ATG

Gesellschaft für
Versicherungswissenschaft
und -gestaltung e.V.

GVG

Management-Papier "*Pseudonymisierung / Anonymisierung*"

Aktionsforum Telematik im Gesundheitswesen

Köln, den 19.08.2003

Kontakt:
Jürgen Dolle (Koordinierung), GVG
mailto: j.dolle@gvg-koeln.de

Autoren-Team:

Albert, Jürgen	VdAK /AEV
David, Dr. Dagmar M.	BÄK/ÄKNo
Langerfeld, Carsten	IKK-BV
Keil, Werner	ABDA/WuV
Rey, Heinz-Theo	KBV
Schöner, Claus	BKK-BV
Siebert, Irmgard	KZBV

Inhaltsverzeichnis

1 Zusammenfassung (Management Summary)

Ausgehend von der in der Einleitung beschriebenen Datenlage im deutschen Gesundheitssystem und dem daraus erwachsenen Transparenzbedürfnis in Verbindung mit den Vorgaben des Datenschutzes wird die Notwendigkeit von Pseudonymisierungs- und Anonymisierungsverfahren erkannt. Das vorliegende Management-Papier enthält jedoch keine Beschreibung technischer Details, sondern die Erläuterung geeigneter Strukturen und Verfahrensweisen zum Schutz der Persönlichkeitsrechte des Einzelnen bei der Realisierung von Datentransparenzverfahren (Kapitel 2).

Dazu werden zunächst die verwendeten Begriffe im Sinne des vorliegenden Dokumentes definiert (Kapitel 3). Daran schließt sich eine beispielhafte Beschreibung existenter Verfahren an, in denen bereits heute Anonymisierung oder Pseudonymisierung eingesetzt wird (Kapitel 4).

Aus dieser Betrachtung der Ist-Situation sowie der vorliegenden Rahmenbedingungen – insbesondere der aktuellen Gesetzgebungsbestrebungen im GMG zur Datentransparenz – werden Lösungsansätze abgeleitet. Dazu werden zunächst die möglichen Verfahren in Form von Modellen einzeln betrachtet und anschließend jedes Verfahren für sich nach definierten Kriterien bewertet. Abschließend werden die einzelnen Verfahren in einer Tabelle gegenübergestellt und in Relation bewertet (Kapitel 5 und Anhang).

Im Ergebnis gibt das Papier konkrete Empfehlungen ab, welche Verfahren nach Auffassung des ATG zu favorisieren sind (Kapitel 6).

Das ATG betont hierzu jedoch, dass generell kein spezielles Verfahren als geeignet erklärt werden kann, sondern dass ausgehend von differenzierten Anforderungen konzipiert werden muss, welche Daten zur Erfüllung der speziellen Zwecke notwendig sind, wo diese Daten vorliegen und auf welche Weise sie bereitgestellt werden können.

Des weiteren appelliert das ATG mit Blick auf das aktuelle Gesetzgebungsverfahren an das zuständige Bundesministerium für Gesundheit und Soziale Sicherung, die jeweiligen Positionen zur Erreichung von Datentransparenz zeitnah mit der Selbstverwaltung auszutauschen und die Ergebnisse der Gespräche noch in den aktuellen Gesetzesentwurf einfließen zu lassen (ebenfalls Kapitel 6).

2 Einleitung

Die Beteiligten im Gesundheitswesen verfügen auf Grund unterschiedlicher Rechtsnormen über Versorgungsdaten, die sie nach den ihren Aufgaben entsprechenden Erfordernissen unter Berücksichtigung der geltenden datenschutzrechtlichen Bestimmungen verarbeiten. Dazu zählt auch Datentransport, der z. T. bereits heute pseudonymisiert bzw. anonymisiert erfolgt.

In den vom ATG in der Vergangenheit erarbeiteten Managementpapieren zu den Themen Sicherheitsinfrastruktur, Elektronisches Rezept, Elektronischer Arztbrief und Europäische und internationale Perspektiven von Telematik im Gesundheitswesen wurden zum einen Aspekte der Pseudonymisierung / Anonymisierung bewusst ausgeklammert, zum anderen eine zeit- und sektorübergreifende Gesundheitsberichterstattung gefordert. Hinzu kommt, dass auf europäischer Ebene eine umfassende Gesundheitsberichterstattung der Mitgliedsstaaten gefordert wird, um eine Vergleichbarkeit der Systeme und ihrer Versorgung (Public Health) zu ermöglichen.

Konkrete Vorgaben zur Erreichung von Datentransparenz macht die Bundesregierung derzeit in ihrem aktuellen Gesetzgebungsverfahren zum Gesundheitssystemmodernisierungsgesetz (GMG). Diese sind in Teilbereichen mit den Vorstellungen des ATG zu vereinbaren, teilweise aber auch konträr zu ihnen.

Szenarien zur Pseudonymisierung und Anonymisierung werden beschrieben.

