



## BUSINESS WHITEPAPER

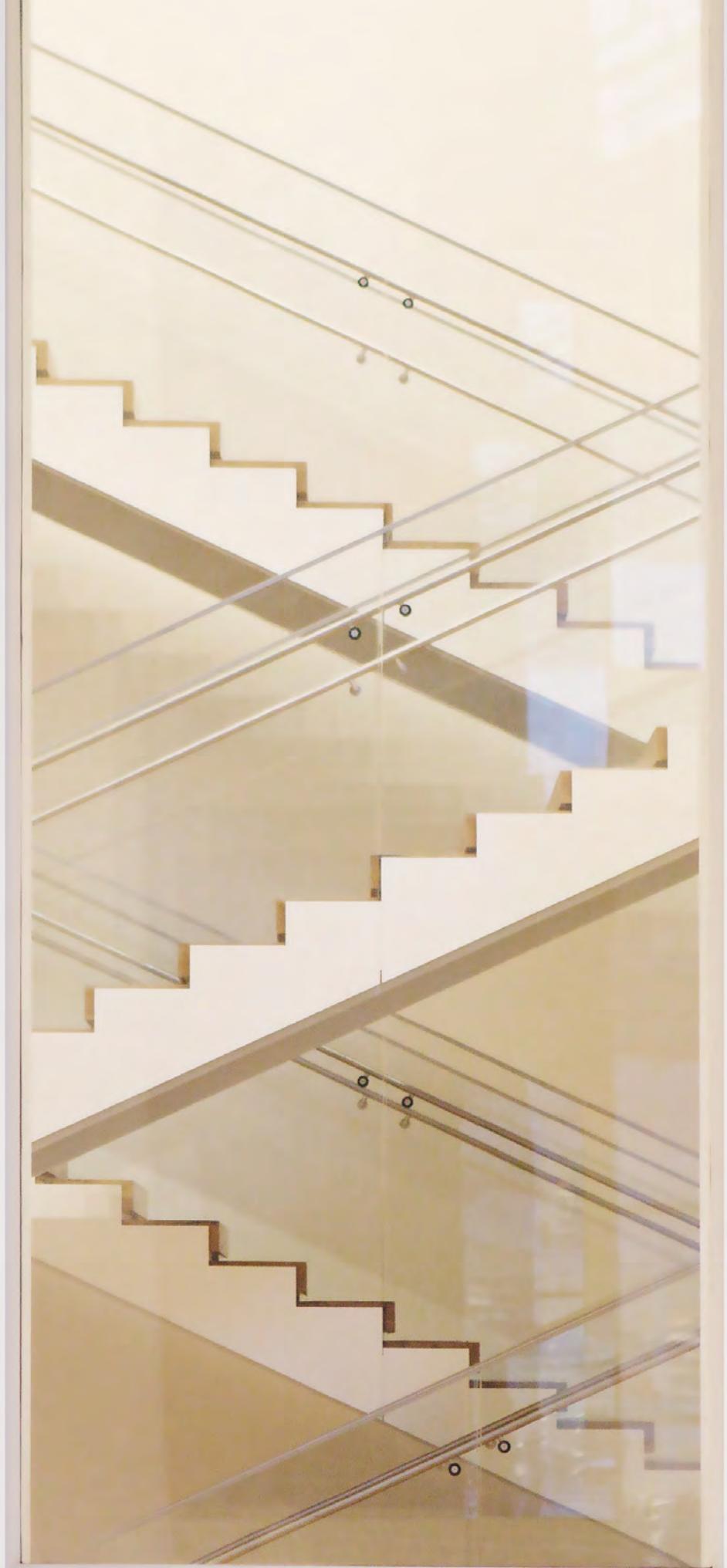
# DATA QUALITY IST MEHR ALS ADRESSVALIDIERUNG

Datenqualität ist weit mehr als Adressvalidierung und Dubletten-Bereinigung. Wir zeigen Ihnen vier unterschiedlichen Stufen auf, die bei der Qualität der Daten beachtet werden müssen, damit sie am Ende als echtes Asset für Ihr Unternehmen gelten.



GROUND  
TRUTH  
by UNISERV





# Inhalt

VORWORT	04
STUFE 1: DIE TECHNISCHE UND FACHLICHE EBENE DER DATENQUALITÄT	07
STUFE 2: DIE INHALTLICHE RICHTIGKEIT DER DATEN	09
STUFE 3: DIE GESCHÄFTLICHE NUTZBARKEIT DER DATEN	10
STUFE 4: UNTERNEHMENSKRITISCHE DATENQUALITÄT MIT DEM GROUND TRUTH	12
EIN BEISPIEL: DATENQUALITÄT IN EINEM TELEKOMMUNIKATIONSUNTERNEHMEN	14
EINWÄNDE, DIE BEI TELEKOMMUNIKATIONSUNTERNEHMEN OFT GENANNT WERDEN	14
ZUSAMMENFASSUNG	16
DATENQUALITÄT BEWERTEN: DATA QUALITY SCORECARD	16



Datenqualität ist weit mehr als nur das Prüfen und gegebenenfalls Korrigieren von Adressen und das Zusammenfinden von ähnlichen Datensätzen.

## Vorwort

In Gesprächen über Datenqualität werden schnell die Begriffe Adressvalidierung und Dubletten-Bereinigung oder Matching genannt. Aber Datenqualität ist weit mehr als nur das Prüfen und gegebenenfalls Korrigieren von Adressen und das Zusammenfinden von ähnlichen Datensätzen.

