

IMPROVING INFORMATION QUALITY AND INFORMATION TECHNOLOGY SYSTEMS IN THE 21ST CENTURY

Frank M. Guess

Department of Statistics
University of Tennessee
Knoxville, TN 37996-0532

Key Words: data quality strategies; leveraging; management and research agendas; cyberspace as an alias for dataspace or infospace.

ABSTRACT

Improving and deploying information quality, as a deeply practiced organizational culture via information technology systems, are important issues for highly successful, adaptive, modern corporations. This paper briefly discusses several key books, papers, and web sites to demonstrate the importance of integrating information quality into the culture and into all information technology systems. The importance of quality integration increases as the information technology system become larger. E-business and consumer concerns will aggressively drive corporations to such key information quality approaches. This paper explores some general needs and concerns in information quality for 21st century corporations, government & public organizations. Much more work by statisticians, information technologists, etc. in this area is highly encouraged.

The paper further stresses the tremendous opportunities for leveraging with both an IT culture and information quality.

1. INTRODUCTION

The long time leader in business and management issues, Drucker (1995) states how crucial the leveraging of information will be in the future for companies and, indeed, countries that make up a global community. His emphasis is not just to have IT systems, but valuable information deployed as a deeply practiced culture within a group, organization, country, etc. He states explicitly that statistics will be important in developing that information. We are very fortunate in our Statistics community to have individuals of Peter Drucker's caliber singing our cause, wisely applied, in information cultures around the world.

For those who have a statistical process improvement background or just a statistics background, but want to learn more practical approaches for information quality, the book of Redman (1997) is an extremely useful resource. This is a book written for the applied person in the trenches of business. Also, see the helpful website www.dataqualitysolutions.com/ developed by Thomas Redman, Robert Pautke, and others.

Scholars at or associated with MIT's Total Data Quality Management Center have developed a useful website web.mit.edu/TDQM/. The MIT group conducts a small elite Information Quality Conference held each year in October, which has moved this area forward greatly. A third insightful web repository is www.dataquality.com/ spearheaded by the tremendous energy of Jim Hurysz, Helena Zinkham, and others. It, also, includes access to the journal *Data Quality*.

For those with an IT or accounting type background, another book that I highly recommend is Huang, Lee, and Wang (1999). It is by two scholars from MIT's Center and

an IBM Vice President and former Director of Knowledge Management. The website Amazon.com comments on Huang, Lee, and Wang (1999),

“The Nov. 9, 1999 issue of Fortune (circulation 762,700) cites the book in discussing knowledge products:

‘... Take what you know about strategy for ordinary, tangible products, and map that onto what you can learn about the knowledge products you produce and sell. In a forthcoming book, Richard Wang, Co-director for MIT's Total Data Quality Management Program, offers four principles for managing information products: Understand consumers' information needs; create a well-defined production process; stay on top of the life cycle of information products; and appoint product managers. Those principles apply to all kinds of knowledge products; applying them can help you build a strong pillar for an Information Age business.’

The Feb. 1, 1999 issue of Computerworld (circulation 170,000) recommends this book as one of five when looking for books to navigate corporate waters as well as E-commerce seas:

‘Computers churn out tons of data daily, but why is so little of it truly useful to anyone? The authors believe that information should be managed as a product and knowledge as an asset. ... This book is of value to IT managers who are trying to provide an overall framework that incorporates a Web site, intranet, data warehousing, data marts and executive information systems that truly attempt to turn information into knowledge.’ ”

These books above with their related references and the three websites (web.mit.edu/TDQM/, www.dataqualitysolutions.com/, www.dataquality.com/) are helpful starting points to visit for those beginning their work in information quality (IQ). Both of these books and the websites, well read, will suggest many avenues for further research agendas and management strategies for academics, leaders, managers and others interested in furthering development of IQ.

Data quality problems can cause incredible financial problems for organizations. Firth (1996) notes how information quality problems:

- made a fiber-optics manufacturer lose \$500,000 when a mislabeled shipment caused the wrong cable to be laid along the bottom of a lake
- cost a brokerage firm to lose \$500 million when a dealer entered an incorrect exchange rate
- resulted in the U.S. government to losing over \$2 billion in federal loan monies.

See, also, Bowen, Fuhrer, and Guess (1998) and Collins and Guess (2000) for further discussions and examples.