Das Management-Papier Pseudonymisierung / Anonymisierung befasst sich mit den ordnungspolitischen Aspekten der Verfahren Pseudonymisierung und Anonymisierung unter den möglichen Szenarien zum Umgang mit personenbezogenen Daten im Gesundheitswesen. Dabei soll ein gemeinsames Verständnis der Problemlage und der möglichen Lösungsansätze bei den Beteiligten herbeigeführt, sowie eine gemeinsame Vorgehensweise erarbeitet werden. Das ATG will keine technischen Details der Verfahren beschreiben oder über den Einsatz spezieller Verfahren bei einzelnen Datensätzen entscheiden. Es will mögliche Strukturen aufzeigen und den dazu erforderlichen Handlungsbedarf beschreiben.

Der Schutz der Persönlichkeitsrechte hat Priorität.

Mit dem vorliegenden Management-Papier sollen Empfehlungen gegeben werden, wie durch technische und organisatorische Maßnahmen im Umgang mit personenbezogenen Daten der Schutz der Persönlichkeitsrechte des Einzelnen gewahrt werden kann. Dabei sollen nicht dezidierte technische Verfahren zur Pseudonymisierung bzw. Anonymisierung beschrieben werden, sondern Ziel ist, Richtlinien für globale Verfahrensweisen bei der Pseu-

donymisierung/Anonymisierung von Daten aufzuzeigen (z. B. Zeitpunkt der Pseudonymisierung, zentrale/dezentrale Pseudonymisierung).

Keine Beschreibung von technischen Details, sondern Richtlinien zur Auswahl und Anwendung geeigneter Verfahren.

3 Definitionen und Abgrenzung des Themas

3.1 Zielsetzung/Abgrenzung

Wie bereits in der Einleitung angesprochen, sieht das Management-Papier "Pseudonymisierung/Anonymisierung" seinen Anspruch nicht in der Beschreibung von Verfahren zur Erreichung von Datentransparenz.

Keine Beschreibung von Verfahren zur Erreichung von Datentransparenz.

Das ATG ist der Auffassung, dass ausgehend von differenzierten Anforderungen konzipiert werden muss, welche Daten zur Erfüllung der speziellen Zwecke notwendig sind, wo diese Daten vorliegen und auf welche Weise sie bereitgestellt werden können.

3.2 Definitionen

3.2.1 Anonymisierung

§3, Abs. 6 BDSG: „*Anonymisieren ist das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können.*“

Mit Hilfe von Anonymisierungsverfahren werden personenbezogene Daten, derart verändert, dass sie sich weder auf eine bestimmte Person beziehen noch eine solche erkennen lassen (z.B. durch Weglassen oder Änderung von personenidentifizierenden Merkmalen). Ein Wiederherstellen des ursprünglichen Personenbezuges ist nicht möglich.

Anonymisierung ist z. B. erforderlich, wenn die Personenidentitäten oder die Herkunft der Daten für die zu ermittelnden Ergebnisse nicht relevant sind oder sein sollen.

3.2.2 Pseudonymisierung

§3, Abs. 6a BDSG: „*Pseudonymisieren ist das Ersetzen des Namens und anderer Identifikationsmerkmale durch ein Kennzeichen zu dem Zweck, die Bestimmung des Betroffenen auszuschließen oder wesentlich zu erschweren.*“

Mittels Pseudonymisierungsverfahren werden also personenbezogene Daten, durch eine Regel derart eineindeutig verändert, dass sie sich weder auf eine bestimmte Person beziehen noch eine solche erkennen lassen. Durch eine inverse Rechenvorschrift oder eine Zuordnungstabelle ist eine Depseudonymisierung möglich.

Es existieren jedoch auch sogenannte "Einweg-Pseudonymisierungsverfahren", bei deren Anwendung eine Depseudonymisierung nahezu unmöglich ist oder zumindest deutlich erschwert wird, da es keine inverse Rechenvorschrift gibt.

Pseudonymisierung ist eine abgeschwächte Form der Anonymisierung.

Pseudonymisierung ist z. B. erforderlich, wenn ein individueller Verlauf beobachtet werden soll, oder die Zusammenführung unterschiedlicher Daten zu einer Person erfolgen soll, ohne dass die Personenidentität bekannt sein muss.

3.2.3 Datenaggregation

Unter Datenaggregation versteht man Verfahren, bei denen durch Bildung von (entsprechend großen) Gruppen der Personenbezug einer Datenmenge entfernt wird. Ein Wiederherstellen des ursprünglichen Personenbezuges ist grundsätzlich nicht möglich.

3.2.4 Datenquelle

Als Datenquelle im Sinne dieses Dokumentes werden alle Stellen bezeichnet, die Daten in ein Verfahren einspeisen. Dies können sowohl einzelne Leistungserbringer als auch deren Interessenvertreter, Verbände, Dienstleister etc. sein.

3.2.5 Vertrauensstelle

Eine Vertrauensstelle ist eine unabhängige Institution, die zwischen Datenquelle und Datensammelstelle positioniert werden kann. Sie hat die Aufgabe, den Datenfluss zu pseudonymisieren, zu anonymisieren bzw. zu aggregieren. Sie hat insbesondere keine Berechtigung, auf die in der Datensammelstelle vorgehaltenen Daten zuzugreifen.

