

# Serviceorientiertes Datenqualitätsmanagement auf Basis von Open Source Tools

Julia Klingner, David Ramón Engelhardt, Holger Hinrichs

Fachbereich Elektrotechnik und Informatik  
Fachhochschule Lübeck  
Mönkhofer Weg 239  
23562 Lübeck

julia.klingner@stud.fh-luebeck.de  
david.ramon.engelhardt@stud.fh-luebeck.de  
hinrichs@fh-luebeck.de

**Abstract:** Unternehmen sind heute mehr denn je auf qualitativ hochwertige Datenbestände angewiesen, sei es zur effizienten Abwicklung des Tagesgeschäfts oder zur taktischen und strategischen Entscheidungsunterstützung. Unvollständige, inkonsistente oder redundante Daten stören den Prozessablauf, führen zu Nachbearbeitungsaufwänden und gefährden Managemententscheidungen. Umso wichtiger ist es, Datenqualitätsmängel möglichst frühzeitig zu erkennen. Wendet man den klassischen PDCA-Zyklus nach Deming [De82] auf den Bereich Datenqualität an, müssen zunächst in einer Planungsphase (Plan) aktuelle Qualitätsanforderungen an Daten sowie Messinstrumente spezifiziert werden. Bei der Ausführung von Geschäftsprozessen (Do) wird laufend die Datenqualität gemessen. Die anschließende Analyse von Messergebnissen (Check) führt – je nach Bedarf – zum Ergreifen von Verbesserungsmaßnahmen (Act), z. B. einer Datenbereinigung oder Ursachenbekämpfung.

In dem studentischen Projekt DServ (Serviceorientiertes Datenqualitätsmanagement) wurde eine Softwarelösung konzipiert und implementiert, die insbesondere die Phasen Plan, Do und Check unterstützt. Die Messung der Datenqualität erfolgt dabei über einen Satz von Web Services, die mit dem Framework Apache CXF implementiert sind und über einen Apache Tomcat Server bereitgestellt werden. In einem Planungswerkzeug lassen sich Qualitätsanforderungen beschreiben und mit den Web Services assoziieren. Als Beispielszenario dient ein ETL-Prozess, der mit Pentaho Data Integration (Kettle) erstellt wurde. Im Rahmen von DServ wurde ein Plug-In für Kettle entwickelt, welches es erlaubt, die o. a. Qualitätsplanung (XML) einzulesen, mit dem ETL-Datenstrom zu verknüpfen und dynamisch die in der Planung vorgesehenen Web Services zur Qualitätsmessung aufzurufen. Messergebnisse werden in XML-Dateien abgelegt, die wiederum in ein Dashboard-Werkzeug eingelesen werden. Dort erfolgt eine graphische Aufbereitung, die es dem/der Qualitätsverantwortlichen erlaubt, zeitnah auf visualisierte Datenqualitätsmängel zu reagieren. Die Softwarekomponenten sind in Java geschrieben und verwenden das Binding-Werkzeug JiBX zur Abbildung von XML auf Java-Objekte und umgekehrt.