Die Welt ist digital geworden. Wir informieren uns im Internet. Twitter erweist sich als der schnellste Nachrichtenkanal. Wir kaufen in zunehmendem Maße Produkte und Dienstleistungen im Internet. Soziale Medien produzieren Informationen in bisher nicht bekanntem Volumen. Das mobile Internet macht mit Smartphones und Tablets Informationen allgegenwärtig. Die Information ist zum Motor der Welt geworden – ob beim Einkaufen bei Amazon, beim Steuern von Produktionsstraßen, beim Handeln im Finanzwesen, beim Netzwerken auf LinkedIn und Facebook oder beim Entscheiden im Unternehmen.

Worauf es heute ganz besonders ankommt? Die richtige Information in der richtigen Situation am richtigen Ort rechtzeitig zur Verfügung zu haben. Das spart Kosten, steigert Umsätze und erhöht die Wettbewerbskraft. Die Information

ist nach Arbeit und Kapital zum „dritten Produktionsfaktor“ geworden. Daher betrachtet man heute Informationen als Anlagegut („information as an Asset“). Konsequenterweise müssen wir Informationen im Unternehmen entsprechend behandeln. Dabei geht es nicht mehr nur um Adressvalidierung und Dubletten-Bereinigung. Um das Beste aus den Daten herauszuholen, müssen diese über ihren gesamten Lebenszyklus hinweg gepflegt, geschützt und überwacht werden.

Die Zielsetzung von effektivem Information Management ist es, vertrauenswürdige und belastbare Daten zu schaffen, die Ihr Business erfordert – der Ground Truth.

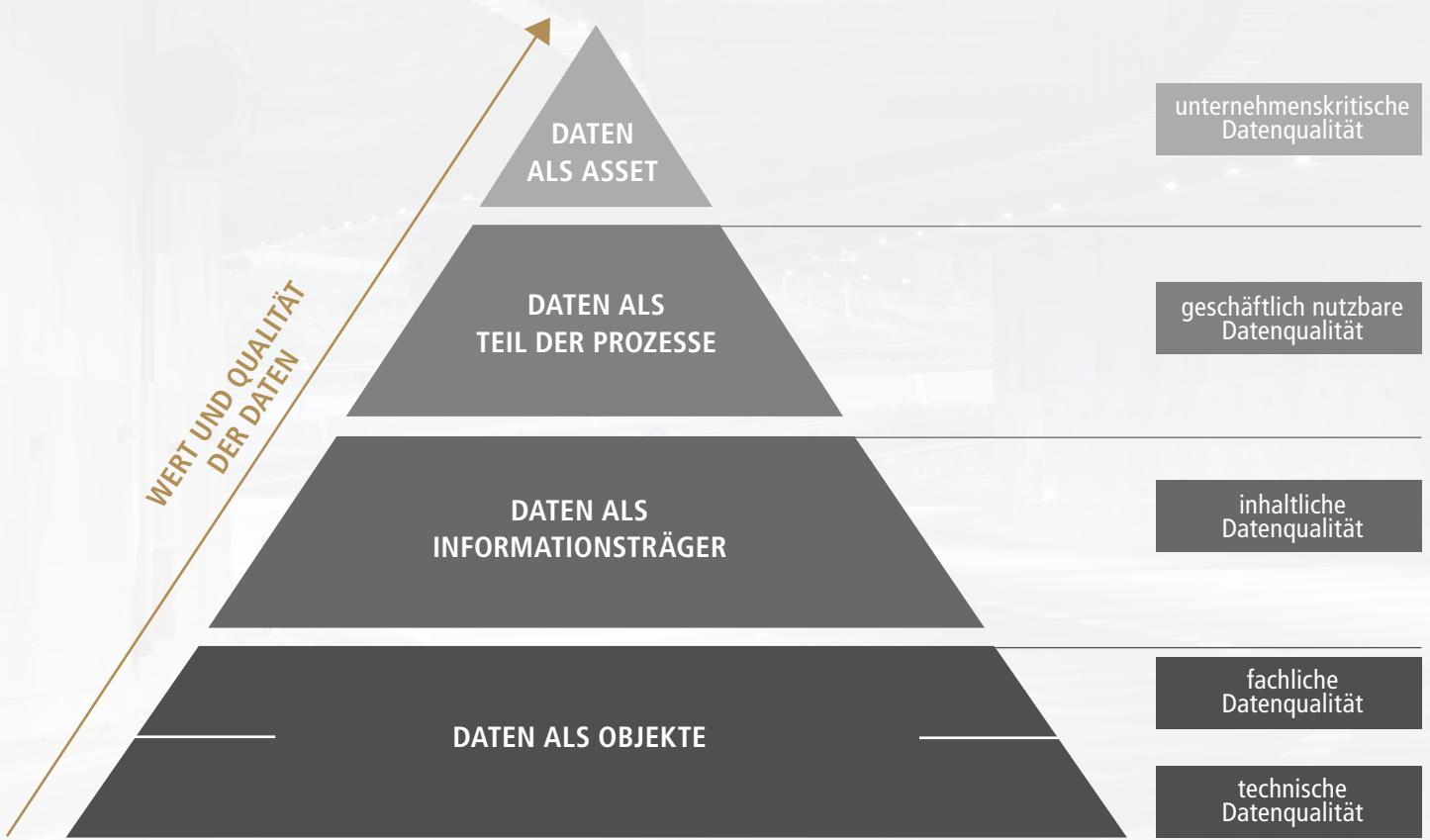
Die Uniserv GmbH kann mit ihrer Produktpalette Unternehmen dabei helfen diesen Ground Truth zu erreichen, indem Ihre Daten geprüft, bewertet, bereinigt und Ergebnisse auch sichtbar gemacht werden. Ob die so gewonnene Datenqualität dann auch dazu führt, dass Ihr Unternehmen aus den qualitativ hochwertigen Daten einen Mehrwert gewinnt, hängt davon ab, ob diese Daten in dem richtigen Kontext benutzt werden. Was nutzt es einem Unternehmen, in seinen

Daten geprüfte und qualitätsgesicherte E-Mail-Adressen von Kunden zu haben, wenn in der Wertschöpfungskette des Unternehmens Emailadressen keine Rolle spielen?

