



---

## **IN FÜNF SCHRITTEN ZU BESSEREN UNTERNEHMENS DATEN**

Datenqualität in der Praxis

---



---

## Inhaltsverzeichnis

---

<b>Einführung</b> .....	<b>4</b>
<b>Die fünf Schritte zu wertvolleren Unternehmensdaten</b> .....	<b>5</b>
Data Profiling .....	6
Datenqualität .....	7
Datenintegration .....	8
Datenanreicherung .....	9
Datenkontrolle .....	10
<b>Zusammenfassung</b> .....	<b>11</b>

### Einführung

Jede gute Geschäftsentscheidung basiert auf einer guten Qualität der Daten. Gleich ob betriebliche, analytische oder strategische Entscheidungen, jede Abteilung in einem Unternehmen, von der Vorstandsetage bis zur Vertriebsabteilung, braucht zuverlässige, präzise Informationen, um solche Entscheidungen korrekt zu treffen.

Die genaue Kenntnis darüber, welche Kunden welche Produkte haben, oder welche Kunden potenzielle Chancen für den Kauf eines höherwertigen Produkts bieten (Up-Selling), basiert auf korrekten Daten. Wahr ist aber auch, dass eine schlechte Datenqualität unabwendbar zu ungesicherten Geschäftsentscheidungen, falschen Strategien und mangelhaftem Kundenservice führt.

Wenn Unternehmen datenbezogene Initiativen wie Enterprise Resource Planning (ERP) oder Customer Relationship Management (CRM) ergreifen oder ein Programm zu umfassender Data Governance implementieren wollen, sollten sie daran denken, dass jedes Unternehmen grundsätzlich Datenqualitätsprobleme hat. Die erfolgreiche Lösung dieser Datenqualitätsprobleme – und die Etablierung von Kontrollen zur dauerhaften Sicherung der erreichten Datenqualität – ist für den Erfolg jeder dieser Initiativen entscheidende Voraussetzung.

Es gibt viele Gründe für die explosionsartige Zunahme inkonsistenter oder nutzloser Daten, angefangen von einfachen Fehlern, verursacht durch Menschen, bis hin zu ungenügenden oder gar nicht vorhandenen unternehmensweiten Datenstandards, die alle Systeme, Geschäftsbereiche und gegebenenfalls Konzerngesellschaften erfassen. Um die Datenprobleme zu überwinden, benötigen Unternehmen eine technologische Grundlage, mit der sie die Datenqualitätsprobleme aufdecken und beheben können.

Datenqualitätsprobleme können jedoch nicht durch Technologie allein gelöst werden. Ebenso essentiell ist die Einbindung der richtigen Personen und Prozesse. Dieses White Paper erläutert anhand einer erprobten Datenqualitätsmethodik, bestehend aus fünf aufeinander folgenden Abschnitten, wie Menschen, Prozesse und Technologie in ein effektives Datenmanagement für die Analyse, Verbesserung und Kontrolle der Geschäftsdaten eingebunden werden können.

---

*Datenqualität und Data Governance erfordern die Einbindung von Menschen, Prozessen und Technologie.*

---

## Die fünf Schritte zu wertvolleren Unternehmensdaten

Technologie ist nur ein Teil der Lösung, die zu konsolidierten, hochwertigen Unternehmensinformationen führt. Eine robuste Technologie, die jede Datenqualitätsinitiative unterstützt, ist von zentraler Bedeutung. Noch wichtiger für den Erfolg der Initiative ist jedoch die richtige Methodik für den Einsatz dieser Technologie.

DataFlux hat eine Methodik für Datenqualität und Data Governance in Unternehmen entwickelt, die aus fünf Komponenten oder Teilverfahren besteht. Sie resultiert aus der Erfahrung mit Tausenden von Implementierungen, die in mehr als einem Jahrzehnt gesammelt wurden. Die fünf Verfahren erfassen Data Profiling, Datenqualität, Datenintegration, Datenanreicherung und Datenkontrolle. Zusammen bilden sie einen bewährten, praktischen Ansatz für Data Governance – hin zu umfassendem, kontrolliertem Datenmanagement, das Unternehmen ermöglicht, ihre Daten zu analysieren, zu verbessern und zu kontrollieren.