Findet ein Pseudonymisierungsverfahren mittels einer Vertrauensstelle statt, so stellt sie im Falle einer Depseudonymisierung im Einzelfall aus einem Pseudonym die ursprüngliche Identität wieder her.

3.2.6 Datensammelstelle

Unter einer Datensammelstelle wird eine Stelle verstanden, die die Daten der Datenquellen, die sie ggf. über eine Vertrauensstelle erhält, zusammenführt und für Auswertungen zur Verfügung stellt.

3.2.7 Datenanfordernde Stelle

Unter einer datenanfordernden Stelle wird eine Stelle verstanden, der Daten für eine (End-)Auswertung zur Verfügung gestellt werden. Eine datenanfordernde Stelle kann auch eine Datenquelle sein.

Anmerkung: Generell sind in jedem Modell (vgl. Anhang) datenanfordernde Stellen vorhanden, auch wenn sie nicht explizit benannt werden.

4 Ist-Situation

4.1 Beschreibung realer Verfahren

Es werden zwei beispielhaft ausgewählte Verfahren beschrieben.

Im Folgenden werden zwei Verfahren beschrieben, die bereits heute Pseudonymisierungs- bzw. Anonymisierungsverfahren in konkreten Anwendungen einsetzen. Es handelt sich dabei um beispielhaft ausgewählte Verfahren aus einer Vielzahl von existierenden Anwendungen.

4.1.1 Erzeugung von Pseudonymen am Beispiel des Deutschen Arzneimittelprüfinstituts (DAPI)

Für die apothekeneigene Statistik des DAPI sind individuelle Merkmale und Verhaltensweisen im Zusammenhang mit Medikationen von großem Interesse.

Hierzu zählen unter anderem die Informationen, ob spezielle Medikamente durch andere ersetzt worden sind und welche Co-Medikationen bei bestimmten Präparaten stattgefunden haben.

Die benötigten Arzneimitteldaten von GKV-Patienten werden von den sechs Rechenzentren, die sich zur Gesellschaft für zentrales Datenmanagement und Statistik im Gesundheitswesen (GDSG) zusammengeschlossen haben, der Werbe- und Vertriebsgesellschaft Deutscher Apotheker mbH (WuV) verschlüsselt zur Verfügung gestellt.

Zuerst werden die Verordnungsdaten mit einem nur den Mitgliedern des GDSG bekannten Schlüssel, der für alle Rechenzentren gleich ist, mittels einer Einweg-Funktion pseudonymisiert. Anschließend werden diese pseudonymisierten Daten von den Rechenzentren direkt an die WuV gesendet.

Im zweiten Schritt verschlüsselt die WuV die zuvor erzeugten Pseudonyme mit einem nur ihr bekannten Schlüssel, bevor die Daten zur weiteren Verarbeitung an das Norddeutsche Apotheken-Rechenzentrum Bremen (NARZ) versendet werden, das anschließend die Berechnung der Statistik durchführt.

Aufgrund der Doppelrolle des NARZ Bremen als pseudonymisierende Stelle und als erstellende Institution der Statistik wurde dieses zweistufige Verfahren gewählt. Es ist mit dem Datenschutz abgestimmt.

Bedingt durch die zweistufige Prozedur ist das NARZ Bremen nicht in der Lage, auch nicht durch einen Vergleich, mit den selbst erzeugten Pseudonymen, auf die Identität der Person zu schließen, da die vorliegenden Pseudonyme mit einem dem NARZ Bremen nicht bekannten Schlüssel verschlüsselt wurden.

Auf der anderen Seite ist es der Werbe- und Vertriebsgesellschaft Deutscher Apotheker mbH ebenfalls nicht möglich, die Originaldaten zu ermitteln, da für die Pseudonymisierung eine nicht umkehrbare Einweg-Funktion verwendet wurde.

4.1.2 GKV Arzneimittel Schnellinformation (GAmSi)

Die Spitzenverbände der gesetzlichen Krankenversicherung haben sich darauf verständigt, die Verordnungsdaten zeitnah für gemeinsame Zwecke zusammenzuführen.

Dazu stellen die Spitzenverbände die Bereitstellung der Arzneimitteldaten ihrer Kassenart sicher. Die einzelnen Datenstellen bereiten die Daten mit einem einheitlichen Transformationsprogramm auf. Dabei werden die personenbezogenen Daten anonymisiert. Die so entstandenen Datenpakete werden verschlüsselt (Grundlage ist die von der GKV eingesetzte „Security-Schnittstelle für das Gesundheitswesen“) und mittels DFÜ an die Relaisstelle (Informationstechnische Servicestelle der GKV – ITSG) übermittelt.

Die Relaisstelle nimmt die Anonymisierung der Absender vor und leitet die Datenpakete an die verarbeitende Stelle weiter. Die Relaisstelle betreibt keine Datenhaltung.

4.2 Rahmenbedingungen

Im vorliegenden Dokument wird bewusst nicht auf die ärztliche Schweigepflicht eingegangen, da das ATG davon ausgeht, dass Pseudonymisierung und Anonymisierung nur auf den Transport und die Bereitstellung von Daten angewendet wird. Die Frage, ob Daten überhaupt für weitergehende Zwecke zur Verfügung gestellt werden dürfen, wird im vorliegenden Dokument nicht behandelt.