Dieses Whitepaper schafft Bewusstsein dafür, dass Datenqualität alleine keinen Mehrwert für ein Unternehmen schafft. Wir zeigen Ihnen im Gegenzug, wie Sie in vier Stufen qualitätsgesicherte Daten schaffen, die in Ihren Unternehmensprozessen auch Verwendung finden und damit am Ende auch als echtes „Asset“ für Ihr Unternehmen gelten.

**Walter Passareck,**  
Data Quality Solution  
Architect,  
Uniserv GmbH





DIE PYRAMIDE SOLL DEN WERT DER DATEN SYMBOLISIEREN. JE WEITER OBEN DIE DATEN ANKOMMEN, DESTO HÖHER IST DIE QUALITÄT DER DATEN UND UMSO MEHR WERT HABEN SIE FÜR DAS UNTERNEHMEN.

# Stufe 1: Die technische und fachliche Ebene der Datenqualität

Die nebenstehende Grafik zeigt die vier unterschiedlichen Stufen auf, die bei der Qualität der Daten beachtet werden müssen, damit sie am Ende als echtes „Asset“ für das Unternehmen gelten.

Datenqualität setzt sehr früh an. In der ersten Stufe ergibt sich eine Unterscheidung in zwei Ebenen, der technischen und der fachlichen Ebene. Auf der technischen Ebene geht es erst einmal um die Daten selbst. Gemeint sind die technischen Objekte, ohne schon Inhalte zu zuordnen.

- ▶ Sind die Felder gefüllt?
- ▶ Sind sie mit den Zeichen gefüllt, die erwartet werden?
- ▶ Sind sie richtig formatiert?
- ▶ Ist zum Beispiel eine ID in einer bestimmten Weise aufgebaut? Etwa zwei Großbuchstaben, ein Bindestrich, drei Zahlen, ein Bindestrich und ein Kleinbuchstabe (AA-nnn-a).

Auf der fachlichen Ebene geht es um Inhalte, aber noch ohne geschäftlichen Bezug.

- ▶ Sind Felder mit sinnvollen Inhalten gefüllt?
- ▶ Ist ein Wert an dieser Stelle auch zulässig?
- ▶ Ist ein Datum in einem Feld auch korrekt oder hat sich der 31.02. eingeschlichen?

Auch eine Prüfung von Abhängigkeiten wird auf dieser Ebene durchgeführt. Wenn Feld A mit einem Wert gefüllt ist, dann muss auch Feld B mit einem Wert belegt sein, sonst wäre der Datensatz falsch. Es geht aber noch nicht um den inhaltlichen Bezug der Werte. Nur wenn schon hier die Qualität stimmt, dann kann auch das Datum einen Sinn ergeben und zu einem Mehrwert für ein Unternehmen beitragen.

Die Qualität wird auf dieser Stufe mit regelbasierten Prüfungen automatisiert auf Feldebene durchgeführt. Bei Fehlern können diese unter Umständen automatisch bereinigt werden. In anderen Fällen muss die Überarbeitung manuell durch einen Data Steward oder verantwortliche Personen durchgeführt werden. Auch die Ausschleusung der Daten an den Absender zur Korrektur ist ein mögliches Szenario.



JETZT KOMMEN PRÜFUNGEN WIE DIE ADRESSVALIDIERUNG ODER DUBLETTEN-BEREINIGUNG ZUM EINSATZ, WEIL SIE NUN AUCH SINNVOLL UMGESetzt WERDEN KÖNNEN.

# Stufe 2: Die inhaltliche Richtigkeit der Daten

In der nächsten Stufe, geht es um die Prüfung der inhaltlichen Richtigkeit, also um die Inhalte der Daten. Jetzt kommen Prüfungen wie die Adressvalidierung oder Dubletten-Bereinigung zum Einsatz, weil sie nun auch sinnvoll umgesetzt werden können. Erst wenn fest steht, dass ein Feld für eine Telefonnummer gefüllt ist und auch nur zulässige Zeichen im vorgegebenen Format enthält, ist eine Prüfung auf die Richtigkeit der Telefonnummer sinnvoll und notwendig. Ist eine Adresse korrekt? Ist das eingetragene Produkt tatsächlich definiert? Stimmt die angegebene Email-Adresse? Sind Dubletten vorhanden und wie lassen sie sich zusammenführen?

Sind die Daten auf den ersten beiden Stufen korrekt, können sie schließlich in verbundenen Prozessen genutzt werden. Ein Beispiel sind Produktnummern in einer Bestellung: Sie sind nicht nur in der Bestellung enthalten (gefülltes Feld), sondern auch als Produkte definiert. Also verweist die Produktnummer auf das richtige Produkt. Nur dann

kann die Bestellverarbeitung oder die Versandvorbereitung gestartet werden. Denn ohne korrekte Daten machen die damit verbundenen Prozesse keinen Sinn.

