Digital Business University
of Applied Sciences

Cyber- & IT-Security (M.Sc.)

Datenmodellierung und Datenbanksysteme

Sommersemester 2024

Datenschutz in relationalen Datenbanksystemen: Technische und organisatorische Maßnahmen zur Umsetzung der DSGVO

Studienarbeit

Eingereicht von:
Elias Häussler

Matrikelnummer:
200094

Dozent:
David Lübeck

2. November 2024

Inhaltsverzeichnis

Abkürzungsverzeichnis

BDSG Bundesdatenschutzgesetz

DAC Discretionary Access Control

DBMS Datenbankmanagementsystem

DIN Deutsches Institut für Normung e. V.

DSGVO Datenschutz-Grundverordnung

EG Europäische Gemeinschaft

ErwG Erwägungsgrund

EU Europäische Union

HMAC Hash-based Message Authentication Code

MAC Mandatory Access Control

MAC Message Authentication Code

ML Machine Learning

MLS Multi-level security

NoSQL Not-only-SQL

PDCA Plan–do–check–act

PGP Pretty Good Privacy

PHI Protected Health Information

PII Personally Identifiable Information

RBAC Role Based Access Control

RNG Random Number Generator

RSA Rivest–Shamir–Adleman

SQL Structured Query Language

WDSP Web Data Service Provider

1 Einleitung

Die wachsende Digitalisierung organisatorischer und struktureller Prozesse erfordert schon seit vielen Jahren die Nutzung komplexer Speichersysteme für die Masse an Daten, die dort verarbeitet werden. Eine Vielzahl dieser Daten sind personenbezogen und lassen somit Rückschlüsse auf einzelne Individuen zu. Schon früh wurde erkannt, dass diesen höchst schützenswerten Daten ein besonderes Augenmerk in puncto Sicherheit und Datenschutz zugelegt werden muss.

Da es lange Zeit keine länderübergreifende Regelungen zum Datenschutz personenbezogener Daten gab, haben unterschiedliche Länder eigene Richtlinien entwickelt; in Deutschland beispielsweise durch das im Jahr 1978 erstmalig in Kraft getretene *Bundesdatenschutzgesetz (BDSG)*. Zwar verabschiedete die Europäische Gemeinschaft (EG) im Jahr 1995 die *Richtlinie 95/46/EG* „zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr“ (Voigt & von dem Bussche, 2024, S. 2), diese war jedoch nicht direkt anwendbar, sondern erforderte die Umsetzung in nationales Recht in den einzelnen Mitgliedstaaten (Voigt & von dem Bussche, 2024).

Um das Datenschutzniveau in allen Mitgliedstaaten anzugleichen, wurde schließlich im Jahr 2016 die *Datenschutz-Grundverordnung (DSGVO)* verabschiedet. Zu diesem Zeitpunkt nutzten viele Unternehmen bereits die umfassenden Möglichkeiten zur Datenspeicherung, unter anderem in Form relationaler Datenbanksysteme. Für diese Systeme galt es nun, gemäß der DSGVO entsprechende Maßnahmen zur Sicherstellung des geforderten Datenschutzniveaus umzusetzen. Hierfür forderte die DSGVO die Implementierung technischer und organisatorischer Maßnahmen.

1.1 Problemstellung

Datenschutzmaßnahmen waren grundsätzlich schon vor Verabschiedung der DSGVO vorhanden und durch geltendes Recht bereits ein notwendiger Bestandteil im Datenbank-Management. Die Komplexität, die durch die Anpassung bestehender oder Entwicklung neuer Datenbanksysteme nach der DSGVO entsteht, liegt vor allem in ihren deutlich strengeren Richtlinien und den damit verbundenen hohen Sanktionen bei deren Verletzung. Es erfordert daher umfangreiche technische und organisatorische Maßnahmen für den DSGVO-konformen Betrieb relationaler Datenbanksysteme.

In dieser Arbeit werden folgende Fragestellungen behandelt, die der Problematik eines datenschutzkonformen Designs und Managements relationaler Datenbanksysteme nachgehen:

- Wie können relationale Datenbanksysteme technisch gerüstet werden, um Konzepte wie Datenminimierung, Speicherbegrenzung und Rechenschaftspflicht angemessen zu erfüllen?

- Welche technischen Maßnahmen für Integrität und Vertraulichkeit sind auf relationale Datenbanksysteme anwendbar?
- Welche ergänzenden organisatorischen Maßnahmen sind notwendig, um die Grundsätze für die Verarbeitung personenbezogener Daten nach Art. 5 DSGVO einzuhalten?
- Wie kann sichergestellt werden, dass alle geforderten Maßnahmen auch in den Umgang mit archivierten Datenbank-Backups einfließen?

1.2 Aktueller Forschungsstand

Viele wissenschaftliche Artikel und Literaturveröffentlichungen befassen sich mit den komplexen Maßnahmen an einen datenschutzkonformen Umgang mit personenbezogenen Daten. Voigt und von dem Bussche (2024) stellen einen umfangreichen Praxisleitfaden zur Umsetzung technischer und organisatorischer Maßnahmen vor, wobei auch auf das in Deutschland verschärfte Recht durch Anwendung des aktualisierten BDSG Bezug genommen wird. In Betrachtung von Marketing und Vertrieb veröffentlichten Krämer und Mauer (2023) einen dreiteiligen Leitfaden zum Status Quo und Ausblick der DSGVO, verbunden mit einer umfangreichen Auswahl konkreter Anwendungsfälle, darunter *Anwendungsfall G: Aufbau und Absicherung von Datenbanken mit personalisierten Daten*. Einen verschärften Blick liefern Pina et al. (2024), die auch auf ethische Überlegungen einzelner Maßnahmen eingehen und diese kritisch bewerten. Darüber hinaus stellen Shyamasundar et al. (2021) verschiedene Modellverfahren vor, mit denen Datenschutz in Datenbanksystemen gewährleistet werden soll.

1.3 Inhalte der Arbeit

In diese Arbeit wird speziell auf die Anwendung der DSGVO in relationalen Datenbanksystemen eingegangen. Hierzu werden in Kapitel 2 zunächst die notwendigen Grundlagen erläutert. Kapitel 3 stellt eine umfassende Auswahl an technischen Maßnahmen vor, die möglichst viele Grundsätze der DSGVO in puncto technischem Datenschutz bedienen sollen. Ergänzend werden in Kapitel 4 organisatorische Maßnahmen für Datenschutz in relationalen Datenbanksystemen vorgestellt. Damit sollen alle technischen Implementierungen durch organisatorische Prozesse abgesichert werden, um Art. 32 DSGVO gerecht zu werden. Kapitel 5 schließt mit einer Zusammenfassung der vorgestellten Maßnahmen, gibt Antworten auf die eingangs gestellten Fragen und beleuchtet künftige Entwicklungsmöglichkeiten.

Implementierungen der vorgestellten Maßnahmen sind nicht Bestandteil dieser Arbeit. An einigen Stellen werden Ansätze für mögliche Implementierungen benannt, aber nicht weiter ausgeführt. Hierfür bietet sich Alternativliteratur an, die spezieller auf die vorgestellten Maßnahmen eingeht.

2 Grundlagen

Zunächst werden nachfolgend die wichtigsten in dieser Arbeit verwendeten Begriffe benannt und erläutert. Zum Verständnis der Arbeit wird darüber hinaus vorausgesetzt, dass in den Bereichen *Datenbanken* und *Datenschutz* ausreichend Vorwissen vorhanden ist, um einzelne vorgestellte Konzepte logisch miteinander verknüpfen zu können.

