Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2008 Proceedings

Americas Conference on Information Systems (AMCIS)

2008

A Framework for Classification of the Data and Information Quality Literature and Preliminart Results (1996-2007)

M. Pamela Neely
Rochester Institute of Technology, pneely@saunders.rit.edu

Jack Cook
Rochester Institute of Technology, jackcook@hotmail.com

A Framework for Classification of the Data and Information Quality Literature and Preliminary Results (1996-2007)

M. Pamela Neely
Rochester Institute of Technology
pneely@saunders.rit.edu

Jack Cook
Rochester Institute of Technology
Jackcook@hotmail.com

ABSTRACT

The value of management decisions, the security of our nation, and the very foundations of our business integrity are all dependent on the quality of data and information. However, the quality of the data and information is dependent on how that data or information will be used. This paper proposes a theory of data quality based on the five principles defined by J. M. Juran for product and service quality and extends Wang et al's 1995 framework for data quality research. It then examines the data and information quality literature from journals within the context of this framework.

Keywords

Literature assessment, data quality, information quality, management information systems (MIS), research framework

INTRODUCTION

Quality data and information form the foundations of the information systems used to manage internal operations and external relationships within the supply chain. Service-oriented architecture (SOA) and multi-channel databases demand high quality data information in order to be of use to decision makers who must rely on these systems [9, 32, 54]. As a result, data quality has increasingly become a focus of both academic research and practitioner concerns [4, 10, 15, 18, 21, 27, 51, 74].

Researchers widely recognize that synthesizing existing studies is an extremely important activity for advancing scholarly understanding of the current state of a discipline, validating theoretical underpinnings and suggesting future research directions. Few have synthesized the data and information quality (DIQ) sub-discipline into an overarching framework. This study classifies the intellectual development of DIQ research utilizing a multi-faceted assessment of DIQ articles published over a period of 11 years (1996-2007).

This paper is organized as follows: The development of the data quality framework will first be discussed. Subsequently, the collection and analysis of data will be described in the methodology section followed by a discussion on the findings of this research. Finally, areas for future research will be described based on the outcome of this literature review.

FRAMEWORK DEVELOPMENT

The distinct differences between data and other goods must be a consideration when determining data and information quality rather than product or service quality [39]. A product or service is exhausted in its use. Data, on the other hand, is not depleted in its use. Data elements can be exploited simultaneously by multiple users and still be available for employment within a different context by subsequent users. This data characteristic is significant in the fitness for use discussion

Fitness for Use- Juran

J. M. Juran is credited with first coining the term "fitness for use" and has developed a short list of inputs that companies, organizations, and individuals can use for determining a product or service's fitness for use. The questions for consideration are [76, 77]:

- Who are the users of the product or service? (Who?)
- What are the economic resources of both the producer and the user? (What?)
- How will the product or service be used? (How?)
- What are the user's specific determinants of a product or service's fitness for use? (Economic Benefits)

- What is the possibility and/or probability of endangering humans? (Human Safety)

  Human safety in some industries such as healthcare and aviation are very important. In other industries, privacy and security are concerns that impact human safety.

Information as a Product- Wang

Wang et al reviewed the early assessments of data quality literature through 1995 [69], building upon research which proposed that information should be treated as if it were a physical product. Their research describes a framework for data quality that is comprised of seven key elements: management responsibilities, operation and assurance cost, research and development, production, distribution, personnel management and legal function. These categories have become the building blocks of a contemporary classification schema for data quality literature, that extends Wang's framework with Juran's fitness for use concepts.

A NEW FRAMEWORK

The data quality framework described by Wang et al provides insight as to the research on data and information quality across the life cycle of the data. By blending Juran's five principles with Wang's framework, a matrix can be built that permits research categorization. As seen in Table 1, Wang's criteria are vertically positioned while Juran's principles are horizontally placed; thus, allowing research to be categorized on two dimensions- the point in the life cycle, and how the user will define quality.

	Who?	**What?**	**How?**	**Economic Resources**	**Human Safety**
Management Responsibilities					
Operations & Assurance Costs - Information Systems					
Operations & Assurance Costs – Database					
Operations & Assurance Costs - Accounting					
Research & Development - Analysis & Design					
Research & Development - Quality in Information Systems Design					
Research & Development - Dimensions & Measurement					
Production					
Distribution					
Personnel Management					
Legal Function					

Table 1- Empty Framework

The primary objective of this research was to systematically code DIQ articles from 1995 to 2007 by utilizing the proposed extended framework to identify areas within the data and information quality realm that have been insufficiently researched. Because the framework has been designed to classify DIQ literature, the criteria for placing articles in the various cells needed to be determined. Thus, the framework cells were populated in a logical, two-phase approach: Various types of research questions that could be asked at each characteristic intersection were determined. The researchers generated a list of questions that evolved from the calls for papers from the Journal of Data and Information Quality, as well as the International

Conference on Information Quality. The researchers also considered questions related to their own prior experience with data and information quality literature. By altering the framework, a more consistent process was created for coding the preliminary set of journal articles. The end result provided a rich repository of potential questions related to data and information quality. The next section briefly describes the broad categories of the framework.

