HIGH QUALITY DATABASE CONTENT: SUCCESS FACTORS AND IMPROVEMENT STRATEGY

Chiemeke S.C and Akpon-Ebiyomare
Department of Computer Science
University of Benin, Benin City, Nigeria

ABSTRACT

The quality of data in an information system is an important issue, be they population census data, voters register, observational records or survey data. There is a merging requirement by many governments around the world for data to be of high quality and be better documented. Too often, data are used uncritically without consideration of the error contained within, and this can lead to erroneous results, misleading information, unwise decisions, loss of revenue, duplication of efforts, increase in cost of processing, poor business relationships and even loss of lives. Corporations, government agencies and not-for-profit groups are all inundated with enormous amounts of data in their information systems databases. This data has the potential to be used to generate greater understanding of a country for proper planning; for an organisation's customers, processes, and the organisation itself. Attention must be paid to the quality of data going into computer-based information systems. The data may not only be "inaccurate" or "wrong" but may also be missing, out-of-date, inconsistent or otherwise inadequate for the specific purposes of the user. This paper explores factors that control data quality in information systems databases. It also discusses those factors that are critical to the improvement of the quality of data in information systems databases and strategies for achieving high quality database content.

Keywords: Cleansing, Database, Dimensions, Improvement, Quality

1.0 INTRODUCTION

Because of the vast amounts of data held by private and government organizations in their information systems, there is the need to develop a strategy for ensuring the quality of the data captured otherwise what we may have could just be databases full of junk. Attention to data quality is a critical issue in all areas of information resources management. For instance, the government analyses data gained by population census to decide, which regions of the country require further investments in the infrastructure, like schools and other educational facilities, because of expected future trends. In other words, inaccurate census statistics could result in wrong allocation of scarce resources. Even business are not spared by poor quality data. An article in the Wall Street Journal (13/7/98) relates the domino effect that occurred when erroneous information was typed into a central data-base. A new airport in Hong Kong suffered catastrophic problems in baggage handling, flight information, and cargo transfer. The ramifications of the dirty data were felt throughout the airport. Flights took off without luggage; airport officials tracked flights with plastic pieces on magnetic boards; and airlines called confused ground staff on cellular phones to let them know where even more confused passengers could find their planes (Arnold, 1998). The airport had been depending on the central database to be accurate. When it wasn't, the airport paid the price in terms of customer satisfaction and trust.

In Nigeria, the Central Bank of Nigeria (CBN) in august 2009 made public in the Guardian newspaper (Nigerian Guardian, 29th July 2009) bank loan defaulters of all banks in Nigeria. Many companies and individuals had issues with the CBN over the figures indicated against their names for different reasons which included;

(i) They never applied for nor obtained loan from the indicated banks.
(ii) Amounts they owed were less than amount indicated against their names.
(iii) Loan obtained had been completely paid up yet published list claimed they were still debtors.
(iv) They were never or no longer directors of the debtor companies as published.

It was later discovered that the database used to generate the debtor list was not updated before use.

While not all of these errors are 100% attributable to data quality, Strong *et al* (1997) notes that the percentage contributed by poor quality are quite high. He also notes the social impact when government organization fail to ensure their data have sufficient quality to make effective decisions. The cost to organisations is far more than merely financial. Trust is lost from valuable customers (both internal and external), potential customers and sales are missed, operational costs increased, workers lose motivation, long-term business strategy is hindered and business re-engineering is impeded (Bowen, Fuhrer, & Guess, 1998; Redman, 1996), (Loshin, 2001). Redman also details how poor data quality affects operational, tactical and strategic decisions (Redman, 1996).

Case studies concerning data quality problems are frequently documented. Data quality problems have been investigated in a substantial body of literature. It is imperative that the issue of data quality be addressed for the data base to beneficial to an organisation.

