Reemers Publishing Services GmbH
O:/Wiley/Reihe_Dummies/9783527714476_Gerken/3d/c01.3d from 08.08.2018 14:03:05
3B2 9.1.580; Page size: 176.00mm x 240.00mm

IN DIESEM KAPITEL

Was sind Daten?

Auswertungen von Dateien und Datenbanken ohne Data Warehouse

Reporting-Tools und analytische Informationssysteme

Warum ein Data Warehouse sinnvoll ist

Kapitel 1
Ein Beispiel zur Einführung

Vereinfacht formuliert ist ein Data Warehouse eine spezielle Datenbank zur Unterstützung betrieblicher Analyse- und Entscheidungsaufgaben, die eine einheitliche Sicht auf ursprünglich unterschiedliche (heterogene) und verteilte Datenbestände ermöglicht. Dieses Kapitel soll dazu dienen, den Sinn und Zweck eines Data Warehouse klarzumachen. Anhand eines Beispiels wird erläutert, wie in einem Unternehmen Datenbestände ausgewertet und analysiert werden und woher diese Daten kommen.

Daten und ihre Verarbeitung

Daten und Datenbanken

Daten, also gespeicherte Fakten über uns selbst, andere und unser Umfeld, begleiten uns unser ganzes Leben. Sie dienen – nach entsprechender Verarbeitung – als Grundlage von Entscheidungen und beeinflussen unser Handeln. Ein Beispiel hierfür ist die Kaufentscheidung für ein Smartphone anhand technischer Kenndaten und Nutzerbewertungen. Insbesondere für Firmen sind Daten immens wichtig, nicht nur bei den täglich anfallenden Geschäftsprozessen, sondern auch bei unternehmerischen Entscheidungen. Daten stellen eine besondere Ressource dar, die gepflegt werden muss. Ohne sie könnten zum Beispiel keine Kundenaufträge bearbeitet, keine Rechnungen geschrieben und keine Gehälter gezahlt werden.

Aber auch für unternehmerische Analysen und Entscheidungen sind Daten notwendig.

- ✔ Wie hoch waren im letzten Monat der Absatz und Umsatz der Produkte in den einzelnen Vertriebsregionen oder bei den einzelnen Kunden?
- ✔ Welche Unterschiede im Kaufverhalten gibt es zwischen den weiblichen und männlichen Kunden?

Reemers Publishing Services GmbH
O:/Wiley/Reihe_Dummies/9783527714476_Gerken/3d/c01.3d from 08.08.2018 14:03:05
3B2 9.1.580; Page size: 176.00mm x 240.00mm

- ✔ Welche Absatzchancen ergeben sich in den einzelnen Vertriebsregionen im kommenden Jahr?
- ✔ Erhöht sich der Deckungsbeitrag, wenn bei Produkt X der Preis um 5 % gesenkt wird?
- ✔ Welche Merkmale zeichnen die umsatzstärksten 10 % aller Kunden aus?

Daten werden in Dateien oder in Datenbanken gespeichert. Letztlich unterscheidet sich das darin, dass man mit Datenbanken versucht, etwas Ordnung in die gespeicherten Daten zu bringen. Meist wird von den Anwendungsprogrammen nicht direkt auf die gespeicherten Daten zugegriffen, sondern über eine Zwischenschicht, das Datenbankmanagementsystem – kurz DBMS. Ein DBMS zusammen mit einer Datenbank wird Datenbanksystem genannt. Der am häufigsten verwendete Typ dabei sind die relationalen Datenbanksysteme, bei denen alle Daten und ihre Beziehungen untereinander in Tabellen, den Relationen, gespeichert sind.

Sie wollen Ihr Wissen über Datenbanksysteme, insbesondere relationale, auffrischen? Kein Problem. Sie können alles Notwendige zu diesem Thema im Buch »Datenbanksysteme für Dummies« genauer nachlesen (Gerken 2018).