Greenman (1999) comments on the term cyberspace being first used in 1984 by science fiction writer William Gibson's novel *Neuromancer* for the idea of an,

“...agglomeration of the world's computer data... .”

Greenman (1999) quotes from his interview with Gibson that other alternatives written down were dataspace, infospace, etc., which clearly did not make the cut in favor of cyberspace. Greenman (1999), also, notes that Gibson's

“... dark, data-clotted worldview has been immensely influential, reflected in everything from television... to film... to music... .”

It is curious the ambivalence of such information connectivity, with its many positives aspects can, also, generate alternative views of concerns for abuses.

Improving and deploying information quality are important issues for highly successful, adaptive, modern corporations. See, for example, the very insightful works of

- (1.) Sviokla (1999), which was a keynote talk about information quality and its rapid deployment for business advantages at the Information

Quality Conference at MIT's Sloan School of Management in 1999
[see his website: www.sviokla.com/ , in general, or more specifically
web.mit.edu/tdqm/www/speech/sld001.htm for his talk's PowerPoint
slides]

and

(2.) Sviokla (1998) on virtual markets.

This invited paper explores briefly some general needs and concerns in information quality for 21st century corporations, government & public organizations. The paper further stresses the tremendous opportunities for leveraging with both an IT culture and information quality (IQ). In Section 2, I discuss the needs of IQ and IT in health issues, with related references for those wanting to explore in that area more. Section 3 covers business, financial and other institutions with their IQ and IT concerns. I make concluding comments in Section 4.

2. HEALTH INFORMATION MANAGEMENT SYSTEMS AND ISSUES

Targeting health database and technology issues, the reader can start with the insightful tomes of Donaldson and Lohr (1994) and Dick, Steen, and Detmer (1997). The later text dwells on computer based patient records and some ways to improve them. For further useful references specifically on health data, research and related topics, I briefly mention Gochman (1997) Volumes 1 to 4, Gochman (1988), Conner and Norman (1996), Cox (1982), Berkman and Breslow (1983), Gelijns and Dawkins (1994) and Anderson, Dumphy, and Wilson (1990). These works range widely in perspective from the patient/end user, social networks and communities supporting that patient, health care providers (MD's nurses, etc), medical researchers, hospital administrators to actual health database

organizations. These works illustrate the tremendous diversity of perspectives, concerns, and even fears each group brings to health and/or data issues.

Dobbins and Guess (1999) stress the importance of management cultures and people in the improvements of information with a medical IT systems. They discuss initial employment of an information quality strategy in HealthMagic, Incorporate with HealthCompass™. HealthMagic has positioned itself as a leading web-based technology company. One of its key products, HealthCompass, is a web-based lifelong health IT system that enables consumers to create a portable and private repository of their entire health history over the Internet. In short, it is an online consumer health information management system in the fullest sense. It allows consumers to not only develop a lifelong health record, but also to access a credentialed health library, and a variety of other interactive health tools.

In February, 1999 HealthMagic publicly announced, "... a multi-year agreement with drkoop.com, the award-winning Internet healthcare portal led by Dr. C. Everett Koop, former U.S. Surgeon General. As part of the agreement drkoop.com will distribute the HealthCompass™ OnSite... ." Other clients of HealthMagic include, for example, Walt Disney World's resident villages.

HealthCompass places consumers at the center of a world of information about their health by bringing the "best of class" health information and products to consumers on-line, in real-time. HealthCompass enables these end users to understand and make decisions about their health, give them access to information relevant to their health, allow communication with their health care team, and remind them to take actions important to their health. Quality and the processes for improving that quality are crucial to such information systems and any decisions based on them. This strategy provides clear advantages to end users having greater trust and confidence in HealthCompass and its information.

HealthMagic wants to do this to understand drivers of customer satisfaction and to segment its customer base by variables that will aid in delivery

of additional information to the customer. Many of you know how Amazon.com has suavely used such data to provide greater value to its customers by providing better matches of additional information to a specific consumer. It is crucial for a health data system to provide ways to initially prevent problems and to continuously improve quality. Data quality here also means collecting information that does not burden the customer unduly. HealthMagic designs and tests its products with these goals as its focus. Also, the crucial needs of security and confidentiality are maintain by digital certificates, the latest encryption and security technology to prevent difficulties. A complete audit trail is available as another safeguard for the users of this IT system.