2.0 WHAT IS DATA QUALITY?

The general definition of data quality is 'data that is fit for use by data consumers (Wang and Strong, 1996). Data quality has many attributes/dimensions. Commonly identified data quality dimensions are:

- **Consistency**: Concerns contradictions and syntactical anomalies.
- **Uniqueness**: Related to the number of duplicates in the data
- **Accuracy**, which occurs when the recorded value is in conformity with the actual value;
- **Timeliness**, which occurs when the recorded value is not out of date;
- **Completeness**, which occurs when all values for a certain variable are recorded

3.0 FACTORS THAT IMPACT ON DATA QUALITY IN INFORMATION SYSTEMS DATABASE

There have been many studies of critical success factors in quality management such as Total Quality Management (TQM) and Just-In-Time (JIT) (Saraph et al 1989; Porter and Parker 1993; Black and Porter 1996; Badri, Davis and Davis 1995; Yusof and Aspinwall 1999). Some of the data quality literature has addressed the critical points and steps for DQ (Firth 1996; Segev 1996; Huang et al 1999; English 1999). Table 5 summarises these factors.

4.0 IMPROVING DATA QUALITY IN INFORMATION SYSTEM DATABASE

English (1999); Redman (1996); Wang *et al*., (1995b) and Ballou and Pazer (2003) all agree that the quality of a real-world data set depends on a number of issues but the source of the data is the crucial factor. Ballou and Pazer (2003) was specific when they said that data entry and acquisition are inherently prone to errors, both simple and complex. Marcus *et al* (2001) says that much effort can be given to improving this front-end process, with respect to reduction in entry errors, but the fact often remains that errors in large data sets are common. English (2003) proposed a data quality improvement cycle (Figure 2).

Table 5 summary the factors affecting the quality of data in databases

	FACTOR	Xu (2001)	English (1999)	Wang (1998,1999)	Firth (1996)	Segev (1996)	Zhu (1995)	Saraph (1989)	Johnson (1981)	Bowen (1993)	Nichols (1987)
1	Role of top management		√	√	√	√	√	√			
2	Data quality polices & standards			√	√						
3	Role of data quality manager	√	√	√	√	√	√	√			
5	Employee/ personnel relations				√		√	√			
6	Performance evaluation and rewards (responsibility for DQ)		√	√						√	
8	Internal control (systems, process)								√		
9	Input control				√						√
10	Continuous improvement	√	√								
11	Training and communication		√								
12	Manage Change		√								

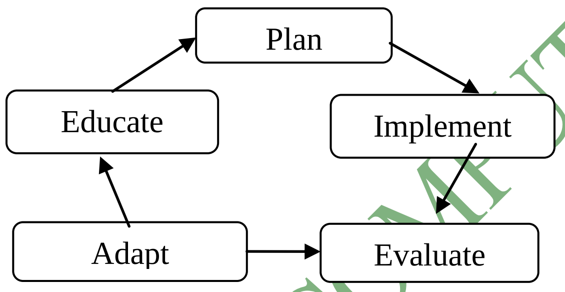

Fig 2: Data Quality Improvement Cycle, English (2003)

Olson (2003) notes that the mission of any data quality improvement programme should be three-fold; to improve, prevent, and monitor. An analysis of the requirements for a data quality improvement programme finds that the data quality practitioners, including English (999a), Wang et al., (2001), Olson (2003) and Loshin (2001), agree that the cause of poor quality data is often found to be human or process error. A programme of work is required by many participants in an organisation and often across business units to implement the above initiatives and Olson (2003) indicated that such a programme requires long term commitment. Embury (2001) notes that the general principles of quality management as applied to products can also be applied to data. This suggests there should be two basic approaches to the improvement of data quality, namely:

- Defect prevention
- Defect detection (and correction)

4.1 Defect Prevention

Prevention, they say, is better than cure. Defect or error prevention is considered to be far superior to error detection, since detection is often costly and can never guarantee to be 100% successful (Dalcin 2004. The cost to input a collection into a database can be substantial (Armstrong 1992) but is only a fraction of the cost of checking and correcting the data at a later date. It is better to prevent errors than to cure them later (Redman 2001) and it is by far the cheaper option. Making corrections retrospectively can also mean that the incorrect data may have already been used in a number of analyses before being corrected, causing downstream costs of decisions made on poor data, or of reconducting the analyses. Defect prevention is considered to be far superior to defect detection, since detection is often costly and cannot guarantee to be 100% successful at any stage. Some data defect prevention methods are:

