Die DSGVO aus Sicht der Business Intelligence

Datenschutz als Frage der Perspektive

Die Entscheidung, welche Daten notwendig sind, ist für operative Systeme leicht zu treffen. Für typische Business-Intelligence-Systeme und Data Warehouses hingegen ist diese Argumentation nicht so einfach, da diese den operativen Systemen nachgelagert sind. Ihre Funktion ist die zentrale und integrierte Speicherung und Verwaltung des Firmenwissens, das zur Analyse von Prozessen und für den Erkenntnisgewinn über die eigene Firma genutzt wird.

von Dr. Emrah Birsin und Dr. Stephan Jennewein

Mit Inkrafttreten der EU-DSGVO mussten sämtliche datenverarbeitenden Systeme überprüft werden. Denn personenbezogene Daten dürfen nicht gespeichert oder verarbeitet werden, außer es liegen besondere Umstände vor. In den meisten Fällen ist dies eine konkrete Einwilligung der Person zur Datenverarbeitung oder eine rechtliche Verpflichtung, siehe Artikel 6 Absatz 1 DSGVO. Bisher wurden alle verfügbaren Firmendaten oft auf elementarer Ebene gespeichert, um aussagekräftige Auswertungen zu ermöglichen. Mithilfe solcher Systeme werden zum Beispiel Fragestellungen aus dem Controlling und Marketing beantwortet und Berichte für sämtliche Bereiche innerhalb einer Firma erstellt. **Abbildung 1** zeigt einen vereinfachten schematischen Aufbau eines solchen Systems. Daten von Quellsystemen werden in den Stage-Bereich des Warehouse geladen um diese zu entkoppeln. Von dort aus werden sie in eine einheitliche Form gebracht, integriert und historisiert im Core abgelegt. Die Datamarts bringen diese integrierten und historisierten Daten in ein performantes Abfrageformat, oft thematisch gegliedert. Parallel dazu werden Metadaten angelegt, die diesen ganzen Transformationsprozess und die verwendeten Daten beschreiben. Auf Basis der in Datamarts bereit-

gestellten Daten werden themenbezogene Reports für unterschiedliche Fachbereiche erstellt, zum Beispiel für Marketing, Controlling, Kundendienst etc. Hierbei stellt sich nun die Frage, wie man das Bedürfnis nach guten und aussagekräftigen Analysen mit den Anforderungen der DSGVO in Einklang bringt. Auch andere Regelungen aus der DSGVO haben Auswirkungen auf den Betrieb eines Data Warehouse.

Dokumentation, Organisation und Archivierung

Als erster Schritt zur Erfüllung der Regelungen aus der DSGVO muss auch bei Data-Warehouse-Systemen der Dokumentationspflicht nachgekommen werden. Hierzu sollte ein Datenverarbeitungskatalog erstellt werden, der sämtliche existierende Daten und Attribute enthält, beschreibt, kategorisiert und darstellt, ob sie Personenbezug haben oder nicht. Dazu kommt, dass auch dokumentiert werden muss, wie personenbezogene Daten verarbeitet werden. Zur Erstellung der Dokumentation ist es hilfreich, eine Metadatenschicht zu pflegen, da mit den richtigen Hilfsmitteln daraus im Idealfall Großteile der Dokumentation inklusive der Data Lineage automatisch generiert werden können.

Alle Daten, die keinen Personenbezug aufweisen, sind von der DSGVO nicht betroffen. In allen anderen Fällen muss geprüft werden, ob die Daten beziehungsweise Attribute für Analysen nötig sind und somit ein begründetes Interesse vorliegt. Falls dies nicht gegeben ist, müssen diese Daten aus dem System entfernt werden. Sind die Attribute begründet gespeichert, müssen Schutzmaßnahmen definiert und dokumentiert werden: Das können zum Beispiel Zugriffbeschränkungen, Verschlüsselung oder bewusste Pseudonymisierung bzw. Anonymisierung sein. Ein alternativer Ansatz ist die Aggregatbildung und somit die Reduzierung des Detailgrads der vorliegenden Daten, wodurch ebenfalls der Personenbezug aufgehoben werden kann. Bei der Aggregatbildung wird die Anzahl an Zeilen in einer Tabelle reduziert, wohingegen bei der Pseudonymisierung bzw. Anonymisierung bestimmte Zellen einer Zeile unkenntlich gemacht werden. Schließlich ist für viele Analysen nur das Wissen über bestimmte Gruppierungen von Interesse, beispielsweise der Umsatz pro Postleitzahl, aber nicht der Umsatz jedes einzelnen Kunden. All diese Maßnahmen sind Teil des Konzepts „Privacy by Design“ und sollten idealerweise bei Neukonzipierung der Analytics-Infrastruktur direkt umgesetzt werden, um langwierige und kostspielige Nachbesserungen zu vermeiden. Darüber hinaus sollten alle beschlossenen Maßnahmen von einem Juristen oder Datenschutzbeauftragten auf ihre Richtigkeit geprüft und erst danach umgesetzt werden.

