.

In the following subsections, RCA is discussed within each methodology giving attention to the sub research questions 1 and 2. In particular, in relation to sub question 1, we discuss where RCA has been included in the methodology and whether it is considered an important part. After reviewing all the RCA instances, the issue of whether RCA is used consistently is discussed.

For sub question 2, the issues of how detailed the RCA guidelines are, what methods are provided for RCA and whether the guidelines have been tailored to DQ problems are also discussed for each methodology.

Methodology reference name	Ref	Placement of RCA	RCA importance	Suggested RCA methods	Level of detail
TQdM	[6]	Informs the corrective improvement actions to implement	RCA results are critical to help develop improvement actions	Pareto chart analysis, Five whys, Cause-effect diagram,	Gives details on cause-effect diagram
TDQM	[48]	Uses the DQ measurement results to identify root causes before improving the Information Product	RCA results are critical to help develop improvement actions	SPC, Pattern recognition, Pareto chart analysis, Introduce dummy records	High level of abstraction
SODQA	[32]	Uses the DQ measurement results to identify root causes before informing what corrective improvement actions to implement	RCA results are critical to help develop improvement actions	none suggested	High level of abstraction
DQFG	[38]	Informs the corrective improvement actions to implement	RCA results are critical to help develop improvement actions	none suggested	High level of abstraction
EDQP	[30]	Informs the corrective improvement actions to implement	RCA is used in multiple steps of the methodology	Five whys, Cause-effect diagram, Track and trace	Most detailed
CSDQ	[11]	Uses the DQ measurement results to identify root causes before improving the Information Product	RCA results are critical to help develop improvement actions	Cause-effect diagram, Causes table	Some detail, but still high level
AMEQ	[42]	Identical to TDQM	Identical to TDQM	none suggested	Very high level

Table 3: DQ Methodologies that recommend root cause analysis

RCA in TQdM

The TQdM methodology actually provides two approaches to DQ improvement: process-driven and data-driven. The data-driven approach does not suggest doing RCA and it is the process-driven approach that includes RCA as one of its steps. The steps in the process-driven approach are shown in Table 4 and the RCA is recommended as part of the 'Develop plan for DQ improvement' step. This is a critical step because it leads to the implementation of the corrective actions, which aims to modify the organisation's data, and the root causes are needed to determine what actions can and should be taken to address the problem(s). Clear guidelines and examples are given showing how to use the cause-effect diagram for DQ problems. The Pareto diagram and the five whys are also mentioned.

Step	Description of Step	Link to study [6]
Select process for DQ improvement	Identify a process where improvements can prevent business problems that cause DQ problems.	p289 step 1
Set up a DQ team	Identify a person responsible for resolving DQ problems, a project sponsor, and team members to facilitate the DQ improvement process.	p292 step 1, point 1, 2 and 3.
Develop plan for DQ improvement	Identify the root causes of a DQ problem and identify corrective actions to eliminate/minimise the causes.	p293 step 2
Implement DQ improvements	Implement improvement actions in a controlled manner to improve DQ to verify that the recommended improvements do solve the real problem.	p298 step 3
Check impact of DQ improvements	Verify the effectiveness of DQ improvement actions.	p299 step 4
Act to standardise DQ improvements	Make DQ improvements a baseline habit.	p300 step 5

Table 4: TQdM process-driven approach

RCA in TDQM

The well known TDQM methodology suggests viewing data as an information product (IP) that is manufactured through various data changing and storage processes. The steps of TDQM are shown in Table 5, and RCA sits between the measurement and improvement steps. The paper describing the approach is pitched at a high level of abstraction and therefore there is no detailed guidance for how to conduct RCA. Statistical process control (SPC), pattern recognition, pareto chart analysis, and introduce dummy records, are recommended as possible approaches that can be used. There is also no explicit mention of how RCA needs to be tailored to determine root causes of DQ problems.

Step	Description of Step	Link to study [48]
Define IP	Define the characteristics for the information product	Define IP p61
Measure IP	Develop and measure DQ metrics	Measure IP p64
Analyse IP	From the measurement results, identify root cause(s) of DQ problems	Analyze IP p64
Improve IP	Identify key areas for improvement such as: (1) aligning information flow and work flow with the corresponding information manufacturing system, and (2) realigning the key characteristics of the IP with business needs	Improve IP p65

Table 5: TDQM steps

RCA in SODQA

In the SODQA methodology, similar to TDQM, the RCA step is between the DQ assessment and the actual improvement of DQ (Table 6). The results from the RCA are fed into the 'improve DQ' step, which executes the necessary DQ improvement actions. For both TDQM and SODQA, RCA appears centrally and directs the DQ improvement actions. RCA is therefore identified as being necessary, but no

specific guidelines are given as to the best practice methods of conducting RCA or whether it needs to be tailored to DQ.