4.1.1 Database design

In a conventional database management system (DBMS), the quality of data has been treated implicitly through functions such as recovery, concurrency, integrity, and security control. However, from the data consumer's perspective, those functions are not sufficient to ensure the quality of data in the database. For example, although there are some essential built-in functions for ensuring data quality in a database like integrity constraints and validity checks, they are often not sufficient to win consumers' confidence on data. In fact, data is used by a range of different organisational functions with different perceptions of what constitutes quality data, and therefore it is difficult to meet all data consumers' quality requirements. Thus, data quality needs to be calibrated in a manner that enables consumers to use their own yardsticks to measure the quality.

In database design, although the primary focus is not on data quality itself, there are many tools that have been developed for the purpose of data quality management. For example, it is recommended to build integrity constraints and use normalization theory to prevent data incompleteness and inconsistencies, as well as through transaction management to prevent data corruption. However, those tools are only related to system design and control. Although they can help for making sure of the quality of data in the system, by themselves they are not sufficient to solve the issue of imperfect data in the real world. Data quality is affected by other factors rather than only by the system, such as whether it reflects real world conditions, and can be easily used and understood by the data user. If the data is not interpretable and accessible by the user, even accurate data is of little value (Wang, Kon & Madnick, 1993b). Therefore, a methodology for designing and representing corporate data models is needed. The use of scenarios, subject areas and design rationale was found to be effective in enhancing understanding of corporate data models (Shanks & Darke, 1999). To prevent data value errors, Redman (1996) gave the tips in Table 7.

4.1.2 Accountability

The assigning of accountability for overall data quality can assist organisations to achieve a consistent level of quality control, provide a point of reference for feedback on errors, and provide a point of contact for documentation and queries.

4.1.3 Education and training

Education and training at all levels of the information chain can lead to vastly improved data quality (Huang *et al.* 1999). This starts with the training and education of collectors in the use of good collection procedures and implementation of the needs of the data users, through training of data input operators and technical staff responsible for the day to day management of the databases, through to education of final users as to the nature of the data, its limitations and potential uses.

4.1.4 Documentation and database design

One of the ways of making sure that error is fully documented is to include it in the early planning stages of database design and construction. Additional data quality/ accuracy fields can then be incorporated.

Table 7: Database Design Tips to Improve Data Quality (from Redman, 1996)

	Database Design Tip	Description of Design Tip
1	Create a data value as few times as possible.	Inconsistencies between multiple values often go unnoticed until they are the source of a problem.
2	Store data in as few databases as possible.	Multiple storage makes it difficult to maintain consistency, especially when data change.
3	Put data in machine-readable form as early in the business process as possible.	Computers and scanners are better than people at activities such as reading and inputting data. However, do not assume that computerized data collection is 100% accurate.
4	Minimize data format changes within the business process.	If format changes are necessary, use computers, not people, to make format changes.
5	When obtaining data for the first time, obtain them just before they are first needed.	Existing data values change rapidly. Capture changes to data values as soon as possible after they change.
6	Discontinue gathering and storing data that are no longer useful.	Plan for periodic review of data needs. When data are no longer useful, they need not be destroyed, simply moved to secondary storage.
7	Employ codes that are easy for data creators and users to understand.	Avoid long, numeric, meaningless coding conventions in favor of short, meaningful words or abbreviations.
8	Place edits as near as possible to data creation or modification.	Use edits as input criteria to a database, as opposed to exit criteria from a database to an application.
9	Employ single-fact data wherever possible.	Single-fact data help reduce code complexity and simplify operators' jobs.