DIQ RESEARCH FRAMEWORK

At the intersection of the Wang (information as product) and Juran (fitness for use) characteristics, research questions that would be described by these characteristics can be identified. Due to space limitations, there will be no further discussion of the research questions. The framework found in Table 1 is distributed over Tables 2 – 6 with a complete listing of appropriate research questions included in these tables.

Management Responsibilities	
Framework Categories	**Description**
Who?	Who is responsible for policies, requirements and the DQ system? Upper management approval (and the necessity of it).
What?	What policies and procedures need to be in place for the system to be effective? Methods of creating policies and procedures; noneconomic impact and role of information quality on business, work processes and strategy; standards and policies for ensuring information integrity for future generations.
How?	How the policies and requirements are implemented. How data quality systems are integrated into the overall corporate structure.
Economic Resources	Budgeting; impact and role on a firm's overall operational or economic performance; strategic cost impact.
Human Safety	Policies and procedures (or lack thereof) specifically dealing with human safety (for example, privacy; how is human safety ensured).

Table 2- Management Responsibilities

Operation & Assurance Costs	
Framework Categories	**Description**
Who?	Who monitors the costs of data quality in system XYZ*?
What?	What are the costs associated with DQ in system XYZ?
How?	How are DQ costs measured; ROI? How are costs gathered and measured?
Economic Resources	Are costs (and the decisions to enhance data quality) affected by economic resources? How do economic resources factor into decision making re: DQ; tradeoffs, cost of PR?
Human Safety	What are the O & A costs associated with human safety? Does the fact that DQ affects human safety make a difference in the O & A costs?
*System XYZ is to be replaced by information system, database or accounting system	

Table 3- Operation and Assurance Costs

Research and Development		
	Framework Categories	**Description**
Semantics & IP	Who?	Who are the players in the DQ of information products (IP)? Who is affected by IP?
	What?	What are information products and conceptual models of IP; semantics? What are measures of IP quality?
	How?	How are IP models implemented? How does DQ relate to IP? ERDs? How are IP metrics used? How is the quality of IP assessed?
	Economic Resources	Cost and benefits of IP models. ROI of IP models.
	Human Safety	Does the identification of IP and the associated quality issues affect human safety?
Verification Tools	Who?	Who is responsible for ensuring that DQ is a part of the IS design. Who are the DQ tool vendors? Who is responsible for evaluation and purchase of DQ tools?
	What?	Data lineage and provenance; new types of database systems that manage data uncertainty (approximate, probabilistic, inexact, incomplete, imprecise, fuzzy, inaccurate data). What DQ abilities should be built into KM, SCM, CRM, extended-enterprise management, global management systems? Technical solutions of IS.
	How?	Entity management; entity resolution; record linking; enterprise architecture deployment; incorporating data quality into KM, SCM, CRM; extended-enterprise management; global management; new ways of understanding; modeling; improving and incorporating information quality.
	Economic Resources	What economic costs are associated with including DQ as part of the IS? What do DQ tools cost? Cost/benefit analysis of DQ tools?
	Human Safety	Will incorporating DQ into the design of IS prevent harm to humans? Data privacy and protection mechanisms; preventing access to systems.
Dimensions & Measurement	Who?	Who cares about dimensions and measurement? Fitness for use?
	What?	What are the dimensions? What measurements have been defined? What causes data errors?
	How?	How are dimensions and measurements used? Is there a relationship between how the data is used and quality dimensions or measurements? How are data errors detected or corrected?
	Economic Resources	Does the necessary economic resources change with the dimension (i.e. is it more costly to be timely than accurate)? What is the relationship between multiple dimensions and cost (i.e. as you add dimensions to be considered is it a straight line increase or exponential)?
	Human Safety	Will the ability to effectively measure DQ affect human safety? Are some dimensions more critical to human safety than others?