Vorsicht bei Selbstbedienung

Besonders das in den letzten Jahren beliebte Thema der Self-Service Business Intelligence müssen wir kritisch betrachten. Hierbei besteht die Möglichkeit für Datenanalysten aus den Fachbereichen, Daten aus Quellen

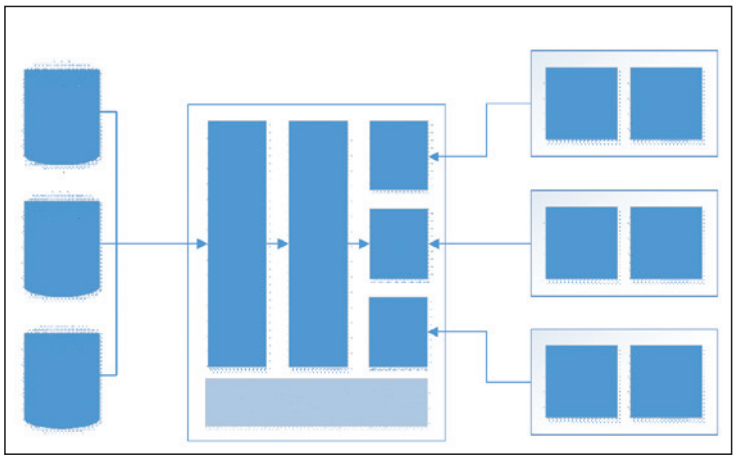

Abb. 1: Schematische Darstellung eines Data-Warehouse-Systems

direkt einzulesen, um damit Fragestellungen flexibler bearbeiten zu können als dies klassisch mit der zentralen Bereitstellung von Analysen und Reports möglich ist. Dabei kann es passieren, dass diese Daten an den dokumentierten Prozessen vorbei in eine Datenbank gelangen. Falls diese Daten einen Personenbezug aufweisen, würde dies den Regelungen der DSGVO zuwiderlaufen. Um das zu verhindern, lassen sich zwei Maßnahmen ergreifen: Einerseits Schulung und Sensibilisierung der Datenanalysten für diese Problematik, andererseits technische Zugriffsbeschränkung auf personenbezogene Daten.

Generell sollte bei der Einführung von Self-Service Business Intelligence darauf geachtet werden, dass diese ganzheitlich in den Auswertungsprozess integriert und nicht nur als Quick Fix für nicht bedachte Prozesse herangezogen wird. Dies erfordert einen guten Überblick über die unterschiedlichen Anforderungen sowohl auf organisatorischer als auch technischer Seite. So lässt sich die Integration sauber durchführen.

Nicht zu vergessen sind personenbezogene Daten, die durch technische Gegebenheiten entstehen. Seien es Logfiles oder technische Tabellen, die automatisch entstehen oder für die Dublettenerkennung notwendig sind. Diese dürfen im gesamten Prozess nicht vergessen werden und müssen entsprechend entfernt oder gepflegt und auf jeden Fall dokumentiert werden. Zu dieser Art von sekundären Daten zählen auch Back-ups und Datensicherungen, die allzu leicht übersehen werden. Je nach Aufbewahrungszeit der Back-ups müssen die dazugehörigen Prozesse angepasst und überarbeitet werden. Gerade das „Recht auf Vergessenwerden“ stellt besondere Anforderungen.

