

**Abstract – Postgraduate Track (doctoral level)**

**MEASURING DATA QUALITY IN DATA WAREHOUSE  
SYSTEMS**

**Markus Helfert**

University of St. Gallen  
Mueller-Friedberg-Strasse 8  
CH-9000 St. Gallen  
Switzerland

Phone: +41-71-224 33 82

Fax: +41-71-224 21 89

Markus.Helfert@unisg.ch

## **Abstract – Postgraduate Track (doctoral level)**

### **MEASURING DATA QUALITY IN DATA WAREHOUSE SYSTEMS**

Data warehousing has captured the attention of practitioners and researchers alike, whereas many projects in practice lack of high quality data (English 1999; Helfert 2000). Thus, quality considerations have accompanied data warehouse research from the beginning, but still there are funded data quality management principles and data quality measurement techniques missing. Current research aims to provide a comprehensive method for data quality management. Based on total quality management (TQM), method-based Data Quality Management (DQM) enfold organizational structure with roles and responsibilities. The process cycle 'define', 'measure', 'analyze', and 'improve' is building the core process organization for ensuring continuous quality improvement. Techniques, tools, standards, guidelines, and rules are supporting this process. Method engineering provides a structure for integrating these elements into a consistent and comprehensive method (Gutzwiller 1994).

Key part of DQM is a model to define and measure data quality requirements. This will provide a base for managing data quality, identifying quality trends, evaluating effects of quality improvements and performing cost benefit analysis. However, no adequate method or model for integrating quality issues in data models or data specifications exists, so far. (Eppler / Wittig 2000). Current approaches in literature are most based on lists of data quality characteristics (Huang / Lee / Wang 1999). An evaluation of these approaches shows that there is still lack of integrated quality models, which could be used for modelling data quality requirements. Furthermore, the quality characteristics are not integrated in a measurement systems with quality indicators and measurement techniques. The most models lack of methods for applying it to company specific situations. One approach, which was developed at the MIT, models quality requirements based on Entity-Relationship-modelling (Wang / Kon / Madnick 1993). Another approach, which was developed within an European research project, is concentrated on technical integration of data quality requirements in the meta data management (Jarke et al. 2000). Even if these approaches have contribute for modelling and measuring data quality requirements, there is still further research needed. On one side, the proposed models are designed on an high conceptual level without any technical representation and on the other side they are on a technical level without any link to end users. However, it is necessary to develop an integrated model, which links the conceptual and the technical level together.

Goal of the PhD-Thesis is to develop an integrated model for defining and measuring data quality within data warehouse systems. Therefore a set of relevant quality characteristics and associated techniques for assigning values to these characteristics have to be selected and developed. Present work adapts appropriate data quality characteristics and integrates these into a suitable quality model. On the top level, subjective data quality requirements are modelled conceptually. To objectify the quality measures, the model is then linked to the logical level, which specify target quality requirements. Assessment techniques, as the physical representation of the measurement system, are then linked to the specification.

## References:

- **English, L.:** Improving Data Warehouse and Business Information Quality, New York: Wiley & Sons, 1999.
- **Eppler, M. J., Wittig, D.:** Conceptualizing Information Quality: A Review of Information Quality Frameworks from the Last Ten Years, in Klein, B. D., Rossin, D. F. (ed.): Proceedings of the 2000 Conference on Information Quality, Cambridge, MA: Massachusetts Institute of Technology, 2000, pp. 83-96.
- **Gutzwiller, T. A.:** Das CC-RIM-Referenzmodell für den Entwurf von betrieblichen, transaktionsorientierten Informationssystemen, Heidelberg: Physica-Verlag, 1994.
- **Helfert, M.:** Eine empirische Untersuchung von Forschungsfragen beim Data Warehousing aus Sicht der Unternehmenspraxis, Working Paper BE HSG/CC DWS/05, Institute of Information Management, University of St. Gallen, 2000
- **Huang, J., Lee Y. W., Wang R. Y.:** Quality Information and Knowledge; Upper Saddle River, NJ: Prentice Hall 1999.
- **Jarke, M., Lenezerini, M., Vassiliou, Y., Vassiliadis, P.:** Fundamentals of Data Warehouses, Berlin et al.: Springer 2000.
- **Wang R. Y. and Kon H. B. and Madnick S. E.:** Data quality requirements analysis and modeling, in Proceedings of the 9th International Conference on Data Engineering (ICDE)}, Vienna, Austria: IEEE Computer Society, 1993, pp. 670-677.