2.1 Terminologie

2.1.1 Relationale Datenbanksysteme

Ein relationales Datenbanksystem bildet strukturierte Daten in Form von Tabellen und Spalten ab. Tabellen repräsentieren jeweils einzelne Entitäten, wobei jede Zeile eine Instanz der Entität darstellt und jede Spalte ein bestimmtes Merkmal (Attribut) beschreibt (Hebing & Manhembué, 2024). Der Begriff *Relation* bezieht sich dabei „auf die relationale Algebra als den Teil der Mathematik [...], der sich wiederum mit Operationen auf Tabellen beschäftigt“ (Hebing & Manhembué, 2024, S. 38).

2.1.2 Structured Query Language (SQL)

Die *Structured Query Language (SQL)* ist eine Sprache zur Interaktion mit relationalen Datenbanken. Sie ermöglicht es, Daten auszulesen (`SELECT`), zu verändern (`UPDATE`) oder zu löschen (`DELETE`). Abfragen mittels SQL werden auch als *Statements* bezeichnet und enthalten konkrete Anweisungen an die Datenbank (Simon, 2023):

```
1 SELECT first_name, last_name
2 FROM employee
3 WHERE employed = 1;
```

Listing 1: Einfaches SQL-`SELECT`-Statement

2.1.3 NoSQL-Datenbanken

Im Gegensatz zu relationalen Datenbanksystemen bilden NoSQL-Datenbanken semistrukturierte und unstrukturierte Daten ab (Hebing & Manhembué, 2024). Die Bezeichnung steht umgangssprachlich für *Not-only-SQL (NoSQL)* und deutet auf die Verwendung nicht-relationaler Speicher- und Sprachfunktionen in Kombination mit SQL hin. Zu den verschiedenen Arten von NoSQL-Datenbanken zählen beispielsweise Key-Value-Speicher, Dokumentenspeicher oder auch Graphen-basierte Datenbanken (Kaufmann & Meier, 2023).

NoSQL-Datenbanken werden häufig mittels *Sharding* auf mehreren Servern verteilt. Dies ermöglicht eine horizontale Skalierung und damit eine hohe Erreichbarkeit und Flexibilität, insbesondere bei der Speicherung großer Datenmengen. Häufig wird dieses Verfahren in Dokumentenspeichern angewendet (Kaufmann & Meier, 2023).

2.1.4 Datenschutz-Grundverordnung (DSGVO)

Im Jahr 2016 verabschiedete die Europäische Union (EU) die sog. *Datenschutz-Grundverordnung (DSGVO)*. Sie sollte „zu mehr Rechtssicherheit innerhalb der EU führen und Hindernisse für den grenzüberschreitenden Austausch personenbezogener Daten beseitigen“ (Voigt & von dem Bussche, 2024, S. 2). Außerdem sollte damit „ein verantwortungsbewusster Umgang mit [...] personenbezogenen Daten sichergestellt werden“ (Voigt & von dem Bussche, 2024, S. 3).

Die Anforderungen der DSGVO an Unternehmen sind vielfältig. Sie reichen von der Rechenschaftspflicht über die Dokumentation und Haftung bei Einbindung eines Auftragverarbeiters bis hin zu den hierfür notwendigen technischen und organisatorischen Maßnahmen. Darüber hinaus legt die DSGVO weitgehende Rechte für betroffene Personen fest, beschreibt umfangreiche Auskunftsrechte und -pflichten und definiert die notwendige Zusammenarbeit mit Aufsichtsbehörden. Nicht zuletzt werden hohe Sanktionen benannt, die bei Nichteinhaltung einzelner Vorschriften fällig werden (Voigt & von dem Bussche, 2024).

Die DSGVO trat erstmals zwei Jahre nach ihrer Verabschiedung am 25. Mai 2018 in Kraft. Seitdem müssen alle Unternehmen, die im Geltungsbereich der Verordnung personenbezogene Daten verarbeiten, die Vorschriften der DSGVO einhalten.

2.2 Zustände von Daten in Datenbanksystemen

Damit eine Betrachtung möglicher technischer und organisatorischer Maßnahmen zum Schutz von Daten in Datenbanksystemen möglich ist, bedarf es zunächst einer Klärung, in welchen Zuständen diese Daten vorliegen können. Grundsätzlich lässt sich der Datenfluss und damit die Möglichkeit des Datenzugriffs in drei Kategorien aufteilen (nach Swanzy et al. (2024)):

1. **Data in motion** bzw. **data in transit** umfasst den Prozess des aktiven Datenaustausches über ein Netz von einem Ort zum anderen. Häufig erfolgt dies durch eine Übertragung zwischen *Client* und *Server*.

2. **Data in use** beschreibt die aktive Verarbeitung von Daten durch eine oder mehrere Anwendungen, zusätzlich zur passiven Speicherung auf einem Speichergerät.

3. **Data at rest** bezeichnet Daten, die nicht aktiv verarbeitet oder abgerufen werden. Häufig handelt es sich dabei um Backups oder archivierte Daten auf Servern, physischen Speichergeräten oder in der Cloud.

Alle genannten Zustände ermöglichen den Zugriff auf personenbezogene und teilweise sehr sensible Daten und müssen daher mittels geeigneter Datenschutzmaßnahmen effektiv abgesichert werden.

2.3 Unterschiedliche Ansätze zwischen relationalen Datenbanksystemen und NoSQL-Datenbanken

Aufgrund des unterschiedlichen strukturellen Aufbaus lassen sich geeignete technische und organisatorische Maßnahmen nicht gleichermaßen auf relationale Datenbanksysteme und NoSQL-Datenbanken anwenden. Techniken, die in relationalen Datenbanksystemen direkt auf einzelnen Spalten ausgeführt werden, erfordern in NoSQL-Datenbanken aufgrund ihres dynamischen Datenbankschemas andere Herangehensweisen (Sahatqija et al., 2018).

Darüber hinaus erfordert die häufig bei NoSQL-Datenbank eingesetzte *Sharding*-Technologie bei vielen Maßnahmen eine deutlich komplexere Prozessgestaltung. Techniken wie Zugriffskontrollen, Verschlüsselung oder Auditings müssen grundlegend anders aufgebaut werden als bei relationalen Datenbanksystemen, um über verteilte Systeme hinweg vollumfänglich anwendbar zu sein (Sahatqija et al., 2018).

In dieser Arbeit werden daher ausschließlich Techniken vorgestellt, die im Rahmen des Einsatzes relationaler Datenbanksysteme als geeignete technische und organisatorische Maßnahmen angewendet werden können. Maßnahmen für NoSQL-Datenbanken werden zum Beispiel bei Gharajeh (2017) und Frick et al. (2023) vorgestellt.

3 Technische Maßnahmen zum Datenschutz relationaler Datenbanken

Die DSGVO benennt zahlreiche Anforderungen zur Sicherstellung des Schutzes personenbezogener Daten. In Datenbanken werden häufig eine Vielzahl solcher personenbezogener Daten gespeichert, wobei es sich teilweise auch um die „Verarbeitung besonderer Kategorien personenbezogener Daten“ nach Art. 9 DSGVO handelt, die besonders schützenswert sind. In Art. 24 Abs. 1 und Art. 25 DSGVO wird daher die Umsetzung geeigneter technischer und organisatorischer Maßnahmen gefordert, um die speziellen Anforderungen der Verordnung zu erfüllen. Darüber hinaus benennt Erwägungsgrund (ErwG) 78 DSGVO die Verpflichtung, bei allen Umsetzungen die Datenschutzkonzepte *data protection by design* und *data protection by default* zu berücksichtigen.