*Data Profiling, Datenqualität, Datenintegration, Datenanreicherung und Datenkontrolle sind die fünf Prozesskomponenten eines Datenmanagementprogramms.*



Abbildung 1: die fünf Komponenten eines Datenmanagement-Programms

Diese Komponenten bieten eine Systematik, mit der sich der gesamte Prozess für eine verbesserte Datenqualität erfolgreich managen lässt. Nutzt man sie konsequent, resultiert diese Methodik in einer einheitlichen Sicht jeder Art von Daten im gesamten Unternehmen, darunter die zentralen Daten zu Kunden, Produkten und Lieferanten.

**Data Profiling**

Der erste Schritt in diesem Ansatz, das Data Profiling, untersucht die Struktur, die Beziehungen und den Inhalt vorhandener Datenquellen, um ein genaues Bild vom Zustand der Daten zu ermöglichen. Die Einschätzung des gegenwärtigen Datenzustands als Ergebnis der Datenanalyse bildet die Grundlage für die Planung der sinnvollsten Wege zur Korrektur und Harmonisierung der Informationsbestände.

Das Data Profiling erfasst drei verschiedene Aspekte der Daten und analysiert sie jeweils auf eigene Weise. Zusammen ergeben sie ein klares Bild von der Art und dem Ausmaß der Datenqualitätsprobleme.

Data Profiling besteht im Wesentlichen aus dem Prozess, die Kernfragen zu den Daten für jeden der folgenden Aspekte zu beantworten:

- Datenstruktur: Entsprechen die Datenmuster den vorgesehenen Mustern?  
Entsprechen die Daten den korrespondierenden Metadaten?
- Dateneinträge: Sind die Datenwerte vollständig, genau und eindeutig?  
Sind die Daten entsprechend den geltenden Regeln standardisiert?
- Datenbeziehungen: Weisen die Daten in allen Spalten und Tabellen die erforderliche Zuordnung zu der spezifizierten Schlüsselbeziehung auf?  
Bestehen abgeleitete Relationen über Spalten, Tabellen und Datenbanken hinweg?  
Gibt es redundante Daten?

Abbildung 2 zeigt die Art der Berichte, die während der Datenanalyse erzeugt werden. Diese Analyse ist ein Beispiel für die Ermittlung von Daten. Sie zeigt für einige Felder in der Datenbank Nullwerte an. Es fällt auf, dass die Gesamtzahl von 5.000 Einträgen konstant bleibt, während einige Felder eine hohe Zahl von Nullwerten aufweisen. Aus der Information, dass hier in einigen Schlüsselfeldern ein hoher Anteil an Werten fehlt – Adresse, Stadt, Staat und besonders Geschlecht –, können Geschäftsprozessanalytiker sofort das Ausmaß und die Art der Datenqualitätsprobleme erkennen.

*Datenanalyse bietet ein genaues Bild des gegenwärtigen Zustands der Unternehmensdaten.*

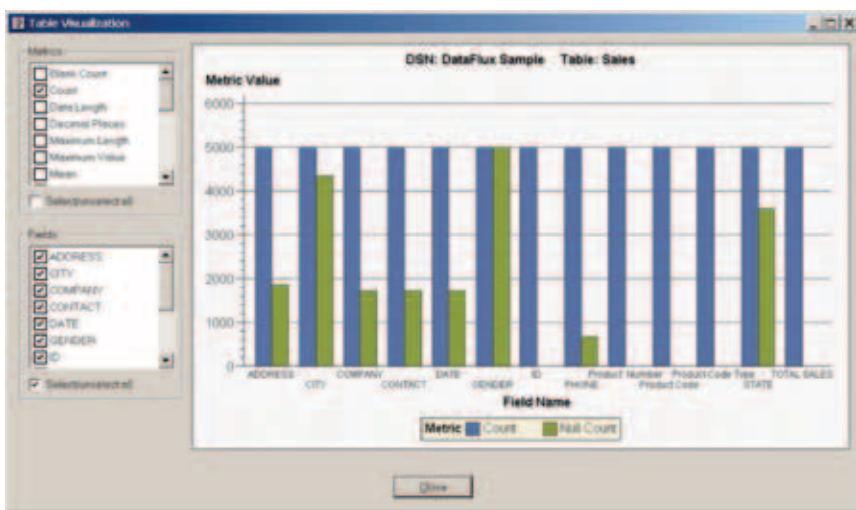


Abbildung 2: Die Grafik zeigt eine Kundenstammtabelle mit einem hohen Anteil an Nullwerten

Data Profiling liefert quantifizierbare, detaillierte Informationen über die Stärken und Schwächen der Unternehmensdaten. Aus diesen Kenntnissen lassen sich die weiteren Maßnahmen für eine Verbesserung der Datenqualität ableiten. Die durch die Datenanalyse gewonnenen Informationen bilden zudem den Ausgangspunkt für das nächste Verfahren, die Datenqualität.

