



Kernkompetenzzentrum  
Finanz- & Informationsmanagement



Projektgruppe  
Wirtschaftsinformatik

## Der unbekannte Kunde - Potenziale der Integration von Kundendaten

von

Henner Gimpel, Fabian Schmied, Anna-Luisa Stöber

in: HMD Praxis der Wirtschaftsinformatik, HMD 319, 55. Jahrgang, Februar 2018

WI-684

Universität Augsburg, D-86135 Augsburg  
Besucher: Universitätsstr. 12, 86159 Augsburg  
Telefon: +49 821 598-4801 (Fax: -4899)

Universität Bayreuth, D-95440 Bayreuth  
Besucher: Wittelsbacherring 10, 95444 Bayreuth  
Telefon: +49 921 55-4710 (Fax: -844710)



# Der unbekannte Kunde – Potenziale der Integration von Kundendaten

Henner Gimpel<sup>1</sup>, Fabian Schmied<sup>1</sup>, Anna-Luisa Stöber<sup>1</sup>

<sup>1</sup> Kernkompetenzzentrum Finanz- & Informationsmanagement, Universität Augsburg  
Universitätsstraße 12, 86159 Augsburg  
{henner.gimpel,fabian.schmied,anna-luisa.stoeber}@fim-rc.de

**Korrespondierender Autor.** Fabian Schmied

**Zusammenfassung.** Kunden haben heutzutage zahlreiche analoge und digitale Möglichkeiten, mit Unternehmen in Kontakt zu treten. Neben dem persönlichen Besuch einer Unternehmensfiliale oder eines Vertriebspartners können Kunden bspw. telefonisch, per E-Mail, über Internetauftritte, durch Online Social Networks oder mithilfe von Mobile Apps Kontakt mit Unternehmen aufnehmen. Häufig werden bei diesen Interaktionen kundenbezogene Daten generiert. In Abhängigkeit des gewählten Kanals und des Anlasses der Interaktion werden die entstandenen Daten bei vielen Unternehmen in verschiedenen Datenbanken innerhalb oder außerhalb des Unternehmens, bspw. bei Tochtergesellschaften oder Vertriebspartnern, gespeichert. Dies liegt zum einen am unterschiedlichen Zweck der Datenbanken, kann aber auch auf datenschutzrechtliche Gründe oder auf das historische Wachstum der Unternehmens-IT zurückzuführen sein. Ein umfassendes Kundenverständnis wird durch diese Fragmentierung der Kundeninformationen deutlich erschwert. Diese Problematik greift der vorliegende Beitrag auf. Anhand eines Praxisbeispiels aus der Automobilindustrie wird aufgezeigt, wie entschieden werden kann, welche im Unternehmen vorhandenen Kundendaten zusammengeführt und welche Daten von Tochtergesellschaften oder externen Partnern integriert oder vom Unternehmen zusätzlich erhoben werden sollten. Ziel der Datenintegration ist es dabei, in den verschiedenen Unternehmensbereichen ein besseres Verständnis der Kundenbedürfnisse zu ermöglichen und letztlich eine durchgängige Customer Experience zu bieten – unabhängig davon, aus welchem Anlass und über welchen Kanal der Kontakt zwischen Unternehmen und Kunden stattfindet.

**Schlüsselwörter.** Kundendaten, Datenintegration, Vorgehensmodell, Markov-Ketten

**Abstract.** Today, customers can use a variety of traditional and digital channels to get in contact with companies, such as physical stores, phone, e-mail, corporate websites, online social networks, or mobile apps. Customer data are generated in almost all customer-firm interactions. Depending on the channel and the reason of an interaction, customer data are stored in different databases inside or outside the company, e.g., at subsidiaries or distribution partners. This phenomenon may be caused by different reasons. On the one hand, most of the databases have been set up for different reasons. On the other hand, also data privacy laws and regulations may have an influence on data storage issues. Due to the fragmentation of customer data, the understanding of customers' needs and preferences is complicated. These issues are addressed in the following article. Based on a real-world example from the automotive industry, the article shows how to decide which customer data from inside and outside the company should be integrated in a customer database. The integration of customer data aims to improve the understanding of customers' needs and preferences in all parts of the company and to create a consistent customer experience across all channels.

**Key Words.** Customer Data, Data Integration, Procedure Model, Markov Chains

## 1. Ausgangslage und Zielsetzung

Für viele Kunden ist es heutzutage selbstverständlich, dass sie eigenständig entscheiden können, wann sie über welchen Kanal Informationen von einem Unternehmen abrufen bzw. mit diesem in Dialog treten möchten. Je nach Aufenthaltsort, zeitlicher Verfügbarkeit oder Anlass der Kontaktaufnahme bevorzugen Kunden mitunter verschiedene Kanäle, um mit einem Unternehmen zu interagieren (Gimpel und Röglinger 2015). Teilweise werden sogar mehrere Kanäle simultan verwendet (Heinemann 2013). Um diesen Bedürfnissen gerecht zu werden, stellen Unternehmen zunehmend unterschiedliche Kontaktmöglichkeiten bereit (Hansen und Sia 2015). So sind viele Unternehmen mit eigenen Auftritten in sozialen Netzwerken vertreten und stellen Mobile Apps zur Verfügung. Zudem

besteht für die Kunden die Möglichkeit, über klassische Kanäle wie beispielsweise Filialen, Niederlassungen oder Agenturen Kontakt aufzunehmen.

In vielen Unternehmen werden kundenbezogene Informationen je nach Anlass und gewähltem Kanal in unterschiedlichen Datenbanken gespeichert. Bestimmte Informationen über Kunden können bislang von Unternehmen nicht gespeichert werden. Dies trifft beispielsweise für Kontaktaufnahmen über soziale Netzwerke zu, da viele Nutzer keine Klarnamen verwenden und die Zuordnung zu realen Kundendaten somit in der Regel nicht zweifelsfrei möglich ist. In zahlreichen Branchen erfolgt der Vertrieb von Produkten über vertraglich gebundene Händler oder Vertriebspartner, beispielsweise in der Automobil- sowie in der Versicherungsbranche, oder über externe Unternehmen, beispielsweise im klassischen Einzelhandel. Somit besteht für viele Unternehmen die zusätzliche Herausforderung darin, dass sie über keine direkte Endkundenschnittstelle verfügen.