Nachfolgend werden einige dieser Maßnahmen vorgestellt und erläutert, inwieweit sie auf die unterschiedlichen Anforderungen der DSGVO einzahlen.

3.1 Anonymisierung

„*Anonymisierung* ist eine Technik zur Veränderung personenbezogener Daten mit dem Ergebnis, dass keine Verbindung der Daten zu einer natürlichen Person (mehr) besteht“ (Voigt & von dem Bussche, 2024, S. 18). Vor allem bei der Speicherung großer Datenmengen kann Anonymisierung ein sinnvolles Mittel sein, um im Sinne der Datenminimierung diejenigen Daten zu anonymisieren, die für die weitere Verarbeitung nicht mehr notwendig sind.

3.1.1 Anonymisierung im Kontext der DSGVO

Wenn Daten vollständig anonymisiert wurden, unterliegen sie nicht mehr den strengen Anforderungen der DSGVO und können bedenkenlos weiterverwendet werden:

> „Die Grundsätze des Datenschutzes sollten [...] nicht für anonyme Informationen gelten, d.h. für Informationen, die sich nicht auf eine identifizierte oder identifizierbare natürliche Person beziehen, oder personenbezogene Daten, die in einer Weise anonymisiert worden sind, dass die betroffene Person nicht oder nicht mehr identifiziert werden kann.“
>
> (*ErwG 26 S. 5 DSGVO*)

3.1.2 Anonymisierungstechniken

Je nach Art der zu anonymisierenden Daten können unterschiedliche Techniken zum Einsatz kommen. Dies ist notwendig, um einen höchstmöglichen Grad der Anonymisierung zu erreichen, damit vollständig anonymisierte Daten keine Re-Identifikation der

Ursprungsdaten ermöglichen. Jeder Anonymisierung haftet ein gewisser Risikofaktor an, weshalb bei der Auswahl der Anonymisierungstechniken die „Schwere und Wahrscheinlichkeit des identifizierten Risikos Berücksichtigung finden“ muss (Voigt & von dem Bussche, 2024, S. 19).

Folgende Anonymisierungstechniken sind vorhanden und lassen sich auf in relationalen Datenbanken gespeicherte personenbezogene Daten anwenden (nach Marques und Bernardino (2020), Pina et al. (2024) und Voigt und von dem Bussche (2024)):

- **Randomisierung**: Bei der auch als *Shuffling* oder *Scrambling* bezeichneten Technik werden Daten nach dem Zufallsprinzip gemischt oder neu angeordnet. Die Werte der ursprünglichen Attribute verbleiben dabei in der Datenbank, können aber ihre Assoziation zum ursprünglichen Datensatz verlieren.

- **Generalisierung**: Mithilfe dieser Technik werden Daten verallgemeinert bzw. verwässert, indem einzelne Attribute bezüglich ihres Maßstabes, ihrer Größenordnung oder ihres Bezugspunktes verändert werden. Hierzu existieren unterschiedliche Ansätze:

  - k-Anonymität: Nach diesem Modell werden k Individuen in Kategorien zusammengefasst, um denselben Kombinationen zu entsprechen.
  - l-Diversität: Als Erweiterung der k-Anonymität stellt diese Technik sicher, dass jede gleichwertige Gruppe und jedes sensible Attribut mindestens l unterschiedliche Werte annimmt, sodass eine ausreichend hohe Variabilität der Attribute gewährleistet ist.

- **Löschung**: Wenn Attribute als irrelevant angesehen werden können, ist eine Entfernung aus dem Datensatz häufig ein geeignetes Mittel zur Anonymisierung. Diese Technik wird auch als *Suppression* bezeichnet.

- **Maskierung**: Einzelne Zeichen können mittels Maskierung durch neutrale Zeichen wie X oder $*$ ersetzt werden, um die Daten effizient zu anonymisieren.

- **Datenstörung/Perturbation**: Die auch als *Noise Addition* bezeichnete Technik verändert einzelne Werte um ein gewisses Maß (*noise level*), um die Präzision der Attribute zu verringern und damit eine Re-Identifikation der Ursprungswerte zu erschweren. Je nach *noise level* kann diese Technik mehr oder weniger wirksam ausfallen.

- **Aggregation**: Durch Standardisierung und Gruppierung werden bei dieser Technik ähnliche Daten zu zusammengefassten Versionen mit weniger Attributen verdichtet, wodurch sie im Vergleich zur Generalisierung aktiv verändert werden.

3.2 Pseudonymisierung

Anders als bei der Anonymisierung von Daten besteht das Ziel der *Pseudonymisierung* darin, dass „Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können“ (Art. 4 Abs. 5 DSGVO). Einzelne Daten oder Attribute werden dabei nicht gänzlich entfernt, sondern durch bestimmte Angaben ersetzt. Die zur vollständigen Identifikation notwendigen zusätzlichen Informationen müssen dabei gesondert aufbewahrt und durch ergänzende Maßnahmen – beispielsweise Verschlüsselung – gesichert werden (Voigt & von dem Bussche, 2024).

Zum Beispiel können sensible Daten wie Name und Geburtsdatum mit einem Hashwert versehen werden. Der Hashwert wird zusätzlich mit den verschlüsselten Daten separat gespeichert, um eine spätere Zuordnung durch autorisierte Personen zu ermöglichen. Durch die Umwandlung in einen Hashwert können die pseudonymisierten Informationen nicht mehr direkt mit einer bestimmten Personen verknüpft werden; zur Identifikation benötigt es die separat gespeicherten Informationen (Pina et al., 2024).

3.2.1 Pseudonymisierung im Kontext der DSGVO

Grundsätzlich lässt sich feststellen, dass Daten, die mithilfe von Anonymisierungstechniken verändert wurden, aber weiterhin eine Re-Identifikation ermöglichen, nicht anonymisiert, sondern lediglich pseudonymisiert wurden (Marques & Bernardino, 2020). Da das Risiko einer Re-Identifikation bei pseudonymisierten Daten entsprechend höher ist als bei anonymisierten Daten, fallen sie weiterhin in den Anwendungsbereich der DSGVO (Voigt & von dem Bussche, 2024):

> „Die Anwendung der Pseudonymisierung auf personenbezogene Daten kann die Risiken für die betroffenen Personen senken und die Verantwortlichen und die Auftragsverarbeiter bei der Einhaltung ihrer Datenschutzpflichten unterstützen. Durch die ausdrückliche Einführung der „Pseudonymisierung“ in dieser Verordnung ist nicht beabsichtigt, andere Datenschutzmaßnahmen auszuschließen.“
>
> (*ErwG 28 DSGVO*)

3.2.2 Pseudonymisierungstechniken

Für die Pseudonymisierung von Daten stehen ebenfalls eine Reihe unterschiedlicher Techniken zur Verfügung. Einzelne Pseudonymisierungstechniken ähneln dabei stark den in Abschnitt 3.1.2 vorgestellten Anonymisierungstechniken. Folgende Techniken können dabei zur Anwendung kommen (nach Jensen et al. (2019) und Pikulík und Štarchoň (2019)):

- **Counter**: Bei dieser simplen Technik wird eine Variable mit einem Wert – beispielsweise `0` – initialisiert und für jeden zu pseudonymisierenden Datenpunkt

hochgezählt. Die Daten werden dann durch die entsprechenden Variablen ersetzt, wobei sich die Werte nicht wiederholen dürfen, um einen eindeutigen Bezug herstellen zu können. Da zur Re-Identifikation eine Mapping-Tabelle erstellt werden muss, eignet sich diese Technik eher für kleinere Datenbanken.

