Datenschutz-Grundverordnung für Datenbank-Administratoren

Alexander Kornbrust, Red Database Security

Am 25. Mai 2018 tritt die europäische Datenschutz-Grundverordnung (DSGVO, engl. GDPR) [1] in Kraft. Diese Verordnung umfasst 99 Artikel und 173 Erwägungsgründe (Abmilderungen/Einschränkungen). Der Artikel konzentriert sich auf das Thema „Administration von Datenbanken", da dort in der Regel personenbezogene Daten abgelegt sind.

Der Autor gibt keinen rechtlichen Beistand. Die nachfolgenden Informationen sind jedoch nach bestem Wissen und Gewissen zusammengestellt. Dabei wird auf folgende Fragen eingegangen:

- Was ist die DSGVO in wenigen Sätzen?
- Was hat ein DBA mit der DSGVO zu tun?
- Warum ist die DSGVO so schwierig umzusetzen?
- Was muss man minimal umsetzen?
- Wie findet man (automatisiert) personenbezogene Daten?
- Wie lässt sich die Abfrage einer Person nach ihren Daten umsetzen?

Man kann über die DSGVO ganze Bücher schreiben. Nachfolgend eine kurze, informelle Zusammenfassung über die Motivation der EU-Behörde und die Zielsetzung hinter dieser Verordnung. Das erste Hauptziel der DSGVO ist es, jedem Bürger das Recht zu geben, die Daten, die über ihn von Firmen/Organisationen in der EU gespeichert sind, zu sehen (Artikel 15), zu ändern/korrigieren (Artikel 16) und zu löschen (Artikel 17), falls diese rechtlich nicht mehr benötigt werden. Zudem kann die Verarbeitung der Daten eingeschränkt werden (Artikel 18).

Das zweite Hauptziel ist die Sicherstellung, dass Hackerangriffe (intern und extern) nicht mehr unter den Tisch gekehrt werden können, da die zuständige Aufsichtsbehörde innerhalb von 72 Stunden nach Bekanntwerden des Vorfalls informiert werden muss (Artikel 33). Zusätzlich sind die vom Datenleck betroffenen Personen (Artikel 34) direkt oder durch öffentliche Bekanntmachung zu informieren.

Um diese beiden Ziele sicherzustellen, wurden die hohen Strafen hinzugefügt (Artikel 83). Gleichzeitig existiert ein Recht auf Schadensersatz bei betroffenen Personen (Artikel 82). Hinzu kommen allerlei Dokumentationspflichten (wie Verfahrensverzeichnis, Backup/Restore-Konzept etc.), die von den Firmen/Organisationen erfüllt werden müssen.

Der Datenbank-Administrator und die DSGVO

Im Grunde hat ein Datenbank-Administrator sehr wenig mit der DSGVO zu tun, da er normalerweise nicht für die (personenbezogenen) Systeminhalte verantwortlich ist. Anforderungen (Artikel 32) wie zeitnahes Restore

betroffener Systeme und Kapazitätsplanung sind in der Regel standardmäßig implementiert; der Datenbank-Administrator könnte sich also gemütlich zurücklehnen. Da er aber normalerweise hilfsbereit, wissend und vorausschauend ist, sollte er die Datenschützer und Anwendungsverantwortlichen nach Möglichkeit mit seinem Datenbank-Know-how bei der Umsetzung der DSGVO unterstützen.