4.1.5 Basic data capture and accuracy checks

The human factor is potentially the greatest threat to the accuracy and reliability of information. It is also the one factor that can ensure both the reliability, and generate an understanding, of the weaknesses inherent in any given data set. The first step in data capture may be done through use of skilled data entry operators or through electronic scanning of information. The level of error due to data entry can often be reduced through double-keying, using learning and training software associated with scanning, and through using experts and supervisors to carry out testing of entry on a sample-basis. Data entry entry error account for 85% of the errors in a database. Maletic acknowledged this when he said that data entry is inherently prone to errors both simple and complex (Maletic and Marcus, 2000). The output of those involved in the entry of data into the system should be monitored by the supervisor. The process involves checking a certain number of each operator's records against the number of errors. Not only does this maintain the quality of the data, it allows the operator to learn and improve from making mistakes. Well-designed procedures such as these can assist in educating the new users should be in place. Conversely if there are no procedures, there is little way of ensuring consistency between operators and between tasks.

4.1.6 Auditability

It is important for custodians to know what data have been checked and when. This helps redundancy and stops data records falling through the cracks and being missed. The best way of doing this is to maintain a documented audit trail of validation.

4.1.7 Edit controls

Edit controls involve business rules that determine the permitted values for a particular field. For example, the value in the month field must be between 1 and 12, the value for day must be between 1 and 31 with the maximum value also dependent upon the month etc. Univariate rules apply to a single field (e.g. the month example, above), bivariate rules apply to two fields (e.g. the combination of day and month).

4.1.8 Minimise duplication and reworking of data

Experience in the real world has shown that the use of information management chain (figure 4) can reduce duplication and re-working of data and lead to a reduction of error rates by up to 50% and reduce costs resulting from the use of poor data by up to two thirds (Redman 2001). This is largely due to efficiency gains through assigning clear responsibilities for data management and quality control, minimising bottlenecks and queue times, minimising duplication through different staff re-doing quality control checks, and improving the identification of better and improved methods of working.

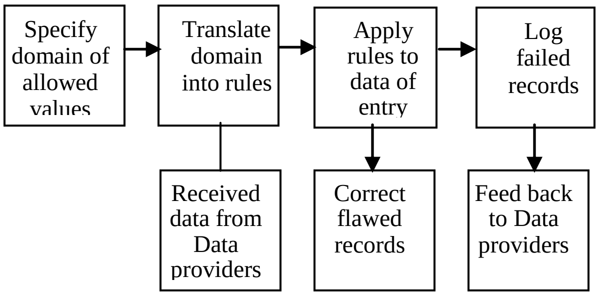

Figure 4: Information mgt chain

4.1.9 Feedback

Users of the data also have a responsibility to data quality. Users need to feed back to custodians information on any errors or omissions they may come across, errors in documentation of the data, and additional information they may need recorded in the future, etc. It is often the user, when looking at the data in the context of other data, who can identify errors and outliers in the data that would otherwise go un-noticed. (Olivieri *et al.* 1995). The user also has a responsibility for determining the fitness of the data for their use, and to not use the data in inappropriate ways.

It is essential that data custodians encourage feedback from users of their data, and take the feedback that they receive seriously. The user often has a far better chance of picking up certain error types through combining data from a range of sources, than does each individual data custodian working in isolation. The development of good feedback mechanisms is not always an easy task. A feedback button can be placed on the query interface page, or an attachment sent to users at the time of downloading data setting out methods for feeding back data errors and comments to the custodians..

4.1.10 User-interfaces

The development of a specific data-entry User Interface can also be a way of decreasing data-entry errors. Many institutions use unskilled staff or volunteers as data-entry operators and the development of a simple (non-technical) user interface that data entry operators feel comfortable with can increase the accuracy of entry. Such an interface can help data input by being able to quickly search authority fields, existing entries in the database, other related databases, and even use search engines such as Google that can help an operator decide on the correct spelling or terminology where they may have difficulty reading a label, or determining what should and shouldn't go into particular fields. In some cases this can be applied through database design that incorporates Authorities tables and drop-down menus (pick lists) that precludes unskilled data-input personnel having to make decisions about names, localities, etc.