Datenverarbeitung durch Dritte

Zu den allgemeinen Anforderungen der DSGVO gesellt sich im Falle der Nutzung von Software oder Plattform as a Service (SaaS/PaaS) noch die Thematik der Auftragsverarbeitung. Hierbei handelt es sich zum Beispiel um Hosting- oder Cloud-Anbieter, die in der DSGVO als „Auftragsverarbeiter“ bezeichnet werden. Anbietern dieser Dienste kommen mehr Pflichten als bisher zu: Als

Auch wenn der Auftragnehmer bereits die Daten verschlüsselt, muss der Auftraggeber seiner Sorgfaltspflicht nachkommen und ebenfalls Maßnahmen zur Zugriffsbeschränkung einführen.

Erstes ist hier der Abschluss eines Vertrags zur Auftragsverarbeitung nach Artikel 28 der DSGVO zu erwähnen, da ohne einen schriftlichen Vertrag keine Rechtsgrundlage zur Auftragsverarbeitung besteht. Dabei muss dieser Vertrag gewisse Mindestanforderungen erfüllen, die in Artikel 28 Absatz 3 DSGVO festgelegt sind.

Artikel 4 Nummer 2 DSGVO listet reichlich Beispiele auf, was unter Verarbeitung zu verstehen ist: Erheben, Erfassen, Organisieren, Ordnen, Speichern, Anpassen oder Verändern, Auslesen, Abfragen, Verwenden, Offenlegen, Verbreiten oder Bereitstellen, Abgleichen oder Verknüpfen, Einschränken, Löschen oder Vernichten. Somit verarbeitet selbst ein Storage-Anbieter Daten, wenn diese nur bei ihm abgelegt werden. In allen Fällen bleibt aber der Auftraggeber in der Verantwortung zur rechtskonformen Verarbeitung der ihm anvertrauten Daten. Dies bedeutet allerdings auch, dass sich die Sorgfaltspflicht auf Auftragnehmer und Auftraggeber aufteilt. Auch wenn der Auftragnehmer bereits die Daten verschlüsselt, muss der Auftraggeber dennoch seiner Sorgfaltspflicht nachkommen und ebenfalls Maßnahmen zur Zugriffsbeschränkung einführen. Dabei stehen ihm zum Beispiel Rechtevergabe, zusätzliche Verschlüsselung, Pseudonymisierung oder Anonymisierung zur Verfügung, um seinerseits DSGVO-konform zu handeln.

Pseudonymisierung vs. Anonymisierung

Bisher wurden Pseudonymisierung und Anonymisierung als Möglichkeiten zur Erfüllung der DSGVO genannt. Nun wollen wir klären, worin die genauen Unterschiede bestehen. Unter Pseudonymisierung versteht man das Zuweisen eines Pseudonyms zu jedem personenbezogenen Datensatz und die darauffolgende Ersetzung des originalen Datensatzes durch das Pseudonym. Diese 1:1-Mapping-Tabelle wird separat gespeichert, um im Zweifelsfall die Zuordnung wieder herstellen zu können. Somit sind die Daten ohne Wissen über die Mapping-Tabelle geschützt. Wenn aber sowohl pseudonymisierte Daten als auch Mapping-Tabelle verfügbar sind, ist eine eindeutige Rücktransformation der Daten möglich. Dem gegenüber steht die Anonymisierung als eine unwiederbringliche Datenreduktion. Dabei werden in bestehenden Datensätzen Attribute unkenntlich gemacht oder gelöscht, bis nur noch die absolut notwendigen Informationen übrig sind. Dieser Vorgang darf

Vorname	Name	Straße	Postleitzahl	Stadt
Michael	Meier	Landstraße	12345	Berlin
Michael	Meier	Hauptstraße	14392	Berlin
Michael	Meier	Baumweg	13293	Berlin

Tabelle 1: Beispiel für 1-Anonymität, jede Person ist eindeutig zuordenbar; das „SELECT“-Statement würde Meier = 1, Meier = 1, Meier = 1 ausgeben, damit wäre k = 1.