DSGVO-Projekte sind deshalb so schwierig zu realisieren, weil es in den Köpfen der DSGVO-Verantwortlichen (Datenschutz, Juristen, Projektteam) nur ganz wenige (betroffene) Datenbanken gibt, in denen personenbezogene Daten vorhanden sind. In großen Firmen existieren oft mehr als hundert Datenbanken, in Konzernen sogar zwischen tausend und zehntausend (von relational bis zu Big Data). Diese Datenbanken enthalten nach Erfahrung des Autors zu einem großen Anteil personenbezogene Daten (je nach Definition beispielsweise Vorname, Nachname, E-Mail-Adresse etc.). In diesen Systemen sind Hunderte bis Tausende Tabellen mit personenbezogenen Daten üblich, die alle betrachtet werden müssen – auch wenn es zum Teil unterschiedliche Meinungen gibt („Diese Personendaten zählen nicht", „Das weiß der Kunde sowieso nicht", „Das sind ja nur ein paar Daten", „Wir machen nur Produktionssysteme"). Unter diese Datenbanken fallen natürlich auch die Q/A-, Pre-Live-, Test- und Entwicklungssysteme, sofern diese nicht vollständig anonymisiert sind, was normalerweise selten erfolgt, da man nicht genau weiß, wo überhaupt welche personenbezogenen Daten liegen.

Die minimale Umsetzung

Da bei Erscheinen dieses Artikels nur noch wenig Zeit für die Umsetzung bleibt, sollte man sich über die Priorität der DSGVO-Implementierung Gedanken machen. Wenn man x Leute fragt, wird man x+1 verschiedene Meinungen dazu erhalten. Der Artikel geht auf die Anforderungen ein, die ab 25. Mai 2018 eine Außenwirkung zeigen. Auch innerhalb weniger Wochen kann man viel erreichen, wenn man es wirklich will und einen Plan hat.

Anfrage einer Person nach ihren Daten

Da es sehr wahrscheinlich ab dem 25. Mai 2018 Anfragen zu Personendaten geben wird, sollte man den entsprechenden Workflow einsatzbereit haben. Die Antworten müssen spätestens nach einem Monat (Artikel 12) dem Anfragenden (kostenlos) zugesandt werden. Bei einer größeren Menge von Anfragen gibt es maximal zwei Monate mehr Zeit für die Beantwortung, Anfragen vom 25. Mai 2018 müssen also bis allerspätestens 25. August 2018 beantwortet sein.

Verzeichnis der Verarbeitungstätigkeiten für die Aufsichtsbehörde

Gemäß Artikel 30 ist ein Verzeichnis aller Verarbeitungstätigkeiten schriftlich zu führen, wenn eine Mindestanzahl an Mitarbeitern (mehr als 250, Erwägungsgrund 13) vorliegt. Dieses Verarbeitungsverzeichnis lässt sich beispielsweise aus einem Konfigurationsmanagementsystem (CMDB) erstellen und gemäß DSGVO um die üblicherweise fehlenden Informationen ergänzen (etwa Kategorie personenbezogener Daten, Dauer der Datenhaltung etc.).

Verhalten nach einem Hackerangriff

Zumindest auf dem Papier sollte definiert sein, wer im Falle eines Falles in welcher Zeit informieren sollte, welche (Forensik-)Daten von wem und wo zu sammeln sind und wer für die Kommunikation mit den Aufsichtsbehörden zuständig ist. Sollte es bereits einen Incident-Management-Prozess geben, empfiehlt es sich, die Datenschutz-Grundverordnung einzubauen.

Audit einer Wirtschaftsprüfungsgesellschaft

Da eine maximale Strafe von vier Prozent des Gesamtumsatzes ein (großes) Risiko für ein Unternehmen darstellt, werden die Wirtschaftsprüfungsgesellschaften zu diesem Thema prüfen. Da solche Prozesse oftmals Papierprüfungen sind, sollte man die entsprechenden Dokumente vorhalten beziehungsweise im Vorhinein erstellen (Artikel 32: Sicherheit der Verarbeitung, Backup/Recovery, regelmäßige Überprüfung etc.). Da Audits nicht gleich im Mai 2018 stattfinden werden, hat man bei diesem Punkt mehr Zeit.