4.1.11 Storage of data

The storage of data can have an effect on data quality in a number of ways. Many of these are not obvious, but need to be considered both in the design of the storage vessel (database) and as a unit in the data quality chain.

4.1.12 Backup of data

The regular backup of data helps ensure consistent quality levels. It essential that organisations maintain current disaster recovery and back-up procedures. Whenever data are lost or corrupted, there is a concomitant loss in quality.

4.1.13 Archiving

Archiving (including obsolescence and disposal) of data is an area of data and risk management that needs more attention. Data archiving, in particular by universities, should be a priority data management issue.

4.2 Defect Detection and Correction – Data Cleansing

Prevention of errors does nothing for errors already in the database, however, data validation and cleaning remains an important part of the data quality process). Error detection has a particularly important role to play when dealing with legacy collections (Chapman and Busby 1994). The cleanup process is important in identifying the causes of the errors that have already been incorporated into the database and should then lead to procedures that ensure those errors aren't repeated. Cleanup must not occur in isolation though; otherwise the problems will never disappear. The two operations, data cleaning and error prevention, must run concurrently. To decide to clean the data first and worry about prevention later, usually means that error prevention never gets satisfactorily carried out and in the meantime more and more errors are added to the database. Rahm and Do (2000) states that the term Data cleansing, also called data improvement or *scrubbing*, are used synonymously to mean detecting and removing errors and inconsistencies from data in order to improve its quality. For Lee (2004), the process of cleansing is to improve the quality of data within the existing data structures. Figure 5 shows the process of data cleansing.

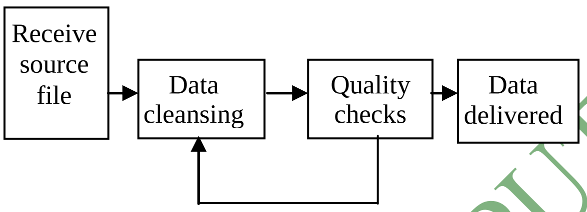

Fig 5: The process of data cleansing

Cleansing data from impurities is an integral part of data processing and maintenance. This has lead to the development of a broad range of methods intending to enhance the accuracy and thereby the usability of existing data. This means standardizing non-standard data values and domains, filling in missing data, correcting incorrect data, and consolidating duplicate occurrences.

The general framework for data cleaning (after Maletic and Marcus 2000) is:

(i) Define and determine error types
(ii) Search and identify error instances
(iii) Correct the errors
(iv) Document error instances and error types
(v) Modify data entry procedures to reduce future errors

The actual process of data cleansing may involve removing typographical errors or validating and correcting values against a known list of entities. The validation may be strict (such as rejecting any address that does not have a valid street code) or fuzzy (such as correcting records that partially match existing, known records). It is often better to retain both the old (original data) and the new (corrected data) side by side in the database so that if mistakes are made in the cleaning process, the original information can be recovered, Chapman (2005). The process of manual cleaning of data is a laborious and time consuming one, and is in itself prone to errors (Maletic and Marcus 2000). A number of tools and guidelines have been produced in recent years to assist with the process of data validation and data cleaning of data.

4.2.1 Data cleansing tools

Data quality tools are available to enhance the quality of the data at several stages in the process of developing a data warehouse. Cleansing tools can be useful in automating many of the activities that are involved in cleansing the data- parsing, standardizing, correction, matching, transformation and householding. Many of the tools specialize in auditing the data, detecting patterns in the data, and comparing the data to business rules. The tools that may be used to extract/transform/clean the source data or to measure/control the quality of the inserted data can be grouped in the following categories (Orli 1997):

- Data Extraction.
- Data Transformation.
- Data Migration.
- Data Cleaning and Scrubbing.
- Data Quality Analysis.