Durch die Verbindung von Produkten mit dem Internet sorgt das Internet der Dinge darüber hinaus in vielen Branchen (z.B. Haushaltsgeräte, Gebäudetechnik, Fahrzeuge) für eine völlig neue Art von Kundendaten. Die generierten Daten können – sofern nach der jeweiligen Gesetzgebung zulässig und falls erforderlich mit Zustimmung der Kunden – von Unternehmen ausgewertet werden, um Rückschlüsse auf das Nutzungsverhalten des Kunden zu gewinnen. Welches Potenzial derartige Daten bieten, zeigen die sogenannten Telematik-Tarife von Kfz-Versicherungen (Maier und Todte 2013). Mithilfe moderner Technik wird das Fahrverhalten des Kunden ausgewertet. Riskantes Fahren birgt eine erhöhte Gefahr von Unfällen und in der Folge eine erhöhte Inanspruchnahme von Versicherungsleistungen. Um dieses Risiko zu kompensieren, können die Versicherungsunternehmen individuell an das Fahrverhalten des Kunden angepasste Risikoprämien berechnen.

Neben den Daten, die durch die Interaktion mit den Kunden oder durch die Nutzung der Produkte und Dienstleistungen entstehen, ist es in einigen Staaten zudem möglich, bei öffentlichen Einrichtungen hinterlegte Daten über eine Person einzusehen und auszuwerten. So sind beispielsweise in Skandinavien die Steuerdaten aller Bürger frei zugänglich im Internet verfügbar. Dadurch kann beispielsweise die Einkommenssituation eines Kunden nachvollzogen werden. In den USA können unter anderem die Zulassungsdaten von Kraftfahrzeugen von den Behörden erworben werden.

Die heterogene Datenlandschaft, individuelle Erwartungen der Kunden sowie die uneinheitliche nationale Datenschutzgesetzgebung stellen insbesondere global agierende Konzerne vor eine große Herausforderung, auch wenn die EU-Datenschutz-Grundverordnung (DSGVO) ab 2018 den Datenschutz innerhalb der Europäischen Union vereinheitlichen wird. Die obigen Beispiele zeigen auf, dass die Menge von Daten und Informationen über Kunden stetig zunehmen. Häufig ist in diesem Zusammenhang vom Umgang mit „Big Data“ die Rede. Die Nutzbarkeit der Daten ist jedoch abhängig von organisationalen, technischen und rechtlichen Faktoren. Die generierten Daten werden in großen Konzernen oftmals in den Datenbanken verschiedener Gesellschaften (z.B. Landesgesellschaften, Tochterunternehmen) und Organisationseinheiten gespeichert. Teilweise haben Unternehmen keinen Zugriff auf die Daten ihrer Kunden, da sich diese im Besitz externer Partner, beispielsweise selbständige Händler und Vertriebspartner sowie Anbieter von unternehmensübergreifenden Loyalitätsprogrammen, befinden. Die vom Unternehmen genutzten Datenbanken greifen zudem meist auf unterschiedliche technische Plattformen zurück. Somit ist die Struktur der Daten in der Regel nicht einheitlich. Darüber hinaus müssen bei der Speicherung und Verarbeitung der Daten die nationalen Datenschutzgesetze und Selbstverpflichtungen beachtet werden. Der verantwortungsbewusste Umgang mit persönlichen Daten gewinnt insbesondere dadurch an Bedeutung, dass Kunden zunehmend sensibel auf die Missachtung von Datenschutzrichtlinien und -gesetzen reagieren (Kleindienst et al. 2017). Teilweise gehen ihre Erwartungen an den Umgang mit persönlichen Daten sogar über die gesetzlichen Standards hinaus. Dabei wird insbesondere die Integration von Daten aus verschiedenen Quellen kritisch beäugt (Kleindienst et al. 2017). Zu beachten ist außerdem, dass die Einwilligung von Kunden zur Nutzung ihrer Daten meist nur für einen bestimmten Zweck gegeben ist.

Laut einer branchenübergreifenden Studie mit 50 Unternehmen ist der richtige Umgang mit der Integration von Daten eine der wesentlichen Herausforderungen der Digitalisierung (Gimpel und Röglinger 2015). Vor diesem Hintergrund erarbeiten derzeit viele Unternehmen eine neue Daten- und IT-Architektur. Laut einer aktuellen Studie denken 77% der befragten Unternehmen darüber nach, erstmals eine Datenmanagement- und Architekturlösung zu implementieren oder bestehende Lösungen zu ersetzen (Deloitte und Informatica 2016). Ziel ist es dabei häufig, die Daten auf einen „Single Point of Truth“ auszurichten. Demnach sollen Daten nicht mehr in unabhängigen Silos, sondern in einem allumfassenden „Data Lake“ gespeichert werden, um eine Vielzahl von Analysen und Auswertungen vornehmen und Zusammenhänge erkennen zu können. Als Data Lake wird ein Sammelbecken (Repository) bezeichnet, in dem große Mengen an Daten in ihrer ursprünglichen Form gespeichert werden (pmOne 2017). Folglich werden die Daten vor der Speicherung weder bereinigt noch in ein einheitliches Format überführt. Allerdings ist dieses Konzept auch mit Risiken verbunden. So fällt es aufgrund der großen Datenmenge aus unterschiedlichen Quellen wesentlich schwerer, innerhalb des Data Lake die geforderte Datenqualität sicherzustellen.

Der Data Lake kann somit leicht zu einem unübersichtlichen Datensumpf werden, der korrekte Auswertungen erschwert (Gartner 2015). Zudem ist es kaum möglich, Zugriffsrechte für sensible Daten auf einzelne Mitarbeiter oder Analysezwecke einzuschränken. Dies kann unter anderem bei der Erfüllung von datenschutzrechtlichen Anforderungen problematisch werden.

