

Data Quality Problems and Proactive Data Quality Management in Data-Warehouse-Systems

Research Paper

Markus Helfert

University of St. Gallen

Institute of Information

Management

Mueller-Friedbergstrasse 8

CH-9000 St. Gallen

(Switzerland)

phone: ++41-71-224 33 82

fax: ++41-71-224 21 89

email: markus.helfert@unisg.ch

Gregor Zellner

University of St. Gallen

Institute of Information

Management

Mueller-Friedbergstrasse 8

CH-9000 St. Gallen

(Switzerland)

phone: ++41-71-224 33 48

fax: ++41-71-224 21 89

email: gregor.zellner@unisg.ch

Carlos Sousa

University College Dublin

The Michael Smurfit

Graduate School of Business

Dublin

(Ireland)

phone +353-1-716 8811

fax +353-1-716 8993

email: carlos.sousa@ucd.ie

Keywords:

Data Quality, Data Quality Management, Data Warehouse Systems

1 Abstract

The following article describes major Data Quality Problems, requirements and common strategies to manage data quality in data warehouse systems. The results are based on a survey among large German and Swiss companies, which has been carried out in 2001. The first section of the article describes different views of data quality. The second section of the article explores data quality in Data-Warehouse-Systems. Nearly all companies have problems with data quality. The problems are usually tackled through data cleansing (reactive) as well as data quality management (proactive). Data quality itself is very complex and comprehends the quality of data modelling, the quality of data values and the quality of the overall system (inclusive software Quality). The results show that data quality characteristics can be classified in quality of design and quality of execution. Consistency, completeness and accuracy are the main characteristics of quality of conformance. Besides these timeliness, referential integrity and syntactical accuracy are important as well. Quality of design deals with the semantic of data, descriptions and modelling standards. Here, consistent descriptions and detailed documentation are very important. The article finishes with an approach for managing data quality in data warehouse systems, the proactive Data Quality Management.