

Influence of Cultural Issues on Data Quality Dimensions

TATJANA WELZER, MARKO HÖLBL, LILI NEMEC ZLATOLAS, MARJAN DRUZOVEC, University of Maribor

Successful software – like information system is meaning nothing if the work is not supported by correct and high quality data, what means, that the data are an important part of software or application especially data applications. Increasing quantity of data and demand for integrated and complex data applications requires high quality in data modelling and data. New concepts, tools and techniques for a database modelling, development and retrieval are required with a final goal: better data and information quality. One of the possible solutions for higher data quality is the integration of cultural aspects. Cultural aspects include different viewpoints, including country dependent parameters and business and domain dependent cultural issues. As a consequence, data quality as well as information quality of applications - software improves if the mentioned approach is applied. In this paper the influence of cultural issues on the data quality dimensions and Deming's Fourteen Points will be presented and discussed.

Categories and Subject Descriptors: **H.2.1 [Database Management]:** Logical Design—*Data models*; **K.6.4 [Management of Computing and Information System]:** System Management—*Quality assurance*;

General Terms: Information Quality

Additional Key Words and Phrases: Data Quality, Data Modelling Quality, Culture, Cultural Issues, Deming's Fourteen Points

1. INTRODUCTION

To improve data quality demands in today increasing quantity of data, various approaches are used. Unfortunately quite often these approaches only vaguely take into consideration that the prerequisite for the high data quality are data quality dimensions. The majority of such approaches do not take into account the influence of cultural issues.

Information technology has automated many operations and made data available to more applications and people. However, the progress of information technology also had an impact on data quality by worsening it. Because users often assume that digital data are correct, the guilt is often put on data and its incorrectness. These problems can grow out of proportions especially in data warehouses and big data environments, as well as on the Internet [Welzer 1998], [Welzer 2013]. The data quality dimension needs to get a new presentation and understanding which includes cultural issues.

In general data quality is multidimensional and complex, and involves not only data management and modelling but also analysis, quality control and assurance, storage and presentation. As stated by Strong et al. [Strong 1997], data quality is related to a specific use case and cannot be assessed independently of a specific domain and / or user. In a database the data does not have actual quality or value [Dalcin 2004] it only has potential value which is harvested when data is used. In addition, English has introduced information (data) quality as data's ability to satisfy customers and to meet customers' needs [English 1999], whereas Redman, suggested that data must be accessible, accurate, timely, complete, consistent with other sources, relevant, comprehensive, provide a proper level of detail, be easy to read and easy to interpret [Redman 2001]. In such a sense a data administrator needs to consider what may need to be done with the data to increase its usability, increase its potential use and relevance, and make it suitable for a wider range of purposes and users [Chapman 2005].

To fulfil Chapman's statement and the before discussed finding we have to give the conceptual view of data much more attentions by incorporating data quality dimensions and cultural issues. The later have to be incorporated and business and domain aspects have to be considered.

In Chapter 2 we will briefly present some notions, including data and information quality, quality dimensions and Deming's Fourteen Points. An overview of cultural issues is described in chapter 3. The main goal of chapter 4 is to introduce the influence of cultural issues on data quality dimensions and

Author's address: T. Welzer, 2000 Maribor, Slovenia, email: tatjana.welzer@um.si, tel: +386-2-220-7299; M. Hölbl, 2000 Maribor, Slovenia, email: marko.holbl@um.si; L. Nemeč Zlatolas, 2000 Maribor, Slovenia, email: llili.nemeczlatolas@um.si; M. Druzovec, 2000 Maribor, Slovenia, email: marjan.druzovec@um.si

Copyright © by the paper's authors. Copying permitted only for private and academic purposes.

In: Z. Budimac, M. Heričko (eds.): Proceedings of the 4th Workshop of Software Quality, Analysis, Monitoring, Improvement, and Applications (SQAMIA 2015), Maribor, Slovenia, 8.-10.6.2015. Also published online by CEUR Workshop Proceedings (CEUR-WS.org, ISSN 1613-0073)

Deming's Fourteen Points. Finally we will conclude with a summary of the proposed concepts and future research in chapter 5.

2. DATA QUALITY, INFORMATION QUALITY AND DATA QUALITY DIMENSIONS

The concept of quality is difficult to describe because of its amorphous nature and various definitions presented by different authors. This results in the facts that different authors tend to emphasize different aspects of quality [Fox 1997]. When the quality concept was defined, the emphasis was given on how to achieve quality and how to make it compliant with a standard or a specification. Rapid changes in later years have led to new definitions of quality. One of the most well-known and recognized definitions is the IEEE standard definition [IEEE 1998] in which quality is defined as the totality of features and characteristics of a product or service that bears on its ability to satisfy given users' needs.

For further discussion the most important definitions are those of data quality and information quality, as well as the definition of quality dimensions, which are as well presented by different authors.