Hier kommen regelbasierte Prüfungen zum Einsatz, allerdings werden nun die Inhalte der Daten gegen ihre Referenzen geprüft. Die Regeln sind also mit einer fachlichen Logik versehen und deshalb sollten sie in enger Abstimmung oder direkt von den Fachabteilungen entwickelt werden. Entsprechende Rollen sind zu definieren und Zuständigkeiten zuzuweisen. Auch spezielle Prüfprozesse, wie die Adressprüfung oder die Prüfung auf Dubletten können jetzt ausgeführt werden. Fehlerhafte Datensätze, die nicht automatisiert oder manuell bereinigt werden konnten, müssen ausgeschleust werden, damit sie in den Folgeprozessen nicht zu Fehlern, Mehraufwand oder gar falschen Entscheidungen führen.



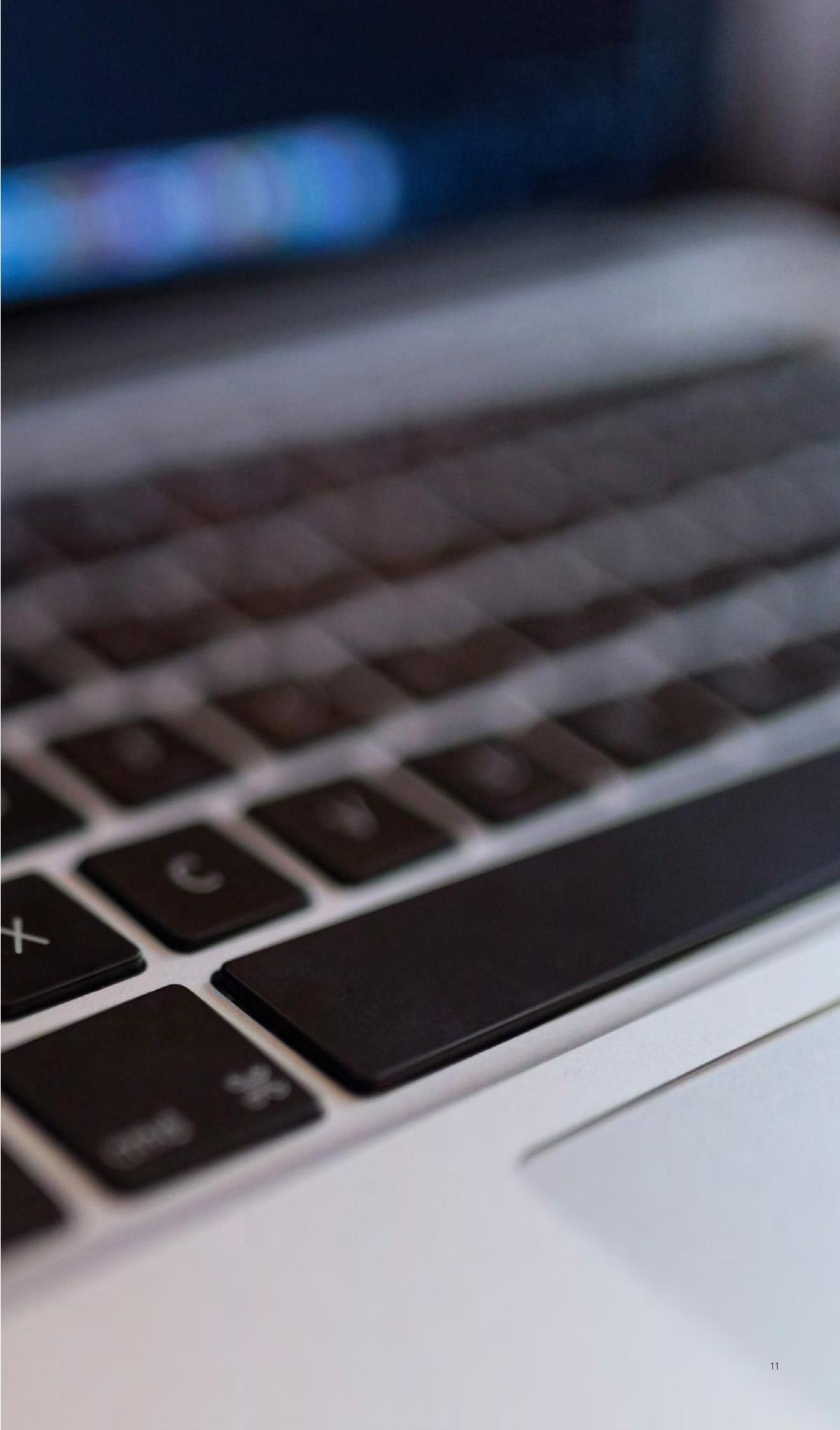
# Stufe 3: Die geschäftliche Nutzbarkeit der Daten

In einer dritten Stufe kann nun auf Basis der geschaffenen Datenqualität das Thema der geschäftlichen Verwendbarkeit angegangen werden. Die Verwendung der Daten kann dabei unterschiedlichster Natur sein und von Unternehmen zu Unternehmen unterschiedliche Anforderungen abdecken:

- ▶ Kann eine Bestellung angenommen und verarbeitet werden?
- ▶ Wie viele ungültige Bestellungen gibt es?
- ▶ Sind die Produkte mit der gewählten Produktnummern im Lager verfügbar (Referenz zum Lager) und in welchen Stückzahlen?
- ▶ Welche Produkte haben sich in dem letzten Monat besonders gut verkauft?
- ▶ Sind bestimmte Produkte auffallend häufig in bestimmten Regionen bestellt oder verkauft worden?
- ▶ Welche Bestellungen konnte man nicht erfüllen, weil Produkte fehlten?
- ▶ Welcher Umsatz ging damit verloren?
- ▶ Bei manchen Unternehmen ist es wichtig zu erfahren, ob ein Besteller schon 18 Jahre alt sind?
- ▶ Bei anderen Unternehmen muss man wissen, ob die bestellten Produkte oder Empfänger auf einer Embargolist zu finden sind?
- ▶ Lassen sich sogenannte Up- und Cross-Selling Potentiale erkennen?

