

Kundenindividuelle Lösungen für ein ganzheitliches Datenqualitätsmanagement

Interview von Frau Katja Hossenfelder, Human Inference GmbH mit Herrn Jan Hüfner, Geschäftsführer der TIQ Solutions GmbH.

Hossenfelder: Herr Hüfner, in letzter Zeit wird der Begriff „Ganzheitliches Datenqualitätsmanagement“ immer wieder im Zusammenhang mit den Erfordernissen zur Verbesserung der Wettbewerbsfähigkeit von Unternehmen in die Diskussion gebracht. Was genau verbirgt sich hinter diesem Begriff?

Hüfner: Die Idee eines ganzheitlichen Datenqualitätsmanagements basiert auf den Konzepten des Total Quality Management (TQM). TQM versteht sich ja als ein System, in dem Qualitätskontrolle nicht auf die Prüfung des Endproduktes reduziert wird. Vielmehr soll die komplette Organisation von qualitätssichernden Maßnahmen durchdrungen sein – vom Einkauf des Rohmaterials bis zum Verkauf bzw. zur Auslieferung des Endprodukts.

Auf der Basis solcher Überlegungen wurde bereits Anfang der neunziger Jahre von einer Forschergruppe des Massachusetts Institute of Technology (MIT) in Boston das Konzept des Total Data Quality Management (TDQM) abgeleitet. Informationen werden in diesem Ansatz als Produkte betrachtet, die es von innerbetrieblichen Akteuren (Datenerfasser, Datenverwalter bzw. Datennutzer) entsprechend zu managen gilt. Eine hohe Datenqualität wird sichergestellt durch einen wohldefinierten „Produktionsprozess“, gemanagt über den kompletten Lebenszyklus der „Informationsprodukte“.

Sie merken schon, hier geht es nicht mehr nur um die schlichte qualitative Aufbereitung von Daten, die – wenn überhaupt – heute oft noch ohne jede Methodik und nicht selten per Hand bewerkstelligt wird. Vielmehr wird eine der Bedeutung der Informationen adäquate Unternehmenskultur, eine permanente Kontrolle bzw. Prävention erforderlich, um wirklich dem Anspruch auf Ganzheitlichkeit gerecht zu werden.

Hossenfelder: Warum erlangt das Thema Datenqualität in den Unternehmen gerade heute zunehmend an Gewicht?

Hüfner: Begünstigt durch die zunehmende Verfügbarkeit leistungsfähiger Rechnersysteme hoher Speicherkapazität werden in den Unternehmen seit einigen Jahren Datensammlungen aufgebaut, deren Volumina oft im Terabyte-Bereich liegen. Die unternehmensrelevanten Daten haben ja längst einen strategischen Stellenwert erhalten. Dabei spricht man derzeit von einer Zunahme der Datenvolumina zwischen 50 und 150% pro Jahr. Nach einer Studie der Gartner Group wird das Datenaufkommen eines mittleren Unternehmens im Jahr 2005 schon bis zu 500 Terabyte betragen. Experten schätzen, dass bis zu 20% der gespeicherten Daten fehlerhaft sind. Die Meta Group geht in ihrer aktuellen Studie deshalb von einem Bedeutungszuwachs des Datenqualitätsmanagements von 20-30% für die nächsten zwei bis vier Jahre aus.

Die Wettbewerbsposition der Unternehmen ist immer stärker von der Korrektheit und Verfügbarkeit, aber auch von der Aktualität und vom richtigen Verständnis des jeweils relevanten elektronischen Datenmaterials abhängig. Es dient in den verschiedenen Unternehmensbereichen als Grundlage für strategische und taktische Entscheidungen sowie für das operative Geschäft. Qualitativ schlechtes Datenmaterial führt zu kostenträchtigen Fehlentscheidungen oder auch zu Störungen in den Betriebsabläufen. Konkreter: Es kann beispielsweise zu massiven Beeinträchtigungen in den Kundenbeziehungen führen, in besonders heiklen Fällen können strategische Chancen der Unternehmensausrichtung verpasst werden, in jedem Fall leidet irgendwann auch der Unternehmenswert darunter.

Hossenfelder: Gibt es für Datenqualitätsmanagement verlässliche Kosten-Nutzen-Schätzungen?