T.C. Redman defines the data quality in its broadest sense. In his book *Data Quality for Information Age* [Redman 1996] he implies to data quality definition the relevance of intended uses and sufficient details as well as quality with a high degree of accuracy and completeness, consistent with other sources and presented in appropriate ways.

Giri Kumar Tayi and Donald P. Bollou as guest editors of *Examining Data Quality in the Communications of the ACM*, have defined the term data quality as fitness for use which implies that the concept of data quality is relative [Tayi 1998]. Data appropriate for one use may not possess sufficient quality for another use. Or opposite, already used data comply with some kind of quality. A related problem with multiple users of data is also that of semantics. The data designer and/or initial user may fully agree with same definitions regarding the meaning of the various data items, but probably other users will not share the same view. Such problems are becoming increasingly critical as organizations implement data warehouses, using big data or taking into account different cultural aspects, according to business and expert areas. At the same time the conceptual view on data including cultural issues is more and more important owing to the facts that it can a possible solution for the mentioned problems.

The data quality definition of Ken Orr [Orr 1998] introduces a kind of measurement view on the term. It is defined as a measure of the agreement between the data views presented by an information system and the understanding of the same data by the user. Data administrator wants to ensure that data is accurate enough, timely and consistent for the enterprise to survive and make reasonable decisions. However, the most significant problem of data quality is the facts that it often changes. Data in a database is mostly static, but in the real world it is rapidly changing. One reason more to apply a conceptual view influenced by cultural issues.

If defining and understanding data and data quality is difficult and it varies, then defining and understanding information is a hornet's nest. In some environments the term information mistakenly refers to both data and information [Strong 1997]. Data usually refers to information at their early stages of processing and information to the product at a later stage when the meaning is added. Rather than switching between the terms information is used to refer to data or information values at any point in the process. But still we must bear in mind that different information definitions depend on different points of view. For example:

- From the information management point of view, information is simply processed data [Redman 1996].
- From the information theory point of view, information is the non-redundant part of a message [Redman 1996].
- From the information technology for management point of view, information is data that has been organized in a way that is has meaning to the user [Turban 1996]

However once a point of view is fixed, no conflict should arise. Once again it is important to emphasize that the prerequisite for information quality is data quality.

But to get a better view on data quality, particularly from the conceptual point of view, data quality dimensions have to be introduced. Redman defined for the users' perspectives 15 characteristics of an ideal view [Redman 1996]:

- Relevance – data that is needed by the application;
- Obtainability – values should be easily obtained;
- Clarity of definition – all terms should be clearly defined;
- Comprehensiveness – all required data should be available and included;
- Essentialness – unneeded data should not be included;
- Attribute granularity – right level of definitions and abstractions for data;
- Domain precision – appropriates of domains;
- Naturalness – existence in the real world;
- Occurrence identifiability – identification of entities (data);
- Homogeneity – minimization of unnecessary attributes;
- Minimum redundancy;
- Semantic consistency – clear and consistent view;
- Structural consistency;
- Robustness – wide view;
- Flexibility – easy to change.

We also have to emphasize Deming's Fourteen Points. Deming defined his 14 points or key principles with the intention to make easier implementation of changes in companies, departments and teams. They are a guide to the importance of building users awareness. And from that point of view it is important to introduce cultural issues into data quality dimensions.

3. CULTURAL ISUESS

One of the most familiar words in any community is culture. The word itself is used in different combinations and meanings, which leads to many definitions of culture. Culture can be presented as an artistic activity, as a social, philological or anthropological concept or as a culture of groups, societies and countries. The world itself has grown over the centuries to reach its currently broad understanding [Baldwin 2004]. However, culture is not something that we simply absorb; it is something that we have to learn although the common knowledge is mostly opposite.

The area of culture has been studied by well-known researchers and we are faced with different definitions and explanations that are showing us authors' point of view on the topic [Welzer 2011]. Hofstede for example defined culture as a collective phenomenon, because it is shared with people who live or lived within the same social environment. According to Hofstede culture consists of unwritten rules of social game. It is the collective programming of the mind that distinguishes the member of one group or category of people from others [Hofstede 2001]. Lewis, another important researcher, explains that the culture is an integrated pattern of human knowledge, a core belief, and a behaviour that depends upon the capacity for symbolic thought and social learning [Lewis 2007]. Culture also refers to the cumulative deposit of knowledge, experience, beliefs, values, meanings, hierarchies, religion, notions of time, roles, spatial relations, concepts of the universe and material objects and possessions acquired by a group of people in the course of a generation through individual and group striving [Schneider 2003]. If people adjust to cultural differences, they can better face challenges and become better in their own profession [Welzer 2010].