- **Zufallszahlen**: Ähnlich wie bei der Counter-Technik werden durch einen *Random Number Generator (RNG)* Zahlen erzeugt und einzelnen Datenpunkten zugewiesen. Der Unterschied liegt darin, dass die Zahlen zufällig gewählt werden – je nach Verfahren durch einen echten Zufallszahlengenerator oder einen kryptografischen Pseudo-Zufallszahlengenerator (*CPRNG*). Die Zahlen dienen als Pseudonym und müssen in einer Mapping-Tabelle gesichert werden.

- **Scrambling**: *vgl. „Randomisierung" unter Anonymisierungstechniken*

- **Verschlüsselung**: Durch Verschlüsselung veränderte Daten stellen ein nach Art. 32 Abs. 1 DSGVO mögliches Konzept der Pseudonymisierung von Daten dar. Wichtig ist hierbei, dass der Schlüssel zur Entschlüsselung getrennt von den pseudonymisierten Daten aufbewahrt wird.

- **Maskierung**: *siehe Anonymisierungstechniken*

- **Tokenisierung**: Bei dieser Technik werden sensible durch nicht sensible Daten (sog. *Token*) ersetzt, die keine besondere Bedeutung oder Wert haben. Länge und Typ der Daten bleiben dabei unverändert, weshalb diese Technik anfällig für auf Längen- und Typmerkmale sensibilisierte Systeme ist.

- **Blurring**: Hierbei werden Näherungswerte der ursprünglichen Daten verwendet, um ihre ursprüngliche Bedeutung zu verschleiern. Dadurch kann eine Re-Identifikation ebenfalls erschwert werden, wenngleich die Technik als nicht allzu sicher gilt.

- **Hashing**: Wie anfangs beispielhaft beschrieben, kann auch die Berechnung von Hashwerten eine nützliche Pseudonymisierungstechnik sein. Hierzu werden kryptografische Hashfunktionen eingesetzt.

- **Message Authentication Code (MAC)**: Bei dieser Methode wird ein zusätzlicher Geheimschlüssel für die Generierung des Pseudonyms verwendet. Als Alternative zum Hashing wird hierzu häufig der sog. *Hash-based Message Authentication Code (HMAC)* eingesetzt.

3.3 Verschlüsselung

Verschlüsselungstechniken zählen in vielen Bereichen der IT-Sicherheit zu einem gängigen Mittel, um Vertraulichkeit und Integrität von Daten sicherzustellen. Auch im

Bereich des Datenschutzes in relationalen Datenbanken ist es eine essenzielle Technik, um sensible Informationen zu schützen.

3.3.1 Verschlüsselung im Kontext der DSGVO

In der DSGVO wird Verschlüsselung ebenfalls als wichtiges Mittel zur sicheren und rechtmäßigen Verarbeitung personenbezogener Daten aufgeführt:

> „Zur Aufrechterhaltung der Sicherheit und zur Vorbeugung gegen eine gegen diese Verordnung verstoßende Verarbeitung sollte der Verantwortliche oder der Auftragsverarbeiter die mit der Verarbeitung verbundenen Risiken ermitteln und Maßnahmen zu ihrer Eindämmung, wie etwa eine Verschlüsselung, treffen."
>
> (*ErwG 83 DSGVO*)

In Art. 34 Abs. 3(a) DSGVO wird darüber hinaus ausgeführt, dass mittels Verschlüsselung gesicherte personenbezogene Daten bei einer Datenschutzverletzung keine Benachrichtigung der betroffenen Personen erfordern. Das wiederum erläutert die Wichtigkeit und Tragweite beim Einsatz von Verschlüsselungstechnologien.

3.3.2 Wahl des Verschlüsselungsalgorithmus

Die Wahl eines geeigneten Verschlüsselungsalgorithmus und damit die Wirksamkeit der Verschlüsselung hängt von mehreren Faktoren ab. Zum einen erfordert sie eine sorgfältige Bewertung der für die Verschlüsselung infrage kommenden Datentypen, wie etwa *Personally Identifiable Information (PII)* oder *Protected Health Information (PHI)*. Darüber hinaus sollte der gewählte Algorithmus ein Gleichgewicht zwischen Sicherheit und Performance herstellen, um einerseits die Daten größtmöglich abzusichern, andererseits aber auch praktikabel nutzbar zu sein. Nicht zuletzt ist es von essenzieller Bedeutung, auf welcher Ebene der Verschlüsselungsalgorithmus eingesetzt wird (Pina et al., 2024).

3.3.3 Verschlüsselungsebenen in Datenbanksystemen

Je nach Angriffspunkt müssen Verschlüsselungsverfahren an unterschiedlichen Ebenen in Datenbanksystemen implementiert werden. Grundsätzlich lassen sich diese Ebenen wie folgt unterteilen (nach Shmueli et al. (2009)):

- **Dateisystemebene**: Der physische Speicher der Datenbank kann durch Verschlüsselung des Dateisystems gänzlich abgesichert werden. Da die gesamte Datenbank in diesem Fall mit nur einem einzigen Schlüssel wieder entschlüsselt werden kann, ist bei dieser Methode allerdings keine individuelle Vergabe von Zugangsberechtigungen möglich.

- **Datenbankmanagementsystem (DBMS)-Ebene**: Eine direkte Verschlüsselung einzelner Datenbanktabellen, -zeilen oder -spalten ist ebenfalls möglich, erfordert aber die Wahl eines ausgeklügelten Verfahrens, um ein gutes Gleichgewicht zwischen Sicherheit und Performance zu gewährleisten. Beispielsweise existieren Verfahren zur Verschlüsselung auf Zeilenebene und Entschlüsselung auf Zellebene oder alternativ die spalten- oder zeilenorientierte Verschlüsselung mithilfe der *Rivest–Shamir–Adleman (RSA)*-Verschlüsselung.

  Das DBMS *PostgreSQL* stellt zum Beispiel mit der *pgcrypto*-Erweiterung eine Vielzahl kryptografischer Datenbankfunktionen zur Verfügung. Neben Ver- und Entschlüsselungsfunktionen bietet die Erweiterung auch Funktionen zum Hashing oder Umgang mit *Pretty Good Privacy (PGP)*-verschlüsselten Daten.

- **Anwendungsebene**: Auf Anwendungsseite kann mithilfe einer *Web Data Service Provider (WDSP)*-Middleware jede Datenbankabfrage so übersetzt werden, dass sie von einem verschlüsselten DBMS ausgeführt werden kann. Dadurch entsteht ein regelbasierter Übertragungsweg direkt auf der Transportschicht, um den Zugriff auf sichere Daten noch früher zu reglementieren.

3.4 Zugriffskontrollen

Eine weitere wichtige Maßnahme für Datenschutz in relationalen Datenbanksystemen stellen Zugriffskontrollen dar. Dabei handelt es sich allerdings weniger um konkrete Maßnahmen für den Datenschutz, sondern eher um ein Verfahren im Bereich der Datensicherheit. Durch vorkonfigurierte Kontrollmechanismen soll sichergestellt werden, dass bestimmte Bereiche von Datenbanken – also etwa ausgewählte Datenbanktabellen oder -spalten – nur von autorisierten Datenbanknutzern les- und schreibbar sind. Damit Anwender auf diese Bereiche zugreifen können, müssen sie sich mittels eines konfigurierten Datenbankbenutzers autorisieren und erhalten dadurch die notwendigen Zugriffsrechte.