Table 4- Research and Development

Production and Distribution		
	Framework Categories	**Description**
Production	Who?	Who is responsible for the production of data? Who is responsible for assigning data tags? Who determines if the process produces correct data?
	What?	What are the production processes? What processes are needed to ensure that data is correct? What data tags are meaningful? What is data integrity? What ETL tools exist? What causes data errors?
	How?	How are processes created and put in place to ensure correct data as output? How is data cleaning achieved? How do you implement data integrity? How should an ETL tool be controlled and executed? How are data errors detected and corrected?
	Economic Resources	Cost/benefit analysis of data production. Is more (data) always better? Cost/benefit analysis of data tags? How expensive is an ETL tool?
	Human Safety	If the processes for data production are incorrect, will it impact human safety?
Distribution	Who?	Who decides what metadata is needed? Who is responsible for overseeing distribution? Who is responsible for data integration- producer or consumer?
	What?	What metadata is needed? What should be in a metadata tool? What processes/procedures should be a part of the distribution system? What measures exist for distribution?
	How?	Data integration processes. How should metadata be captured? How does the context affect metadata; technical layers of networks and communication? How to document how data flows through the system. How often should data sources be integrated or refreshed? How should metadata be used to model an ETL tool?
	Economic Resources	Cost/benefit of metadata collection; cost of metadata tools.
	Human Safety	If data does not move through the system properly, is there an impact on human safety? Can the metadata impact human safety

Table 5- Information as Product

Personnel Management and Legal		
	Framework Categories	**Description**
Personnel Management	Who?	Who will be the data quality champions? What qualifications do they need? Who needs to be trained on DQ procedures?
	What?	What training is necessary?
	How?	How is training accomplished? How are qualifications determined? How do you motivate employees to pay attention to DQ issues?
	Economic Resources	What does it cost to train people in DQ procedures/issues? What salary level will a DQ champion command?
	Human Safety	Will lack of training in DQ issues/procedures result in human safety issues?
L e	Who?	Who is responsible for compliance?

	What?	What legal requirements exist? (e.g., HIPPA, SOX, GLB)
	How?	How is compliance with legal requirements accomplished?
	Economic Resources	What is the cost of compliance? What are the potential legal liabilities of non-compliance?
	Human Safety	Non-compliance and harm to groups, organizations and society.

Table 6- Personnel and Legal

In summary, we have identified research questions at the juncture of each Wang and Juran characteristic. This provided the foundation for classification of the reviewed journal articles and allowed us to provide a snapshot of the density of articles in various cells of the framework.

METHODOLOGY

This study objectively examines a representative sample from the DIQ research literature, using a classification based approach. This section explains what criteria and processes were used to select keywords for determining representative articles, the coding procedure and the inter-coder reliability.

Gathering the data

In order to classify the data and information literature, we searched the Proquest- ABI/Inform and EBSCO databases for peer reviewed journal articles that contained the keywords "data quality" or "information quality". The focus was on articles written since 1995 in order to avoid duplication of the references in the Wang et al article. Extending the journal list through examination of article references resulted in additional articles. Finally, authors who had written data quality articles were searched to ascertain that all articles by these authors were included. Preliminarily, 74 articles were deemed to be within the realm of data and information quality, based on these criteria.

Each of the 74 journal articles was coded for inclusion in the framework described above. This coding was done by the two researchers and an assistant. Each article was coded into the framework, using as many codifications as a coder felt were relevant. To ensure consistency in coding, the raters discussed the articles in sets of five for the first 25 articles to refine the framework and achieve inter-rater reliability. Once the coding was completed, the percentage of inter-rater agreement overall was 83.1%.

Due to space restrictions, articles are not analyzed individually in this paper. The articles reviewed can be found at the end of this paper [1-68, 70-75].

RESULTS

Results fall into two general categories: the data and information quality literature as it is categorized within the framework and an analysis of the outlets for data and information quality literature.

Literature Categorization- Framework Cells

Given that an article can be coded into multiple cells, there were 219 different codings for the 74 articles, as can be seen in Table 7. Note that "Operation & Assurance Costs- Accounting Information Systems" does not have any articles coded. Although there are a number of articles in the field of AIS, none of them deal specifically with data quality and the cost of maintaining data quality.