A survey of data quality tools by Barateiro J. and Galhardas reveal that there are hundreds of tools for improving the quality of data in a database. It has been observed the generalization of a new kind of software: ETL tools, which allow the optimization, through user-friendly interfaces, of the alimentation process. Recently, some editors have started to offer tools dedicated to data quality management

4.3 Data Quality Mprovement Strategy

Because of the vast amounts of data held by private and government organizations information systems, there is a need to develop a strategy for capturing and checking of the data. A good strategy to follow (for both data entry and quality control) is to set short, intermediate and long-term goals. For example (after Chapman and Busby 1994):

- **Short term.** Data that can be assembled and checked over a 6-12-month period (usually includes data that are already in a database and new data that require less quality checking).
- **Intermediate.** Data that can be entered into a database over about an 18-month period with only a small investment of resources and data that can be checked for quality using simple, in-house methods.
- **Long term.** Data that can be entered and/or checked over a longer time frame using collaborative arrangements, more sophisticated checking methods, etc. May involve working through the collection systematically.

One goal of any information specialist is to avoid needless error. By directly recognizing error, it may be possible to confine it to acceptable limits.

5.0 CONCLUSION

Data start out as attributes of the real world. They are extracted through some measurement, lab test, or examination; recorded either directly on paper or in a computer system; or stored in human memory prior to recording. The process of recording data may require coding, applying terminology, or other error-prone transformations. The data are collected, aggregated, stored, and manipulated by various systems. Finally, the data are extracted and turned into information in some form of report or statistic. Quality—or the lack thereof—results from the overall performance of these processes.

Several principles of data quality improvement are universal. Data quality must be designed into the data production process, not added after the fact. The quality improvement cycle from the manufacturing industry applies equally well to data production and data quality improvement. Data quality improvement depends on continuous feedback to the processes producing the data. Continuous feedback is best accomplished by putting each data element to as many uses as possible, ideally as a central part of the data collectors' day-to-day work.

Data quality must be designed into systems using proven engineering principles. Data quality is too often left to chance or given only superficial attention in the design of information systems. While good engineering principles are sometimes applied to software development, data quality is usually left up to the end user. Applying engineering principles to data quality involves understanding the factors that affect the creation and maintenance of quality data. It is helpful to look at data as the output of a data manufacturing process.

6.0 REFERENCES

Ballou D.P. and Tayi G.K. (1999). Enhancing data quality in data Warehouse Environments. Communications of the ACM, 42(1),3-78.

Ballou D.P., Wang R., Pazer H. and Tayi G.K. (1993). Modelling Information Manufacturing Systems to Determine Information Product Quality. Management Science, 462-484.

Ballou D.P. and Pazer H. (2003). Modelling Completeness versus Consistency Tradeoffs in Information Decision Contexts. Transactions on Knowledge and Data Engineering, 15(1), 240-243.

Barateiro J. and Galhardas H. (2005). *A Survey of Data Quality Tools*, Datenbank Spectrum 14.

Black S.A. and Porter L.J. (1996). 'Identification of the Critical Factors of TQM', *Decision Sciences*, Vol. 27, pp 21

Bowen P. (1993). *Managing Data Quality Accounting Information Systems: A Stochastic Clearing System Approach*, Unpublished Ph.D. Dissertation, University of Tennessee.

Bowen P., Fuhrer D.A., and Guess F.M. (1998). Continuously Improving Data Quality in Persistent Databases. Data Quality, 4(1).

Chapman A.D. (2005). Environmental Data Quality – b. Data Cleaning Tools. Campinas, Brazil: CRIA 57 pp. http://splink.cria.org.br/docs/appendix_i.pdf [Accessed 14 Jul. 2004]

Chapman A.D. and Busby J.R. (1994). Linking Plant Species Information to Continental Biodiversity Inventory, Climate and Environmental Monitoring 177-195 **in** Miller, R.I. (ed.). *Mapping the Diversity of Nature*. London: Chapman and Hall.