Data quality is a key strategy for the company. Ways to prevent problems, not just “fix” them is the emphasis. High quality data are important to all parties. Developing a culture of valuing information quality and continuous improvements are aggressively affirmed with HealthMagic. See Dobbins and Guess (1999) for more details.

Since many health groups and businesses are currently locating shop fronts on the WEB this approach will be of general interest. E-business and consumer concerns will aggressively drive corporations to developing data quality approaches.

3. BUSINESS, FINANCIAL & OTHER IT SYSTEMS

In Redman (1997), Blan Godfrey comments on Redman’s emphasis of data quality as an important competitive strategy for corporations concerned with excellence and thriving in the market place. Compare, also, Caby, Pautke, and Redman (1995), Redman (1995) and Redman (1992) for related comments on IQ. Bowen, Fuhrer, and Guess (1998) discuss usage of a continuous process improvement approach of data quality in an accounting and auditing setting. The

importance of a corporate culture and techniques of continuously improving information quality is their emphasis, also.

Recall Firth (1996) comments about poor IQ creating the loss of \$2 billion in federal loans. Note that Clapp, Green, Poitrowski and Bell (1997) work on the prevention of fraud in various government programs (e.g., welfare, Medicare, etc.) by improving data processing and checks.

We mention other areas and papers dealing with IQ issues. Redman (1994) considers the need on IQ in telecommunications. For IQ issues in client server environment see Bowen, Schneider, and Fields (1995). Compare, also, Bowen (1992) on IQ in business and other accounting systems. Another excellent book, for an accounting or IT audience is Weber (1999). Collins and Guess (2000) discuss further the current banking data management on loan approval processes. These processes have great room for improvements of information quality and prevention of data problems in general, but especially with regards to fair lending and fair pricing practices. They first review briefly typical data collection protocols deployed at many financial institutions for loan approval and loan pricing. Portions of these data protocols are mandated by federal regulation. In discussing the data capture and analysis for fair lending, they illustrate some initial key steps currently needed in improving information quality for all parties involved.

Regarding client or customer information quality improvements, Collins (2000) points out that the regulators and enforcement agents have escalated testing for data quality in loan underwriting, including tests for patterns in missing data elements by ethnicity, gender, age and income strata. In this compliance environment, the only means by which to reduce the likelihood of unfavorable examination findings for this type of test is to electronically capture the needed data in the underwriting process so as to monitor and improve its quality.

As support for the value of collecting these additional data elements, Collins, Harvey & Nigro (2000) show that data loss, particularly among credit

underwriting factors, works to the detriment of low-to-moderate income borrowers relative to upper-income borrowers for home improvement credits. That is, neglecting some qualitative underwriting factors in favor of a “streamlined” credit scoring process raises the denial rate for low-to-moderate income borrowers more than for upper-income borrowers. This disparity in denial rates generates a potential fair lending violation due to poor data warehousing. See, also, Harvey, Collins, Nigro, and Robinson (2000) for more on disparities in lending data.

4. CONCLUSIONS

Improving and deploying information quality are important issues for highly successful, adaptive, modern corporations. Effective and reliable IT systems are the back bone of highly successful organizations, from airlines and safe air traffic controls, internationally focused stock markets and financial institutions, medical and health care providers, modern militaries, universities, to telecommunications.

Clearly, IT systems will grow in both usage and needs of high quality information that help leverage a group’s advantages. Guess and Bowen (2001) discuss briefly some of the many issues in improving the reliability of IT systems. Drucker (1995) provides insightful views that the key will be for information that is real with a culture that will implement that information. Also, I strongly encourage more applied and theoretical research for improving IQ and IT deployment within industry, government and academic communities.

ACKNOWLEDGEMENTS

I gratefully acknowledges financial support from the University of Tennessee College of Business Administration Scholarly Research Grant

Program, from the system-wide University of Tennessee Professional Development Award Program, and matching funding from the Department of Statistics at UT. Also, I express cheerful appreciation to Ned Clapp, Oak Ridge National Laboratories, for reading and comments on an earlier draft of this paper. Last, I thank Professor Ramesh Gupta for inviting me to speak and write on this for the International Conference on Statistics in the 21st Century.