Hüfner: Offen gestanden: Dies ist ein schwieriges Thema, das mich schon lange beschäftigt. Wie stelle ich einem Kunden im



Herr Jan Hüfner

Vorfeld eines Projektes den Nutzen einer möglichen Verbesserung seiner Datenqualität genauer dar?

Bei Kunden- bzw. Adressdaten mag das noch relativ einfach gelingen, indem ich mir überlege, was es kostet, wenn ich z.B. einen Katalog mehrfach an die gleiche Adresse versende, etwa weil ich in meiner Kundendatenbank Dubletten besitze, ohne es zu wissen bzw. eine Haushaltsverknüpfung nicht existiert. Oder wenn ich den Katalog zurückerhalte und mich bei entsprechenden Stellen über die richtige Adresse informieren muss, bevor ich ihn wieder rausschicke.

Bei Millionen von Kunden und vielen Mehrfacherfassungen oder falschen Adressen ist der Nutzen schon enorm. Vielleicht ist der Bereich Kundenbeziehungsmanagement in Datenqualitätsfragen deshalb so weit fortgeschritten, weil man hier Kosten und Nutzen vergleichsweise einfach gegenüberstellen kann? Aber auch Imagefragen spielen dabei als weiche Kostenfaktoren eine wichtige Rolle, die so genau gar nicht bezifferbar sind. Das kann dann nur ex post gelingen, wenn man die relevanten Einflussfaktoren entsprechend herauszufiltern vermag.

Mangelnde Datenqualität beeinflusst aber auf jeden Fall die Kosten in allen Bereichen einer Unternehmung in denen die verschiedenen Arten von elektronisch gehaltenen Daten wie Produktions-, Finanz- oder Adressdaten eine Rolle spielen. Sie stellt letztlich eine erhebliche Vergeudung von Ressourcen sowie Umsatz- und Gewinnpotenzialen dar und betrifft Verwaltungs-, Kommunikations-, Produktions- und/oder Analysekosten in nicht unerheblichem Maße. Experten gehen bei diesen Kosten von signifikanten Größenordnungen aus, die zwischen 8 und 12% des Umsatzes liegen sollen. Laut Data Warehouse Institute ergeben sich in diesem Bereich allein für die US-amerikanische Industrie und Verwaltung jährliche Schäden in Höhe von 600 Mrd. Dollar. Und selbst in Deutschland haben nur noch 18% der Unternehmen Vertrauen in ihre Daten.

Inzwischen kommt in der Argumentation auch Unterstützung von ganz renommierten Beratungshäusern. Z.B. geht der unlängst herausgegebene Butler-Report zu Fragen der „Datenqualität und Datenintegration“ von einem beträchtlichen Return on Investment für Datenqualitätsmaßnahmen im hier beschriebenen ganzheitlichen Sinne aus. Beispielhaft hat die Firma Microsoft in einem internen Projekt gezeigt, dass entsprechende Investitionen in die Verbesserung der Qualität des Datenbestandes schon im gleichen Jahr ein Vielfaches an Einsparungen ermöglichen, so dass der Nutzen sofort auf der Hand liegt.

Hossenfelder: Für welche Anwendungssysteme und welche Branchen ist Datenqualitätsmanagement ein besonders wichtiges Thema?

Hüfner: Ganzheitliches Datenqualitätsmanagement ist ein Querschnittsthema. Das heißt, in allen Firmen, die mit (mehreren) gro-

ßen Datenquellen arbeiten, bieten adäquate DQM-Maßnahmen entsprechende Kostenvorteile. Das gilt zunächst einmal unabhängig von den jeweils genutzten Anwendungssystemen und Branchen, wobei ich nicht bestreiten will, dass Daten in bestimmten Branchen sensibler sind als in anderen.

Wenn beispielsweise eine Fusion ansteht, bei der die Informationsressourcen verschiedener Unternehmen miteinander abgeglichen und zusammengeführt werden müssen, um die erhofften Synergieeffekte tatsächlich zu ermöglichen, oder wenn das Risikomanagement im Zuge der Basel II-Anforderungen mehr Datentransparenz verlangt: Immer ist eine Verbesserung der Datenqualität die Voraussetzung dafür, um größere interne und externe Reibungsverluste zu verhindern, IT-Projekte im zeitlichen und finanziellen Rahmen sowie mit dem erwarteten Erfolg umzusetzen. Nach Ansicht der Meta Group stellt eine unzureichende Datenqualität eines der Top-Five-Kriterien für das häufige Scheitern von Projekten zur Einführung von Data Warehouse-Systemen dar. Entsprechende Schlüsse muss man natürlich auch für andere moderne Softwarelösungen in den Bereichen Customer Relationship- und Supply Chain-Management, Enterprise-Resource-Planning usw. ziehen.