Step	Description of Step	Link to study [32]
Perform a subjective and objective data quality assessment	The subjective and objective assessments of a specific DQ dimension are compared	p215 Assessments in Practice 1st bullet point
Determine root causes of discrepancies	Comparing the results of the assessments, identifying discrepancies, and determining root causes of discrepancies	p215 Assessments in Practice 2nd bullet point
Improve DQ	Determine what the improvement options are and taking necessary actions for improvement	p215 Assessments in Practice 3rd bullet point

Table 6: SODQA steps

RCA in DQFG

RCA in the DQFG methodology is considered an important part to inform the development of solutions to the DQ problems, similar to the previously described methodologies above. On a more detailed level, the RCA step is positioned directly after setting up a DQ team (Table 7), and this is exactly the same as in TQdM. DQFG suggests that this team should be built with people who know and are involved with the 'problem processes' causing DQ problems, and that people who are close to the problem will be in the best position to determine root causes.

Step	Definition of Step	Link to study [38]
Select project	Identify and select a project that the DQ improvement will focus on.	p133 fig22.2 step 1
Form and charter project team	Setup a team of people to carry out the DQ project.	p133 fig22.2 step 2
Conduct root cause analysis	Find the root causes of the DQ problems	p133 fig22.2 step 3
Identify solutions to the DQ problem	Identify simple DQ improvement solutions as a starting point.	p133 fig22.2 step 4
Trial simple solutions to the DQ problem	Trial simple solutions with the aim of demonstrating that the solutions work before rolling out a full implementation.	p133 fig22.2 step 4
Implement solution	Roll out the solution once it has been proven successful	p133 fig22.2 step 5
Hold the gains	Confirm that the solution works and ensure that the problem does not recur	p133 fig22.2 step 6

Table 7: DQFG steps

RCA in EDQP

The steps of the EDQP methodology are shown in Table 8. RCA in EDQP comes after both the assess DQ and assess business impact steps, which can be done in parallel and provide input to the root cause analysis step. As with the other methodologies, RCA is followed by the development and execution of the improvement plans. RCA features heavily in this methodology and the results of RCA are used in multiple steps. Similar to TQdM, this methodology contrasts a process approach with a data-driven approach, and the process-driven step (prevent future data errors) includes RCA, whereas the data-driven step (correct current data errors) does not. This methodology gives the most detailed guidelines and makes sensible suggestions on how to carry out three RCA methods for DQ problems. The three suggested methods are five whys, track and trace, and cause-effect diagrams. Five whys and the cause-effect diagrams are commonly used methods. Track and trace is defined as "a way to identify the location of the problem by tracking data through the information life cycle and determining root causes where the problem first appears" [30].

Step	**Description of step**	**Link to study [30]**
Define business need and approach	Determine why the DQ improvement is important to the business and plan the project	p57 fig 2.14 step 1
Analyse information environment	Understand the environment so that future steps benefit from increased knowledge of the context.	p57 fig 2.14 step 2
Assess DQ	Provide a picture of the actual quality of the data using suitable DQ dimensions	p57 fig 2.14 step 3
Assess business impact	Used to determine the impact of the DQ problems on the business	p57 fig 2.14 step 4
Identify root causes	Identify root cause(s) of DQ problems	p57 fig 2.14 step 5
Develop improvement plans	Develop alternative DQ improvement options/remedies. For example, an option might be to update the company database more frequently or distribute the updates to remote sites more often. Another option could be to perform data cleansing on the database at selected time intervals. These could be data-oriented or process-oriented approaches	p57 fig 2.14 step 6
Prevent future data errors	Implement appropriate solutions that address the root causes of the DQ problems	p57 fig 2.14 step 7
Correct current data errors	Implement solutions that correct the existing data errors	p57 fig 2.14 step 8
Implement controls	Implement ongoing monitoring and metrics, and verify the improvements that were implemented	p57 fig 2.14 step 9

Table 8: EDQP steps

RCA in CSDQ and AMEQ

CSDQ is an extension of TDQM that incorporates a user-centered approach to improving the quality of customer support data; CSDQ therefore also includes the RCA step. This step has not been moved or modified and still sits between the analyse and improve steps of TDQM (see Figure 2 in [11]). As well as suggesting the use of cause-effect diagrams, the authors of this approach suggest a new method for RCA called the causes table (see Table 9). This is a simple method that aims to be lightweight and fast to ap-

ply. In the 'caused by' column of Table 9, a double tick indicates the most significant cause.