Automatisiert personenbezogene Daten finden

Um die Anfrage einer Person nach ihren Daten beantworten zu können, sollte man sich einen Workflow überlegen. Dieser könnte wie folgt aussehen:

- Anfrage einer Person (per E-Mail, Webseite etc.)
- Identität der Person überprüfen (etwa anhand von Vertragsdaten, Video-Ident-Verfahren, Personalausweis [2], durch persönlichen Besuch etc.)
- Suche nach Daten dieser Person in (allen) Datenbanken
- Erstellung eines (maschinenlesbaren) Berichts
- Zusenden des Berichts zum Anfragenden

Dieser Workflow ist relativ einfach zu verstehen und in der Theorie einfach umzusetzen. Sobald man ihn jedoch realisieren soll, stößt man auf ein paar Probleme:

Deutsch	Englisch	Französisch
Vorname	Firstname	prenom
Nachname	Lastname	nom
Straße	Street	rue
PLZ	Zip	codepostal
Ort	Town	ville
Stadt	City	
Gebdat	Dob	ne
Geburtsdatum	Dateofbirth	datenaissance
...	...	...

Tabelle 1

```
Select * from acc.t_inv_osuser where vorna='Alexander' and
nachn='Kornbrust';
Select * from damrepo.osusers where vorname='Alexander' and
nachname='Kornbrust';
Select * from dam.kunden where firstname='Alexander' and
lastname='Kornbrust';
...
```

Listing 1

- Was sind personenbezogene Daten?
- Wo sind diese Daten abgelegt?
- Wie komme ich an diese Daten?

Die Definition der personenbezogenen Daten muss jede Firma/Organisation für sich selbst festlegen und auch regelmäßig aktualisieren. Hier gibt es zwar einen kleinsten gemeinsamen Nenner (Vorname, Nachname, Adresse, Geburtsdaten etc.), die Ansicht über weitergehende personenbezogene Daten (GPS-Daten, IP-Adressen etc.) ist innerhalb einer Firma allerdings oft unterschiedlich.

Sobald diese Daten definiert sind, sollte man eine Liste dieser Begriffe in unterschiedlichen Sprachen anlegen, da Datenbank-Entwickler oft die eigene Muttersprache zur Bezeichnung verwenden *(siehe Tabelle 1)*. Dabei ist zu beachten, dass es oft mehrere Synonyme für einen Begriff gibt (Lastname, namelast, lname, familyname, surname etc.), die alle zu überprüfen sind. Ob ein Begriff üblicherweise verwendet wird, lässt sich gut mithilfe einer Suchmaschine ausprobieren (beispielsweise durch die Suche nach „create table" „vorname varchar").

Sobald die personenbezogenen Daten definiert sind, kann man die Metadaten (Tabellen und Spaltennamen) der Datenbanken (Oracle, MSSQL, MySQL, SAP etc.) nach diesen Schlüsselwörtern durchsuchen. Dies kann mit Programmen wie Oracle Security Assessment Tool (nur Oracle, englisch) [3] oder gdprscan (mehrere Plattformen, mehrsprachig) [4] erfolgen. Der automatisierte Ansatz ist hier der manuellen Anfrage an die unterschiedlichen Fachabteilungen vorzuziehen, da er schneller und oft zuverlässiger ist.

Die Fundstellen in (nicht-leeren) Tabellen gilt es zu identifizieren; anschließend die Orte, in denen sich personenbezogene Daten befinden; also eine nicht-leere Tabelle mit Vorname, Nachname, Handynummer enthält in der Regel personenbezogene Daten. Normalerweise sind die Kunden überrascht, was alles in ihren Datenbanken gefunden wurde. Zum einen findet man am Anfang oft False-Positives (Suche nach „%ort%" findet auch „Port", „%RASSE%" auch „Strasse" etc.), was sich jedoch durch Finetuning/Ausnahmen verbessern lässt. Zum anderen findet man personenbezogene Daten oft in den privaten Schemata der Entwickler (Kopie der Produktionstabelle), in Backup-Tabellen („EMP_BCK", „EMP_03012011" etc.) oder im Oracle Recylce-Bin (das weiterhin zugreifbar bleibt). Diese Fundstellen sollten von den entsprechenden Fachabteilungen/Verantwortlichen kontrolliert und verbessert werden. Nicht notwendige Tabellen mit personenbezogenen Daten (Backup, Recycle Bin, private Kopien etc.) sind am besten gleich zu löschen.