Dalcin E.C. (2004). Data Quality Concepts and Techniques Applied to Taxonomic Databases. Thesis for the Degree of Ph.D. School of Biological Sciences, University of Southampton. http://www.dalcin.org/eduardo/downloads/edalcin_thesis_submission.pdf. November 2004. 266 pp. [Accessed [Accessed 7 Jan. 2004].

Davis R.E., Foote F.S., Anderson J.M., Mikhail E.M. (1981). *Surveying: Theory and Practice*, Sixth Edition: McGraw-Hill.

English L (1999). Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits. New York, NY: Wiley Computer Publishing.

English L. (2003). Plain English about information quality: How to save $576 925,000 through IQ management. Retrieved 21.08.06, from http://www.dmreview.com/article sub.cfm?articleId=6823

Firth C. (1996). 'Data Quality in practice: Experience from the Frontline'. In: *Proceedings of Conference of Information Quality*.

Heinrich B., Kaiser M. and Klier M.F. (*2007).* "How to measure Data Quality - a Metric Based Approach" *In appraisal for: International Conference on Information Systems, Montréal,*

Huang K.T., Lee Yang W., Wang R. (1998). Quality Information and Knowledge Prentice Hall

Lee Y.W. (2004). Crafting Rules: Context Reflective Data Quality Problem Solving. Journal of Management Information Systems, 20(3), 92-119.

Loshin D. (2001). Enterprise Knowledge Management. The Data Quality Approach. California: Academic Press.

Loshin David. (2003). *Business Intelligence*. San Francisco: Morgan Kaufmann Publishers.

Maletic J.I. and Marcus A. (2000). Data Cleansing: Beyond Integrity Analysis pp. 200-209 In: *Proceedings of the Conference on Information Quality (IQ2000)*. Boston: Massachusetts Institute of Technology. http://www.cs.wayne.edu/~amarcus/papers/IQ2000.pdf [Accessed 21 November, 2003].

Olivieri S., Harrison J. and Busby J.R. (1995). Data and Information Management and Communication. pp. 607–670 **in** Heywood, V.H. (ed.) *Global Biodiversity Assessment*. London: Cambridge University Press. 1140pp

Olson J.E. (2003). Data Quality. The Accuracy Dimension. San Francisco: Morgan Kaufmann Publishers.

Orli R.J. (1997). Data Extraction, Transformation, and Migration Tools. Kismet Analytic Corp, http://www.kismeta.com/ex2.htm.

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

Porter L.J. and Parker A.J. (1993). 'Total Quality Management - the Critical Success Factors', *Total Quality Management*, no. 4, pp. 13-22.

Redman T.C. (1996). Data Quality in the Information Age (pp. 303): Artech House.

Redman T.C. (2001). Data Quality. The Field Guide. Boston: Digital Press.

Segev A. (1999). Data Quality Challenges in Enabling e-Business Transformation. Paper presented at the Seventh International Conference on Information Quality, Boston.

Saraph J.V., Benson P.G. and Schroeder R.G. (1989). 'An Instrument for Measuring the Critical Factors of Quality Management', *Decision Sciences*, Vol. 120, No. 4, pp. 457-78.

Shanks G. and Darke P. (1998). 'Understanding Data Quality in Data Warehousing: A Semiotic Approach', In: *Proceeding of the 1998 Conference on Information Quality*, Boston, Massachusetts.

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

Vassiliadis P. (2000). Data Warehouse Modeling and Quality Issues

Wang R.Y., Kon H.B. and Madnick S.E. (1993). 'Data Quality Requirements Analysis and Modeling'. In: *Proceedings of the Ninth International Conference of Data Engineering*, IEEE, Computer Society Press, Vienna, Austria.

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

Wang R.Y. and D. Strong (1996). Beyond accuracy: What data quality means to Data Consumers. *Journal of Management Information Systems* 12 (4): 5-34.

Yusof S.M. and Aspinwall E. (1999). 'Critical Success Factors for Total Quality Management.

Implementation in Small and Medium Enterprises', *Total Quality Manage-ment*, p. 803.