BIBLIOGRAPHY

- Anderson, K. M., Dumphy, D. C., and Wilson, P. W. F. (1990). "Management of data quality in a long-term epidemiologic study: The Framingham heart study," *Data quality control: Theory and pragmatics*, Eds. Liepens and Uppuluri, 57-68. New York: Marcel-Dekker.
- Bowen, P. L. (1992). *Managing data quality in accounting information systems: A stochastic clearing system approach*. Ph.D. Dissertation, University of Tennessee.
- Bowen, P. L., Schneider, G. P., and Fields, K. T. (1995). "Managing data quality in client/server environments," *IS Audit & Control Journal* 4: 28-35.
- Bowen, P., Fuhrer, D., and Guess, F. (1998). "Continuously improving data quality in persistent databases," *Data Quality @* www.dataquality.com/998bowen.htm. (See www.dataquality.com/ in general.)
- Caby, E. C., Pautke, P. W., and Redman, T. C. (1995). "Strategies for improving data quality," *Data Quality* 1, no. 1 (March): 4-12.
- Clapp, N.E., Green, P.L., Poitrowski, S.T., and Bell, D. (1997). *Automated tools for cross-referencing large databases*. Springfield: National Technical Information Service.
- Collins, M. (2000). "A note on risk-based pricing," www.bankinfo.com/compliance/compliance.html. (See www.bankinfo.com in general)
- Collins, M. and Guess, F. (2000). "Improving information quality in loan approval processes for fair lending and fair pricing," *Information Quality*

Conference in 2000 at Massachusetts Institute of Technology's Sloan School of Management Proceedings, pp. 45-53.

Collins, M., Harvey, K., and Nigro, P. (2000). "The influence of bureau scores, customized scores and judgmental review on the bank underwriting decision making process," Working Paper, The Office of the Comptroller of the Currency and The University of Tennessee.

Dobbins, J. G. and Guess, F. (1999). "Developing a data quality strategy in a provider of Web based health information systems," *Information Quality Conference in 1999 at Massachusetts Institute of Technology's Sloan School of Management Proceedings*, pp. 176-184.

Drucker, P. F. (1995). *Managing in a Time of Great Change*. New York: Truman Talley Books/Dutton.

Firth, C. P. (1996). "Data quality in practice: Experience from the frontline," *Information Quality Conference in 1999 at Massachusetts Institute of Technology's Sloan School of Management Proceedings The 1996 Information Quality Conference*, Massachusetts Institute of Technology, October 25-26. (Also see website sunflower.singnet.com.sg/~cfirth/dataquality/. Special viewing authorization may be needed.)

Guess, F., and Bowen, P. (2001). "Growth and Reliability of Information Technology Systems," *International Journal of Reliability and Application*, invited paper (forthcoming).

Harvey, K., Collins, M. C., Nigro, P., and Robinson, B. (2000). "Determinants of Denial Disparities in Home Mortgage Lending: Bank Performance, Macro-economic Factors and Regulatory Influence," *Journal of Real Estate Finance & Economics* (forthcoming).

Huang, K-T., Lee, Y. L. and Wang, R. Y. (1999). *Quality Information and Knowledge*. New York: Prentice Hall

Redman, T. C. (1992). *Data Quality, Management and Technology*. New York: Bantam Books.

Redman, T. C. (1994). "Data quality for telecommunications," *IEEE Journal On Selected Areas In Communications*, February 12, no. 2: 306-312.

Redman, T. C. (1995). "Improve data quality for competitive advantage," *Sloan Management Review* 36 (2), 99-107.

Redman, T. C. (1997). *Data Quality for the Information Age*. Artech House Computer Science Library.

Sviokla, J. J. (1998). "Virtual Value and the Birth of Virtual Markets," Chapter 10, *Sense & Respond: Capturing Value in the Network Era*, coedited by Stephen P. Bradley & Richard L. Nolan, Boston: Harvard Business School Press.

Sviokla, J. J. (1999). Keynote speech at the Information Quality Conference at Massachusetts Institute of Technology's Sloan School of Management: "Customer Data and eCommerce," [See web.mit.edu/tdqm/www/speech/sld001.htm for his talk's PowerPoint slides or, in general, his web site: www.sviokla.com/.] *Information Quality Conference in 1999 at Massachusetts Institute of Technology's Sloan School of Management Proceedings*

Weber, R. (1998). *Information Systems Control and Audit*. Upper Saddle River, New Jersey: Prentice Hall.