4.2.1 Datentransparenz

Durch die Gesundheitsreform 1993 wurde der Datenaustausch mit Leistungserbringern (§§ 294 ff SGB V) gesetzlich normiert. Ziel war es unter anderem, für mehr Transparenz im Vertrags- und Leistungsgeschehen im Gesundheitswesen zu sorgen. Die Leistungs- und Abrechnungsdaten liegen derzeit personenbezogen und verteilt bei verschiedenen Stellen des Gesundheitswesens vor.

Übergreifende Auswertungen, fundierte Gesamtbetrachtungen zum Versorgungsgeschehen (beispielsweise bei bestimmten Krankheitsbildern, Disease-Management-Programme), die Entwicklung des morbiditätsorientierten Risikostrukturausgleichs oder die Abbildung der medizinischen Versorgung in Städten und Regionen bzw. durch einzelne Leistungserbringer sind derzeit nur unter größtem Aufwand und nicht systematisch möglich.

Die Verfolgung dieser Ziele erfordert aus datenschutzrechtlichen Gründen gesicherte Pseudonymisierungs- und Anonymisierungsverfahren, die den Vorschriften des BDSG entsprechen.

Datentransparenz erfordert Pseudonymisierungs-/ Anonymisierungsverfahren.

4.2.2 Datenschutz

Im Bereich öffentlicher Stellen ist die Datenerhebung, -verarbeitung und -nutzung nur zulässig, wenn eine Rechtsnorm dies ausdrücklich erlaubt/vorsieht, oder die schriftliche Einwilligung des Betroffenen vorliegt (§§ 4 u. 4a BDSG). Hinzu kommt das Prinzip der Datenvermeidung und Datensparsamkeit (§ 3a BDSG). Besondere Schutzwürdigkeit wird u. a. personenbezogenen Daten mit Informationen zur Gesundheit zugeschrieben (§ 3 Abs. 9 BDSG).

Als zentraler Paragraf für den Umgang mit personenbezogenen Daten im Gesundheitswesen unter dem besonderen Aspekt der Aggregation und Auswertung wird vom ATG § 3a BDSG Datenvermeidung und Datensparsamkeit angesehen.

Datenvermeidung und Datensparsamkeit sind zentrale Paragrafen für den Umgang mit personenbezogenen Daten.

Gestaltung und Auswahl von Datenverarbeitungssystemen haben sich danach an dem Ziel auszurichten, keine oder so wenig personenbezogene Daten wie möglich zu erheben, zu verarbeiten oder zu nutzen. Insbesondere ist von den Möglichkeiten der Anonymisierung und Pseudonymisierung Gebrauch zu machen, soweit dies möglich ist und der Aufwand in einem angemessenen Verhältnis zu dem angestrebten Schutzzweck steht.

4.2.3 Gesetzgebung

Im aktuellen Gesetzgebungsverfahren zum GMG werden bereits konkrete Vorgaben zur Erreichung von Datentransparenz gemacht[1]. So sollen die Selbstverwaltungsorganisationen verpflichtet werden, entsprechende Institutionen (Arbeitsgemeinschaft, Beirat, Vertrauensstelle, Datenaufbereitungsstelle) zu gründen und zu betreiben. Es werden bereits Festlegungen zu den anzuwendenden Verfahren getroffen (einstufige Pseudonymisierung), während Zweckbestimmung der Daten und Zugriffsberechtigungen sehr allgemein beschrieben werden bzw. erst durch eine Arbeitsgemeinschaft formuliert werden sollen.

4.3 Bewertung

Aus den gesetzlichen Vorgaben des Bundesdatenschutzgesetzes (Datensparsamkeit) und dem Wunsch, Transparenz zu schaffen, ergibt sich die Notwendigkeit, geeignete Verfahren der Anonymisierung/Pseudonymisierung bzw. Aggregation zu definieren und einzusetzen. Welche Verfahren das sein können, muss aus den einzelnen Analyseanforderungen abgeleitet werden, wie die unter 4.1.1 und 4.1.2 beschriebenen Projekte beispielhaft zeigen.

Die Auswahl der Verfahren ist abhängig vom Auswertungszweck.

Die vom Gesetzgeber im GMG-Entwurf formulierte Vorgehensweise, zunächst die organisatorische Struktur aufzubauen, die einzusetzenden Verfahren festzulegen und erst anschließend durch eine Arbeitsgemeinschaft Verwendungszwecke und notwendiges Datenmaterial erarbeiten zu lassen, kann vom ATG nicht bestätigt werden.

GMG-Vorgaben zur Datentransparenz müssen sich auf Zielvorgaben beschränken.

[1] Referenziert wird der GMG-Entwurf vom 02.06.2003, siehe dort §303a-f

Des weiteren wird seitens des ATG die zweckunabhängige Vorgabe eines einzigen Pseudonymisierungsverfahrens datenschutzrechtlich kritisch gesehen. Anonymisierungsverfahren oder mehrstufige Pseudonymisierungsverfahren werden im GMG nicht erwähnt.