	Count	Who?	What?	How?	Economic Resources	Human Safety
Management Responsibilities	38	10	12	10	5	1
Operation & Assurance Cost- Information Systems	1		1			
Operation & Assurance Cost- Database	4		1	1	2	
Research &Development- Semantics	34	6	12	13	3	
Research & Development Verification Tools	22	2	8	10	2	
Research & Development- Dimensions & Measurements	45	10	19	14	1	1
Production	21	2	9	9		1
Distribution	38	3	14	18	2	1
Personnel Management	11	2	4	4		1
Legal Function	5	1	2	1	1	
Total	219	36	82	80	16	5

Table 7- Count of articles represented in each framework cell

Literature Categorization- Overarching Characteristics

In the Wang et al classification schema [69], articles were coded with a major focus and minor points. That process was followed in this paper by assigning an overarching Wang and Juran classification to each paper, in addition to the classifications within the framework cells. Thus an article such as Fisher and Kingma's article, "Criticality of data quality as exemplified in two disasters" [18] which deals with the Challenger and USS Vincennes disasters and how data quality played a role in these disasters, would be given an overall Wang classification of production (the production of the data caused the data quality problem) and a Juran classification of human safety (the article clearly shows that data quality problems led to these disasters, which involved loss of human life.)

As can be seen in Figure 1, the densest number of articles by Wang classification can be found in "R&D- Dimensions" and "Management Responsibilities", followed closely by "Distribution" and "R&D- Semantics". Figure 2 shows us that the majority of the articles address the questions of "What" and "How" within the Juran classification.