Die Verarbeitung von Daten

Die Speicherung von Daten ist kein Selbstzweck. Daten sind eine notwendige Ressource zur Durchführung einerseits operativer und andererseits dispositiver, analytischer betrieblicher Aufgaben. Doch was unterscheidet diese beiden Aufgabengruppen? Zum Beispiel eine Auftragsbearbeitung (operative Aufgabe) von einer Absatzanalyse (dispositive Aufgabe)? Die Bearbeitung eines Auftrags ist ein wohldefinierter Geschäftsprozess, der immer gleich abläuft und Daten auch verändert. Dagegen kann eine Absatzanalyse unterschiedlich ablaufen und erfordert je nach Zielrichtung der Analyse unterschiedliche Daten, eventuell auch in unterschiedlichem Detaillierungsgrad, und greift nur lesend auf die Dateien und Datenbanken zu. Siehe dazu auch die Tabelle 1.1.

Operative Aufgaben findet man vor allem im täglichen Routinegeschäft eines Unternehmens; man nennt das aus Informatik-Sicht Online Transaction Processing, kurz OLTP. Dispositive Aufgaben werden als Online Analytical Processing (OLAP) bezeichnet. Doch dazu mehr in Kapitel 10.

Aufgabentyp	operativ	dispositiv, analytisch
Aufgabenstruktur	fest	flexibel
Zugriff auf Daten	lesend und verändernd	nur lesend
Umfang der benötigten Daten	bekannt und fest	variabel

Tabelle 1.1: Bedarf an Daten für operative und dispositive, analytische Aufgaben

Diese Unterscheidung legt nahe, die Daten für operative und dispositive, analytische betriebliche Aufgaben zu trennen. Fällt Ihnen in diesem Zusammenhang der Begriff »Data Warehouse« ein? Der folgende Abschnitt beschreibt, wie eine klassische analytische Aufgabenstellung aussieht.

Reemers Publishing Services GmbH
O:/Wiley/Reihe_Dummies/9783527714476_Gerken/3d/c01.3d from 08.08.2018 14:03:07
3B2 9.1.580; Page size: 176.00mm x 240.00mm

Analyse von Absatzmengen und Planzahlen als Beispiel

Nehmen wir einmal an, im Datenbestand eines Unternehmens gibt es eine Tabelle mit Verkaufszahlen, aus der ersichtlich ist, welche Kunden wo und wann etwas gekauft haben.

KundenNr	Name	Ort	ArtikelNr	Datum	Menge
1001	Clara	Hamburg	1010	01.10.2017	200
101	Clara	Berlin	1010	20.06.2018	300
100	Klaus	Hamburg	1010	23.02.2018	200
100	Klaus	Hamburg	2005	17.04.2018	100
101	Clara	Berlin	1010	19.04.2018	500
102	Julia	Hannover	2005	10.10.2018	100
...					

Tabelle 1.2: Tabelle mit Verkaufszahlen

Ferner soll es eine Tabelle mit den Planzahlen für die geplanten jährlichen Absätze geben.

ArtikelNr	Jahr	Planabsatz
1010	2018	2500
1011	2018	2000
2005	2018	1000
...		

Tabelle 1.3: Tabelle mit Planabsatzzahlen

Für das Jahr 2018 sollen jetzt die Verkaufszahlen ausgewertet werden.

- ✔ Welche Artikel wurden von welchen Kunden gekauft?
- ✔ Wie sieht die Gegenüberstellung der Ist- und Planabsätze bei den einzelnen Artikeln aus?

Bei sehr kleinen Unternehmen mag es noch möglich sein, die Auswertungen mit einem Tabellenkalkulationsprogramm zu machen. Wenn ein Unternehmen wächst, stößt man damit aber schnell an die Grenzen. Für die Planzahlen mag das noch gehen, denn Controller lieben EXCEL; die Absatzzahlen werden aber in der Datenbank des Onlineshops oder der Verkaufssoftware gespeichert sein.

Für die Auswertung kann man dann eines der auf dem Softwaremarkt verfügbaren Analysetools verwenden. Diese Werkzeuge können auf unterschiedliche Datenbanken und Dateien zugreifen und diese analysieren; natürlich auch auf ein **Data Warehouse**, wie Sie später noch sehen werden. Damit lassen sich unter anderem sogenannte Dashboards zur übersichtlichen, oft mit Grafiken versehenen Darstellung von Informationen erzeugen, die die gewünschten Auswertungen beinhalten.

Reemers Publishing Services GmbH
O:/Wiley/Reihe_Dummies/9783527714476_Gerken/3d/c01.3d from 08.08.2018 14:03:08
3B2 9.1.580; Page size: 176.00mm x 240.00mm

Derartige Tools werden als Business-Intelligence-Tools (BI-Tool) bezeichnet. Die genaue Einführung des Begriffs erfolgt in Kapitel 3. Hier belassen wir es erst einmal bei dieser knappen Erläuterung.