Die Liste der Beispiele ließe sich noch weiter fortsetzen, denken wir nur an die großen Herausforderungen im E-Commerce. Und wie schon betont, hat Datenqualitätsmanagement letztlich eine organisatorische, ja kulturelle Dimension, die ohnehin über jedes Anwendungssystem hinaus geht. Daraus ergibt sich auch ganz klar, dass das DQM nicht in der Verantwortung des IT-Bereiches liegen kann. Es geht um eine strategische Gesamtentscheidung, die entsprechende Unterstützung durch das Top-Management verlangt, um alle Organisationsprozesse durchdringen zu können, um also dem Faktor „Information“ wirklich zu der Bedeutung zu verhelfen, die ihm im so oft heraufbeschworenen Informationszeitalter für den Erfolg des Unternehmens längst zusteht.

Hossenfelder: Was muss man sich unter solchen DQ-Maßnahmen konkret vorstellen?

Hüfner: Selbstverständlich gehen wir auch beim Total Data Quality Management von einem permanenten Regelkreis aus, in dessen Rahmen die DQ-spezifischen Probleme definiert, gemessen, bewertet und gegebenenfalls gelöst werden. Aber natürlich haben verschiedene Daten für das Unternehmen unterschiedliche Bedeutung und werden deshalb auch im Rahmen eines effizienten DQM differenziert behandelt. Unterliegen Daten beispielsweise nur geringen Änderungshäufigkeiten, reicht es oft, wenn man in bestimmten Abständen automatisierte Qualitätsanalysen und gegebenenfalls Qualitätsverbesserungen im Rahmen von Batchläufen umsetzt. Vorausgesetzt, die Daten haben eine entsprechend hohe Bedeutung für das Unternehmen, so dass sich solche Maßnahmen auch ökonomisch rechtfertigen lassen.

Unterliegen sie hingegen einer hohen Änderungshäufigkeit, dann wird es auf Dauer sicher effizienter sein, wenn man sich darüber

Gedanken macht, wie man die verschiedenen menschlichen bzw. technischen Fehlerquellen von vorn herein eliminieren kann. Ganz klar: Die Ursachen sollen beseitigt werden und nicht die Symptome! Es macht keinen Sinn, im Nachhinein und gegebenenfalls mit ineffizienten Mitteln, immer wieder die gleichen Mängel beheben zu wollen. Hier sind Präventivstrategien gefragt, die durch die unmittelbare Einbindung von Referenzdaten bzw. die Hinterlegung von Plausibilitätsregeln unterstützt werden können.

Für letzteres wiederum müssen aber erst einmal die fachlichen Zusammenhänge der jeweiligen Datenattribute bekannt sein. Damit wären wir bei der Bedeutung von anwendungsspezifischem Wissen für das DQM, was meine Bemerkung von einem Querschnittsthema in gewisser Weise relativiert. Ohne dieses Fachwissen ist nämlich ein DQM in dem dargelegten ganzheitlichen Ansatz gar nicht umsetzbar. Das heißt, in solchen Projekten ist das Know-how des jeweiligen Kunden unverzichtbar. Viele fachliche Zusammenhänge in den Datenstrukturen müssen aber auch erst gewonnen werden, z.B. durch den Einsatz von Data Mining-Werkzeugen.

Hossfelder: Wie kann man das Datenqualitätsmanagement Ihrer Meinung nach am besten durch Systemlösungen optimieren?

Hüfner: Wie ich bereits ausführte, ist das Thema Datenqualitätsmanagement gerade im CRM-Bereich sehr weit entwickelt. In diesem Zusammenhang existieren umfassende werkzeuggestützte Lösungsansätze mit mehreren aufeinander abgestimmten Modulen, die den Anwendern einen echten Mehrwert durch die Aufbereitung von Kunden- und Adressdaten bieten und damit einen wichtigen Beitrag im Rahmen einer ganzheitlichen DQM-Strategie leisten können.