Unternehmen müssen vor diesem Hintergrund entscheiden, in welchem Umfang Kundendaten zusammengeführt werden sollten, um ein möglichst gutes Verständnis der Kundenbedürfnisse zu ermöglichen. Gleichzeitig gilt es jedoch, die oben genannten Risiken zu berücksichtigen. Die vollständige Integration der Daten in einen zentralen Data Lake ist daher nicht immer sinnvoll. Um eine ökonomisch fundierte Entscheidung hinsichtlich des Integrationsgrads treffen zu können, sollten sowohl Kosten als auch Nutzen der Datenintegration betrachtet werden (Goodhue et al. 1992). Kosten entstehen nicht nur durch die technische Umsetzung der Datenintegration und organisatorischen Änderungen (z.B. in der Zusammenarbeit mit Händlern), sondern auch durch die potenzielle Abwanderung von Kunden, die mit dem veränderten Umgang mit ihren Daten nicht einverstanden sind. In diesem Zusammenhang besteht zudem die Gefahr von Rechtsstreitigkeiten. Auf der anderen Seite sollte klar aufgezeigt werden, in welcher Weise Daten zum Geschäftserfolg beitragen (Forrester 2015). So können beispielsweise durch eine bessere Customer Experience und individuelle Angebote erzielte Zusatzerlöse, die direkt auf die gewonnenen Erkenntnisse aus den zusammengeführten Daten zurückzuführen sind, als Nutzen der Datenintegration angesehen werden.

Ziel dieses Beitrags ist es, ein Vorgehensmodell zu entwickeln, das Unternehmen dabei unterstützt, die zunehmenden Herausforderungen der Integration von Kundendaten strukturiert zu bewältigen, indem das Potenzial durch die Integration von Kundendaten aus verschiedenen Datenquellen bewertet wird. Dafür wird in Kapitel 2 zunächst ein Vorgehensmodell vorgestellt. In Kapitel 3 wird die beispielhafte Anwendung des Vorgehensmodells bei einem internationalen Automobilkonzern beschrieben. Abschließend werden in Kapitel 4 Handlungsempfehlungen für Unternehmen abgeleitet.

## 2. Vorgehensmodell zur Bewertung des Potenzials durch Integration von Kundendaten

Unter Datenintegration versteht man die Kombination heterogener Daten aus unterschiedlichen Quellen in einer physischen oder virtuellen Datenbank sowie die einheitliche Darstellung dieser Daten (Lenzerini 2002). Die Integration von Kundendaten bietet Chancen, ist allerdings auch mit Kosten verbunden. Zur Ermittlung der Kosten der Datenintegration kann auf bewährte Verfahren zurückgegriffen werden. So können beispielsweise mithilfe des Konzepts der Total Cost of Ownership (TCO) die tatsächlich aus einer IT-Investition resultierenden Kosten ermittelt werden (Krcmar 2015). Darüber hinaus hilft beispielsweise das Constructive Cost Model (COCOMO) dabei, Kosten von Softwareentwicklungsprojekten abzuschätzen (Boehm et al. 1995). Hierbei sollten, wie oben beschrieben, nicht nur Kosten der technischen Umsetzung berücksichtigt werden, sondern auch die der organisatorischen Umsetzung, der möglichen Abwanderung von Kunden sowie möglicher rechtlicher Folgen. Der Fokus dieses Beitrags liegt daher nicht auf dieser Kostenbetrachtung, sondern auf der Bewertung potenzieller Nutzen einer Datenintegration. Zu diesem Zweck wird im Folgenden ein mehrstufiges Vorgehensmodell zur Potenzialbewertung der Integration von Kundendaten vorgestellt (Abbildung 1). Die initiale Version des Vorgehensmodells wurde auf Basis der Erkenntnisse aus mehreren Experteninterviews und -workshops mit Vertretern verschiedener Unternehmen von den Autoren abgeleitet. Im Rahmen der Interviews wurden Experten aus mehreren Unternehmensbereichen (Geschäftsführung, Marketing, Vertrieb, IT) befragt. Das initial von den Autoren ausgearbeitete Vorgehensmodell wurde anschließend mit Firmenvertretern diskutiert und auf Basis der Anmerkungen weiterentwickelt. Aus den Gesprächen mit den Experten resultiert insbesondere die Erkenntnis, dass es sinnvoll ist, einen prototypischen integrierten Datenbestand zu erstellen und auf dieser Grundlage den Nutzen der Datenintegration zu ermitteln. Dadurch kann vermieden werden, dass eine aufwendige und kostspielige Integration von Daten erfolgt, durch die kein erkennbarer Nutzen entsteht. Abschließend wurden die einzelnen Schritte des Vorgehensmodells im Kontext eines internationalen Automobilkonzerns exemplarisch umgesetzt (siehe Kapitel 3).



Abbildung 1: Vorgehensmodell zur Potenzialbewertung der Datenintegration

Ein integrierter Bestand von Kundendaten stellt für sich genommen keinen Nutzen für das Unternehmen dar. Der Nutzen wird generiert, sofern durch die Zusammenführung der Daten für die geschäftliche Entwicklung relevante Fragestellungen besser beantwortet und daraus entsprechende Handlungen abgeleitet werden können. Daher gilt es in der ersten Phase zunächst, derartige relevante Fragestellungen zu identifizieren. Die Fragestellungen sollten sich an Strategie und Geschäftsmodell des Unternehmens orientieren, können beispielsweise in Workshops unter Beteiligung von Mitarbeitern aus unterschiedlichen Fachabteilungen und Landesgesellschaften erarbeitet werden und mit Anwendungsfällen angereichert werden, die bei Wettbewerbern zu beobachten sind, von Technologieanbietern und Beratungen proklamiert oder in der Wissenschaft diskutiert werden. Mögliche Anwendungsfälle sind beispielsweise eine bessere Kundensegmentierung, die bessere Personalisierung der Kundenkommunikation, eine genauere Tarifierung oder die Berechnung von Storno- und Abwanderungswahrscheinlichkeiten. Die identifizierten Fragestellungen und Anwendungsfälle dienen später dazu, den Nutzen der Datenintegration zu bewerten.