In einem **Dashboard** werden hochaggregierte Unternehmenskennzahlen, die oft aus unterschiedlichen Quellen kommen, in übersichtlicher Form dargestellt. Damit kann man zum Beispiel sehen, wie hoch der gestrige Umsatz war oder wie oft der Onlineshop besucht worden ist.

Häufig zu finden ist die Darstellung von Kennzahlen in Ampel- oder Tachometerform

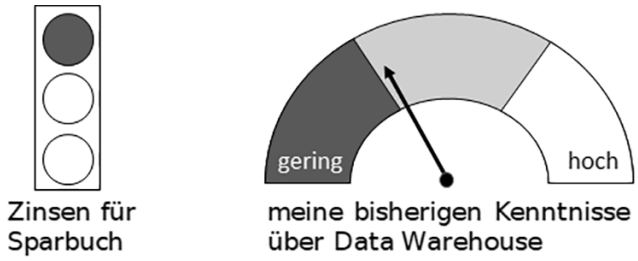

Abbildung 1.1: Ampel- und Tachometerdarstellung

Damit kann man sehr schnell erkennen, ob noch alles »im grünen Bereich« ist. Wir werden später erneut darauf zurückkommen, wenn wir die grundsätzlichen Analysemöglichkeiten von Data-Warehouse-Daten betrachten (Teil III).

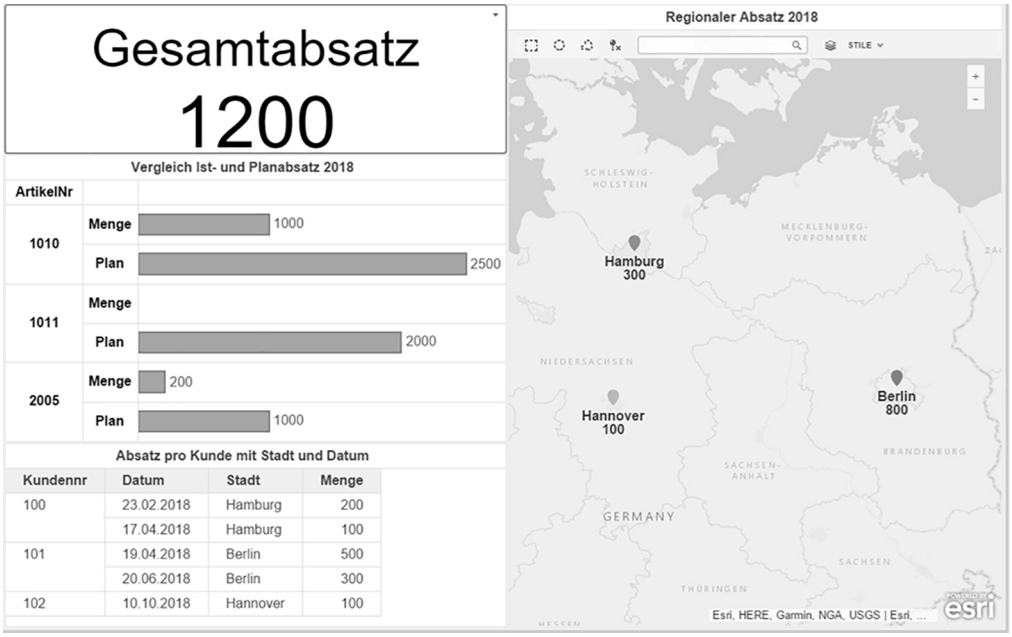

Abbildung 1.2: Beispiel für Dashboard
© 2017 MicroStrategy Incorporated. All Rights Reserved

Reemers Publishing Services GmbH
O:/Wiley/Reihe_Dummies/9783527714476_Gerken/3d/c01.3d from 08.08.2018 14:03:09
3B2 9.1.580; Page size: 176.00mm x 240.00mm

Das Dashboard für die Absatzanalyse soll aus vier Bereichen bestehen:

1. Der Gesamtabsatz 2018 für alle Artikel zusammen
2. Der Absatz pro Verkaufsort, dargestellt auf einer Landkarte
3. Ein Vergleich zwischen dem Ist- und dem Planabsatz durch ein Balkendiagramm
4. Eine tabellarische Detaildarstellung der Absatzmengen je Artikel, Kunde und Datum.