3.4.1 Zugriffskontrollen im Kontext der DSGVO

In der DSGVO ist aufgeführt, dass geeignete Vorkehrungen getroffen werden müssen, um den Zugang zu personenbezogenen Daten einzuschränken, damit Datenschutz gewährleistet ist. Dort heißt es:

> **Sicherheit der Verarbeitung**
>
> „Bei der Beurteilung des angemessenen Schutzniveaus sind insbesondere die Risiken zu berücksichtigen, die mit der Verarbeitung verbunden sind, insbesondere durch – ob unbeabsichtigt oder unrechtmäßig – Vernichtung,

Verlust, Veränderung oder unbefugte Offenlegung von beziehungsweise unbefugten Zugang zu personenbezogenen Daten, die übermittelt, gespeichert oder auf andere Weise verarbeitet wurden."

(*Art. 32 Abs. 2 DSGVO*)

3.4.2 Arten von Zugriffskontrollen

Je nach gewähltem DBMS stehen unterschiedliche Formen für die Durchführung von Zugriffskontrollen zur Verfügung (nach Bertino und Sandhu (2005)):

- **Discretionary Access Control (DAC)**: Bei dieser Form der Zugriffskontrolle werden Lese- und Schreibrechte basierend auf den Identitäten von Subjekt (Benutzer) und Objekt (Daten) vergeben. Bei jedem Datenzugriff wird auf das Vorhandensein entsprechender Berechtigungen geprüft und anhand der Zugriffsparameter entschieden, ob ein Zugriff gestattet wird.

  Die einzelnen Zugriffsregeln werden je nach gewähltem Administrationsverfahren (sog. *authorization administration policy*) unterschiedlich vergeben:

  - *Zentralisierte Verwaltung*: Lediglich ausgewählte privilegierte Benutzer haben die Möglichkeit, Zugriffsrechte zu vergeben und zu entziehen.
  - *Eigentümer-Verwaltung*: Der Eigentümer bzw. Ersteller von Daten kann Zugriffsrechte verteilen und entziehen. Er kann diese Administrationsrechte auch an weitere Benutzer im Rahmen der *administration delegation* weitergeben, wodurch eine dezentralisierte Verwaltung geschaffen wird.

  Die Art der diskretionären Rechtevergabe macht diese Form der Zugriffskontrolle zu einem flexiblen Mechanismus, der heutzutage von vielen DBMS eingesetzt wird.

- **Mandatory Access Control (MAC)**: Diese Form der Zugriffskontrolle basiert auf vordefinierten Klassifizierungen von Subjekten und Objekten. Dabei erhält jeder Benutzer eine Zugriffskennzeichnung, die aus einer Kategorie und einer zugehörigen Sicherheitsstufe besteht. Die Zugriffskennzeichnungen sind hierarchisch angeordnet und ermöglichen eine feingranulare Zugriffsregelung, indem ein Benutzer nur auf Daten zugreifen kann, deren Klassifizierung seine eigene Sicherheitsstufe nicht übersteigt. Dabei werden im Wesentlichen zwei grundlegende Prinzipien angewandt, die verhindern sollen, dass sensible Informationen in niedrigere Sicherheitsstufen gelangen:

  - *No read-up*: Einem Subjekt wird kein Lesezugriff auf höher klassifizierte Objekte ermöglicht.

– *No write-down*: Einem Subjekt wird kein Schreibzugriff auf niedriger klassifizierte Objekte ermöglicht.

Für die Umsetzung in relationalen Datenbanksystemen werden häufig sog. *Multi-level security (MLS)*-Modelle eingesetzt, mithilfe derer verschiedenen Tabellenzeilen unterschiedliche Sicherheitsstufen zugeordnet werden können. Dadurch entstehen komplexe Zugriffs- und Verwaltungsszenarien.

- **Role Based Access Control (RBAC)**: Bei dieser weit verbreiteten Form der Zugriffskontrolle wird der Zugriff auf einzelne Daten durch Rollen bestimmt. Eine Rolle bildet meist eine bestimmten Funktion innerhalb einer Organisation ab und umfasst dabei eine Menge von Aufgaben und Verantwortlichkeiten, die dieser Funktion zugeordnet sind. Benutzer, die Mitglied einer bestimmten Rolle sind, erhalten automatisch alle dieser Rolle zugewiesenen Berechtigungen, wodurch eine direkte Zuweisung von Berechtigungen an einzelne Benutzer vermieden wird.

3.4.3 Anwendungsbeispiel

Im nachfolgenden Beispiel werden in einer relationalen *PostgreSQL*-Datenbank Rollen nach dem RBAC-Modell vergeben. Das DBMS stellt hierzu geeignete `CREATE ROLE`-Statements zur Verfügung:

```
1 CREATE ROLE admin WITH LOGIN SUPERUSER CREATEROLE;
2 CREATE ROLE employee WITH LOGIN;
```

Listing 2: Anlegen von Rollen in PostgreSQL

Den erstellten Rollen sind zunächst keine besonderen Berechtigungen auf Ebene des Datenbankschemas erteilt. Diese können nun mithilfe von `GRANT`-Statements einzeln zugewiesen werden:

```
4 GRANT ALL PRIVILEGES ON ALL TABLES IN SCHEMA public TO admin;
5 GRANT SELECT, INSERT, UPDATE ON customers TO employee;
```

Listing 3: Zuweisung von Rollenberechtigungen in PostgreSQL

Analog dazu können einzelnen Rollen jederzeit bestimmte Berechtigungen wieder entzogen werden. Hierfür stellt PostgreSQL `REVOKE`-Statements bereit:

```
7 REVOKE INSERT ON customers FROM employee;
```

Listing 4: Entzug von Rollenberechtigungen in PostgreSQL

3.5 Sicherung & Wiederherstellung

Alle zuvor genannten technischen Maßnahmen spielen eine wichtige Rolle für den Betrieb relationaler Datenbanksysteme. Damit die Verfügbarkeit und Integrität dieser

Systeme gewährleistet werden kann, führen Unternehmen darüber hinaus Sicherungs- und Wiederherstellungsprozesse durch, bei denen ebenfalls sichergestellt werden muss, dass sie datenschutzkonform gestaltet sind.

3.5.1 Sicherung & Wiederherstellung im Kontext der DSGVO

Prozesse zur Sicherung und Wiederherstellung von Daten werden in der DSGVO nicht explizit gefordert. Jedoch sollen geeignete technische und organisatorische Maßnahmen umgesetzt werden, um „[...] die Vertraulichkeit, Integrität, Verfügbarkeit und Belastbarkeit der Systeme und Dienste im Zusammenhang mit der Verarbeitung auf Dauer sicherzustellen“ (Art. 32 Abs. 1(c) DSGVO). Darüber hinaus werden angemessene Speicher- und Löschfristen gefordert, um personenbezogene Daten nicht länger als notwendig vorzuhalten (vgl. ErwG 39 DSGVO).

3.5.2 Datenschutzkonforme Speicherung von Datenbank-Backups

Alle bisher genannten technischen Maßnahmen stellen auch für die Speicherung von Datenbank-Backups datenschutzkonforme Lösungen dar. Es ist allerdings darauf zu achten, dass Daten in einer Form gespeichert werden, die eine spätere Wiederherstellung der Ursprungsdaten möglich machen. Dies ist etwa bei Anonymisierung und Pseudonymisierung nicht immer möglich, weshalb diese Techniken unter Umständen nur auf Daten angewendet werden sollten, die nicht direkt zur Wiederherstellung spezifischer Identitätsmerkmale benötigt werden.