## Datenqualität

Das Datenqualitätsverfahren nutzt die Resultate der Datenanalyse, um mit dem Aufbau hochwertigerer Daten zu beginnen. In diesem Schritt werden Fehler korrigiert, Informationen standardisiert und Daten unternehmensweit validiert. Es können verschiedene Taktiken verfolgt werden. Die am häufigsten verwendeten Methoden sind:

- Die Datenstandardisierung, um vielfältige Permutationen von Daten herauszufiltern und zu korrigieren. Ein Beispiel: „ACME Manufacturing Corporation“ kann in derselben Datenquelle als „Acme Mftg Corp“, „ACME“ und „ACME Manufacturing“ repräsentiert sein. Intelligentes Fuzzy Matching dient dazu, diese Varianten zu standardisieren und eine einheitliche Nomenklatur zu schaffen.
- Die Musterstandardisierung, um für alle Tabellen und Spalten geltende Datenmuster zu erstellen. Manche Datenbestandteile wie Telefonnummern oder Sozialversicherungsnummern sind in Musterabweichungen leicht zu finden. Andere, wie Produkt- oder Teiledaten, können in Unternehmen oder Branchen unterschiedlichen Standards folgen. Die Musterstandardisierung kann Informationen aus nicht standardisierten Formaten in ein gewünschtes Standardformat umwandeln.
- Die Adressverifizierung, um sicherzustellen, dass Adressen gültig und nutzbar sind. Beispielsweise kann die DataFlux Software erkennen, wenn die Stadt und das Land nicht mit der Postleitzahl übereinstimmen. Die Software sucht nach postalischen Daten, findet die richtige Postleitzahl und ändert den Eintrag in eine korrekte Adresse.

*SAS ermöglicht eine faktenbasierte Diskussion über die Verteilung knapper Marketingressourcen und hilft dem Marketing, den Investitionscharakter des eigenen Budgets nachzuweisen.*

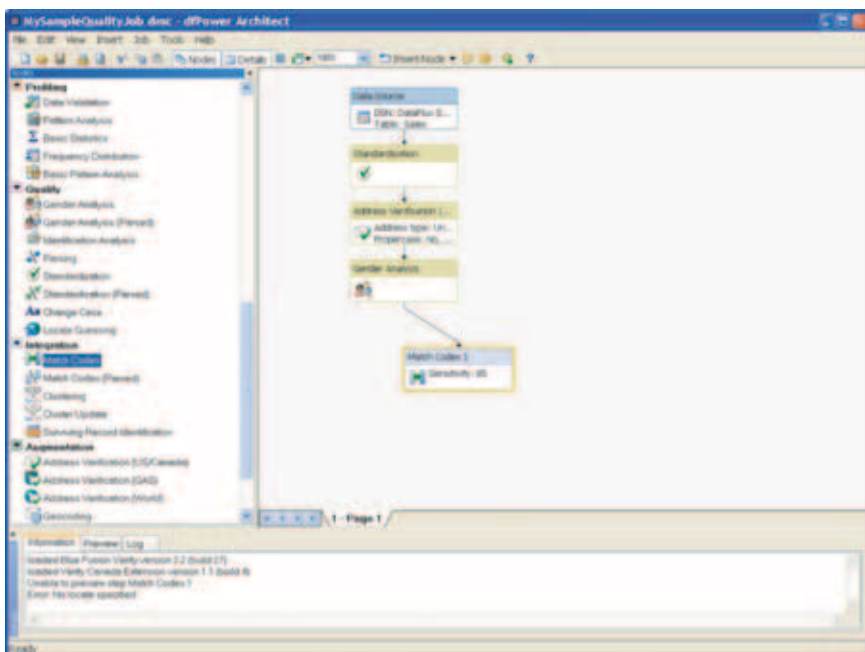


Abbildung 3: Beispiel für das Datenqualitätsverfahren

Abbildung 3: Die Technologie führt durch die Schritte zum Aufbau von Geschäftsregeln für hochwertige Daten. In diesem Beispiel beginnt der Ablauf mit der Quelltafel, setzt sich über die Standardisierung fort und führt die Verifizierung und weitere Routinen aus. Die vorausgegangene Data Profiling-Routine bringt drei Probleme ans Licht, für deren Lösung der Anwender die jeweiligen Schritte einleitet.