Das Beispiel in Abbildung 1.2, welches auf den Tabellen 1.2 und 1.3 basiert, wurde mit dem BI-Tool MicroStrategy Desktop® erzeugt; siehe dazu das Kapitel 20, »10 Schritte auf dem Weg zu Ihrem ersten Dashboard« im Top-10-Teil. Damit können Sie dieses Beispiel selbst nachvollziehen, was ich Ihnen empfehle.

Besonderheiten analytischer Aufgabenstellungen

Das obige Beispiel war insofern sehr einfach, weil alle Daten als EXCEL-Tabelle vorliegen, was bei sehr kleinen Unternehmen noch zutreffen mag. In der Praxis gibt es aber viele Datenquellen, die eventuell übergreifend ausgewertet werden müssen. Für Produktempfehlungen benötigt man zum Beispiel Absatzzahlen aus der Auftragsbearbeitung und zusätzlich Tracking-Daten aus dem Webserver des Onlineshops. Vielleicht hat das Unternehmen ja auch mehrere ausländische Tochtergesellschaften mit unterschiedlicher Software. Das macht es nur noch komplizierter. Generell sind in einem Unternehmen Daten auf viele verschiedene Arten gespeichert; insbesondere in folgenden Datenhaltungssystemen:

- ✔ Tabellenkalkulation
- ✔ Relationale Datenbank
- ✔ NoSQL-Datenbank
- ✔ Sonstige Dateien

Damit ergibt sich für ein BI-Tool, das auf Verkaufsdaten (eventuell mehrerer Filialen), Planungsdaten, Stammdaten von Kunden und Lieferanten usw. zugreifen muss, beispielsweise das in Abbildung 1.3 zu sehende Bild.

Die Integration der Daten erfolgt im BI-Tool, und dabei kann es viele Probleme geben:

- ✔ Die Schnittstelle zur Datenbank muss zur Verfügung stehen, insbesondere müssen die Berechtigungen zum Lesen der Daten vorhanden sein. Eventuell muss eine eigene Benutzersicht (VIEW) bereitgestellt werden.
- ✔ Es gibt unterschiedliche Bezeichnungen für dieselben Daten.
- ✔ Es gibt gleiche Bezeichnungen für unterschiedliche Daten.

Reemers Publishing Services GmbH
O:/Wiley/Reihe_Dummies/9783527714476_Gerken/3d/c01.3d from 08.08.2018 14:03:10
3B2 9.1.580; Page size: 176.00mm x 240.00mm

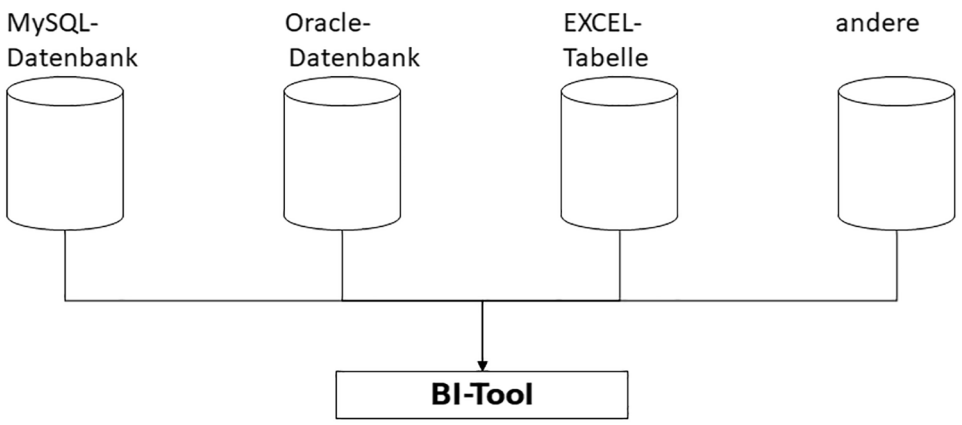

Abbildung 1.3: Datenanalyse mit mehreren Datenquellen

- ✔ Daten sind unterschiedlich und teilweise nicht vollständig verschlüsselt, zum Beispiel das Geschlecht einmal als (1, 2) einmal als (m, w, i).
- ✔ Es sind inkonsistente Werte vorhanden.