Können die Fragestellungen eines Unternehmens aus den vorhandenen Daten nicht beantwortet werden, kann dies an falschen oder fehlenden Daten liegen. Dieser Mangel muss dann in den darunter liegenden Stufen 1 und 2, den Anforderungen entsprechend, abgestellt werden.

DIE VERWENDUNG DER DATEN KANN UNTERSCHIEDLICHER NATUR SEIN UND VON UNTERNEHMEN ZU UNTERNEHMEN UNTERSCHIEDLICHE ANFORDERUNGEN ABDECKEN.



# Stufe 4: Unternehmenskritische Datenqualität mit dem Ground Truth

In der vierten Stufe ist die Data Quality so hoch, dass unternehmenskritische Entscheidungen auf Basis dieser Daten getroffen werden können. Diese allumfassende Datenwahrheit – Ground Truth – stellt sicher, dass Voraussagen genauer werden und Investitionen zur Steigerung der Wertschöpfung im Unternehmen beitragen können. Ohne die Qualität der Daten besteht das Risiko, Investitionen fehlzuleiten und damit möglicherweise Verluste für das Unternehmen in Kauf zu nehmen. Zwar kann nicht garantiert werden, dass alle Entscheidungen zwangsläufig auch zu einer Wertsteigerung führen werden. Dennoch verringert eine hohe Datenqualität dieses Risiko signifikant. Auch der gesamte Bereich der „predictive Analytics“ macht nur auf qualitätsgesicherten Daten Sinn. Denn nur hoch qualitative Daten können zu einem „Asset“, einem echten Mehr-Wert für das Unternehmen werden.

Der Weg zum „Ground Truth“ führt über fünf Schritte (siehe Abbildung rechts): Innerhalb des ersten Schrittes werden die verschiedenen Datenquellen in ein zentrales Repository – dem Smart Customer MDM – integriert. Das Ergebnis ist der so genannte „Golden Record“, also der Stammdatensatz, in dem alle Daten identischer Personen zusammengefasst werden.

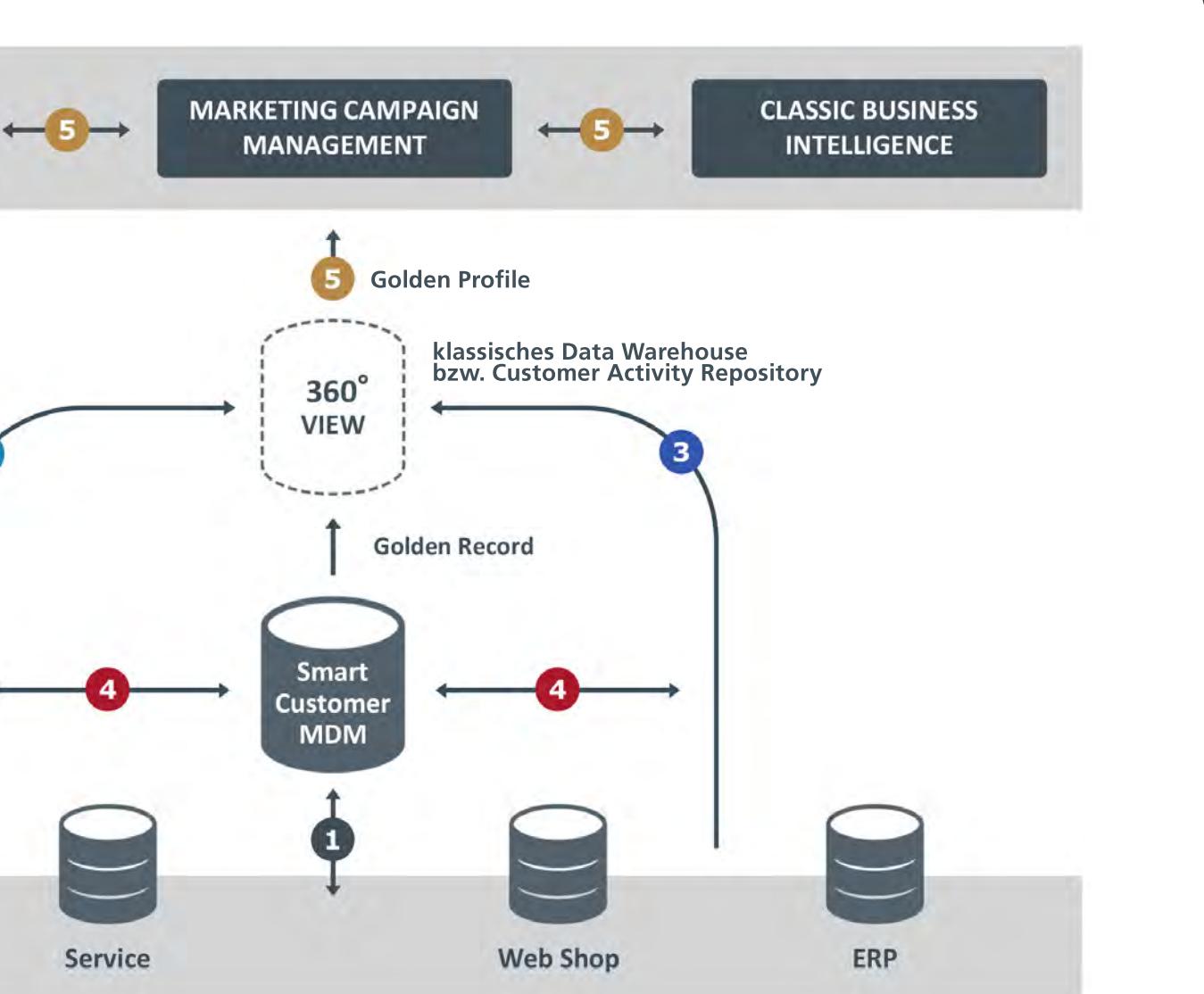
Angereichert mit Interaktions- und Transaktionsdaten und einem Abgleich der Transformationsquellen-ID mit der Golden Record-ID entsteht dann das „Golden Profile“ des Kunden. Erst damit ist ein vollständiger, präziser und aktueller 360 Grad-Blick auf den Kunden möglich. Und erst damit erhält das Unternehmen die korrekte Grundlage für den Predictive-Analytics-Prozess und damit für erfolgreiche Marketingkampagnen und Business-Intelligence-Analysen.