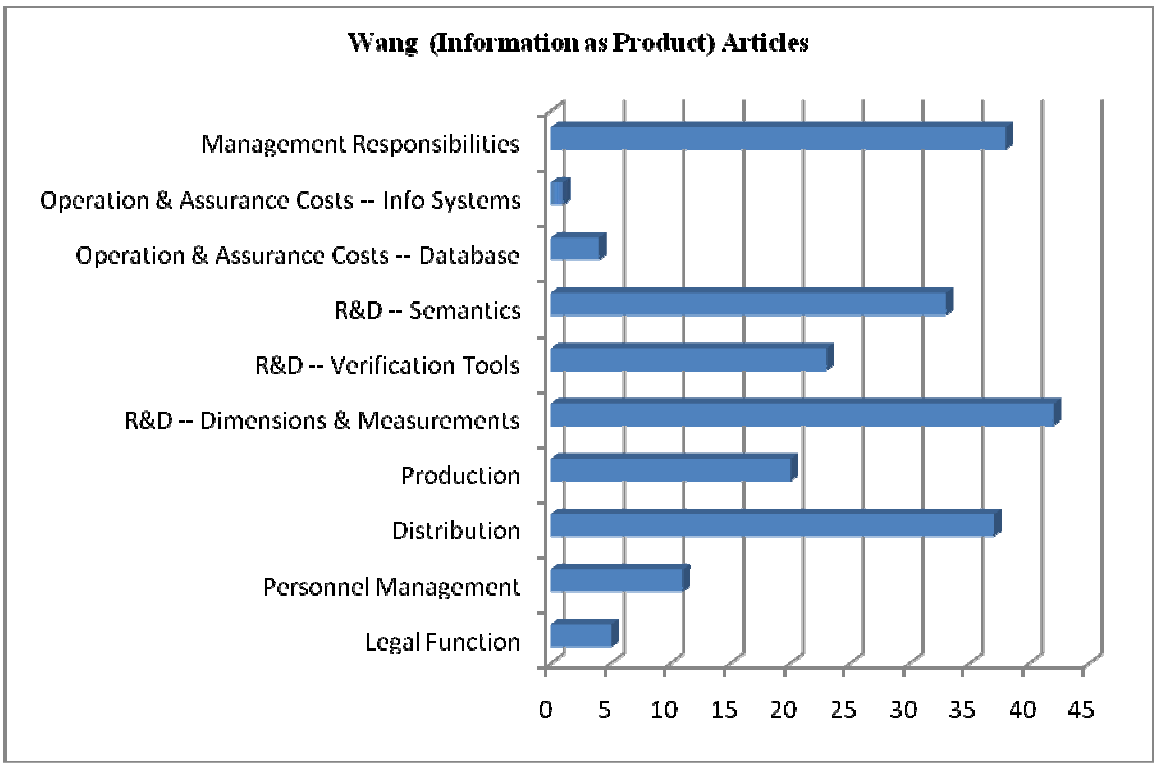

Figure 1- Wang Articles

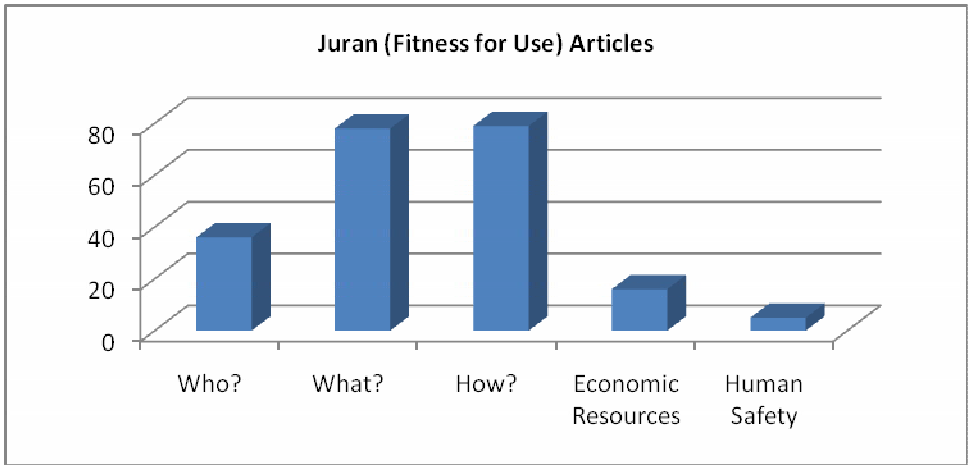

Figure 2- Juran Articles

Publication Outlets

The most popular outlets for data and information quality papers can be seen in Table 8. The top four journals (Communications of the ACM (CACM), Information Systems (IS), Journal of Management Information Systems (JMIS) and Journal of Database Management (JDBM)) were further analyzed to determine if these four journals published articles that were similar to each other or published a wide variety of articles within the realm of data and information quality. These results can be seen in tables 9 and 10.

Journal	Count
Communications of the ACM	12
Information Systems	9
Journal of Management Information Systems	5
Journal of Database Management	5
Information Resources Management Journal	4
Management Science	4
IEEE Transactions on Knowledge and Data Engineering	4
Sloan Management Review	3
Database for Advances in Information Systems	2
Decision Support Systems	2
Information & Management	2
MIS Quarterly	2
Journal of Computer Information Systems	2

Table 8- Publication outlets with more than one paper

The top four journals were all ranked in the top 25 journals on the ISWorld MIS Journal Rankings website (http://isworld.org/csaunders/rankings.htm), a system frequently using for promotion and tenure decisions. CACM has an average ranking of 2.75, JMIS has a ranking of 4.86, Journal of Database Management comes in at 19.67 and Information Systems places 20.

Top 4 Publication Outlets – Wang Characterisitics					
	CACM	**IS**	**JMIS**	**JDBM**	**Total**
Management Responsibilities	1			1	2
Operations & Assurance Costs - Information Systems					0
Operations & Assurance Costs – Database	1				1
Operations & Assurance Costs - Accounting					0
Research & Development - Analysis & Design	2			1	3
Research & Development - Quality in Information Systems Design	1	1	1		3
Research & Development - Dimensions & Measurement	5		2	1	8

Production	1	4	1	1	7
Distribution	1	4	1	1	7
Personnel Management					0
Legal Function					0
Total	12	9	5	5	31
CACM = Communications of the ACM IS =Information Systems JMIS = Journal of Management Information Systems JDBM = Journal of Database Management					

Table 9- Top 4 Publication Outlets- Information as Product

.

Top 4 Publication Outlets – Juran Characteristics					
	CACM	**IS**	**JMIS**	**JDBM**	**Total**
Who					0
What	3		2	3	8
How	7	9	3	2	21
Economic Resources	2				2
Human Safety					0
Total	12	9	5	5	31

Table 10- Top 4 Journal Outlets- Fitness for Use

DISCUSSION AND CONCULUSION

This paper has described a framework for the classification of data and information quality research. We have identified over-researched areas in this realm (dimensions, semantics and distribution with respect to what and how) as well as under-researched areas (research associated with costs, personnel management and the legal implications along all fitness for use characteristics). In addition, we have identified historical publication outlets for data and information quality research. It should be noted that there are now two journals devoted to data and information quality- the International Journal of Information Quality and the new ACM Journal of Data and Information Quality. Papers in these journals are being reviewed for an expanded version of this paper.

Perhaps even more importantly, we have identified numerous potential research questions within this area of research. Data and information quality are rapidly being recognized as critical success factors in business. The academic research can, and should, be a leader in ways to help businesses address this important need. By providing a framework for the existing literature, as well as a set of questions that need to be addressed, we believe that this paper will help researchers to better focus their research in areas that can make a real difference to the business community.