In der Abbildung 1.2 sehen Sie, dass ein Artikel überhaupt nicht verkauft worden ist. Das kann möglich sein. Wie sieht es aber aus, wenn es Artikel ohne Planzahl gibt? Und was ist, wenn sich Daten, die aus verschiedenen Quellen kommen, widersprechen?

Natürlich kann man es den einzelnen Anwendern unter dem Stichwort »Self Service Business Intelligence« überlassen, die Daten aus verschiedenen Quellsystemen zu lesen, zu harmonisieren und auszuwerten. Es stellt sich aber dann die Frage, ob das in einem Unternehmen zu einem einheitlichen Verständnis der Unternehmensdaten führt. Hinzu kommt, dass die Anwender sehr kreativ mit ihren Anforderungen und Analysewünschen sind:

- ✔ Vielleicht wollen sie zusätzlich die Absatzzahlen pro Kunde nach Verkaufsort aufsummieren?
- ✔ Vielleicht sollen die Ist- und Planabsätze nicht nur je Artikel, sondern auch je Artikelgruppe gegenübergestellt werden?
- ✔ Vielleicht sollen die Istabsätze für die folgenden Monate prognostiziert werden, wozu man die Absatzhistorie benötigt?

Zum einen sind solche zusätzlichen Fragestellungen insbesondere für Mitarbeiter in Führungspositionen teilweise nur mit großem Detailwissen über das Tool und die zur Verfügung stehenden Daten zu beantworten, und zum anderen müssen eventuell weitere Datenquellen eingebunden werden, wodurch die Komplexität des Systems steigt. Und was ist, wenn das Unternehmen im Ausland eine Tochtergesellschaft gründet, die eine andere Software mit anders strukturierter Datenbank verwendet, und deren Daten in die Auswertung integriert werden sollen?

Reemers Publishing Services GmbH
O:/Wiley/Reihe_Dummies/9783527714476_Gerken/3d/c01.3d from 08.08.2018 14:03:10
3B2 9.1.580; Page size: 176.00mm x 240.00mm

Zusammenfassend lässt sich festhalten, dass es sicherlich sinnvoll wäre, wenn:

- ✔ die Daten nur aus einer Datenquelle (in Abbildung 1.4 als Data Warehouse bezeichnet) vom BI-Tool gelesen werden müssen und keine Attributverknüpfungen durch Endanwender mehr notwendig sind,
- ✔ die Datenquelle so strukturiert ist, dass Analysen mit unterschiedlichem Detaillierungsgrad problemlos möglich sind, zum Beispiel Umsätze pro Kunde oder Kundengruppe und pro Artikel oder Artikelgruppe, und
- ✔ auch historische Daten bei Bedarf für Analysen zur Verfügung stehen.

Prägen Sie sich die Abbildung 1.4 gut ein, denn davon handelt letztendlich dieses Buch.