The summary of all these different, but also similar definitions, is the definition given by Rossinski, which understand culture in the frame of a group as a set of unique characteristics that distinguishes its members from another group [Rossinski 2003]. This definition can be easily applied to nations and subgroups, as well as business environments (business and corporate culture) [Welzer 2010].

In addition to other important terms, there is also a very important term: cultural awareness. None or poor cultural awareness means a poor understanding of cross-cultural dialogue, which can lead to blunders and damaging consequences, especially in business, management and advertising, where cultural awareness seems to be of key importance for success. However, engineering, medicine and many other areas are also not immune to it [Hofstede 2004]. According to the definition, cultural awareness is the foundation of communication, and it involves the ability of observing our cultural values, beliefs and perceptions from the outside [Hofstede 2004]. Cultural awareness is important in communication with people from other cultures, and we have to understand that people from different cultural societies might see, interpret and evaluate things in different ways.

A good illustration of culture and cultural awareness can also be found in cultural levels, as defined by Alvesson and Berg [Alverson 1992]. They introduced different levels within the concept of culture [Alverson 1992]: culture in societies and nations, regional and local, business cultures, organizational and corporate, functional subcultures at the organizational level, social groups in the organization, professional and functional cultures. The numbered levels contain subgroups, or rather, specific cultures according to social life, geographical location and business domain, including enterprise and organizational culture [Welzer 2011]. Such a definition of culture is probably more comprehensible to engineers and other business and technical groups because they are more familiar with this presentation of culture.

4. CULTURAL ISSUES AND DATA QUALITY DIMENSIONS

Quality and culture on their own represent two big areas of research and bear high importance in development of new products as well as in behaviour of users. In this paper we would like to introduce the influence of cultural issues on data quality dimensions. In Chapter 2, 15 characteristics and Deming's Fourteen Points adapted for data are presented and are an important introduction to this chapter.

According to 15 characteristics as well as fourteen points an introduction the possible cultural issue was done:

- Relevance – data needed by the application – *different cultural point of view can provide different data needed*
- Obtainability – values should be easily obtained – *according to cultures and flowing laws, the obtaining of some data could be not as easy as expected*
- Clarity of definition – all terms should be clearly defined – *we can find different definitions for same or similar data in different environments*
- Comprehensiveness – all data needed should be available and included – *some cultures may not allow the collection of some data*
- Attribute granularity – right level of definitions and abstractions for data – *it must be clear what are sensitive data for which cultures*
- Domain precision – appropriates of domains – *culture sensitives of domains*
- Naturalness – existing in real world – *counterpart in the real word can differentiate from culture to culture*
- Occurrence identifiability – identification of entities (data) – *different points of view in culture*
- Homogeneity – minimization of unnecessary attributes – *some of the attributes can be results of cultural differences*
- Semantic consistency – clear and consistent view – *strong cultural influence*
- Robustness – wide view – *generalization in the cultural point of view*
- Flexibility – easy change – *limitations from cultural point of view are possible.*

In addition, points of view of influence of cultural issues on data quality dimensions is given by Deming's Fourteen Points for quality management, adapted for data [Redman 1996]. Through this the new philosophy of understanding quality from cultural point of view is becoming even stronger [Redman 1996], [Welzer 1998], [Welzer 2013]:

- Point 1 – Recognize the importance of data and information to the enterprise. *Same data has different meanings and importance in different enterprises and so also cultures.*
- Point 2 – Adopt new philosophy. The enterprise can no longer live with currently accepted levels of data quality. *Introducing cultural issues is confirming this point.*
- Point 3 – Cease dependence on error detection. Eliminate the need for error detection by building accuracy and other quality attributes into processes that create data. *Culturally influenced detection of errors.*
- Point 5 – Constantly improving the systems by which data are produced and used to create value for customers, the enterprise and its stakeholders. *Introducing the cultural issues is an improvement.*
- Point 6 – Institute job training. *Cultural awareness help individuals and organizations to understand how the cultural issues impact data.*

- Point 7 – Teach and institute leadership for supervisors on workers, who produce data. Managers of organizations that produce data must become responsible for quality, not simply numeric production. The entire enterprise productivity will improve with improved data. *Cultural point of view has to be introduced in those activities to confirm improvement for the entire enterprise.*
- Point 9 – Break down barriers between organizations. *Application, functional and business domain ensure a free flow of cultural awareness across the organizational boundaries.*
- Point 12 – Remove barriers standing between data products and their rights to pride in their work. *Cross cultural awareness motivates designers to different solutions and models.*
- Point 13 – Institute training on data and information, their roles in the enterprise and how they may be improved. *The training has to be supported by cultural issues and cultural awareness.*
- Point 14 – Create a structure in top management that recognizes the importance of data and information and their relationships to the rest of the business. *Cultural issues supports this recognition and the top management can always find a support for understanding data in existing models and applications.*

5. CONCLUSION

There is no doubt that data quality is needed, but as with many other activities we have to agree on adequate measures and this is highly demanding activity.