REFERENCES

1. Aiken, P.H., *Reverse Engineering of Data.* IBM Systems Journal, 1998. **37**(2): p. 246-259.
2. Audini, B., A. Pearce, and P. Lelliott, *Accuracy, completeness and relevance of Department of Health returns on provision of mental health residential accommodation: A data quality audit.* Journal of Mental Health, 2000. **9**(4): p. 365-370.
3. Ballou, D., et al., *Modeling Information Manufacturing Systems to Determine Information Product Quality.* Management Science, 1998. **44**(4): p. 462-484.
4. Ballou, D.P. and G.K. Tayi, *Enhancing data quality in data warehouse environments.* Communications of the ACM, 1999. **42**(1): p. 73-80.
5. Ballou, D.P. and H.L. Pazer, *Modeling Completeness versus Consistency Tradeoffs in Information Decision Contexts.* IEEE Transactions on Knowledge & Data Engineering, 2003. **15**(1): p. 240-243.
6. Ballou, D.P. and I.N. Chengalur-Smith, *Sample-Based Quality Estimation of Query Results in Relational Database Enironments.* IEEE Transactions on Knowledge and Data Engineering, 2006. **18**(5): p. 639-650.
7. Birman, S., *Control the Data, Control the Costs.* Quality, 2003. **42**(1): p. 50-56.
8. Bricker, J. and A. Maydanchik, *Data quality assurance: Plan redesign affords an opportunity to consider an automated approach to cleansing employee records.* Compensation & Benefits Management, 1999. **15**(4): p. 49-55.
9. Cappiello, C., C. Francalanci, and B. Pernici, *Time-Related Factors of Data Quality in Multichannel Information Systems.* Journal of Management Information Systems, 2003. **20**(3): p. 71-91.
10. Chengalur-Smith, I., D.P. Ballou, and H.L. Pazer, *The Impact of Data Quality Information on Decision Making: An Exploratory Analysis.* IEEE Transactors on Knowledge and Data Engineering, 1999. **11**(6): p. 853-864.
11. Chiang, R.H.L., E.-P. Lim, and V.C. Storey, *A Framework for Acquiring Domain Semantics and Knowledge for Database Integration.* The DATA BASE for Advances in Information Systems, 2000. **31**(2): p. 46-64.
12. Divorski, S. and M. Scheirer, *Improving Data Quality for Performance Measures: Results From a GAO Study of Verification and Validation.* Evaluation and Program Engineering, 2000. **24**: p. 83-94.
13. Embury, S.M., et al., *Adapting integrity enforcement techniques for data reconciliation.* Information Systems, 2001. **26**(8): p. 657-689.
14. English, L., *Information quality: Critical ingredient for national security.* JOURNAL OF DATABASE MANAGEMENT, 2005. **16**(1): p. 18-32.
15. Even, A. and G. Shankaranarayanan, *Utility-Driven Assessment of Data Quality.* Database for Advances in Information Systems, 2007. **38**(2): p. 75-93.
16. Even, A. and G. Shankaranarayanan, *Economics-Driven Data Management: An Application to the Design of Tabular Data Sets.* IEEE Transactions on Knowledge and Data Engineering, 2007. **19**(6): p. 818-831.
17. Fan, W., et al., *Discovering and reconciling value conflicts for numerical data integration.* Information Systems, 2001. **26**(8): p. 635-656.
18. Fisher, C.W. and B.R. Kingma, *Criticality of data quality as exemplified in two disasters.* Information & Management, 2001. **39**(2): p. 109.
19. Fisher, C.W., I. Chengalur-Smith, and D.P. Ballou, *The Impact of Experience and Time on the Use of Data Quality Information in Decision Making.* Information Systems Research, 2003. **14**(2): p. 170-189.
20. Henderson, I. and D. Murray, *Prioritising and deploying data quality improvement activity.* Journal of Database Marketing & Customer Strategy Management, 2005. **12**(2): p. 113-119.
21. Kaplan, D., et al., *Assessing Data Quality in Information.* Communications of the ACM, 1998. **41**(2): p. 72-78.
22. Kelly, V.E., C.P. Thomas, and H. Wang, *Managing Data-Based Systems Across Releases Using Historical Data Dictionaries.* Bell Labs Technical Journal, 2000. **5**(2): p. 121-133.

23. Khalil, O. and M. Elkordy, *EIS Information: Use and Quality Determinants.* Information Resources Management Journal, 2005. **18**(2): p. 68-93.

24. Kim, W., *On US homeland security and database technology.* JOURNAL OF DATABASE MANAGEMENT, 2005. **16**(1): p. 1-17.

25. Klein, B.D., *How Do Actuaries Use Data Containing Errors? Models of Error Detection and Error Correction.* Information Resources Management Journal, 1997. **10**(4): p. 27-36.