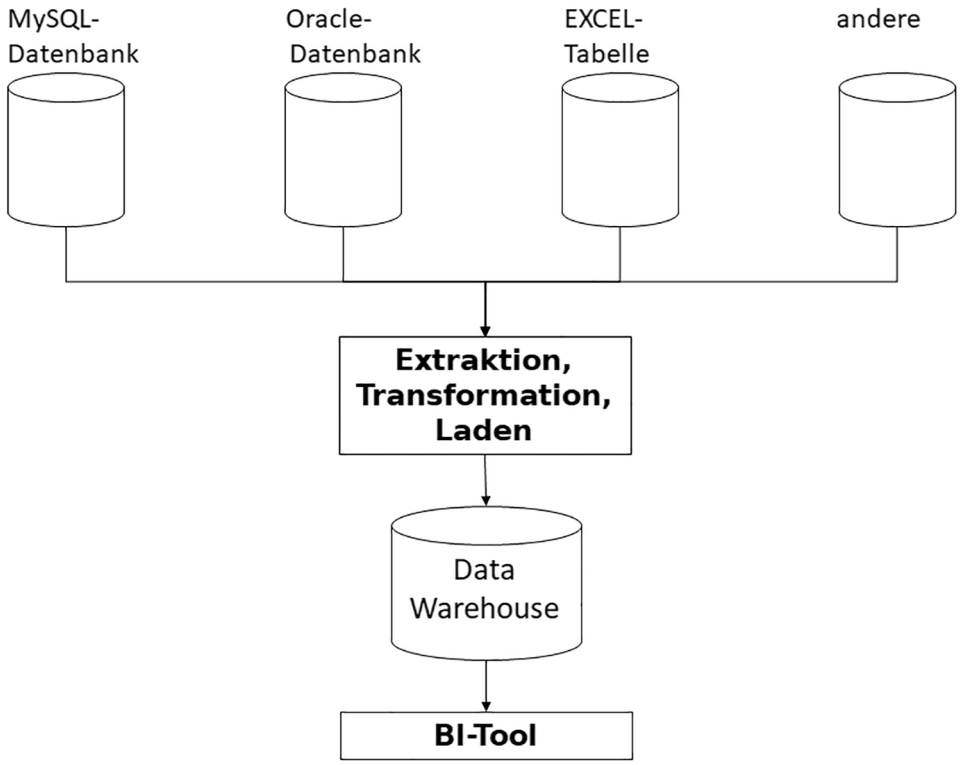

Abbildung 1.4: Datenanalyse mit Data Warehouse

Hier kommt jetzt das Data Warehouse ins Spiel. Dort werden die Daten der operativen Informationssysteme wie zum Beispiel der Auftragsverwaltung eines Onlineshops, des User-Tracking-Systems der Website, des Warenwirtschaftssystems, des Personalmanagements und der Finanzbuchhaltung gesammelt, für Auswertungen und Analysen vielfältiger Art angemessen strukturiert und stehen dann den Anwendern eines BI-Tools für ihre analytischen Aufgabenstellungen zur Verfügung.

Reemers Publishing Services GmbH
O:/Wiley/Reihe_Dummies/9783527714476_Gerken/3d/c01.3d from 08.08.2018 14:03:12
3B2 9.1.580; Page size: 176.00mm x 240.00mm

Vielleicht fragen Sie sich an dieser Stelle, wie denn die Daten der MySQL-Datenbank, Oracle-Datenbank usw. in das Data Warehouse gelangen. Das ist Aufgabe des Prozesses »Extraktion, Transformation, Laden«, auf den in Kapitel 5 detailliert eingegangen wird.

Ohne Data-Warehouse-Systeme sind die Steuerung betrieblicher Geschäftsprozesse und die Durchführung dispositiver und strategischer Entscheidungsprozesse in einem Unternehmen kaum noch denkbar.

Self-Service-BI-Lösungen auf Desktop-Ebene sind für prototypische oder begrenzte Auswertungen durchaus vertretbar. Belastbare, unternehmensweite Lösungen sollten aber auf Basis eines serverbasierten Data Warehouse implementiert werden.

Oder wollen Sie wirklich EXCEL-Tabellen zur Grundlage Ihrer unternehmerischen Entscheidungen machen?

Wenn personenbezogene Daten ins Spiel kommen

Die Notwendigkeit und wirtschaftliche Bedeutung von Datenanalysen mithilfe eines BI-Tools ist unbestritten. Dies gilt sowohl für die primären wertschöpfenden als auch für die sekundären unterstützenden Geschäftsprozesse in einem Unternehmen.

Wenn bei den Datenanalysen kein Personenbezug zum Beispiel zu Kunden, Lieferanten oder Mitarbeitern hergestellt werden kann, ist das alles unproblematisch. Wenn dieser Personenbezug allerdings gegeben ist, muss einiges beachtet werden:

- ✔ Zum einen muss der/die Betroffene der Nutzung zugestimmt haben bzw. muss es Vereinbarungen oder Verträge geben, die das zulassen,
- ✔ und zum anderen muss die Datenanalyse einem vordefinierten Zweck dienen und somit »erforderlich« sein.

Helfen kann in diesem Fall eine Pseudonymisierung oder besser Anonymisierung der Daten, damit keine Rückschlüsse auf die tatsächlich betroffenen Personen mehr bzw. nur mit hohem Aufwand möglich sind. Allerdings können pseudonymisierte Daten eventuell via statistischer Verfahren zurückgerechnet werden.

Beachten Sie bei Ihren Datenanalysen die Regelungen der EU-Datenschutzgrundverordnung (DSGVO) und das neue Bundesdatenschutzgesetz.

In Kapitel 7 kommen wir noch einmal auf dieses Thema zurück.