PREDICTIVE ANALYTICS

- 1 Customer Master Data „bidirectional“
- 2 Interaction Data
- 3 Transaction Data
- 4 Transformation Source ID to Golden Record ID
- 5 Golden Profile



CRM



GROUND TRUTH – IN DER VIERTEN STUFE IST DIE DATA QUALITY SO HOCH, DASS UNTERNEHMENSKRITISCHE ENTSCHEIDUNGEN AUF BASIS DIESER DATEN GETROFFEN WERDEN KÖNNEN.

# Ein Beispiel: Datenqualität in einem Telekommunikationsunternehmen

Bei einem Telekommunikationsunternehmen werden in den Verträgen nicht immer alle Produkte erfasst. Der Kunde hat einen Vertrag und eigentlich Telefon, Internet und TV. Im Vertrag sind aber nur Telefon und Internet angekreuzt. Der Preis im Vertrag gilt für alle drei Anschlüsse und bei der Installation schaltet der Techniker auch alle drei Anschlüsse frei als der Kunde reklamiert. In den Unterlagen wird dies dann aber nicht nachgetragen. Ein Neukunde im gleichen Haus möchte nun zu seinem Telefonanschluss auch noch TV dazu bestellen. Dies wird ihm auch bestätigt, da das System einen freien TV Anschluss ausweist. Am Tage der Installation stellt ein Techniker dann fest, dass kein Anschluss mehr frei ist. Der Neukunde ist verärgert, da er seinen Anschluss nicht am zugesagten Termin bekommt. Der Altkunde wird vielleicht verärgert, weil ihm der Anschluss abgeschaltet und dem Neukunden gegeben wird, weil er laut Vertrag keinen TV besitzt. Das Unternehmen muss neue Anschlüsse erstellen lassen, die Teile sind aber gerade nicht auf Lager und müssen bestellt werden. Also viel Ungemach.

Wären die Daten in Stufe 1 und 2 geprüft worden, wäre folgendes festgestellt worden:

- ▶ der Preis passt nicht zu den angegebenen Produkten
- ▶ der Anschluss wäre dazu konfiguriert worden
- ▶ der Anschluss wäre als belegt ausgewiesen worden
- ▶ eine Erweiterung wäre geplant worden
- ▶ fehlende Teile wären nachbestellt worden

Und wenn jetzt auch noch in Stufe 3 aus Analysen erkannt worden wäre, dass in den letzten zwei Monaten 30% mehr Anfragen aus diesem Straßenzug nach TV-Anschlüssen eingegangen sind, hätte in Stufe 4 eine Entscheidung für eine entsprechende Investition sich sicher gelohnt. Der Aufwand an Zeit und Geld für eine Prüfung und Korrektur der Daten ist im Vergleich zu den entstandenen Kosten oder entgangenen Einnahmen eher gering. Zumal die Prüfungen automatisiert durchgeführt werden können. Bei den Korrekturen sind manuelle Eingriffe notwendig, wenn eine automatisierte Berichtigung nicht möglich ist. Diese sollten von Fachpersonal ausgeführt und eventuell erst nach einer Gegenprüfung freigegeben werden.

## EINWÄNDE, DIE BEIM BEISPIEL DES TELEKOMMUNIKATIONSUNTERNEHMENS OFT GENANNT WERDEN

Der Einwand, die Daten kommen aus Systemen, bei denen in der Eingabemaske schon geprüft wird, ob alle Angaben korrekt vorhanden sind, ist sicher richtig und die Systeme werden immer ausgereifter. Aber Hand aufs Herz, in welchem Unternehmen werden Daten nicht doch noch in Excel-Tabellen oder per Hand gepflegten Datenbanken eingestellt und vorgehalten?



DER AUFWAND AN ZEIT UND GELD FÜR EINE PRÜFUNG UND KORREKTUR DER DATEN IST IM VERGLEICH ZU DEN ENTSTANDENEN KOSTEN DER FALSCHEN AUFTRAGSBEARBEITUNG UND ENTGANGENEN EINNAHMEN EHER GERING.