Vorname	Name	Straße	Postleitzahl	Stadt
Michael	Meier	###	1####	Berlin
Michael	Meier	###	1####	Berlin
Michael	Meier	###	1####	Berlin

Tabelle 2: 3-Anonymität: Die Personen sind nicht mehr zu unterscheiden; das „SELECT“-Statement würde Meier = 3 ausgeben, es wäre k = 3 und zugleich k = n, da es insgesamt nur drei Datensätze gibt

Vorname	Name	Straße	Postleitzahl	Stadt
Michael	Meier	###	1####	Berlin
Michael	Meier	###	1####	Berlin
Michael	Meier	###	1####	Berlin
Michelle	Müller	###	1####	Berlin

Tabelle 3: 1-Anonymität durch Eindeutigkeit eines einzelnen Datensatzes; das „SELECT“-Statement würde Meier = 3 und Müller = 1 ausgeben, damit wäre k = 1

nicht umkehrbar sein, da es sich sonst wieder nur um eine Pseudonymisierung handeln würde.

In beiden Fällen bleibt die Anzahl an Zeilen pro Tabelle gleich, allerdings wird der Detailgrad in den Spalten reduziert. Als Alternative dazu kann auch der Detailgrad in Form von Reduktion der Zeilen (Aggregation) durchgeführt werden, um den Personenbezug zu entfernen. Mit dem Konzept der k-Anonymität lässt sich beurteilen, wie stark der Personenbezug entfernt wurde.

Wie anonym ist anonym?

Diese Frage ist vermutlich eine der schwierigsten Fragen, die man sich im Kontext der DSGVO stellen kann, da ein direkter Personenbezug laut Artikel 4 Absatz 1 DSGVO bereits dann vorliegt, wenn eine Person direkt oder indirekt durch jegliche Art von Information identifizierbar ist. In Tabelle 1 ist der Fall von direkter Zuordenbarkeit exemplarisch dargestellt.

Um sich zu veranschaulichen, wie stark anonymisiert ein Datensatz oder eine Tabelle ist, gibt es das Konzept der k-Anonymität [1]. Dabei stellt man sich die Frage, wie oft ein identischer Datensatz innerhalb einer Tabelle existiert. Dies bestimmt man für jeden Datensatz innerhalb einer Tabelle, und die niedrigste Anzahl von mehrfachen Vorkommen bestimmt das k. Wenn in der Tabelle also alle Einträge doppelt vorkommen, liegt k = 2, also 2-Anonymität, vor. Das bedeutet, dass alle Informationen in dieser Tabelle bis auf zwei Personen genau zugeordnet werden können, im Fall von k = 3 oder k = 4 auf drei bzw. vier Personen genau. Wenn k den Wert der Anzahl aller Datensätze n in einer Tabelle annimmt, also k = n, ist die maximal mögliche Anonymität erreicht, jede Person verschwindet in der Masse.

Sobald allerdings ein einziger Eintrag genau einer Person zuordenbar ist spricht man von k = 1, also einer direkten Zuordenbarkeit (Tabellen 1-3). Die entsprechenden Werte für k und damit auch das minimale k kann man mit dem folgenden SQL-Statement herausfinden. Dieses Statement gruppiert identische Datensätze des Beispiels und gibt die Anzahl der Vorkommen an: *SELECT Name, count(Name) FROM tabelle GROUP BY Vorname, Name, Straße, Postleitzahl, Stadt.*

Datensparsamkeit

Zur Einhaltung der Zweckbindung, Datenminimierung (Datensparsamkeit) und Speicherbegrenzung sollten die Daten in Business-Intelligence-Systemen kontinuierlich darauf überprüft werden, ob sie noch gebraucht werden. Ist schon beim Anlegen der Daten ein Verfallsdatum für den Zweck vorherzusehen, sollten die Daten mit einem Verfallsdatum versehen werden. Wenn möglich, ist eine automatische Entfernung vorzuziehen, sei es durch explizite Löschung oder Anonymisierung. Aber auch ein Verfahren, bei dem vor der Löschung die Bestätigung einer verantwortlichen Person eingeholt wird, ist möglich. Da je nach erhobenen Daten verschiedene Zeiträume in Frage kommen könnten, sollte das Verfallsdatum auf Datensatzebene gesetzt werden.

Für Daten, deren Verfallsdatum an Ereignisse gekoppelt ist, deren Eintrittszeitpunkt nicht klar ist, sollten, wenn möglich, Trigger hinterlegt werden, die beim Eintreten des Ereignisses entsprechendes Personal darauf hinweisen und gegebenenfalls ein Verfallsdatum vergeben, zum Beispiel im Fall der möglichen zukünftigen Kündigung.