26. Klein, B.D., D.L. Goodhue, and G.B. Davis, *Can Humans Detect Errors in Data? Impact of Base Rates, Incentives, and Goals.* MIS Quarterly, 1997. **21**(2): p. 169-194.

27. Klein, B.D., *Detection of Data Errors in the Practice of Inventory Management.* The Journal of Computer Information Systems, 1999. **40**(2): p. 34-40.

28. Klein, B.D. and D.F. Rossin, *Data quality in neural network models: Effect of error rate and magnitude of error on predictive accuracy.* Omega, 1999. **27**(5): p. 569.

29. Klein, B.D., *The Detection of Data Errors in Computer Information Systems: Field Interviews with Municipal Bond Analysts.* Information Resources Management Journal, 2000. **13**(3): p. 23-32.

30. Klein, B.D., *User perceptions of data quality: Internet and traditional text sources.* The Journal of Computer Information Systems, 2001. **41**(4): p. 9-15.

31. Klein, B.D., *Internet data quality: Perceptions of graduate and undergraduate business students.* Journal of Business and Management, 2002. **8**(4): p. 425-432.

32. Lam, M. and R.K.H. Ching, *Information Integration in Multidimensional Databases: A Case Study.* Information Systems Management, 1998. **15**(4): p. 36-45.

33. Lee, T.Y. and Y. Yang, *Constraint-based wrapper specification and verification for cooperative information systems.* Information Systems, 2004. **29**(7): p. 617-636.

34. Lee, Y., et al., *AIMQ: a methodology for information quality assessment.* Information and Management, 2002. **40**(2): p. 133-146.

35. Lee, Y., *Crafting Rules: Context-Reflective Data Quality Problem Solving.* Journal of Management Information Systems, 2003. **20**(3): p. 93-119.

36. Lee, Y.W. and D.M. Strong, *Knowing-Why About Data Processes and Data Quality.* Journal of Management Information Systems, 2003. **20**(3): p. 13-39.

37. Lee, Y.W., et al., *Process-Embedded Data Integrity.* Journal of Database Management, 2004. **15**(1): p. 87-104.

38. Levitin, A. and T. Redman, *Quality dimensions of a conceptual view.* Information Processing & Management, 1994. **31**(1): p. 81-88.

39. Levitin, A.V. and T.C. Redman, *Data as a resource: Properties, implications, and prescriptions.* MIT Sloan Management Review, 1998. **40**(1): p. 89-102.

40. Low, W.L., M.L. Lee, and T.W. Ling, *A knowledge-based approach for duplicate elimination in data cleaning.* Information Systems, 2001. **26**(8): p. 585-606.

41. Madnick, S., R. Wang, and R. Xiang Xian, *The Design and Implementation of a Corporate Householding Knowledge Processor to Improve Data Quality.* Journal of Management Information Systems, 2003. **20**(3): p. 41-70.

42. Marsh, R., *Drowning in dirty data? It's time to sink or swim: A four-stage methodology for total data quality management.* Journal of Database Marketing & Customer Strategy Management., 2005. **12**(2): p. 105-112.

43. Naumann, F., J.-C. Freytag, and U. Leser, *Completeness of integrated information sources.* Information Systems, 2004. **29**(7): p. 583-615.

44. Nord, G., J. Nord, and H. Xu, *An investigation of the impact of organization size on data quality issues.* JOURNAL OF DATABASE MANAGEMENT, 2005. **16**(3): p. 105-112.

45. Orr, K., *Data Quality and Systems Theory.* Communications of the ACM, 1998. **41**(2): p. 66-71.

46. Parssian, A., S. Sarkar, and V.S. Jacob, *Assessing Data Quality for Information Products: Impact of Selection, Projection, and Cartesian Product.* Management Science, 2004. **50**(7): p. 967-982.

47. Pierce, E.M., *ASSESSING DATA QUALITY WITH CONTROL MATRICES.* CACM, 2004. **47**(2): p. 82-86.

48. Pinto, M., *Data representation factors and dimensions from the quality function deployment (QFD) perspective.* Management Science, 2006. **32**(2): p. 116-130.

49. Pipino, L.L., Y.W. Lee, and R.Y. Wang, *Data Quality Assessment.* Communications of the ACM, 2002. **45**(4): p. 211-218.

50. Price, R. and G. Shanks, *A semiotic information quality framework: development and comparative analysis.* JOURNAL OF INFORMATION TECHNOLOGY, 2005. **20**(2): p. 88-102.

51. Redman, T.C., *The Impact of Poor Data Quality on the Typical Enterprise.* Communications of the ACM, 1998. **41**(2): p. 79-82.