5 Lösungsansätze und Bewertung

5.1 Einleitung

Beschreibung organisatorischer Verfahren. Vorhandene technische Verfahren werden als sicher vorausgesetzt.

Das ATG vertritt die Auffassung, dass die heute verfügbaren technischen Verfahren zur Erzeugung von Pseudonymen und Anonymen grundsätzlich sicher sind. Die verschiedenen organisatorischen Verfahren der Pseudonymisierung/Anonymisierung lassen sich modellhaft als Datenflussdiagramme in ihrem grundsätzlichen Aufbau beschreiben.

Für die Entwicklung der Modellsystematik gelten folgende Prämissen:

Prämissen: Auswahl und Ausgestaltung der Verfahren sind abhängig vom Erfüllungszweck bzw. vom Zweck der Datenauswertung.

- Es wird bewusst keine Aussage über einzelne Felder gemacht, die im Rahmen der Erhebung, der Verarbeitung und der Nutzung personenbezogener Daten zu anonymisieren/pseudonymisieren sind. Dies ist abhängig vom Erfüllungszweck der Daten und wird in der Verantwortlichkeit der dazu von den Beteiligten zu schließenden Vereinbarung gesehen.

- Für die Entscheidung, ob Pseudonymisierungs- oder Anonymisierungsverfahren anzuwenden sind, geht das ATG davon aus, dass sich alle Überlegungen zur Datenspeicherung in zentrale und dezentrale Datenhaltung unterscheiden lassen. Diese wiederum lassen sich weiter untergliedern in verschiedene Varianten einstufiger oder zweistufiger Pseudonymisierung bzw. Anonymisierung.

- Auch die Wahl des Pseudonymisierungsverfahrens ist abhängig vom Zweck der Datenauswertung. So muss z. B. bei Langzeitbeobachtungen trotz Veränderung von personenidentifizierenden Daten (z. B. Kassenwechsel, Namensänderung) sichergestellt sein, dass für die betrachtete Person weiterhin dasselbe Pseudonym erzeugt wird. Außerdem ist bei der Wahl des Verfahrens zu bewerten, ob die Notwendigkeit der Depseudonymisierung im Einzelfall be-

steht (falls nicht, sind Einweg-Pseudonymisierungsverfahren zu bevorzugen).

Die ausführliche Darstellung und Erläuterung, sowie Bewertungskriterien der einzelnen Modelle findet sich im Anhang. Im Anschluss an die Einzelbetrachtung der untersuchten Modelle sind diese in einer zusammenfassenden Bewertungsmatrix gegenübergestellt.

5.2 Korrumpierung der Verfahren

5.2.1 Pseudonymisierung

5.2.1.1 Wahrscheinlichkeit einer Korrumpierung

Einweg-Pseudonymisierungsverfahren sind schwerer zu korrumpieren.

Grundsätzlich ist die Korrumpierung von Einweg-Pseudonymisierungsverfahren wesentlich aufwändiger und damit weniger wahrscheinlich, da zur Herstellung eines Personenbezuges nicht nur das "Geheimnis" (Algorithmus, Schlüssel, Zuordnungstabelle) bekannt sein muss, sondern außerdem sämtliche der Pseudonymisierung zugrunde liegenden Personendaten von allen in die Auswertung einbezogenen Personen.

Die Wahrscheinlichkeit einer Korrumpierung ist aber auch abhängig von der Anzahl der pseudonymisierenden Stellen. Erfolgt die Pseudonymisierung an einer zentralen Stelle, so ist die Korrumpierung weniger wahrscheinlich, als bei einer Pseudonymisierung an mehreren Stellen. Je mehr Stellen in Kenntnis desselben Algorithmus und Schlüssels sind, um so angreifbarer wird das Verfahren.

Zentrale sowie zweistufige Pseudonymisierungsverfahren verringern die Korrumpierungswahrscheinlichkeit; höherstufige Verfahren rechtfertigen die entstehenden Kosten jedoch nicht.

Die Wahrscheinlichkeit einer Korrumpierung sinkt, wenn zweistufige Pseudonymisierungsverfahren zum Einsatz kommen. Darüber hinausgehende Pseudonymisierungsstufen sind organisatorisch und wirtschaftlich im Verhältnis zum Nutzen nur mit unverhältnismäßig hohem Aufwand realisierbar und verringern die Wahrscheinlichkeit der Korrumpierung aus Sicht des ATG nicht wesentlich.

Eine zusätzliche Möglichkeit, Rückschlüsse auf Personenidentitäten zu erschweren, wäre die Einführung von sog. Session-Keys (d.h. Einmal-Schlüssel) bei der Abfrage von Daten.

Die Gefahr der Depseudonymisierung aufgrund von Alleinstellungsmerkmalen einzelner Personen ist grundsätzlich unabhängig von der Anzahl der eingezogenen Pseudonymisierungsstufen oder der Verwendung von Session-Keys.

5.2.1.2 Folgen der Korrumpierung

Eine Korrumpierung erfordert die Löschung gesammelter pseudonymisierter Daten und erfordert die Generierung eines neuen Pseudonymisierungsschlüssels.