In manchen Fällen mag es sein, dass zwecks statistischer Analysen gewisse Daten über einen Zeitraum gespeichert werden sollen, der weit über den eigentlichen Verwendungszeitraum hinausgeht, etwa um die Entwicklung der Kundenzahlen über die Jahre hinweg zu analysieren. In diesem Fall muss sichergestellt werden, dass die Daten nicht einer Person zugeordnet werden können, sprich keine Pseudonymisierungsdaten mehr vorliegen und die Daten so stark wie möglich anonymisiert sind.

Sowohl die Menge als auch der Informationsgehalt der gespeicherten Daten sollte auf ein Minimum reduziert sein. Soll beispielsweise nur die Anzahl von

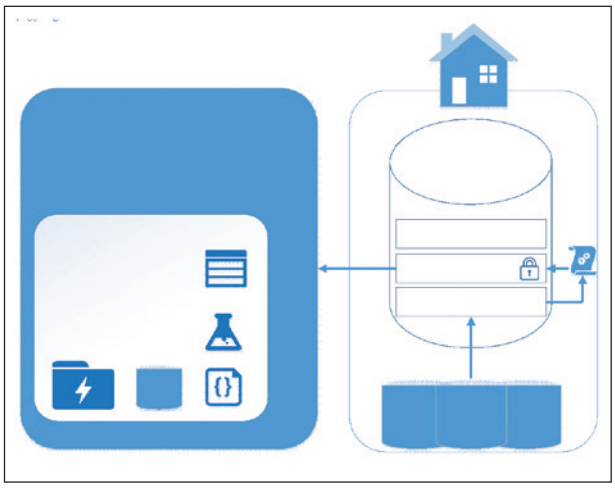

Abb. 2: Schematische Darstellung des Offloadings mit vorhergehender Pseudonymisierung

Kunden über die Jahre analysiert werden, so reichen Jahresangaben für den Kundenkontakt und eine interne Kunden-ID, die keine Zuordnung zu einer natürlichen Person ermöglicht.

Szenario 1: Umzug in die Cloud

Es gibt viele gute Gründe für die Nutzung einer Cloud. Einer davon ist die Modernisierung des Data Warehouse und dabei der Umstieg von einer Data-Warehouse- auf eine Data-Lake-Architektur [2], [3], [4]. Besonders Daten mit nichtrelationalen Strukturen für zukünftige Auswertungen speichern zu können, ist eine besondere Stärke einer solchen Data-Lake-Architektur. Hierbei ist Vorsicht geboten, da eine zweckfreie Speicherung von personenbezogenen Daten nicht erlaubt ist. Ein weiteres Szenario besteht darin, dass ein bereits etabliertes Data Warehouse mit Inkrafttreten der DSGVO möglicherweise den regulatorischen Anforderungen nicht mehr genügt. Ein kompletter Neuaufbau bzw. eine Restrukturierung des bestehenden Systems dauert lange und droht entsprechend kostenintensiv zu werden. Hier kann das Offloading in einen Data Lake, der den entsprechenden Regelungen entspricht, bei gleichzeitiger sehr strikter Zugriffsbeschränkung auf das bestehende Data Warehouse das Problem abfedern. Gleichzeitig lässt sich damit Zeit für notwendige Anpassungen des Data Warehouse erkaufen. Ein Data-Lake-Konstrukt ermöglicht es, die Daten relativ unabhängig von im vorhandenen Data Warehouse existierenden Strukturen und Mechanismen zu extrahieren, um sie mit individuellen und flexiblen Algorithmen rechtskonform übertragen zu können. Gleichzeitig schafft man die Möglichkeit, einen weiteren Schritt in Richtung einer Zentralisierung der Unternehmensdaten zu gehen und ermöglicht die Integration weiterer Datenquellen wie Streamingdienste oder Sensordaten.