Anschließend werden in der zweiten Phase potenziell relevante Kundendaten für diese Anwendungsfälle identifiziert. Um die Qualität der Datenidentifizierung zu verbessern, ist eine zweistufige Herangehensweise sinnvoll. Zunächst sollen Mitarbeiter die Kundendaten identifizieren, die zur Beantwortung der zuvor erarbeiteten Fragestellungen benötigt werden. Um sich von möglichen unternehmensspezifischen Restriktionen zu lösen, ist es bei diesem Vorgehen besonders wichtig, sich nicht von den Daten leiten zu lassen, die im Unternehmen bereits digital vorliegen. Vielmehr ist es das Ziel, alle Daten zu identifizieren, die zur Beantwortung der Fragestellungen beitragen können – unabhängig davon, ob diese Daten derzeit im Unternehmen analog oder digital zur Verfügung stehen, bei anderen Stellen außerhalb des Unternehmens verfügbar sind oder neu erhoben werden müssen. Hierfür können verschiedene Kreativitätstechniken verwendet werden. Erst im zweiten ergänzenden Schritt erfolgt dann explizit der Blick auf bereits vorhandene Daten oder solche, auf die bei Tochtergesellschaften, Partnerunternehmen oder dem Kunden selbst leicht zugegriffen werden kann. Hier gilt es, weitere für die Beantwortung der Fragestellung relevante Kundendaten zu ermitteln, die mithilfe der Kreativitätstechniken im ersten Schritt eventuell noch nicht identifiziert wurden. Da einzelnen Mitarbeitern möglicherweise der Gesamtüberblick über vorhandene Kundendaten fehlt, ist es auch in diesem Schritt sinnvoll, Mitarbeiter aus verschiedenen Fachabteilungen und Landesgesellschaften zu beteiligen.

Ziel der dritten Phase ist es, einen prototypischen integrierten Datenbestand zu erstellen. Dieser sollte alle zuvor identifizierten relevanten Kundendaten enthalten, um in der Folge Auswertungen über den Nutzen des integrierten Datenbestands und jedes einzelnen Datums im Zusammenspiel mit den anderen Daten zu ermöglichen. Da der Nutzen der einzelnen Daten noch nicht bekannt ist, ist zu diesem Zeitpunkt noch keine vollumfängliche und kostenintensive Integration der Kundendaten sinnvoll. Stattdessen sollte auf die Daten zurückgegriffen werden, die bereits im Unternehmen vorliegen, mit geringem Aufwand einmalig gespiegelt und ohne die Verletzung rechtlicher Restriktionen zusammengeführt werden können, beispielsweise aufgrund von Pseudonymisierung. Eine prototypische Ergänzung dieses Datenbestands kann beispielsweise durch den Einkauf externer Daten, die Zusammenarbeit mit Wissenschaftlern, die Durchführung von Kundenbefragungen oder die Kooperation mit einzelnen Vertriebspartnern erreicht werden.

Auf Basis des prototypischen Datenbestands erfolgt in der vierten Phase eine analytische Auswertung der Daten hinsichtlich der zuvor identifizierten Fragestellungen. Je nach Fragestellung können dabei unterschiedliche Auswertungsmethoden aus Statistik, Data Mining und maschinellem Lernen erforderlich sein. Besonders geeignet sind dabei unter anderem Scoring-Modelle, Cluster-Analysen, Regressionen, Entscheidungsbäume und Markov-Ketten.

Abschließend sollen die Daten in der fünften Phase hinsichtlich ihrer Relevanz für die Beantwortung der spezifischen Fragestellung bewertet werden. Auf Basis der Bewertung kann entschieden werden, welche Daten später in den integrierten Datenbestand aufgenommen werden sollten. Die Bewertung greift in diesem Schritt auf die Verbesserung der statistischen Modelle durch einzelne Datenbestände und -typen zurück und bewertet ökonomisch, welchen Wertbeitrag die jeweilige Verbesserung einer Analyse in einem Anwendungsfall für das Unternehmen liefert, beispielsweise durch Erhöhung des Customer Lifetime Values (CLV) aufgrund einer verbesserten Customer Experience. In einem abschließenden Schritt wird diese Nutzenbewertung den erwarteten Kosten der Integration gegenübergestellt und eine gesamthafte Bewertung vorgenommen.

### **3. Beispielhafte Anwendung des Vorgehensmodells in der Automobilbranche**

Im folgenden Abschnitt wird das oben beschriebene Vorgehensmodell beispielhaft auf einen global agierenden Automobilhersteller angewendet. Innerhalb des betrachteten Unternehmens werden Automobile mehrerer Marken

entwickelt, produziert und verkauft. Das Unternehmen ist in verschiedene zentrale Konzerneinheiten und Landesgesellschaften untergliedert. Im Fokus steht im folgenden Beispiel der deutsche Markt. Verkauft werden die Fahrzeuge über unternehmenseigene Niederlassungen und selbständige Händler. Zudem gibt es in der Automobilbranche einen großen Markt für gebrauchte Fahrzeuge. Folglich gibt es zahlreiche Fahrzeugbesitzer, die dem Automobilhersteller nicht als direkte Kunden bekannt sind und über die somit keine Kundendaten vorliegen. Auf der anderen Seite sorgt insbesondere die zunehmende Vernetzung von Produkten mit dem Internet auch in der Automobilbranche für eine stark steigende Verfügbarkeit von Nutzungsdaten. Das Unternehmen wickelt zusätzlich Finanzierungs- und Leasinggeschäfte ab. Auch aus diesem Geschäftsbereich liegen umfangreiche Kundendaten vor, die beispielsweise zur Kreditwürdigkeitsprüfung herangezogen werden. Neben dem Autoverkauf und der zugehörigen Finanzierungsdienstleistung werden beispielsweise bei Probefahrten mit Neufahrzeugen, bei der Bestellung von Werbematerialien, durch Gewinnspiele und Zufriedenheitsbefragungen oder im Rahmen von Werkstattbesuchen und Ersatzteilbeschaffungen Kundendaten erhoben. Teilweise gehören die Daten rechtlich unabhängigen Händlern, die dem Konzern nur geringen Einblick gewähren (Abbildung 2).