Der Ansatz ist variabel und kann jeweils an den Datensatztyp angepasst werden. Die Entscheidung für die Wahl des Ansatzes obliegt dem Fachbereich, der für die jeweiligen Daten verantwortlich ist. Zu berücksichtigen ist, dass das Datenqualitätsverfahren die Qualität vorhandener Daten verbessert, nicht jedoch an die Wurzel des Datenproblems geht. Für eine nachhaltige Verbesserung der Datenqualität muss ein Unternehmen auch die Gründe ermitteln, die zu den Fehlern in den Daten führten, und geeignete Maßnahmen ergreifen, um die Daten kontinuierlich auf einem höheren Qualitätsniveau zu halten. Dies können Veränderungen in den Geschäftsprozessen, aber auch das Honorieren des korrekten Umgangs mit den Daten sein.

Ist der Datenqualitätsprozess abgeschlossen, verfügt eine Organisation über die Mittel, die Daten jeder Quelle hinsichtlich ihrer Vollständigkeit und Präzision auf das angestrebte Niveau zu bringen. Der nächste Schritt besteht in der Integration vielfältiger Quellen.

### Datenintegration

Für jeden Datentyp existieren häufig verschiedene Datensätze in mehreren Datenbanken, die dasselbe Objekt repräsentieren. Die Datenintegration dient dazu, die Daten aus verschiedenen Quellen zu optimieren. Datenintegration ermöglicht die Konsolidierung dieser voneinander abweichenden Einträge zu einem einzigen Stammdatensatz.

Beispielsweise pflegt eine Firma zwei Datenquellen für ihre Produkte: die Originalproduktdatenbank, extrahiert aus ihrer Enterprise Resource Planning(ERP)-Lösung in den USA, und eine Produktdatenbank in Europa. Die Firma verkauft das gleiche Produkt in beiden Regionen, die Produktbezeichnungen in den Datenbanken sind jedoch unterschiedlich. Zusätzlich wurden die Produktnamen, Fabrikatsbezeichnungen und Produktbeschreibungen in jeder der Datenquellen von unterschiedlichem Personal eingegeben.

Die erste Aufgabe der Datenintegration besteht darin, ein Produkt zu finden, das in beiden Quellen repräsentiert ist. Diesen Vorgang nennt man Linking. Die zweite Aufgabe besteht darin, die Daten in eine einzige Sicht des Produkts zu konsolidieren.

In Kundendaten befindet sich häufig ein in allen Datensätzen verwendetes Feld, beispielsweise die Steueridentifikationsnummer, das beim Auffinden entsprechender Einträge hilft. In diesem Fall lassen sich Mehrfacheinträge für denselben Kunden schnell und leicht finden.

Bei Produktdaten im obigen Beispiel verhilft die Datenintegrationstechnologie dazu, die Produktinformationen zu finden, sie zu verbinden und die Daten zu konsolidieren. Intelligentes Fuzzy Matching identifizierte eine Produktbeschreibung, die in beiden Quellen übereinstimmte. Die Datenbank in den USA enthielt die Beschreibung, das Fabrikat und den Produktnamen zusammen in einem Feld mit verschiedenen Mustern. Die Datenbank in Europa enthielt nur die Produktbeschreibungen, in unterschied-

---

*Intelligente Fuzzy Matching Technologie kann Datenintegrationsprojekte deutlich vereinfachen.*

---

---

*Die Anreicherung von Unternehmensdaten durch Ergänzung um Daten aus vertrauenswürdigen externen Quellen kann den Wert der Daten steigern.*

---



lichen Mustern und Abkürzungen. Zur Konsolidierung der Einträge nutzte die Integrationstechnologie die folgenden Abläufe:

- Identifikation und Isolierung (Parsen) der produktspezifischen Attribute und der Fabrikatsbezeichnung aus der Beschreibung in der Datenbank in den USA
- Abgleich der Unterschiede in den Fabrikatsbezeichnungen
- Abgleich der Unterschiede in den Produktattributen (Kurzformen, Abkürzungen etc.)
- phonetische Anpassung der verglichenen Daten
- Erstellung eines Berichts zum Matching der Produkte

Nach der Datenintegration besitzt das Unternehmen jetzt die komplette Sicht auf alle Daten zu einem Objekt. Die gleichen Prinzipien und Techniken können auf jeden Datentyp angewendet werden, um integrierte Stammdatensätze zu erstellen.