Schon vor der Übertragung aus dem alten System in den Data Lake in der Cloud sollten die Daten pseudonymisiert oder anonymisiert werden, damit auf dem neuen System von Vornherein kein Personenbezug existiert. Wenn Daten aus mehreren Quellen für die Verarbeitung einer bestimmten Person kombiniert werden müssen, sollte die Person durch ein Pseudonym repräsentiert werden. Doch eine Pseudonymisierung allein reicht zum Schutz der Daten nicht aus. Wenn ein Dritter die Daten mit Daten aus anderen Quellen verbindet, könnte es sein, dass Personen identifiziert werden können [4], [5]. Daher müssen personenbezogene Daten verschlüsselt gelagert und transferiert werden, damit im Fall eines Datenlecks keine personenbezogenen Informationen aus den Daten extrahiert werden können.

Wenn möglich, sollten Daten, die für verschiedene Analysen genutzt werden und nicht miteinander verknüpft werden müssen, auch mit einer separaten Pseudonymisierung versehen werden, um die Verknüpfung von Datensätzen durch Dritte zu erschweren. Pseudonymisierungsdaten, also Daten, die Informationen enthalten, auf welche Art pseudonymisiert wurde, und nicht anonymisierte Daten sollten nur in dem System gelagert werden, von dem aus das Offloading stattgefunden hat, und eine eigene Verschlüsselung erhalten, um die Sicherheit der personenbezogenen Daten weiter zu steigern. Gegebenenfalls sollten diese Daten auch physisch separat gespeichert und treuhänderisch verwaltet werden.

Der Zugriff auf Daten sollte durch ein Rechtevergabesystem auf Grundlage der Verantwortung der einzelnen Mitarbeiter geregelt sein, dazu kann der Data Lake in verschiedene Bereiche eingeteilt werden, in denen jeweils nur bestimmte Daten abgelegt werden, um die Menge an Zugriffen zu minimieren. Besonders der Zugriff auf pseudonymisierte und nicht anonymisierte Daten sollte möglichst gering gehalten werden. Auch hier kann eine treuhänderische Verwaltung in Betracht gezogen werden.

Außerdem sollte bedacht werden, dass pseudonymisierte Daten nach wie vor einen Personenbezug besitzen und somit im Fall der Ausübung des Widerspruchs- oder Einschränkungsrechts zeitweise eingeschränkt oder gar nicht verarbeitet werden dürfen. Daher sollte die Information, ob bestimmte Daten überhaupt verarbeitet werden dürfen, hinterlegt sein, am besten auf der Datensatzebene. Analysetools können dann Daten, die nicht zur Verfügung stehen, automatisch filtern und von der Verarbeitung ausschließen.

Ein Offloading mit gleichzeitiger Pseudonymisierung bzw. Anonymisierung lässt sich in relativ kurzer Zeit und mit hohem Automatisierungsgrad realisieren. Eine vereinfachte schematische Darstellung dieses Szenarios ist in **Abbildung 2** zu sehen.

Szenario 2: Recht auf Vergessenwerden

Wenn Daten gelöscht werden sollen – sei es, weil der Zweck der Datenerhebung erfüllt wurde oder aufgrund eines Löschantrags – muss darauf geachtet werden, dass sämtliche Daten, die mit der Person in Bezug stehen, ebenfalls gelöscht werden. Wobei „gelöscht“ in den meisten Fällen bedeutet, dass die Daten komplett anonymisiert werden, da ein mögliches Interesse an

Daten, die ihren Verwendungszweck erfüllt haben, sollten automatisch vollständig anonymisiert werden, wobei darauf zu achten ist, dass keine indirekte Identifizierung möglich ist.

den Daten zwecks historischer Datenanalyse besteht. Dabei wird der Inhalt beziehungsweise der Personenbezug unwiderruflich zerstört. Allerdings behält man die Information, dass es diesen Datensatz gab. Daraus lassen sich weiterhin Aggregate bilden, wie Anzahl oder Gruppierungen.