In the case of data quality we have the support different guidelines like Data Quality Dimension Characteristics or Deming's Fourteen Points. Deming's Fourteen Points for Quality Management were adapted for data by Redman [Redman 1996] and in this paper we have suggested an adaptation of Deming's Fourteen Points as well as 15 characteristics, taking cultural issues into consideration. In this paper we have presented characteristics and points that could be adapted for cultural issues. Additionally, we presented cultural issues and comments concerning them.

Cultural issues as well as comments are based on experiences and previous research [Welzer 2010], [Welzer 2011], [Welzer 2013] about data (same or similar data) in different environments and cultures. For example some cultures (business environments) are using maiden name (a possible attribute), while some others do not use it at all and operate only with the family name (an attribute as well) from which the maiden name can be derived, but only in some environments and cultures, what do not guarantee an easy obtainability, neither clear definitions and availability. Also attribute granularity, domain precision and naturalness as well as semantic consistency and robustness are effected in this case. To obtain the cultural issues in numbered characteristics as guidelines the Deming's Fourteen Points adapted for data can be used.

In further research policy and strategy has to be involved to make the guidelines clear and suitable for testing in daily processes connected to data and quality.

REFERENCES

- Alvesson, M., Berg, P.-O. 1992. Corporate Culture and Organizational Symbolism
- Baldwin, E. et al. 2004. Introducing Cultural Studies. Prentice Hall, Harlow. W de Gruyter, New York.
- Chapman, A. D. 2005. Principles of Data Quality, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen.
- Dalcin, E.C. 2004. Data Quality Concepts and Techniques Applied to Taxonomic Databases. Thesis for the degree of Doctor of Philosophy, School of Biological Sciences, Faculty of Medicine, Health and Life Sciences, University of Southampton. November 2004. 266 pp. http://www.dalcin.org/eduardo/downloads/edalcin_thesis_submission.pdf [Accessed June 1, 2015].
- English, L.P. 1999. Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits. New York: John Wiley & Sons, Inc.
- Fox, C. and Frakes, W. 1997. Elements of the Quality Paradigm. Communications of the ACM, vol. 40, no. 6, pp. 26. IEEE (1989).
- Hofstede, G. 2001. Culture's Consequences, Comparing Values, Behaviors, Institutions and Organizations Across Nations, Sage Publications, Thousand Oaks.
- Hofstede, G., Hofstede, G.J. 2004. Cultures and Organizations: Software of the Mind: Intercultural Cooperation and its Importance for Survival, McGraw-Hill, New York.
- IEEE Standards. 1998 – Third Edition. New York: IEEE.
- Lewis, R.D. 2007. When Cultures Collide, Managing Successfully Across Cultures, Nicholas Brealey Publishing, London.
- Orr, K. 1998. Data Quality and Systems Theory. Communications of the ACM, vol. 41, no. 2, 66-71.

- Redman, T.C. 1996. *Data Quality for the Information Age*. Boston, London: Artech House.
- Redman, T.C. 2001. *Data Quality: The Field Guide*. Boston, MA: Digital Press.
- Rosinski, P. 2003. *Coaching Across Cultures*, Nicholas Brealey Publishing, London.
- Schneider, S.C., Barsoux J-L. 2003. *Managing Across Cultures*, Prentice Hall, Harlow
- Strong, D.M., Lee, Y.W. and Wang, R.Y. 1997. 10 Potholes in the Road to Information Quality. *IEEE Computer*, vol. 30, no. 10, 38-46.
- Strong, D.M., Lee, Y.W. and Wang, R.W. 1997. Data quality in context. *Communications of ACM*, vol. 40, no. 5, 103-110.
- Tayi, G.K. and Ballou, W. 1998. Examining Data Quality. *Communications of the ACM*, vol. 41, no. 2, 54-57.
- Turban, E.; MvLean, E. and Wetherbe, J. 1996. *Information Technology for Management*, New York: John Wiley & Sons.
- Welzer, T. and Rozman, I. 1998. *Information Quality by Meta Model*. *Proceedings Software Quality Management VI*. London: Springer, 81-88.
- Welzer, T., Družovec M., Cafnik, P., Zorič Venuti M., Jaakkola H. 2010. Awareness of Culture in e-learning. *ITHET 2010, IEEE*, pp. 312-315.
- Welzer, T., Holbl, M., Družovec, M., Brumen, B., 2011 *Cultural awareness in social media*. *DETECT 2011, New York: ACM, 2011*, pp. 1-5.
- Welzer, T., Družovec, M., Holbl, M., 2013 *Cultural issues as a components of data modelling quality : lecture*, presented on *SQAMIA 2013, 2nd Workshop on software quality, analysis, monitoring, improvement and applications*, Novi Sad, September 2013.