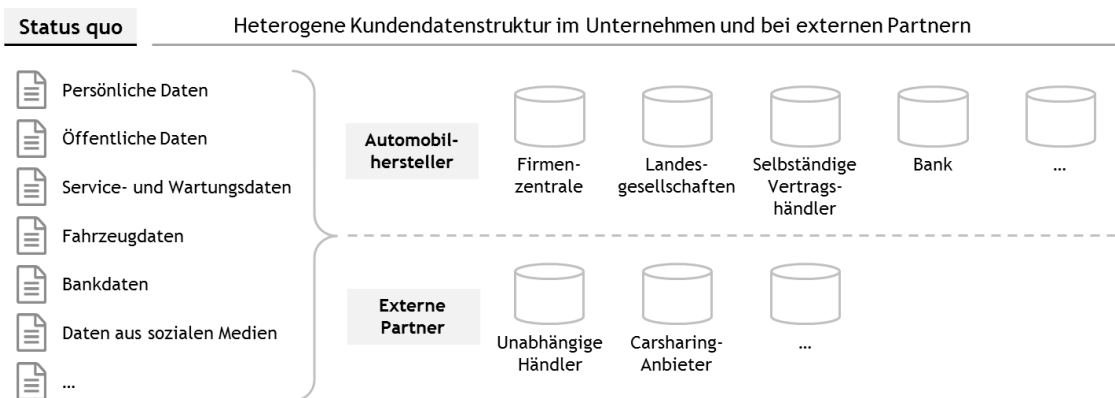


Abbildung 2: Heterogene Kundendatenstruktur

In der ersten Phase des Vorgehensmodells wurden in Workshops Fragestellungen identifiziert, die auf Basis der aktuell in den unterschiedlichen Fachabteilungen verfügbaren Kundendaten von keinem der beteiligten Experten beantwortet werden können. Teilnehmer der Workshops waren Mitarbeiter aus Fachabteilungen des Automobilherstellers (Marketing, Vertrieb, IT). Einige der Personen konnten zudem Erfahrungen aus mehreren Landesgesellschaften des Unternehmens einbringen. Zusätzlich wurden Experteninterviews mit selbständigen Händlern geführt. Im Rahmen der Workshops und Interviews zeigte sich, dass innerhalb des Unternehmens kaum Informationen über das Kundenverhalten über einen längeren Zeitraum vorliegen. So ist beispielsweise nicht bekannt, welche Fahrzeuge von einem Kunden vor seinem aktuellen Fahrzeug genutzt wurden. Informationen zur Kundenmigration zwischen Modellen, Marken und Automobilherstellern könnten jedoch hilfreich sein, um daraus Prognosen über das zukünftige Verhalten des Kunden abzuleiten und Marketingmaßnahmen zielgerichtet auf den Kunden abzustimmen. Dies ist einer von zahlreichen Anwendungsfällen.

In der zweiten Phase wurden Gespräche mit den Mitarbeitern zur Identifizierung relevanter Kundendaten für diesen Anwendungsfall geführt. Dabei zeigte sich, dass es zahlreiche Daten gibt, die zum besseren Verständnis des Kundenverhaltens beitragen könnten. Hierzu zählen Informationen über das Verhalten während des Anschaffungsprozesses, Nutzungs- und Nutzungsgewohnheiten, die Historie von genutzten Fahrzeugen und persönliche Informationen zum Kunden. Diese Daten liegen zum Teil bereits in unternehmenseigenen Datenbanken vor. Insbesondere Daten zur Anschaffung und zur Wartung von Fahrzeugen sind jedoch in der Regel Eigentum der selbständigen Händler. Zudem gibt es Informationen, beispielsweise über Nutzungsgewohnheiten von öffentlichen Verkehrsmitteln oder Carsharing-Angeboten, die derzeit weder beim Automobilhersteller selbst noch bei Händlern vorliegen und somit zunächst erhoben werden müssten.

In der dritten Phase wurde aufgrund der heterogenen Datenstruktur und vorhandener rechtlicher Restriktionen ein prototypisch integrierter Datenbestand auf Basis einer Kundenbefragung erstellt. Insbesondere um die potenziell aufschlussreichen Migrationsbewegungen der Kunden zwischen Automobilherstellern besser nachvollziehen zu können und einen Wettbewerbsvergleich zu ermöglichen, wurden hierbei Kunden verschiedener Automobilhersteller befragt. Ziel der Befragung war es, zu ermitteln, welche Fahrzeuge von der befragten Person zuletzt über einen Zeitraum von mindestens sechs Monaten regelmäßig für private Zwecke genutzt wurden. Erfasst wurden unter anderem Informationen zum Fahrzeug (Marke, Fahrzeugklasse), zum Halter des Fahrzeugs (Privatfahrzeug, Firmenfahrzeug) und zum Anschaffungsvorgang (Leasing, Finanzierung, Barzahlung). Darüber hinaus wurden

demografische Daten sowie Informationen zum persönlichen Nutzungsverhalten erhoben. Dazu zählen beispielsweise die übliche Nutzungsdauer von eigenen Fahrzeugen, die Durchführung von Wartungs- und Reparaturarbeiten oder die regelmäßige Nutzung von Mietwagen- oder Carsharing-Angeboten. Ziel ist hierbei nicht, die Daten kontinuierlich per Kundenbefragung zu gewinnen, sondern im Rahmen des Projekts ihren Wert zu ermitteln und auf dieser Basis zu entscheiden, welche Daten dauerhaft erhoben und integriert werden sollten. Insgesamt wurden für dieses Projekt im Rahmen einer Onlinebefragung Daten von 611 deutschsprachigen Teilnehmern erfasst. Die erhobenen Fahrzeugdaten weisen deutliche Parallelen zu den offiziellen Zulassungszahlen des Kraftfahrtbundesamtes auf. So befinden sich neun der zehn meistgenannten Automobilmarken unter den zehn Marken mit den deutschlandweit meisten Zulassungen, so dass diesbezüglich eine gewisse Repräsentativität angenommen werden kann.