Daten, die ihren Verwendungszweck erfüllt haben, sollten automatisch einer vollständigen Anonymisierung unterzogen werden, wobei natürlich darauf zu achten ist, dass keine indirekte Identifizierung aufgrund verknüpfter Daten durchführbar ist. Bei einem Löschungsantrag müssen alle Daten der Person identifiziert und entsprechend anonymisiert werden. Pseudonymisierungsinformationen (ID) müssen gelöscht werden. Selbst zum Zweck der Dokumentation darf keine Information erhalten bleiben, ob jemand Teil der Datenbank war, da selbst die Information, dass jemand Teil einer Datenbank war, eine personenbezogene Information ist. Es ist auch darauf zu achten, dass die Löschung in allen Kopien und Replikationen unverzüglich durchzuführen ist, was Back-ups miteinschließt. Daher sollten nach der Löschung der Daten im Produktivsystem alte Back-ups aktualisiert werden. Da die Sicherheit der Daten anderer Personen hier eine Rolle spielt, kann das Back-up so lange verschoben werden, bis die Stabilität des Systems und die Sicherheit der Daten anderer Personen gewährleistet sind. Kriterien zur Identifizierung dieses Zeitpunkts sollten in entsprechenden Dokumenten festgelegt sein.

Ein gewisser Konflikt entsteht hier bei der Benachrichtigung von Empfängern der Daten, zum Beispiel Auftragsverarbeitern. Wenn die Löschung der Daten durchgeführt werden soll, gilt es, alle Verarbeiter zu informieren, dass sie ebenfalls diese Daten löschen müssen. Hierzu sollte nur mittels Pseudonym kommuniziert werden, welcher Datensatz zu löschen beziehungsweise zu anonymisieren ist, da ansonsten jede Speicherung des Löschantrags wieder einen Personenbezug aufweisen würde und somit zu löschen wäre. Natürlich muss am Ende durch Löschung des Personen-zu-Pseudonym-Mappings eine Anonymisierung stattfinden.

Ein weiterer Konflikt entsteht, falls ein Unternehmen bilanzierungspflichtig ist und als Datenquelle ein Data Warehouse oder einen Data Lake verwendet. Das Bilanzrecht erlaubt keine nachträgliche Veränderung der Quelldaten, die für die Bilanz herangezogen wurden. Demgegenüber steht das Recht auf Vergessenwerden entsprechend der DSGVO. Allerdings sind in Artikel 17, Absatz 3 DSGVO Ausnahmen definiert, die das Löschrecht hinfällig machen. Um sich darauf berufen zu können, müssen Einzelfälle geprüft und dokumentiert werden.

Fazit

Der Datenschutz von personenbezogenen Daten ist auch in analytischen Systemen ein komplexes Thema. Analytics-Systeme und Data Warehouses müssen immer daraufhin überprüft werden, ob sie auf dem neuesten technischen Stand sind. Prozesse wie Pseudonymisierung und Anonymisierung sowie Zugriffsrechte und Verschlüsselung der Daten bilden eine technische Grundlage zum Schutz der Daten. Allerding sind auch eine Sensibilisierung der Mitarbeiter und eine genaue Betrachtung der Umstände, unter denen die Daten erhoben, gespeichert und verarbeitet werden, unabdingbar.

Dr. Stephan Jennewein ist Business & IT Analyst bei OPITZ CONSULTING Deutschland GmbH im Bereich der Business Intelligence. Durch seine vielfältigen Einblicke in unterschiedlichste Setups erlangte er ein fundiertes Wissen, um technische Abläufe zu analysieren.

Dr. Emrah Birsin ist Business Intelligence & Analytics Developer bei OPITZ CONSULTING Deutschland GmbH. Im Rahmen seiner mehrjährigen Erfahrung mit klinischen Studien erwarb er Kompetenzen im Umgang mit personenbezogenen Daten.

Links & Literatur

[1] Sweeney, LaTanya: „k-Anonymity: A Model for Protecting Privacy“; in: International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 2002

[2] Chroust, Tomas; Nemecek, Josef; Kürschner, Sandro: „Data Lakes – Möglichkeiten und Herausforderungen für eine effiziente Erkenntnisgewinnung“; in: St. Galler Trendmonitor für Risiko- und Finanzmärkte, 2015

[3] Stein, Brian; Morrison, Alan: „The enterprise data lake: Better integration and deeper analytics”; in: Technology Forecast: Rethinking integration, 2014

[4] Vallaey, Matthias: „Why do I need a Data Lake”: http://info.bigindustries.be/why-do-i-need-a-data-lake

[5] Narayanan, Arvind; Shmatikov, Vitaly: „Robust De-anonymization of Large Sparse Datasets“; in: IEEE Symposium on Security and Privacy, 2008