AMEQ also contains the same steps as TDQM, but does not expand RCA from the TDQM methodology or describe any methods to conduct RCA.

DATA QUALITY ATTRIBUTE	PRIMARY OR SECONDARY	ATTRIBUTE IMPORTANCE	CAUSED BY		
			People	Tools	Process
Believability	Secondary				
Accuracy	Primary	High	✓	✓✓	
Reputation	Secondary				
Relevancy	Secondary				
Value-added	Secondary				
Completeness	Primary	High	✓	✓	✓✓
Appropriate amount of data	Primary	Medium	✓✓	✓	✓
Interpretability	Secondary				
Ease of under-standing	Secondary				
Consistent rep-resentation	Primary	Low		✓	
Accessibility	Primary	High		✓	

Table 9: Causes table (reproduced from [11])

DISCUSSION/RESULTS

The above findings are used to answer the proposed research question as presented below:

To what extent is RCA emphasised in DQ improvement methodologies and frameworks?

In total, 7 data quality methodologies were selected as examples of the most influential methodologies. Among these DQ methodologies, RCA has been emphasised as an important activity to enable the improvement actions. Typically, RCA uses the DQ measurement results to identify root causes before informing what corrective improvement actions to implement. The Cause-effect diagram is suggested across many methodologies as the most common RCA method. It must be pointed out that in the TQdM, RCA is a critical part of the process-driven approach (to uncover the real causes of DQ problems).

Sub-question 2: Are there sufficient guidelines available that describe how to conduct DQ RCA?

In addition to the most commonly suggested cause-effect, some common RCA methods are also recommended including: Pareto chart analysis, Five whys, Cause-effect diagram, Statistical process control (SPC), Pattern recognition, Introduce dummy records, Track and trace and Causes table.
TQdM and EDQP provide relatively detailed guidelines and examples that relate to DQ. All the selected methodologies have not provided clarifications on how to determine and priories the actual DQ root causes from the potentially large number of possible data quality causes found at each organisational level and function.

It must be noted that the current guidelines as summarised in the previous section are very generic, without being tailored specifically to the DQ problems. Whether or not this is required will be investigated in future research.

CONCLUSION

This preliminary study has identified the most influential DQ methodologies and frameworks which highlight the importance of data quality RCA and it forms a key part of the DQ methodology. However, given that it is such an important area for DQ improvement, there has been little attention focussed on some critical aspects, including:

- Determining whether RCA in the DQ context is any different to other contexts where it has been used traditionally, and what issues this raises;
- Differentiating the generic problem context and the DQ problem context, which can make the selection of the appropriate RCA difficult (e.g. control chart may not be very useful for DQ problems);
- DQ problem related elements and aspects (e.g. different DQ problem scope may lead to different root causes (e.g. do certain DQ dimensions always lead to similar root causes, or require specific methods to be used?);
- Detailed guidelines on how DQ RCA should be conducted and what the root causes should look like;
- When to apply different methods (e.g. the selection of a RCA technique may result in different root causes).

This study has found that there are limited case studies on how organisations thoroughly conduct data quality RCA in real practices. Thus, future research will seek verifications on the above mentioned issues with industry case studies. It is felt that in any future academic study on data quality frameworks and methodologies, developing sufficient guidelines for RCA will enhance the research outcome and applicability.

ACKNOWLEDGEMENTS

This research has received supported from EPSRC, project reference number EP/G038171/1.

REFERENCES

[1] Alkhattabi, M., Neagu, D., Cullen "An Information quality framework for e-learning systems." *Knowledge Management & E-Learning: An International Journal*, 2 (4). 2010. pp. 340-362.

[2] Augustin, S., Reminger, B. "Trotz Datenflut jede Menge Informationsdefizit! - Ist das erfolgreiche JIT-Konzept auch in der Info-Welt realisierbar." *Der informierte Manager*, ed. H. Bäck. TÜV Rheinland, Köln, Germany. 1990. pp. 73-82.

[3] Ballou, D. P., Wang, R., Pazer, H., Tayi, G. K. "Modeling information manufacturing systems to determine information product quality." *Management Science*, 44(4). 1998, pp. 462- 484.