Die vierte Phase fokussiert die analytische Auswertung des prototypisch integrierten Datenbestands. Dies wird im Folgenden exemplarisch anhand der Daten zu den zuletzt genutzten Fahrzeugen gezeigt. Ziel ist es dabei zu überprüfen, ob Informationen zur bisherigen Nutzungshistorie der Kunden Vorhersagen auf ihr zukünftiges Verhalten ermöglichen. Lässt sich diese Vermutung bestätigen, wäre es für den Automobilhersteller empfehlenswert, Daten zur Historie der Kunden zu erheben und zu speichern. Statistisch kann dies mithilfe des Konzepts der Markov-Ketten überprüft werden. Eine Markov-Kette ist ein stochastischer Prozess. Die grundlegende Annahme lautet, dass sich eine Beobachtung zu einem beliebigen Zeitpunkt einer Kategorie (Zustand) zuordnen lässt. Eine Markov-Kette erster Ordnung wird beschrieben durch einen Ausgangszustand sowie einer Menge von Übergangswahrscheinlichkeiten (Anderson und Goodman 1957). Die Übergangswahrscheinlichkeiten geben an, welche Zustände die Beobachtung zum nächsten Zeitpunkt mit welcher Wahrscheinlichkeit annimmt. Markov-Ketten sind in vielen Anwendungsgebieten verbreitet. Im Marketing werden Markov-Ketten beispielsweise verwendet, um Bewegungen von Kunden zwischen verschiedenen Segmenten oder Wiederkaufwahrscheinlichkeiten in zukünftigen Zeitpunkten darzustellen und zu analysieren (Ha et al. 2002, Pfeifer und Carraway 2000).

Im Zuge der Auswertung der Befragungsdaten wurden zunächst zehn Zustände definiert. Diese geben an, welche Fahrzeugkategorie vom Kunden primär genutzt wird. Die Kategorien orientieren sich dabei an der vom Kraftfahrtbundesamt angewandten Segmentierung von Personenkraftwagen sowie an der in der Automobilbranche üblichen Unterscheidung zwischen Premium- und Volumenherstellern. Einige Segmente wurden dabei zu größeren Kategorien zusammengefasst (Tabelle 1). So umfasst beispielsweise die neu definierte Kategorie „Kleinwagen“ Fahrzeuge der vom Kraftfahrtbundesamt definierten Segmente „Minis“ und „Kleinwagen“. Die Kategorie „Oberklasse“ fasst die Segmente „Obere Mittelklasse“, „Oberklasse“, „SUVs“, „Geländewagen“ und „Sportwagen“ zusammen, da die Fahrzeuge dieser Segmente – relativ zu den anderen Segmenten – hochpreisig sind und von vielen Käufern auch aufgrund des vermittelten Gefühls von Luxus ausgewählt werden.

Kategorie	Beispiele
Premium Kleinwagen	Smart Fortwo, Mini Cooper, Audi A1
Premium Kompaktklasse	Mercedes A-Klasse, BMW 1er, Audi A3
Premium Mittelklasse	Mercedes C-Klasse, BMW 3er, Audi A4
Premium Oberklasse	Mercedes E-Klasse, Mercedes S-Klasse, BMW 5er, BMW 7er, Audi A6, Audi A8, Porsche Cayenne
Premium Vans	Mercedes-Benz V-Klasse
Volumen Kleinwagen	VW Polo, Opel Corsa, Renault Clio, Seat Ibiza
Volumen Kompaktklasse	VW Golf, Opel Astra, Mazda 3, Toyota Corolla
Volumen Mittelklasse	VW Passat, Opel Insignia, Renault Laguna
Volumen Oberklasse	VW Phaeton, VW Touareg, Opel GT
Volumen Vans	VW Touran, Opel Zafira, Seat Alhambra

Tabelle 1: Kategorisierung der Fahrzeuge

Die befragten Personen haben insgesamt Daten zu 1.815 Fahrzeugen angegeben. Dies entspricht durchschnittlich drei Fahrzeugen pro Teilnehmer, also beispielsweise das aktuell genutzte Fahrzeug sowie zwei vorherige Fahrzeuge. Auf Grundlage der Angaben zu den genutzten Fahrzeugen wurden die Übergangswahrscheinlichkeiten für die zuvor definierten Kategorien berechnet (Tabelle 2).



Von	Zu	Premium Kleinwagen	Premium Kompaktklasse	Premium Mittelklasse	Premium Oberklasse	Premium Vans	Volumen Kleinwagen	Volumen Kompaktklasse	Volumen Mittelklasse	Volumen Oberklasse	Volumen Vans	Total
Premium Kleinwagen		36%	7%	7%	7%		7%	11%	11%	11%	4%	28
Premium Kompaktklasse		3%	35%	14%	11%		14%	14%	8%	3%		37
Premium Mittelklasse		1%	8%	41%	12%		13%	13%	7%	4%	1%	95
Premium Oberklasse			4%	17%	30%		9%	9%	13%	14%	3%	99
Premium Vans		17%		17%	17%			17%	17%	17%		6
Volumen Kleinwagen		4%	5%	7%	4%		25%	27%	17%	8%	2%	275
Volumen Kompaktklasse		1%	4%	8%	4%	1%	11%	38%	18%	14%	2%	318
Volumen Mittelklasse		1%	2%	8%	5%	1%	12%	21%	34%	12%	7%	164
Volumen Oberklasse		2%	2%	4%	7%		7%	16%	19%	42%	2%	117
Volumen Vans		2%	2%	6%	8%		17%	25%	11%	17%	13%	53
<b>Total</b>		<b>31</b>	<b>59</b>	<b>129</b>	<b>93</b>	<b>4</b>	<b>167</b>	<b>292</b>	<b>213</b>	<b>168</b>	<b>36</b>	<b>1192</b>