Darüber hinaus existieren unter anderem folgende Maßnahmen für datenschutzkonforme Datenbank-Backups:

- **Datenlöschkonzept**: Nach Art. 17 DSGVO steht betroffenen Personen ein Recht auf Löschung ihrer personenbezogenen Daten zu. Die hierfür bereitgestellten Löschkonzepte gelten gleichermaßen auch für Datenbank-Backups. Unternehmen müssen sicherstellen, dass Daten nicht länger als notwendig vorgehalten werden (vgl. Art. 5 Abs. 1(e) DSGVO) und nach Ablauf der notwendigen Aufbewahrungsdauer automatisch aus Datenbank-Backups gelöscht werden.

- **Wiederherstellbarkeit**: Es müssen regelmäßige Integritätsprüfungen durchgeführt werden, um sicherzustellen, dass Datenbank-Backups vollständig wiederhergestellt werden können.

- **Datenverschleierung**: Der auch als *data obfuscation* zusammengefasste Überbegriff von Maßnahmen zur Verschleierung personenbezogener Daten findet vor allem bei Datenbank-Backups, die für Test- und Entwicklungszwecke verwendet werden, Anwendung. Ziel dieser Maßnahmen ist der temporäre Schutz personenbezogener Daten, wenn diese für bestimmte Zwecke wie die Weiterentwicklung einer zugehörigen Software-Anwendung nicht von Relevanz sind.

4 Organisatorische Maßnahmen für relationale Datenbanksysteme

Mithilfe geeigneter technischer Maßnahmen werden bereits viele Anforderungen der DSGVO an einen datenschutzkonformen Betrieb relationaler Datenbanksysteme erfüllt. Die DSGVO schreibt jedoch explizit vor, dass auch geeignete organisatorische Maßnahmen für einen datenschutzkonformen Betrieb erforderlich sind:

> **Verantwortung des für die Verarbeitung Verantwortlichen**
>
> „Der Verantwortliche setzt [...] geeignete technische und organisatorische Maßnahmen um, um sicherzustellen und den Nachweis dafür erbringen zu können, dass die Verarbeitung gemäß dieser Verordnung erfolgt."
>
> (*Art. 24 Abs. 1 DSGVO*)

Nachfolgend werden daher zwei organisatorische Maßnahmen vorgestellt, die speziell für den Betrieb relationaler Datenbanksysteme von Relevanz sind.

4.1 Prüf- & Löschkonzepte

4.1.1 Regelmäßige Überprüfung der Wirksamkeit von Maßnahmen

Art. 32 Abs. 1(d) DSGVO fordert „ein Verfahren zur regelmäßigen Überprüfung, Bewertung und Evaluierung der Wirksamkeit der technischen und organisatorischen Maßnahmen zur Gewährleistung der Sicherheit der Verarbeitung". Dies ist auch im Rahmen des Betriebs relationaler Datenbanksysteme von Bedeutung.

So ist es etwa notwendig, die mittels RBAC zugewiesenen Rollen und Rollenberechtigungen regelmäßig zu überprüfen, um unnötige oder veraltete Berechtigungen zu entfernen. Dies trägt zum Datenschutz bei, indem der Zugang zu den Daten aktuell gehalten und stets auf das notwendige Minimum reduziert wird (*Principle of least privilege*).

Solche Konzepte können auch die Umwandlung einzelner Berechtigungen in gruppenbasierte Berechtigungen beinhalten. Dadurch erhalten einzelne Benutzer keine individuellen Rechte mehr zugewiesen, sondern werden in Gruppen mit gemeinsamen Aufgabenbereichen zusammengefasst. Regelmäßige Evaluationen dieser Gruppenberechtigungen stellen sicher, dass nur autorisierte Benutzer Zugriff auf sensible Daten erhalten.

Prüfkonzepte zur fortlaufenden Evaluation eingesetzter Maßnahmen können auch im Rahmen eines *Plan–do–check–act (PDCA)-Zyklus* umgesetzt werden. Dieser kann im Rahmen eines übergeordneten Datenschutz-Managementsystems eingesetzt werden, das dazu beiträgt, „angemessene technische und organisatorische Maßnahmen zur Erreichung des von der DSGVO vorgesehenen Datenschutzniveaus umzusetzen" (Voigt & von dem Bussche, 2024, S. 53).

4.1.2 Konzepte zur Prüfung auf Einhaltung definierter Löschfristen

Wie bereits in Abschnitt 3.5.2 vorgestellt, ist es insbesondere bei Datenbank-Backups von besonderer Bedeutung, Löschkonzepte zur Erreichung der in Art. 5 Abs. 1(e) DSGVO geforderten Speicherbegrenzung zu implementieren. Darüber hinaus heißt es:

> „Um sicherzustellen, dass die personenbezogenen Daten nicht länger als nötig gespeichert werden, sollte der Verantwortliche Fristen für ihre Löschung oder regelmäßige Überprüfung vorsehen. Es sollten alle vertretbaren Schritte unternommen werden, damit unrichtige personenbezogene Daten gelöscht oder berichtigt werden."
>
> (*ErwG 39 S. 10-11 DSGVO*)

Eine organisatorische Maßnahme für die Erfüllung eines geeigneten Löschkonzepts sieht vor, dass in regelmäßigen Abständen geprüft wird, „ob die Speicherdauer und die damit verbundenen Löschfristen der verarbeiteten Daten eingehalten werden" (Voigt & von dem Bussche, 2024, S. 275). Als Orientierung hat das *Deutsches Institut für Normung e. V. (DIN)* die Leitlinie *DIN 66398 („Leitlinie zur Entwicklung eines Löschkonzepts mit Ableitung von Löschfristen für personenbezogene Daten")* entwickelt, das auch für den Einsatz in relationalen Datenbanksystemen tauglich ist (Voigt & von dem Bussche, 2024). Gemäß dieser Leitlinie sollten folgende Aspekte bei der Entwicklung organisatorischer Maßnahmen beachtet werden (nach Hammer (2016)):

- **Löschregeln**: Es müssen Regeln gebildet werden, die definieren, welche Daten zu welchem Zeitpunkt gelöscht werden sollen. Hierzu werden sogenannte *Löschklassen* verwendet, die je nach Art der gespeicherten Daten genau festlegen, wann eine Löschfrist erreicht wird (sog. *Regellöschfrist*).
- **Umsetzungsvorgaben**: Damit Löschregeln auch tatsächlich anhand konkreter Maßnahmen umgesetzt und Verantwortliche zur Löschung verpflichtet werden, sieht die Leitlinie die Definition bestimmter Vorgaben zur Implementierung von Löschkonzepten vor. Darüber hinaus enthält sie Vorschläge zum Umgang mit Ausnahmesituationen, wenn etwa Daten länger benötigt werden oder technische Störungen auftreten.
- **Verantwortlichkeiten**: Um die Kontinuität und Stabilität eines einmal erstellten Löschkonzepts zu gewährleisten, müssen für einzelne Aufgaben Verantwortlichkeiten festgelegt werden. Darunter fallen etwa die Pflege der einzelnen Löschregeln sowie die Entwicklung und Weiterführung der Umsetzungsvorgaben.