[4] Blechar, M., Friedman, T. *Improve Application and Data Quality via Application Development Best Practices*. Gartner. 21 November, 2005, Accessed on 11 June, 2012. Available at: http://www.gartner.com/id=486846

[5] Baškarada, S., Koronios, A., Gao, J. "Towards a Capability Maturity Model for Information Quality Management: A TDQM Approach." *Eleventh International Conference on Information Quality*. Cambridge, MA, USA. 2006.

[6] Batini, C., Cappiello, C., Francalanci, C., Maurino, A. "Methodologies for Data Quality Assessment and Improvement." *ACM Computing Surveys*, 41 (3). 2009. pp.1–52.

[7] Caballero, I., Piattini, M. "CALDEA: A Data Quality Model Based on Maturity Levels." *Third International Conference on Quality Software*. Dallas, Texas, USA. 2003. pp. 380-387.

[8] Caro, A., Calero, C., Piattini, M. "A Portal Data Quality Model For Users And Developers." 12th *International Conference on Information Quality*. MIT, Cambridge, Massachusetts, USA. 2007. pp. 462-476.

[9] Corradini, F., Hinkelmann, K., Polini, A., Polzonetti, B. "C2ST: A Quality Framework to Evaluate e-

Government Service Delivery**.”** *8th International EGOV Conference.* Linz, Austria. 2009, pp. 74 – 84.

[10] Davenport, T. *Information Ecology: Mastering the Information and Knowledge Environment.* Oxford University Press, Oxford. 1997.

[11] English, L. *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits*. John Wiley & Sons. 1999.

[12] Eppler, M. J. “A Generic Framework for Information Quality in Knowledge-Intensive Processes.” *6th International Conference on Information Quality*. MIT, Cambridge, Massachusetts, USA. 2001. pp.329-346.

[13] Eppler, M. “Information oder Konfusion – Neue Kriterien für die betriebliche Kommunikation.” *io management*, 5. 1997. pp. 38-41.

[14] Even, A., Shankaranarayanan, G. “Value-driven data quality assessment.” *10th International Conference on Information Quality*. MIT, Cambridge, Massachusetts, USA. 2005. pp. 221-236.

[15] Firth, C. P. “Data quality in practice: Experience from the frontline.” *Conference on Information Quality*. MIT, Cambridge, Massachusetts, USA. 1996.

[16] Friedman T. *Data quality methodologies: Blueprints for data quality success*. Gartner. 26 July, 2005. Accessed on 11 June, 2012. Available at: http://www.gartner.com/id=483843

[17] Friedman, T. *Best Practices for Data Stewardship*. Gartner. 3 December, 2007. Accessed on 11 June, 2012. Available at: http://www.gartner.com/id=554646

[18] Gano, D. L. *Apollo Root Cause Analysis: A New Way of Thinking*, 3rd Edition. Apollonian Publications, USA. 2008. Accessed on 25 June, 2012. Available at: http://www.realitycharting.com/_public/site/files/pdf/ARCA_Appendix.pdf

[19] Gao, J., Koronios, A., Kennet, S., Scott, H. “Business Rule Discovery through Data Mining Methods”. *Engineering Asset Management Review: Definitions, Concepts and Scope of Engineering Asset Management*, ed. J.E. Amadi-Echendu, K. Brown, R. Willet & J. Mathew, Springer, London. 2010. pp. 159-162.

[20] Giannoccaro, A., Shanks, G., Darke, P. “Stakeholder perceptions of data quality in a data warehouse environment.” *Australian Computer Journal*, 31(4). 1999. pp. 110-117.

[21] Harris, K., Fleming, M. *KM and Content Quality: What Can You Trust.* Gartner Group Research Note. 29 December, 1998.

[22] Jarke, M., Jeusfeld, M.A., Quix, C., Vassiliadis, P. “Architecture and Quality in Data Warehouses.” *10th International Conference on Advanced Information Systems Engineering*. Pisa, Italy. 1998. pp. I93-113.

[23] Kahn, B. K., Strong, D. M. “Product and Service Performance Model for Information Quality: An Update.” *Conference on Information Quality*. MIT, Cambridge, Massachusetts, USA. 1998.

[24] Kahn, B., Strong, D.M., Wang, R.Y. “Information Quality Benchmarks: Product and Service Performance.” *Communications of the ACM*, 45(4). 2002. pp. 184-192.

[25] Keenan, S.L., Simmons, T. “CSDQ: A User-Centered Approach to Improving the Quality of Customer Support Data.” *10th International Conference on Information Quality*. MIT, Cambridge, Massachusetts, USA. 2005.