Im nächsten Schritt soll der Wert der Daten durch Ergänzung um zusätzliche Informationen erhöht werden.

### **Datenanreicherung**

Bei der Datenanreicherung werden die Datensätze ergänzt, um den Informationswert und die Nützlichkeit der Daten zu erhöhen. Dies kann auf verschiedene Art und Weise erfolgen. So können beispielsweise die eigenen Daten durch Daten von Partnern oder anderen Firmen ergänzt und so genauere Informationen über Kunden aufgebaut werden – etwa zu ihrem Einkaufspotenzial oder ihrer Loyalität. Durch sogenannte Geokodierung können inzwischen geografische Informationen integriert werden. Auf diese Weise lässt sich etwa feststellen, dass alle Häuser einer bestimmten Postleitzahl in Kombination mit einer bestimmten geografischen Region nach 1980 gebaut wurden und einen bestimmten Immobilienwert übersteigen. Daraus lassen sich sehr spezifische Produktangebote für potenzielle Kunden in diesem Areal ableiten.

Eine weitere Möglichkeit der Datenanreicherung ist, die vorhandenen Kundendaten mit den Daten über das Kundenverhalten zu kombinieren. Letztere lassen sich durch die Analyse bestimmter Attribute ermitteln. Dies ermöglicht effektivere Kundensegmentierung und das Herausfiltern spezieller Vertriebschancen.

Abbildung 4 zeigt ein Beispiel für die Ergänzung einer einfachen Hausanschrift um wichtige Daten aus einer externen Quelle – mittels Matching im Rahmen der Technologie zur Verbesserung der Datenqualität.

---

*Datenanreicherung kann zur Ergänzung um internationale Warenwirtschaftsstandards wie eCI@ss oder UNSPSC dienen.*

*Durch die Integration von „Watch Lists“ und „Do-not-Call“-Verzeichnissen kann Datenanreicherung nützlich für Compliance-Initiativen sein.*

---

Submitted Data	Standardized Data	
DataFlux	DataFlux Corporation	
940 Cary Pkwy	940 Cary Pkwy, Ste. 201	
Cary, NC	Cary, NC 27513-2792	
New Data Points Added		
	Carrier Route ID	34
	Delivery Point ID	1
	County Number	183
	County Name	Wake
	Congressional District	4
	Latitude	35.8306
	Longitude	-78.7841
	Census Block Group	371830535122

Abbildung 4: Anreicherung der Daten einer Hausanschrift

Datenanreicherung lässt sich auf mehr als nur Adressdaten anwenden. Viele Arten von Daten, darunter Produkt-, Inventar- und Finanzdaten, können durch die Verbindung zu externen Quellen ergänzt und dadurch aufgewertet werden. So können internationale Warenwirtschaftsregister wie UNSPSC oder eCI@ss dazu dienen, die Produkte oder Lieferantendaten zu standardisieren und damit Einkauf und Verkauf zu rationalisieren. Die Integration dieser universellen Klassifizierung unterstützt die Genauigkeit der Lieferantendaten, was wiederum die Effizienz der Bestellungen erhöht und den Weg für reibungslose internationale Geschäfte ebnet.

Datenanreicherung unterstützt auch das von häufigen Neuerungen betroffene Compliance Management. Es ermöglicht den Vergleich von Kundenlisten oder Transaktionen mit externen Listen in Bezug auf Übereinstimmung mit den gesetzlichen und industriellen Vorschriften. Diese Listen können beispielsweise dazu dienen, der Gefahr von Geldwäsche oder Korruption vorzubeugen.