Legende	1% - 10%	11% - 20%	21% - 30%	> 30%
---------	----------	-----------	-----------	-------

Tabelle 2: Übergangswahrscheinlichkeiten

Übergänge können nur ausgewertet werden, sofern von einem Teilnehmer Informationen zu mindestens zwei Fahrzeugen eingetragen wurden. Hat ein Teilnehmer beispielsweise Informationen zu vier Fahrzeugen angegeben, können drei Übergänge ausgewertet werden. Insgesamt konnten auf diese Weise 1.192 Übergänge betrachtet werden. Auf den ersten Blick zeigt sich, dass eine relativ hohe Wahrscheinlichkeit besteht, dass Automobilkunden bei einer Neuanschaffung der bisherigen Fahrzeugkategorie treu bleiben. So entscheiden sich beispielsweise 41% der Kunden, die zuletzt ein Auto der Premium Mittelklasse genutzt haben, auch beim nächsten Fahrzeug für ein Modell der Kategorie Premium Mittelklasse. Mithilfe des Chi-Quadrat-Tests, für den die in Tabelle 2 dargestellten Übergangswahrscheinlichkeiten als Inputparameter verwendet werden, kann gezeigt werden, dass ein signifikanter Zusammenhang zwischen dem aktuell genutzten Fahrzeug und dem zuletzt zurückliegenden Fahrzeug besteht ( $p$ -Wert < 0,001).

Betrachtet man Markov-Ketten höherer Ordnung, hat nicht nur der aktuelle Zustand einer Beobachtung ( $t = 0$ ) Einfluss auf den zukünftigen Zustand ( $t = 1$ ), sondern auch mindestens ein zurückliegender Zustand ( $t = -1$ ). Betrachtet man also den Einfluss des vorletzten Fahrzeugs auf das aktuell genutzte Fahrzeug (Markov-Kette 2. Ordnung), so lässt sich feststellen, dass es einige Kategorien gibt, bei denen ein derartiger statistischer Zusammenhang besteht. Dies ist beispielsweise für die Kategorien Premium Mittelklasse und Volumen Kleinwagen zu einem Signifikanzniveau von 5% gegeben. So greifen beispielweise 32% der Kunden, die unmittelbar aufeinanderfolgend zwei Fahrzeuge der Kategorie „Volumen Kleinwagen“ genutzt haben, auch aktuell auf ein Fahrzeug dieser Kategorie zurück. Bei einigen Kategorien (z.B. Premium Kompaktklasse, Premium Oberklasse) ist im Gegensatz dazu die Information über das letzte Fahrzeug ausreichend und eine längere Historie nicht erforderlich. Zudem ist der Chi-Quadrat-Test für Markov-Ketten 2. Ordnung auf Basis der vorhandenen Befragungsdaten aufgrund der vergleichsweise geringen Stichprobengröße in einzelnen Kategorien nicht anwendbar. Aus Gründen der Übersichtlichkeit wurde die Analyse in diesem Beitrag auf der Ebene von Fahrzeugkategorien dargestellt. Im Projekt wurden zudem konkrete einzelne Fahrzeugmodelle des beteiligten Unternehmens analysiert.

Mithilfe der statistischen Auswertung konnte der Nutzen der Fahrzeughistorie der Kunden aufgezeigt werden. Daten zum aktuell genutzten Fahrzeug sollten dem Automobilhersteller bekannt sein, da sie Rückschlüsse auf zukünftige Fahrzeuge zulassen. Ebenso konnte gezeigt werden, dass auch Fahrzeuge, die in der Vergangenheit vom Kunden genutzt wurden, aussagekräftig sein können.

In Verbindung mit demografischen Daten ermöglichen die Informationen zur Nutzungshistorie zudem Auswertungen hinsichtlich der Produktloyalität. Als loyale Kunden werden im Sinne der Produktloyalität Kunden angesehen, die beim Übergang zwischen zwei Fahrzeugen weder die Fahrzeugmarke noch das Fahrzeugsegment wechseln. Folgt man dieser Definition haben sich die Teilnehmer der Befragung bei 21% der Fahrzeugwechsel loyal verhalten. Hinsichtlich des Alters der Teilnehmer sind keine signifikanten Unterschiede erkennbar. Betrachtet man den Wohnort der Teilnehmer, so lässt sich feststellen, dass insbesondere Kunden aus Großstädten (18%) weniger loyal sind als andere Kunden. Frauen (24%) verhalten sich loyaler als Männer (20%). Einen wesentlichen Einfluss auf die Produktloyalität hat zudem das generelle Mobilitätsverhalten der Kunden. So sind Kunden, die regelmäßig Carsharing-Angebote oder Mietwagen nutzen, deutlich weniger loyal (13%) als die anderen Kunden (22%). Die Auswertung zur Produktloyalität bestätigt den Nutzen der Informationen zur Nutzungshistorie eines Kunden. Um einzelne Kundensegmente gezielt adressieren zu können, ist es darüber hinaus aus Unternehmenssicht sinnvoll, demografische Daten sowie Daten zum Mobilitätsverhalten der Kunden zu erheben.



Bei der praktischen Anwendung der vorgestellten Methode in einem Unternehmen sollte als fünfter und abschließender Schritt zudem der monetäre Nutzen der Daten berechnet werden, der dann den Kosten der Datenintegration gegenübergestellt werden kann. Um diesen zu ermitteln, kann beispielsweise die Entwicklung des Customer Lifetime Values (CLV) einer Testkundengruppe betrachtet werden, bei der die Marketingmaßnahmen auf Grundlage der integrierten Kundendatenbasis optimiert wurden (Voigt und Hinz 2016). Die ökonomische Bewertung der besseren Prognosen durch bessere Datenverfügbarkeit wird hier nicht dargestellt.