Bei Korrumpierung eines Pseudonymisierungsschlüssels müssen sich alle Stellen, die diesen Schlüssel angewendet haben, auf einen neuen Schlüssel einigen. Alle Pseudonyme, die bisher mit dem korrumpierten Schlüssel generiert wurden, sind damit als solche unbrauchbar. Die zugehörigen Daten sind aus datenschutzrechtlichen Gründen zu löschen, weil ein direkter Personenbezug herstellbar sein kann. Dies bedeutet bei zentraler Datenhaltung, dass alle bisherigen Daten mit neuen Pseudonymen wieder geliefert werden müssen. Sollte die Pseudonymisierung an einer zentralen Stelle erfolgen, so kann – sofern kein Einweg-Pseudonymisierungsverfahren benutzt wurde – bei Korrumpierung der Pseudonyme die zentrale Stelle diese in neue Pseudonyme überführen, ohne den gesamten Datenbestand löschen zu müssen.

5.2.2 Anonymisierung

Eine Korrumpierung von Anonymisierungsverfahren ist grundsätzlich nicht möglich.

Die Anonymisierungsverfahren werden hier nicht betrachtet, da eine Korrumpierung gemäß Definition (BDSG) nicht möglich ist (Vergl. 3.2.1). Zwar mag in Einzelfällen aufgrund von Alleinstellungsmerkmalen eine Identität herstellbar sein, das Verfahren insgesamt bleibt hiervon jedoch unberührt.

5.3 Bewertung / Fazit

Grundsätzlich lassen sich folgende Aussagen treffen:

Die Auswahl des Verfahrens ist abhängig vom Auswertungsinteresse.

- Die Auswahl des Verfahrens ist abhängig vom Auswertungsinteresse.

Anonymisierung ist sicherer als Pseudonymisierung; wenn Pseudonymisierung, dann Einweg-Pseudonyme bevorzugen.

- Anonymisierung ist sicherer als Pseudonymisierung.
- Unter den Pseudonymisierungsverfahren sind Einweg-Pseudonymisierungsverfahren aufgrund der geringeren Korrumpierungsgefahr zu bevorzugen, sofern keine Notwendigkeit zur Depseudonymisierung besteht.

Bezüglich der Sicherheit, des organisatorischen und des administrativen Aufwandes trifft das ATG folgende Aussagen zur Bewertung:

Sicherheit:

Anonymisierung idealerweise an der Datenquelle.

- Eine Anonymisierung/Pseudonymisierung sollte so früh wie möglich durchgeführt werden, bei Anonymisierung idealerweise an der Datenquelle.

Pseudonymisierung außerhalb der Datenquelle nur durch Vertrauensstelle.

- Eine Anonymisierung/Pseudonymisierung außerhalb der Datenquelle sollte durch eine Vertrauensstelle vorgenommen werden.
- Durch Kenntnis der Datenquelle sollte kein Rückschluss auf Personendaten möglich sein (z. B. bei geringem Datenvolumen einer bestimmten Datenquelle).
- Mehrstufige Verfahren sind sicherer als einstufige Verfahren.
- Verfahren mit zentraler Pseudonymisierung sind sicherer als solche mit dezentraler Pseudonymisierung. (Bei dezentraler Pseudonymisierung steigt die Gefahr, dass der Schlüssel bekannt wird mit der Anzahl der pseudonymisierenden Stellen.)
- Verfahren mit dezentraler Anonymisierung sind sicherer als solche mit zentraler Anonymisierung.

Organisatorischer Aufwand:

- Der organisatorische Aufwand steigt mit dem Aufbau neuer Organisationen (Vertrauensstellen).
- Die Implementierung einer Online-Abfragefähigkeit von einheitlich bereitzustellenden Daten bei zahlreichen Datenquellen verursacht erheblichen Aufwand.

Administrativer Aufwand

- Der Betrieb von zusätzlichen Organisationen (Vertrauensstellen) verursacht einen zusätzlichen administrativen Aufwand
- Je mehr Stellen Pseudonymisieren/Anonymisieren desto höher ist der administrative Aufwand.
- Die Online-Abfrage von einheitlich bereitgestellten Daten bei zahlreichen Datenquellen verursacht administrativen Aufwand.

6 Anhang: Modelle der Pseudonymisierung / Anonymisierung

In diesem Anhang soll der Entscheidungsprozess des ATG bei der Bewertung der beschriebenen Pseudonymisierungs- bzw. Anonymisierungsverfahren transparent gemacht werden, daher sind diese Verfahren bewusst in den Anhang aufgenommen worden. Die diskutierten Verfahren werden anhand von "Datenflussplänen" dargestellt, außerdem wurde eine Beurteilung und Gewichtung zueinander hinsichtlich der Sicherheit und des organisatorischen bzw. administrativen Aufwands vorgenommen, auch wenn diese Modelle vom ATG nicht empfohlen werden. Hierbei erhebt das ATG keinen Anspruch auf Vollständigkeit.

Zusätzlich zur Unterscheidung der grundlegenden Methode zur Unkenntlichmachung von personenbezogenen Daten wurde bei den Modellen eine Unterscheidung nach zentraler und dezentraler Datenhaltung getroffen.