Das Finden der Daten einer anfragenden Person mit manuellen Prozessen ist enorm zeitaufwendig – wenn man beispielsweise in zehn unterschiedlichen Abteilungen mit insgesamt hundert verschiedenen Anwendungen nachfragt, ob es Daten über einen „Hans Meier, geb. 23.12.1977" gibt. Diese Daten sind zu sammeln und weiterzuleiten. Hier entstehen pro Anfrage zum Teil Aufwände von mehreren Personentagen.

Alternativ kann man die bei der Analyse der Metadaten gewonnenen Daten mithilfe von dynamischen SQL-Befehlen abfragen, da man nun weiß, welche personenbezogenen Daten in welchen Tabellen beziehungsweise Spalten abgelegt sind. Dabei werden für jede Datenbank und jede Tabelle dynamische SQL-Befehle ausgeführt. *Listing 1* zeigt ein Beispiel.

Die Suche kann dann über mehrere Datenbanken erfolgen. Gegenüber der manuellen Abfrage von hundert Anwendungen bedeutet dies eine erhebliche Erleichterung und Zeitersparnis.

Fazit

Die europäische Datenschutz-Grundverordnung stellt enorme Anforderungen an jede Firma/Organisation. Anstatt den Kopf in den Sand zu stecken, sollte man sich auf die dringendsten Probleme wie Anfrage-Prozess und das Verarbeitungsverzeichnis konzentrieren, um ab dem 25. Mai 2018 nicht in den Fokus der zuständigen Datenschutzbehörde zu geraten. Datenbank-Administratoren können die Datenschützer beziehungsweise das DSGVO-Projektteam dabei stark unterstützen, indem sie sowohl das Finden von personenbezogenen Daten (per Metadaten-Analyse) als auch die Suche nach Daten einer bestimmten Person (per dynamischen SQL) automatisieren.

Weiterführende Links

[1] DSGVO: *https://dsgvo-gesetz.de*
[2] Personalausweis: *https://www.datenschutz-praxis.de/wp-content/uploads/s_November_14_web1.pdf*
[3] Oracle DSAT: *https://docs.oracle.com/cd/E76178_01/SATUG/toc.htm*
[4] GDPRSCAN: *http://www.gdprscan.de*

Alexander Kornbrust
ak@red-database-security.com

Data Analytics 2018: Der Zukunft ein Stück näher

Am 19. und 20. März 2018 fand die gemeinsame Data Analytics Konferenz von Oracle und der DOAG im Phantasialand in Brühl statt. Unter dem Motto „Daten als Motor der Digitalisierung" übertraf die 13. Data Warehouse Konferenz mit rund 250 Besuchern und 14 Ausstellern die Erwartungen.

Zukünftige Innovationen und Digitalisierung zogen sich durch die gesamte Veranstaltung wie ein roter Faden; besonders erfreulich: Neben Oracle-Experten und Großunternehmen hielten Kunden knapp 50 Prozent der Vorträge. Praxisnah berichteten beispielsweise Dominic Marx und Andreas Howanietz von DB Cargo über ihre Reise in die Cloud und welche Stolpersteine es auf dem Weg zu überwinden galt. Sie teilten ihre Erkenntnisse und gaben wertvolle Tipps und Tricks. Professor Hartmut Westenberger von der TH Köln stellte seine Studie zum Thema BI-Strategie und Industrialisierung vor. Die darauffolgende Frage- und Diskussionsrunde sprengte fast den Zeitrahmen und spiegelte das große Interesse wider.

Das Programm war vielfältig: Neben klassischen, technischen Themen fanden auch aktuelle Fragestellungen zu Machine Learning, Design Thinking, Geodaten, Datenschutzgrundverordnung und viele weitere Gehör. In der Eröffnungs-Keynote stellte Sohan de Mel, Vice President of Product Strategy und Business Development bei Oracle, das autonome Datenmanagement der Zukunft vor und sprach somit gleich das Hauptthema der Veranstaltung an.