# Zusammenfassung

Will man Datenqualität erreichen, muss bei einer Anlieferung oder der Erzeugung von Daten darauf geachtet werden, dass die Daten vollständig, korrekt und mit allen Abhängigkeiten gefüllt sind. Es sind also Prüfungen auf die Befüllung der relevanten Felder einzurichten. Außerdem sind die Syntax und Inhalte zu prüfen. Und zuletzt auch die Relationen und Plausibilitäten mit anderen Feldern. Erst wenn die Prüfungen auf der technischen und fachlichen Ebene der ersten Stufe bessere Datenqualität geliefert haben, machen weitere Prüfungen zum Inhalt und dann zum geschäftlichen Kontext Sinn. Alle diese Prüfungen lassen sich automatisiert, wiederholbar mit konfigurierbaren Regeln durchführen. Sie müssen aber mit entsprechenden Maßnahmen zur Korrektur fehlerhafter Daten begleitet werden.

Hierfür ist die Rolle des Data Quality Steward vorgesehen. Diese Rolle sollte in einem Unternehmen auch eingeführt und gelebt werden. Der DQ Steward muss dabei mit entsprechenden Befugnissen versehen werden, sonst sind notwendige Anpassungen nicht durchsetzbar.

Zusätzlich sollte darauf geachtet werden, dass auch bestehende Datenmengen periodisch wieder überprüft werden,

da sich auch hier im Laufe der Zeit Qualitäten verändern können. Änderungen von Straßennamen, Zusammenlegungen von Firmen oder Änderungen des Familienstandes sollten eventuell nachgepflegt werden. In einem Data Warehouse werden zum Beispiel Strukturen für neue Läufe zur Datenintegration geändert, aber die schon vorhandenen Daten werden nicht mit Inhalten für diese neuen Strukturen ergänzt. Es kann zu einer Verfälschung bei Analysen oder Prognosen führen, wenn diese Daten dann mit einbezogen, die Inhalte aber nicht ergänzt wären.

## DATENQUALITÄT BEWERTEN: DATA QUALITY SCORECARD

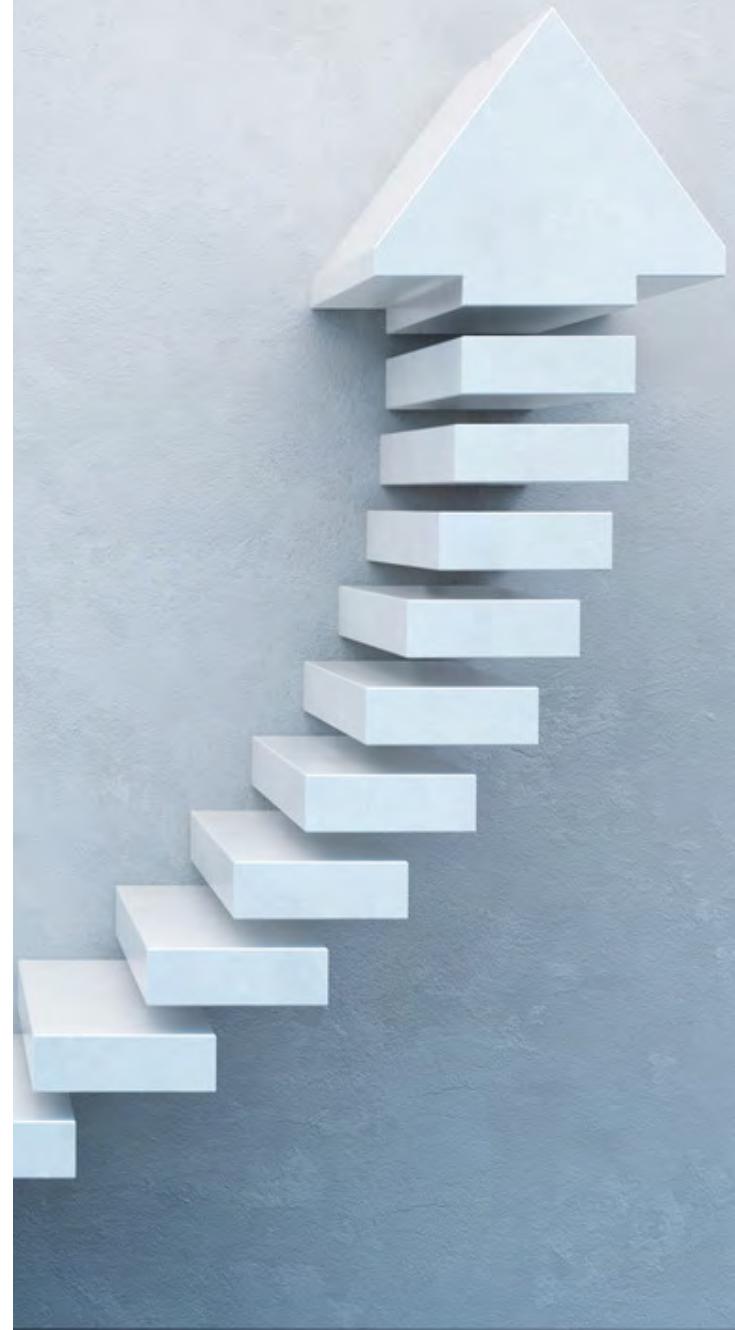
Die 4-stufige Pyramide und das Beispiel des Telekommunikationsunternehmens zeigen, wie wichtig eine adäquate Datenqualität ist. So stellt sich regelmäßig die Frage, wie denn die Datenqualität zu bewerten ist. Und zwar im unternehmerischen Kontext.

Es sind Kennzahlen notwendig, die es ermöglichen eine Aussage zur Datenqualität zu treffen. Mit der DQ Scorecard lassen sich Datenqualitäten, die zu bestimmten Prozessen gehören, mit verschiedensten Regeln prüfen. Der DQ Score

errechnet sich dann aus den gewichteten Prüfergebnissen der einzelnen Regeln. Dabei können die Gewichtungen für den Anteil an der Berechnung konfiguriert werden, damit der Bedeutung eines Datums für ein Unternehmen Rechnung getragen werden kann.