Beim Modell der zentralen Speicherung werden alle Daten des Gesundheitswesens, die für Analysezwecke verfügbar gemacht werden sollen, von den Beteiligten an ein zentrales System übermittelt. Hier stehen diese Daten dann gesammelt in vollständiger Form zur Verfügung. Hierbei wird auch dann von zentraler Datenhaltung gesprochen, wenn die Daten zwar physikalisch und/oder räumlich verteilt, aber unter der Hoheit nur einer Institution gespeichert werden.

Kennzeichen der zentralen Speicherung ist die Übermittlung aller Daten an eine zentrale Sammelstelle.

Bei einem dezentralen Modell verbleiben die Daten bei den jeweiligen Beteiligten des Verfahrens, es wird jedoch eine Möglichkeit geschaffen, die Daten unter Berücksichtigung von Zugriffsrechten abzurufen. Dies setzt ggf. geeignete Schnittstellenlogiken voraus.

Bei dezentraler Speicherung werden die Daten zum Zeitpunkt der Auswertung von den beteiligten Stellen abgerufen.

Für die Entscheidung ob Pseudonymisierungs- oder Anonymisierungsverfahren anzuwenden sind, geht das ATG davon aus, dass sich alle Überlegungen zur Datenspeicherung in zentrale und dezentrale Datenhaltung unterscheiden lassen. Diese wiederum lassen sich weiter untergliedern in verschiedene Varianten einstufiger oder zweistufiger Pseudonymisierung bzw. Anonymisierung. Zur Erleichterung der Orientierung im Text soll die nachfolgende Grafik dienen.

Im vorliegenden Papier werden zentrale und dezentrale Datenhaltung sowie ein– und zweistufige Verfahren unterschieden.

Die Modelle setzen voraus, dass Inhalt und Umfang der bereitgestellten Daten vorher vereinbart oder gesetzlich vorgeschrieben wurden. Dabei müssen die Belange der Auswertungsziele gegenüber den datenschutzrechtlichen

Belangen abgewogen worden sein, da über Alleinstellungsmerkmale eine Identifizierung einer Person dennoch möglich ist.

So könnte etwa bei einer Anonymisierung allein der Versichertennummer über Alleinstellungsmerkmale wie Geschlecht, Körpergröße, Wohnort, Blutgruppe etc. der Versicherte immer noch identifiziert werden.

Sofern im Falle von Pseudonymisierung eine Depseudonymisierung im Einzelfall bewusst ausgeschlossen werden kann, sind aufgrund der geringeren Korrumpierungsgefahr als technische Verfahren Einweg-Pseudonymisierungsverfahren anzuwenden. Die der Verfahrensanalyse zugrunde liegende Modellsystematik ändert sich dadurch nicht, da die dargestellten Datenflusspläne hiervon nicht beinflusst werden.

Alle gezeigten Modelle gehen von einer geeigneten und nicht explizit erwähnten Transportsicherung der Daten aus.

Bei der Übermittlung der Daten ist das unberechtigte Lesen, Verändern oder Löschen der Daten zu verhindern (Transportsicherung, z. B. mit Hilfe einer Public Key Infrastruktur). Die Transportsicherung wird vorausgesetzt und nicht gesondert gekennzeichnet.

Bei den Modellen, bei denen zusätzlich eine Vertrauensstelle vorgesehen ist, wird eine erweiterte Sicherung vorgesehen. Dabei werden die Nutzdaten zunächst verschlüsselt und dann transportgesichert an die nachgeordnete Stelle weitergeleitet. Die Entschlüsselung der Nutzdaten kann nicht von zwischengelagerten Stellen vorgenommen werden. Damit wird erreicht, dass eine Vertrauensstelle zu keiner Zeit Einblick in die Nutzdaten erhält.

Die erweiterte Sicherung wird in den Grafiken "verschlüsselte Daten" genannt.