#### **4. Handlungsempfehlungen**

Eine strukturierte Herangehensweise an die Integration von Kundendaten ist sinnvoll und notwendig, insbesondere wenn Unternehmen über verschiedene Unternehmenseinheiten (z.B. Tochtergesellschaften) und zahlreiche Kontaktpunkte zum Kunden verfügen. In den meisten Fällen ist es nicht sinnvoll, alle Datenbanken zu integrieren, auf die das Unternehmen direkt oder über einen Dritten zugreifen kann. Vielmehr sollten nur Daten zusammengeführt werden, die helfen, relevante Fragestellungen innerhalb des Unternehmens zu beantworten. Folglich ist es sinnvoll, die Integration von Kundendaten im Rahmen eines unternehmensweiten Projekts durchzuführen, um relevante Fragestellungen, bereits vorhandene Kundendaten und Informationsbedarfe innerhalb des Unternehmens bestmöglich erkennen und beurteilen zu können. Hierbei hilft das im Beitrag vorgestellte Vorgehensmodell. Im Sinne einer wertorientierten Unternehmensführung sollten neben den potenziellen Nutzen der Datenintegration jedoch stets auch Kosten und potenzielle Risiken im Blick behalten werden, um somit die Auswirkung des Projekts auf den Unternehmenserfolg bestimmen zu können (Buhl et al. 2011). Hierzu zählen neben den Kosten für die technische und organisatorische Umsetzung der Maßnahmen auch die Reaktionen von Kunden, die möglicherweise mit der Zusammenführung ihrer Daten nicht einverstanden sind. Erfolgt die Integration der Kundendaten in einem ökonomisch sinnvollen Maß, eröffnen sich dem Unternehmen zahlreiche Möglichkeiten. Auf Basis eines integrierten Datenbestands können Kundenbedürfnisse besser erkannt und das zukünftige Kaufverhalten vorhergesagt werden. Durch die Nutzung der gewonnenen Erkenntnisse können die Interaktion mit bestehenden und potentiellen Kunden und deren Customer Experience deutlich verbessert werden. Die Ansprache des Kunden mit geeigneten und individuell auf ihn ausgerichteten Produktinformationen und Zusatzangeboten kann folglich einen wesentlichen Beitrag zur Bindung des Kunden an das Unternehmen leisten.

#### **5. Literatur**

Anderson TW, Goodman LA (1957) Statistical Inference about Markov Chains. The Annals of Mathematical Statistics. S. 89-110.

Boehm B, Clark B, Horowitz E, Westland C, Madachy R, Selby R (1995) Cost models for future software life cycle processes: COCOMO 2.0. Annals of Software Engineering 1:57-94.

Buhl HU, Röglinger M, Stöckl S, Braunwarth KS (2011) Value orientation in process management. Business & Information Systems Engineering 3:163-172.

Deloitte, Informatica (2016) Big Data and Beyond: Future-Proofing Data Management and Architecture Investments. <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/about-deloitte/us-alliance-big-data-and-beyond.pdf>. Gesehen 09. Mai 2017.

Forrester (2015) Put The Business Back In Your Data Management Business Case. <https://www.forrester.com/report/Put+The+Business+Back+In+Your+Data+Management+Business+Case/-/E-RES126762>. Gesehen 09. Mai 2017.

Gimpel H, Röglinger M (2015) Digital Transformation: Changes and Chances – Insights based on an Empirical Study. Project Group Business and Information Systems Engineering (BISE) of the Fraunhofer Institute for Applied Information Technology FIT, Augsburg/Bayreuth.

Gartner (2015) Look Before Diving Headfirst Into a Data Lake. <http://www.gartner.com/smarterwithgartner/look-before-diving-headfirst-into-a-data-lake-2/>. Gesehen 09. Mai 2017.

Goodhue DL, Wybo MD, Kirsch LJ (1992) The Impact of Data Integration on the Costs and Benefits of Information Systems. MIS Quarterly 16:293-311.

Ha SH, Bae SM, Park SC (2002) Customer's time-variant purchase behavior and corresponding marketing strategies: an online retailer's case. Computers & Industrial Engineering 43:801-820.

- Hansen R, Sia SK (2015) Hummel's Digital Transformation Toward Omnichannel Retailing: Key Lessons Learned. *MIS Quarterly Executive* 14:51-66.
- Heinemann G (2013) No-Line-Systeme als höchste Evolutionsstufe des Multi-Channel-Handels. In: Keuper F, Hamidian K, Verwaayen E, Kalinowski T (Hrsg.): *Digitalisierung und Innovation*, Springer Fachmedien, Wiesbaden. S. 169-184.
- Kleindienst D, Nüske N, Rau D, Schmied F (2017) Beyond Mere Compliance – Delighting Customers by Implementing Data Privacy Measures? In: Leimeister JM, Brenner W (Hrsg): *Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017)*, St. Gallen, S. 807-821.
- Krcmar H (2015) *Informationsmanagement*. 6., überarb. Aufl., Springer, Berlin Heidelberg.
- Lenzerini M (2002) Data integration: A theoretical perspective. In: *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, ACM, S. 233-246.
- Maier SC, Todte H (2013) Telematik – eine Revolution in der Kfz-Versicherung? *Zeitschrift für Versicherungswesen* 23:776-782.
- Pfeifer PE, Carraway RL (2000) Modeling customer relationships as Markov chains. *Journal of Interactive Marketing* 14:43-55.
- pmOne (2017) Big Data Analytics und Data Lake. <https://www.pmone.com/loesungen/big-data/big-data-analytics-und-data-lake/>. Gesehen 31. Juli 2017.
- Voigt S, Hinz O (2016) Making Digital Freemium Business Models a Success: Predicting Customers' Lifetime Value via Initial Purchase Information. *Business & Information Systems Engineering* 58:107-118.