

DQ-Management als Teil der DG (DQ/2)

Data Governance

Robert Wieland, München

Zusammenfassung

Data-Governance-Serie, Teil 5.2: Im letzten Teil haben wir sowohl Unterschiede als auch Zusammenhänge und Gemeinsamkeiten zwischen Datenqualität und Data Governance herausgearbeitet. Hier knüpfen wir an – ebenfalls verweise ich auf den Teil 2 dieser Artikel-Serie zur vertikalen Dimension der DG, wobei hier nicht die technischen Felder bzw. BDEs im Vordergrund stehen, sondern Datenqualitätsregeln, welche auf diese referenzieren bzw. mit ihnen verlinkt sind. Das vorgestellte Metamodell mit den beschriebenen Abstraktionsebenen stellt ein Data-Governance-Hilfsmittel dar, um Datenqualitätsregeln im fachlichen Kontext zu formulieren und nachzuverfolgen.

Ansatz

Formulierung von DQ-Regeln auf mehreren Abstraktionsebenen.

Analog zur bereits vorgestellten „vertikalen Struktur“ führen wir eine Hierarchie der DQ-Regeln ein: Es gibt dann also mehrere Verwaltungs- bzw. Abstraktionsebenen im DQ-Regelwerk, mit denen die verschiedenen Stakeholder typischerweise im BI-Umfeld zu tun haben. In aufsteigender Abstraktionsfolge sind dies also:

DQ-Hierarchie	Objekt-Hierarchie
technische DQ-Regeln	Techn. Felder
fachliche DQ-Regeln	BDEs
fachliche DQ-Regel (High-Level) für eine Geschäftsfunktion	BDOs

Das hierarchische Datenmodell zur Verwaltung bzw. Management von DQ-Regelwerken wird in einem zentralen Metadaten-Repository – einer Teilkomponente des Metadaten-Management Systems – hinterlegt.

Halten Sie hier bitte einmal kurz inne und lassen Sie sich das auf der Zunge zergehen: Ein Metadaten-Repository, das nicht nur auf 3 voll-integrierten Abstraktions-Ebenen die Datenmodelle und fachlichen Objekte beschreibt, sondern parallel dazu auch noch einen kompletten Satz DQ-Regeln verwaltet. Kennen Sie ein System vom Markt, das so ein Vorgehen unterstützt? Mein Co-Autor Thilo Riegel und ich nicht. Dennoch gibt es solche Systeme, individuell entwickelt. Wir werden in einem späteren Teil darauf zurückkommen.

DQ-Regeln werden als Metadaten für die zu messenden Tabellen und Datenelemente (auf allen Ebenen) betrachtet. Beispielsweise kann sich eine DQ-Regel für eine bestimmte Geschäftsfunktion auf mehrere BDEs und noch mehr Tabellen beziehen. Im DQ-Repository werden alle DQ-Regeln zentral angelegt, verwaltet bzw. administriert – an dieser Stelle wird der Regelbaum über alle Abstraktionsebenen abgebildet.

Über Reporting-Tools können die Regelbäume für Ad-hoc Berichte für interne Anforderungen oder gar für Regulierungsbehörden zusammengestellt werden. Das Metadaten-Management System stellt den viel geforderten fachlichen Kontext sowohl für die Datenanforderungen (fachl. Mappings, Data Lineage bezogen auf BDEs & BDOs) aber auch für die Datenqualitäts-Regelwerke her.

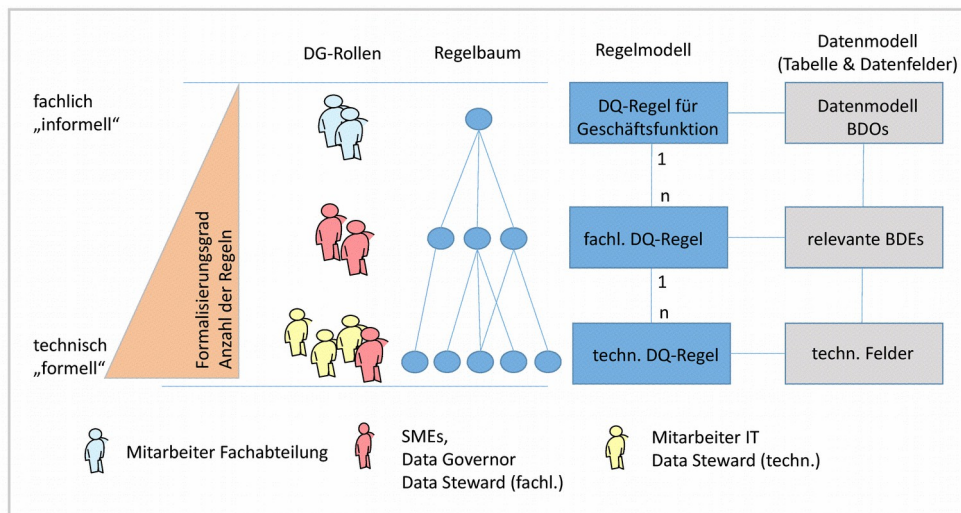


Abbildung 1: Integration der DQ-Regeln in die vertikale Objekt-Struktur

DQ-Regeln existieren auf verschiedenen Ebenen, mit denen eine adäquate Abdeckung der verschiedenen Detailgrade erreicht wird. Der fachliche Kontext von Regeln wird über Entitäten-Levels und BDE-Levels hergestellt.

DQ-Regelwerke werden auf höchster Ebene im Rahmen einer dedizierten Geschäftsfunktion und deren Geschäftsregeln abgeleitet und als Sätze bzw. Formulierungen in Prosa (Freitext) im Metadaten-Repository formuliert. Datenqualitäts-spezifische Kennzahlen (u.a. DQ-Dimensionen) werden ermittelt und auf der niedrigsten Stufe (techn. DQ-Regel) formuliert. Erst auf der technischen Ebene findet der Meßvorgang zur eigentlichen Qualitätverifizierung statt.

Es ist ratsam, eine DQ-Regel immer auf der höchsten Ebene zu formulieren und dann herunter zu brechen. Grundsätzlich ist es aber auch möglich, auf allen Ebenen eine Regel aufzusetzen. Ausschließlich technisch relevante DQ-Regeln (z.B. Validierung eines technischen Schlüsselattributs auf Eindeutigkeit) bringen keinen fachlichen Mehrwert – sie sind deswegen nur auf der untersten Ebene vorhanden.

Praxisbeispiele zeigen oft, dass die Anforderungen bzgl. Überleitungsdokumenten und Datenqualität häufig sehr technisch formuliert sind.

Beispiel: (eine „informelle“ fachliche Regel: „Der Ausfallzeitpunkt muss plausibel sein“ wird konkretisiert auf das „formelle“ SQL-Statement: „DEFAULT_EVENT.EVENT_DTMM < sysdate-1“.

Fazit

Was haben wir damit gewonnen? Nicht nur ein voll-Integriertes Regelwerk, mit dem DQ-Regeln Zielgruppen-gerecht auf verschiedenen Abstraktions-Ebenen verwaltet werden, sondern vielmehr ein komplettes Framework, auf dem sich auf natürliche Art eine Prozessunterstützung im Haus ergibt, die über die einzelne Abteilung hinaus geht, indem es einen gesamthaften Ansatz unterstützt.

So wird es jetzt einfach möglich, die oft so problematische Lücke zwischen IT und Fachlichkeit zu schließen. Endlich ein Ansatz also ohne Medienbrüche!

Bildernachweis

Alle Abbildungen vom Autor.