52. Rieh, S., *Judgment of Information Quality aqnd Cognitive Authority in the Web.* Journal of the American Society for Information Science and Technology, 2002. **53**(2): p. 145-161.

53. Scannapieco, M., et al., *The DaQuinCIS architecture: a platform for exchanging and improving data quality in cooperative information systems.* Information Systems, 2004. **29**(7): p. 551-582.

54. Shankaranarayanan, G., M. Ziad, and R.Y. Wang, *Managing Data Quality in Dynamic Decision Environments: An Information Product Approach.* Journal of Database Management, 2003. **14**(4): p. 14-33.

55. Shankaranarayanan, G. and A. Even, *MANAGING METADATA IN DATA WAREHOUSES: PITFALLS AND POSSIBILITIES.* Communications of the Association for Information Systems (CAIS), 2004. **14**(2004): p. 247-274.

56. Shankaranarayanan, G. and A. Even, *The Metadata Enigma.* Communications of the ACM, 2006. **49**(2): p. 88-94.

57. Shankaranarayanan, G. and A. Even, *Supporting data quality management in decision-making.* Decision Support Systems, 2006. **42**: p. 302-317.

58. Strike, K., K.E. Emam, and N. Madhavji, *Software Cost Estimation with Incomplete Data.* IEEE Transactions on Software Engineering, 2001. **27**(10): p. 890-908.

59. Strong, D.M., Y. Lee, and R. Wang, *10 Potholes in the Road to Information Quality.* IEEE Computer, 1997. **30**(8): p. 38-46.

60. Strong, D.M., Y.W. Lee, and R.L. Wang, *Data Quality in Context.* Communications of the ACM, 1997. **40**(5): p. 103-110.

61. Tee, S., et al., *Factors influencing organizations to improve data quality in their information systems.* Accounting and Finance, 2007. **47**: p. 335-355.

62. Tejadaa, S., C.A. Knoblocka, and S. Mintonb, *Learning object identification rules for information integration.* Information Systems, 2001. **26**(8): p. 607-633.

63. Vassiliadis, P., et al., *Arktos: towards the modeling, design, control and execution of ETL processes.* Information Systems, 2001. **26**(8): p. 537-561.

64. Wallace, W.A., *Assessing the quality of data used for benchmarking and decision-making.* The Journal of Government Financial Management, 2002. **51**(3): p. 16-21.

65. Wand, Y. and R.Y. Wang, *Anchoring Data Quality Dimensions in Ontological Foundations.* Communciations of the ACM, 1996. **39**(11): p. 86-95.

66. Wand, Y., V.C. Storey, and R. Weber, *An Ontological Analysis of the Relationship Construct in Cenceptual Modeling.* ACM Transactions on Database Systems, 1999. **24**(4): p. 494-528.

67. Wang, R., et al., *Manage Your Information as a Product.* Sloan Management Review, 1998. **39**(4): p. 95-105.

68. Wang, R.Y., M. Reddy, and H. Kon, *Toward quality data: An attribute-based approach.* Decision Support Systems, 1995. **13**(3/4): p. 349-372.

69. Wang, R.Y., V.C. Storey, and C.P. Firth, *A Framework for Analysis of Data Quality Research.* IEEE Transactions on Knowledge and Data Engineering, 1995. **7**(4): p. 623-641.

70. Wang, R.Y. and D.M. Strong, *Beyond Accuracy: What Data Quality Means to Data Consumers.* Journal of Management Information Systems (JMIS), 1996. **12**(4): p. 5-34.

71. Wang, R.Y., *A product Perspective on Total Data Quality Management.* Communications of the ACM, 1998. **41**(2): p. 58-65.

72. Winkler, W.E., *Methods for evaluating and creating data quality.* Information Systems, 2004. **29**(7): p. 531-550.

73. Wixom, B.H. and H.J. Watson, *An Empirical Investigation of the Factors Affecting Data Warehousing Success.* MIS Quarterly, 2001. **25**(1): p. 17-38.

74. Xu, H., et al., *Data quality issues in implementing an ERP.* Industrial Management + Data Systems, 2002. **102**(1/2): p. 47-58.

75. Yoon, V., P. Aiken, and T. Guimaraes, *Managing Organizational Data Resources: Quality Dimensions.* Information Resources Management Journal, 2000. **13**(3): p. 5-32.

76. Juran, J.M., *Juran's Quality Control Handbook*. 4 ed, ed. F.M. Gryna. 1988, New York: Mc-Graw-Hill, Inc.

77. Juran, J.M., *Juran on Planning for Quality*. 1988, New York: The Free Press.