[26] Klein, A. “Incorporating quality aspects in sensor data streams.” *First Ph.D. Workshop in conjunction with the Sixteenth ACM Conference on Information and Knowledge Management*. Lisbon, Portugal. 2007. pp: 77-84.

[27] Lee, Y. W., Strong, D. M., Kahn, B. K., Wang, R.Y. “AIMQ: A Methodology for Information Quality Assessment.” *Information & Management*, 40(2). 2002. pp. 133-146.

[28] Lesca, H., Lesca, E. *Gestion de l’information, qualité de l’information et performances de l’entreprise*. Paris. 1995.

[29] Levitin, A.V., Redman, T. C. “Data as a resource: properties, implications, and prescriptions.” *Sloan Management Review*, 40(1). 1998. pp. 89-101.

[30] Loshin, D. *Enterprise Knowledge Management: The Data Quality Approach*. Morgan Kaufmann Pub. 2001.

[31] Madnick, S., Wang, R. Y., Zhang, W. A. "Framework for Corporate Householding." *Seventh International Conference on Information Quality*. MIT, Cambridge, Massachusetts, USA. 2002. pp. 36-40.

[32] McGilvray, D. *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information*. Morgan Kaufmann. 2008.

[33] Miller, H. "The multiple dimensions of information quality." *Information Systems Management*, 13(2). 1996. pp.79-82.

[34] Naumann, F., Freytag, J., Leser, U. "Completeness of Integrated Information Sources." *Information Systems*, 29 (7). 2004. pp. 583-615.

[35] Nauman, F., Roth, M. "Information Quality: How good are off-the-shelf DBMS." *Ninth International Conference on Information Quality*. ed. S. Chengulur-Smith, L. Raschid, J. Long & C. Seko , MIT, Cambridge, Massachusetts, USA. 2004. pp. 260-274.

[36] Nemani, R. R., Konda, R. "A Framework for Data Quality in Data Warehousing." *Third International United Information Systems Conference*, 20(1). Sydney, Australia. 2009, pp. 292-297.

[37] Parker, M.B., Moleshe, V., De la Harpe, R. Wills, G.B. "An evaluation of information quality frameworks for the world wide web." *8th Annual Conference on WWW Applications*. Bloemfontein, Free State Province, South Africa. 2006.

[38] Pipino, L.L., Lee, Y.W., Wang, R.Y. "Data Quality Assessment." *Communications of the ACM*, 45 (4). 2002. pp.211–218.

[39] Price R.J., Shanks, G. "A semiotic information quality framework." *IFIP International Conference on Decision Support Systems: Decision Support in an Uncertain and Complex World*. ed. R. Meredith, G. Shanks, D. Arnott and S. Carlsson, Prato, Italy. 2004. pp. 658-672.

[40] Redman, T. C. *Data Driven: Profiting from Your Most Important Business Asset*. Harvard Business School Press, Boston. 2008.

[41] Redman, T. C. *Data quality for the information age*,1st Edition. Artech House, Boston, MA. 1996.

[42] Redman, T.C. *Data Quality: The Field Guide*. Digital Press, Boston. 2001.

[43] Russ-Mohl, S. *Der I-Faktor*. Osnabrück, Fromm. 1994.

[44] Shanks, G., Darke, P. "Understanding Data Quality in Data Warehousing: A Semiotic Approach." *Third International Conference on Information Quality*. ed. I. Chengilar-Smith and L. Pipino, MIT, Boston. 1998. pp. 247-264.

[45] Strong, D. M., Lee, Y. W., Wang, R. Y. "Data quality in context." *Communications of the ACM*, 40 (1). 1997. pp. 103-110.

[46] Stvilia, B., Gasser, L., Twidale, M.B., Smith, L.C. "A Framework for Information Quality Assessment." *Journal of the American Society for Information Science and Technology*, 58(12). 2007. pp 1720 -1733.

[47] Wand, Y., Wang, R.Y. "Anchoring Data Quality Dimensions in Ontological Foundations." *Communications of the ACM*, 39 (11). 1996. pp. 86-95.

[48] Wang, R.Y. "A Product Perspective on Total Data Quality Management." *Communications of the ACM*, 41 (2). 1998. pp.58–65.

[49] Wang, R.Y., Storey, V., Firth, C. "A framework for analysis of data quality research." *IEEE Transactions on Knowledge and Data Engineering*, 7(4). 1995. pp. 623-640.

[50] Wang, R. Y., Strong, D. M. "Beyond Accuracy: What Data Quality Means to Data Consumers." *Journal of Management Information Systems*, 12 (4). 1996. pp. 5-34.