4.2 Audit-Trails

Ergänzend zu den vorausgehend genannten Konzepten wird in Art. 30 DSGVO („Verzeichnis von Verarbeitungstätigkeiten“) die Notwendigkeit eines Verarbeitungsverzeichnisses benannt. Dieses soll alle Verarbeitungstätigkeiten auflisten, die im Rahmen gespeicherter personenbezogener Daten durchgeführt werden, und damit Maßnahmen zur Ausübung der ebenfalls geforderten Rechenschaftspflicht gemäß Art. 5 Abs. 2 DSGVO aufzeigen.

Als technische Maßnahme im Kontext relationaler Datenbanken bietet sich hierfür die Implementierung sog. *Audit-Trails* an; in PostgreSQL-Datenbanken kann dies zum Beispiel durch den Einsatz der *pgaudit*-Extension erfolgen. Ergänzend zu dieser technischen Maßnahme können darüber hinaus zum Beispiel folgende organisatorische Maßnahmen umgesetzt werden:

- **Überprüfung & Auswertung**: Es muss einen organisatorischen Prozess zur regelmäßigen Überprüfung und Auswertung der Audit-Trails geben. Dieser legt fest, inwiefern Protokolle analysiert werden und wie häufig dies geschehen soll. Ziel ist es, frühzeitig unautorisierte Zugriffe oder sicherheitsrelevante Anomalien zu erkennen und durch entsprechende Maßnahmen darauf zu reagieren.

- **Incident Response**: Werden sicherheitsrelevante Anomalien in Audit-Trails erkannt, müssen durch definierte organisatorische Maßnahmen geeignete Incident-Response-Prozesse angestoßen werden. Dies umfasst etwa Meldungen an die zuständige Datenschutzbehörde (vgl. Art. 33 DSGVO) und betroffene Personen, wenn dies ein hohes Risiko für ihre Rechte und Freiheiten darstellt (vgl. Art. 34 DSGVO).

- **Zugriffsrechte**: Gemäß Art. 30 Abs. 1(d) DSGVO muss genau dokumentiert werden, welche Personengruppen Zugriff auf personenbezogene Daten erhalten. Dies gilt analog für den Zugriff auf Audit-Trails, der nur autorisierten Personen möglich sein soll. Eine organisatorische Maßnahme besteht daher in der kontinuierlichen Verwaltung und Überprüfung dieser Zugriffsrechte.

5 Zusammenfassung

5.1 Ergebnisse

Die DSGVO fordert umfangreiche technische und organisatorische Maßnahmen, um personenbezogene Daten zu schützen. Um welche Maßnahmen es sich dabei konkret handelt, wird größtenteils offen gelassen. Entsprechend groß ist die Herausforderung, geeignete Maßnahmen zu erarbeiten und nach DSGVO-Konformität umzusetzen.

Im Bereich relationaler Datenbanksysteme gibt es bereits eine Vielzahl an möglichen technischen Maßnahmen, die häufig auch schon vor Inkrafttreten der DSGVO in vielen Organisationen Anwendung fanden. Anonymisierungs- und Pseudonymisierungstechniken stellen etwa bekannte Verfahren dar, mit denen personenbezogene Daten entweder vollständig (durch Anonymisierung) oder teilweise (durch Pseudonymisierung) unkenntlich gemacht werden können. Dies ist vor allem im Sinne der Datenminimierung, Speicherbegrenzung und Zweckbindung von Relevanz und wirkt sich beim Einsatz von Datenbank-Backups noch einmal in deutlicherer Form aus. Darüber hinaus stellen Verschlüsselungsmechanismen und Zugriffskontrollen wichtige Maßnahmen zur Gewährleistung der Vertraulichkeit personenbezogener Daten dar. In der DSGVO werden beide Techniken explizit als mögliche Maßnahmen benannt. In relationalen Datenbanksystemen existieren hierfür bereits vielfältige Technologien, wie etwa *Role Based Access Control (RBAC)* in DBMS wie *PostgreSQL*.

Viele der technischen Maßnahmen lassen sich durch organisatorische Maßnahmen ergänzen oder wirken überhaupt erst durch diese zusätzlichen Prozesse. So bieten sich beispielsweise Konzepte zur regelmäßigen Überprüfung der Wirksamkeit technischer Maßnahmen an. Beim RBAC ist es etwa notwendig, die zugewiesenen Rollen regelmäßig neu zu evaluieren, um auch dem *principle of least privilege* gerecht zu werden. Außerdem sollte die Erarbeitung von Löschkonzepten mit definierten Löschregeln und -fristen durchgeführt werden, damit Daten nicht länger als nötig gespeichert werden. Die *DIN 66398*-Norm stellt hierfür einen umfangreichen Leitfaden zur Verfügung, dessen Anwendung bereits viele der geforderten Maßnahmen abdeckt. Speziell für relationale Datenbanken bietet sich darüber hinaus die Implementierung sog. *Audit-Trails* an, um alle Zugriffe auf personenbezogene Daten überwachen zu können. Ergänzende organisatorische Maßnahmen beinhalten hierbei die regelmäßige Auswertung der Audit-Trails sowie Dokumentationen über deren individuelle Zugriffsrechte und möglicherweise notwendige Incident-Response-Prozesse.

5.2 Herausforderungen

Die vorgestellten Techniken und Verfahren zeigen einmal mehr auf, wie wichtig der Einsatz technischer und organisatorischer Maßnahmen für den DSGVO-konformen Betrieb relationaler Datenbanksysteme ist. Techniken, die nach der alten Fassung des BDSG

Anwendung fanden, fallen mitunter nicht unter die strengen Regularien der DSGVO. Eine erneute Evaluierung bisher eingesetzter Maßnahmen ist insofern zwingend notwendig. Gerade bei komplexeren Datenbanksystemen, in denen unterschiedliche Formen personenbezogener Daten gespeichert werden, kann dies eine große Herausforderung darstellen. Dabei ist genau abzuwägen, welche der möglichen Maßnahmen jeweils für einzelne Datenpunkte notwendig sind.

Im Sinne eines strikten Datenschutzes können diese Herausforderungen aber auch eine Chance sein. Eine erneute Evaluierung eingesetzter Maßnahmen kann dazu beitragen, Potenziale zur Datenminimierung und Speicherbegrenzung auszuloten, um etwa gespeicherte, aber tatsächlich nicht genutzte personenbezogene Daten DSGVO-konform zu löschen und künftige Datenspeicherungen zu begrenzen. Im Zuge dieser Maßnahmen kann außerdem eine umfangreiche Dokumentation der eingesetzten Speichermechanismen und Datenkategorien einhergehen, womit automatisch der Weg für bestimmte organisatorische Maßnahmen geebnet wird, um etwa das Auskunftsrecht der betroffenen Personen nach Art. 15 DSGVO zu erfüllen.

5.3 Ausblick

Auch wenn bereits viele Techniken existieren, um wirksame technische und organisatorische Maßnahmen in relationalen Datenbanksystemen umzusetzen, so lohnt sich auch ein Blick auf neue Technologien.

Mit Inkrafttreten der DSGVO stellten Di Cerbo und Trabelsi (2018) im Jahr 2018 ein Verfahren zur automatisierten Erkennung und Anonymisierung personenbezogener Daten mithilfe von *Machine Learning (ML)* vor. Vor allem die erhöhte Flexibilität und Robustheit sowie die vereinfachte Wartung von ML-Modellen sollen dabei helfen, neue Datenkategorien personenbezogener Daten zu erkennen und direkte Maßnahmen wie Anonymisierung oder Pseudonymisierung durchzuführen.