Aber es gibt in den Unternehmen neben Kunden- und Adressdaten auch viele andere, in den Datenbanken strukturierte Informationen, die im Rahmen einer ganzheitlichen Strategie in das Datenqualitätsmanagement einbezogen werden sollten, die aber für die Analyse und Verbesserung ihrer Qualität zum Teil ganz andere Werkzeuge und Lösungsansätze benötigen. Wir befinden uns ja mit dem ganzheitlichen Rahmen ohnehin auf einer anderen Ebene. Und insofern bin ich skeptisch, ob man dies mit einer **standardisierten** Gesamtlösung, wie sie auf dem Markt verschiedentlich angeboten wird, wirklich umsetzen kann. Ich halte jedenfalls nichts von funktionell überfrachteten und oft sehr teuren Pauschalansätzen, die vorgeben, alles zu beherrschen, also für alle unternehmensspezifischen Anwendungsbereiche mit ihren jeweils relevanten Daten eine ideale Gesamtlösung anbieten zu können.

Selbst wenn sie das wirklich könnten: Kein Unternehmen benötigt all diese Funktionalitäten wirklich und sollte daher nach seinen jeweils individuellen Bedürfnissen ein technisch und ökonomisch sinnvolles DQ-Systemkonzept generieren. Dies gilt insbesondere auch deshalb, weil z.B. die bereits im Unternehmen vorhandenen Softwaresysteme oft schon verschiedene Tools enthalten, welche

die Umsetzung einer geeigneten DQM-Strategie bei einem konkreten Problem zumindest teilweise unterstützen können. Sie müssen also nicht im Rahmen einer Pauschallösung noch einmal hinzugekauft werden.

Zum Beispiel sind in den modernen Datenbanksystemen bereits solche Werkzeuge vorhanden, aber auch die sogenannten ETL-Tools (Extraction, Transformation, Loading, z.B. zur Zusammenführung verschiedener Datenquellen und deren Weiterleitung in ein Data Warehouse) besitzen wichtige Funktionalitäten für Datenqualitätsaufgaben, auch wenn sie dieses Themenfeld bei weitem nicht erschöpfen. Werkzeuge für ein effizientes Metadatenmanagement, EAI-Tools (Enterprise Application Integration) und Werkzeuge für die Migration von Daten bringen in dieser Hinsicht ebenfalls geeignete Funktionalitäten mit.

Natürlich werden darüber hinaus zusätzliche Softwaresysteme erforderlich, um ein individuelles Paket schnüren zu können. Dafür bieten sich verschiedene Spezialwerkzeuge an, insbesondere für das Data Profiling zur Feststellung inhaltlicher bzw. struktureller Datenqualitätsmängel oder die bereits erwähnten Cleansing-Tools zur Aufbereitung von Kunden- und Adressdaten. Außerdem wären in diesem Zusammenhang noch Systeme zur Generierung und/oder Hinterlegung von Plausibilitätsregeln zu nennen, wenn eine Präventivstrategie umgesetzt werden soll.

Ein weiterer Vorteil der hier favorisierten Individuallösung ist, dass je nach den konkreten Erfordernissen eines Unternehmens die wirklich geeigneten Werkzeuge aus den verschiedenen Angeboten ausgewählt werden können. Man kann also sehr flexibel entscheiden, zumal sich die Anbieter der jeweiligen Lösungen nicht nur rein qualitativ, sondern auch funktional recht deutlich voneinander abheben.

Die Funktionalitäten der verschiedenen Tools müssen im Rahmen des Gesamtpaketes natürlich so miteinander koordiniert werden, dass das jeweils definierte Vorgehen zumindest einem qualitätspolitischen Imperativ folgt. Einen ISO-Standard gibt es in diesem Zusammenhang noch nicht. Dennoch sollten die Lösungsansätze entsprechende Analogien befolgen, um auch in dieser Hinsicht den Projekterfolg zu sichern.

Hossfelder: Herr Hüfner, ich danke Ihnen für dieses Gespräch! 📌

Über den Interviewpartner

Herr Jan Hüfner ist Geschäftsführer der TIQ Solutions GmbH, die sich auf kundenindividuelle Strategielösungen für ganzheitliches Datenqualitätsmanagement spezialisiert hat. Gemeinsam mit den Kunden werden nachhaltig effiziente Lösungen mit erheblichen Kosteneinsparungspotenzialen entwickelt und umgesetzt.

www.tiq-solutions.de