Durch den Vergleich von DQ Scores lässt sich dann bewerten, ob eingeleitete Maßnahmen zu einer Verbesserung der Datenqualität geführt haben. Warum ist die Datenqualitäts-Kennzahl des DQ Score seit gestern schlechter geworden? Diese Frage kann dann untersucht und hoffentlich auch beantwortet werden, da die Scorecard eine Analyse mittels drill-down bis auf die unterste Regelebene zulässt.

Hilfreich bei der Darstellung der Prüfergebnisse und der ermittelten Kennzahlen sind graphische Aufbereitungen wie sie in Scorecards oder Dashboards zu finden sind. Werden die gewünschten Kennzahlen für die Datenqualität erreicht, können die Daten in der Wertschöpfungskette genutzt werden, um letztendlich daraus den viel beschworenen Mehrwert gewinnen zu können.



# CUSTOMER DATA MANAGEMENT

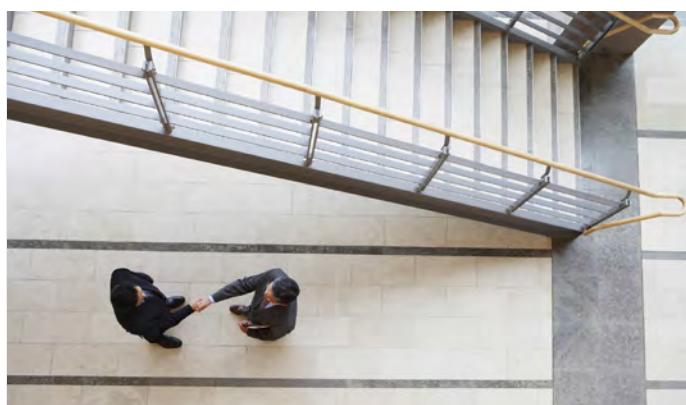
## vom europäischen Marktführer

Uniserv ist Experte für erfolgreiches Kundendatenmanagement. Smart Customer MDM, die MDM-Lösung für Kundenstammdaten, vereint Datenqualitätssicherung und Datenintegration zu einem ganzheitlichen Ansatz. Kundendaten stehen im Mittelpunkt von Initiativen für Master Data Management, Datenqualität, Datenmigration und Data Warehousing, beispielsweise im Umfeld von CRM-Anwendungen, eBusiness, Direct- und Database-Marketing, CDI/MDM-Anwendungen und Business Intelligence.

Mit mehreren Tausend Installationen weltweit bedient Uniserv die Erwartungshaltung einer ganzheitlichen Lösung für alle Geschäfts- und Kundendaten über den gesamten Datenlebenszyklus hinweg. Am Stammsitz in Pforzheim

sowie in den Niederlassungen in Paris, Frankreich, und Amsterdam, Niederlande, beschäftigt das Unternehmen über 130 Mitarbeiter und zählt branchenübergreifend und international zahlreiche renommierte Unternehmen wie beispielsweise Allianz, Deutsche Bank, eBay, EDEKA, E.ON, France Telecom, Lufthansa, Otto, Siemens, Time Warner sowie TUI und VOLKSWAGEN zu seinen Kunden. Der Landesdatenschutzbeauftragte für Baden-Württemberg hat kürzlich bestätigt, dass Uniserv seine Geschäftsprozesse datenschutzgerecht gestaltet.

**Weitere Informationen erhalten Sie unter:**  
[www.uniserv.com](http://www.uniserv.com)



### UNISERV GmbH

Rastatter Str. 13, 75179 Pforzheim, Deutschland

T: +49 7231 936-0

F: +49 7231 936-3002

E: [info@uniserv.com](mailto:info@uniserv.com)

[www.uniserv.com](http://www.uniserv.com)

© Uniserv GmbH, Pforzheim, All rights reserved

TRUE >  
FALSE.

© 2016 Uniserv GmbH oder ein Uniserv-Tochterunternehmen. Alle Rechte vorbehalten. Weitergabe und Vervielfältigung dieser Publikation oder von Teilen daraus sind, zu welchem Zweck und in welcher Form auch immer, ohne die ausdrückliche schriftliche Genehmigung durch die Uniserv GmbH nicht gestattet.

Ground Truth von Uniserv sowie die dazugehörigen Logos sind Marken oder eingetragene Marken der Uniserv GmbH in Deutschland und anderen Ländern. Alle anderen Namen von Produkten und Dienstleistungen sind Marken der jeweiligen Firmen.

Ground Truth ist das Fundament für Umsatzwachstum und neue Geschäftsmodelle vor dem Hintergrund der digitalen Transformation, beispielsweise für die Optimierung von Vertriebs- und Marketingaktivitäten, Kampagnenmanagement, von Blacklist Matching, Compliance oder auch Customer Relationship Management. Im Mittelpunkt steht das Golden Profile eines jeden Kunden, sprich der Golden Record, angereichert mit Interaktions- und Transformationsdaten (Bewegungsdaten). Mit dem Ground Truth bewegen sich Entscheider nicht mehr im luftleeren Raum, sondern haben Bodenhaftung im Decision-Making.

**UNISERV GmbH**

---

Rastatter Str. 13 · 75179 Pforzheim · Deutschland  
T: +49 7231 936 - 0 · E: [info@uniserv.com](mailto:info@uniserv.com) · W: [www.uniserv.com](http://www.uniserv.com)