Insbesondere für ältere Datenbanksysteme, bei denen Tabellen noch nicht vollständig normalisiert oder personenbezogene Daten über mehrere Tabellen ohne erkennbare Beziehungen verteilt sind, stellt das Tool *GDPRizer* eine Möglichkeit zur Optimierung hinsichtlich eines datenschutzkonformen Betriebs dar. Das Tool wurde im Jahr 2021 von Agarwal et al. (2021) vorgestellt und ist in der Lage, automatisch Beziehungen zwischen einzelnen Spalten zu identifizieren und damit bei der in älteren Datenbanken häufig notwendigen manuellen und fehleranfälligen Nacharbeit zur Einhaltung der DSGVO zu unterstützen.

Literaturverzeichnis

Agarwal, A., George, M., Jeyaraj, A., & Schwarzkopf, M. (2021). Retrofitting GDPR compliance onto legacy databases. *Proceedings of the VLDB Endowment*, *15*(4), 958–970.

Bertino, E., & Sandhu, R. (2005). Database Security-Concepts, Approaches, and Challenges. *IEEE Transactions on Dependable and Secure Computing*, *2*(1), 2–19.

Di Cerbo, F., & Trabelsi, S. (2018). Towards Personal Data Identification and Anonymization Using Machine Learning Techniques. *New Trends in Databases and Information Systems*, 118–126.

Frick, D., Kaufmann, J., & Lankes, B. (2023). *Big Data? Frag doch einfach: Klare Antworten aus erster Hand* (1. Aufl.). utb GmbH.

Gharajeh, M. S. (2017). Security Issues and Privacy Challenges of NoSQL Databases. In *NoSQL* (1. Aufl., S. 271–290). CRC Press.

Hammer, V. (2016). DIN 66398: Die Leitlinie Löschkonzept als Norm.

Hebing, M., & Manhembué, M. (2024). *Data Science Management: Vom Ersten Konzept Bis Zur Governance Datengetriebener Organisationen* (1. Aufl.). O'Reilly Verlag GmbH & Co. KG.

Jensen, M., Lauradoux, C., & Limniotis, K. (2019). *Pseudonymisation techniques and best practices* (Techn. Ber.). European Union Agency for Cybersecurity (ENISA).

Kaufmann, M., & Meier, A. (2023). *SQL and NoSQL Databases: Modeling, Languages, Security and Architectures for Big Data Management* (Second edition). Springer.

Krämer, A., & Mauer, R. (2023). Anwendungsfall G: Aufbau und Absicherung von Datenbanken mit personalisierten Daten. In *Datenschutz Für Entscheider in Marketing und Vertrieb*. Springer Vieweg. in Springer Fachmedien Wiesbaden GmbH.

Marques, J. F., & Bernardino, J. (2020). Analysis of Data Anonymization Techniques. *KEOD*, 235–241.

Pikulík, T., & Štarchoň, P. (2019). GDPR compliant methods of data protection. *6th SWS International Scientific Conferences on social sciences 2019 Conference proceedings*, 561–572.

Pina, E., Ramos, J., Jorge, H., Váz, P., Silva, J., Wanzeller, C., Abbasi, M., & Martins, P. (2024). Data Privacy and Ethical Considerations in Database Management. *Journal of Cybersecurity and Privacy*, *4*(3), 494–517.

Sahatqija, K., Ajdari, J., Zenuni, X., Raufi, B., & Ismaili, F. (2018). Comparison between relational and NOSQL databases. *2018 41st international convention on information and communication technology, electronics and microelectronics (MIPRO)*, 0216–0221.

Shmueli, E., Vaisenberg, R., Elovici, Y., & Glezer, C. (2009). Database encryption: an overview of contemporary challenges and design considerations. *SIGMOD record*, *38*(3), 29–34.

Shyamasundar, R. K., Pratiksha, C., Arushi, J., & Aniket, K. (2021). Approaches to Enforce Privacy in Databases: Classical to Information Flow-Based Models. *Information Systems Frontiers*, *23*(4), 811–833.

Simon, M. (2023). *Getting Started with SQL and Databases: Managing and Manipulating Data with SQL* (1. Aufl.). Apress L. P.

Swanzy, P. N., Abukari, A. M., & Ansong, E. D. (2024). Data Security Framework for Protecting Data in Transit and Data at Rest in the Cloud. *Current Journal of Applied Science and Technology*, *43*(6), 61–77.

Voigt, P., & von dem Bussche, A. (2024). *EU-Datenschutz-Grundverordnung (DSGVO): Praktikerhandbuch* (2. Aufl. 2024). Springer Berlin / Heidelberg.

Eigenständigkeitserklärung

Ich trage die Verantwortung für die Qualität des Textes sowie die Auswahl aller Inhalte und habe sichergestellt, dass Informationen und Argumente mit geeigneten wissenschaftlichen Quellen belegt bzw. gestützt werden. Die aus fremden Quellen direkt oder indirekt übernommenen Texte, Gedankengänge, Konzepte, Grafiken usw. in meinen Ausführungen habe ich als solche eindeutig gekennzeichnet und mit vollständigen Verweisen auf die jeweilige Quelle versehen. Alle weiteren Inhalte dieser Arbeit (Textteile, Abbildungen, Tabellen etc.) ohne entsprechende Verweise stammen im urheberrechtlichen Sinn von mir.

Hiermit erkläre ich, dass ich die vorliegende Studienarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Alle sinngemäß und wörtlich übernommenen Textstellen aus fremden Quellen wurden kenntlich gemacht.

Die vorliegende Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Erklärung zu (gen)KI-Tools

Verwendung von (gen)KI-Tools

Ich versichere, dass ich mich (gen)KI-Tools lediglich als Hilfsmittel bedient habe und in der vorliegenden Arbeit mein gestalterischer Einfluss überwiegt. Ich verantworte die Übernahme jeglicher von mir verwendeter Textpassagen vollumfänglich selbst. In der Übersicht verwendeter (gen)KI-Tools habe ich sämtliche eingesetzte (gen)KI-Tools, deren Einsatzform sowie die jeweils betroffenen Teile der Arbeit einzeln aufgeführt. Ich versichere, dass ich keine (gen)KI-Tools verwendet habe, deren Nutzung der Prüfer bzw. die Prüferin explizit schriftlich ausgeschlossen hat.

Hinweis: Sofern die zuständigen Prüfenden bis zum Zeitpunkt der Ausgabe der Aufgabenstellung konkrete (gen)KI-Tools ausdrücklich als nicht anzeige-/kennzeichnungspflichtig benannt haben, müssen diese nicht aufgeführt werden.

Ich erkläre weiterhin, dass ich mich aktiv über die Leistungsfähigkeit und Beschränkungen der unten genannten (gen)KI-Tools informiert habe und überprüft habe, dass die mithilfe der genannten (gen)KI-Tools generierten und von mir übernommenen Inhalte faktisch richtig sind.

Übersicht verwendeter (gen)KI-Tools

Die (gen)KI-Tools habe ich, wie im Folgenden dargestellt, eingesetzt.

(gen)KI-Tool	**Einsatzform**	**Betroffene Teile der Arbeit**
ChatGPT	Generierung von ersten Ideen für eine geeignete Gliederung	Gesamte Arbeit
	Auswahl dem Thema entsprechender geeigneter Artikel der DSGVO	Gesamte Arbeit
DeepL	Übersetzung einzelner Textpassagen englischer Literatur zum besseren Verständnis	Gesamte Arbeit

Münster, 2. November 2024

Elias Häußler