**Datenkontrolle**

Zu einem umfassenden Data Governance-Programm gehört eine kontinuierliche Datenkontrolle. Nur so ist gewährleistet, dass der Datenqualitätserfolg aus den aufwändigen Korrektur- und Aufwertungsmaßnahmen nicht durch die allmähliche Rückkehr von Datenfehlern zunichtegemacht wird. Mit aktiver Datenkontrolle haben Unternehmen Kenntnis vom Zustand ihrer Daten und können die Ursachen für Datenqualitätsprobleme isolieren und korrigieren. Datenkontrolle kann auf verschiedene Weise erfolgen. Die einfachste Datenkontrolle sind automatisierte Data Profiling Reports, die in regelmäßigen Abständen Analysen der Daten in Bezug auf Ausnahmen und nichtstandardisierte Daten liefern. Sie identifizieren beispielsweise Ereignisse wie einen zu großen Prozentsatz an außergewöhnlichen Daten und können gegebenenfalls eine E-Mail oder eine Systemmeldung an zuvor bestimmte Adressaten oder Verantwortliche eines Bereichs auslösen.

---

*Die Geschäftsprozessregeln, die im Verlauf des gesamten Verfahrens entwickelt wurden, können anschließend in der Datenkontrolle zur Sicherung vor dem Absinken der Datenqualität verwendet werden.*

---

Eine dynamischere und gleichzeitig effektivere Methode der kontinuierlichen Datenkontrolle ist die Durchsetzung von Geschäftsprozessregeln in Form von Web Services. Mit dieser Methode können die gleichen Regeln, wie sie für die Bereinigung und Anreicherung der Daten benutzt wurden, dazu dienen, Daten in Echtzeit während der Eingabe oder auf ihrem Weg durch das Unternehmen zu überprüfen. Die Vorteile dieses Ansatzes beinhalten die folgenden Möglichkeiten:

- Es können Data Governance-Regeln aufgestellt und integriert werden. Damit ist gewährleistet, dass die unternehmensspezifischen Regeln für das Datenmanagement befolgt werden.
- Aufgabenkritische Prozesse werden erkennbar, indem Verstöße gegen oder Ausnahmen von Geschäftsregeln in einem Repository protokolliert werden. So lassen sich Trends verfolgen und die Prozesse entsprechend anpassen.
- Es können definierte Ereignisse ausgelöst werden, um die Daten automatisch zu korrigieren und entweder festgelegte Personen durch E-Mails darüber zu informieren oder die Daten in einem Repository zu protokollieren oder einfach die korrigierten Daten in eine Datenbanktabelle einzutragen.

---

## Zusammenfassung

---

Die Qualität jeder Geschäftsanalyse ist nur so gut wie die Daten, die ihre Grundlage bilden. Ohne Daten, die unternehmensweit konsistent, akkurat und zuverlässig sind, können im Unternehmen leicht irreführende, falsche und potenziell geschäftsschädigende Schlussfolgerungen gezogen werden. Für eine erfolgreiche Verbesserung der Datenqualität müssen Unternehmen zunächst eine Methodik und Lösung finden, die alle fünf wesentlichen Bestandteile umfasst: Data Profiling, Datenqualität, Datenintegration, Datenanreicherung und Datenkontrolle.

Es gibt heute ausgefeilte Technologien für Datenqualität und Datenintegration, die es Unternehmen ermöglichen, Datenqualitätsmaßnahmen auf einem hohen Niveau zu realisieren. Das Ziel ist eine einheitliche Sicht der Datenbestände zu Kunden, Produkten, Lieferanten und weiteren Datenarten – in allen möglichen Applikationen im gesamten Unternehmen.

Durch die Kombination einer ausgereiften Methodik mit geeigneter Technologie kann die Datenqualität erheblich verbessert werden. Dies wird zu soliden Entscheidungen führen und mehr Wettbewerbsvorteile schaffen. Denn beides beruht auf der Verfügbarkeit hochwertiger Daten.

---

*Unternehmensdaten müssen analysiert, optimiert und kontrolliert werden.*

---



The BI Company

accantec consulting AG  
Mönckebergstraße 31  
20095 Hamburg

info@accantec.de  
www.accantec.de

Fon 040 - 67 59 59 - 0

Hamburg | Köln | Frankfurt



THE  
POWER  
TO KNOW.

SAS Institute GmbH, In der Neckarhelle 162, 69118 Heidelberg, Germany  
Phone +49 6221 415-123, Fax +49 6221 415-145 [www.sas.de](http://www.sas.de)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies. Copyright © 2012, SAS Institute Inc. All rights reserved.