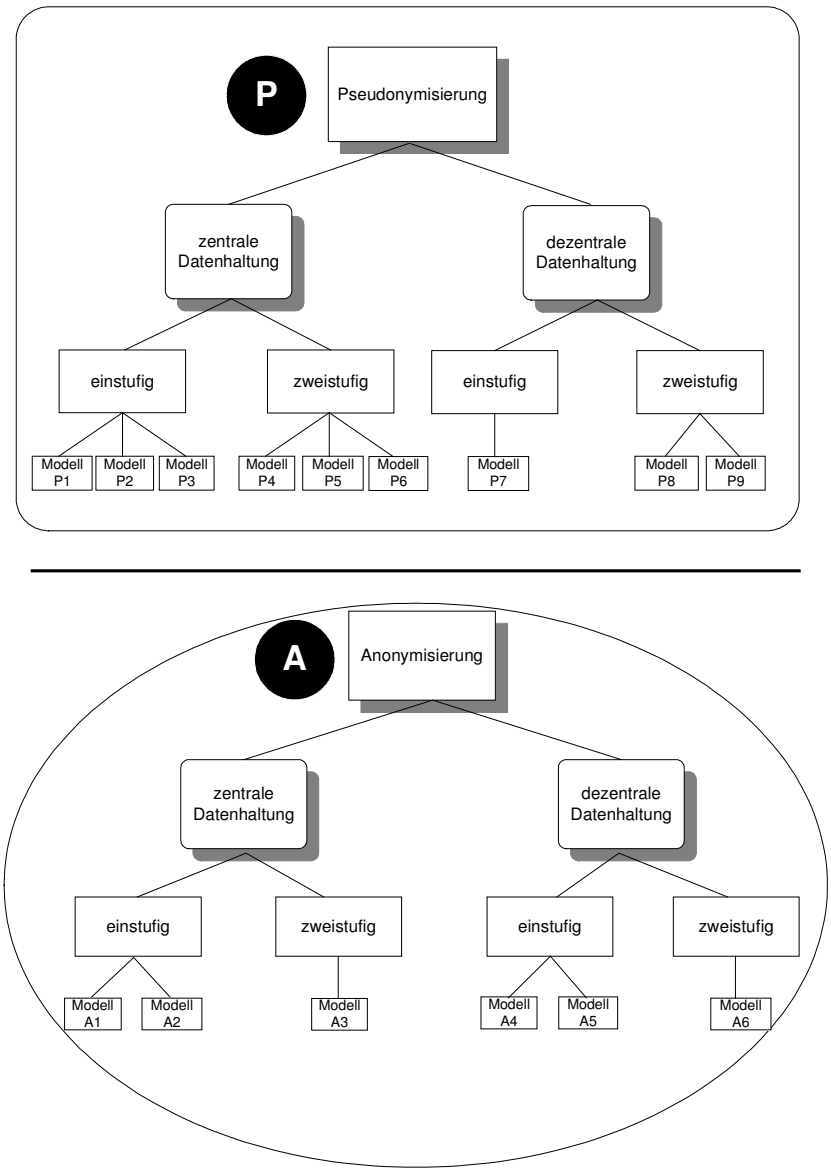


Abbildung 1 Modellsystematik der Pseudonymisierung / Anonymisierung

Am Schluss eines jeden Modells erfolgt eine Bewertung nach verschiedenen Kriterien:

Die einzelnen Verfahren werden bzgl. ihrer Sicherheit, des organisatorischen und des administrativen Aufwandes bewertet.

- Sicherheit
  Zur Bewertung werden hier die Aspekte „Verfahren“, „Korrumpierbarkeit“ und „Stufigkeit“ (s. u.) herangezogen. Sichere Verfahren werden positiv bewertet, unsichere Verfahren negativ.
- Aufwand
  Hierin gehen sowohl die Aspekte des organisatorischen als auch solche des administrativen unter

Berücksichtigung des jeweiligen finanziellen Aufwandes ein.

- Organisatorisch / Verfahrensaufbau

  Hier werden Aufbau und Struktur des Verfahrens bewertet. Aspekte der Systemarchitektur finden in diesem Punkt Berücksichtigung. Ein einfaches Verfahren wird sehr positiv bewertet, während ein vielschichtiges Verfahren eine eher negative Bewertung erfährt.

- Administrativ / Betrieb des Verfahrens

  Ein Verfahren, welches einfach zu betreiben ist, wird positiv bewertet, ein Verfahren, welches mit hohem Aufwand zu betreiben ist, wird negativ bewertet.

Anschließend wird die obige Bewertung jeweils mit folgenden Symbolen klassifiziert:

- ☺☺ sehr positiv
- ☺ positiv
- 😐 indifferent
- ☹ negativ
- ☹☹ sehr negativ

Zur einprägsamen Darstellung wird zu jedem Modell eine „Bewertungszeile“ wie folgt dargestellt (hier als Beispiel):

Modell	**Sicherheit**	**Aufwand**	
		Organisatorisch	**Administrativ**
Xy	☺☺	☺	☹

In Abschnitt 6.6 erfolgt schließlich eine Gesamtübersicht in Form einer Matrix, in der die Modelle gegeneinander bewertet werden.

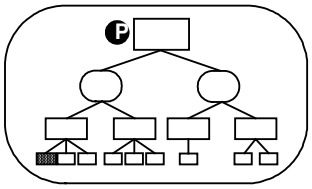
6.1 Pseudonymisierung bei zentraler Datenhaltung

In einem solchen Modell gibt es die Möglichkeit eines einstufigen oder zweistufigen Pseudonymisierungsprozesses. Wird auf dem Weg von den Datenquellen zur zentralen Datensammelstelle nur einmal pseudonymisiert, so spricht man von einem einstufigen Pseudonymisierungsprozess. In allen anderen Fällen spricht man von zweistufigen Pseudonymisierungsprozessen. Die Einteilung in ein- oder zweistufige Verfahren hängt von der Anzahl der Pseudonymisierungsschritte durch verschiedene Stellen ab, nicht von der Gesamtzahl der beteiligten Stellen. Mehrmalige Pseudonymisierung an einer Stelle wird als eine Stufe gewertet, da hierdurch keine höhere Sicherheit erzielt wird.

Die Anzahl der durch verschiedene Stellen durchgeführten Pseudonymisierungsprozesse ergibt die Stufigkeit des Verfahrens.

6.1.1 Einstufige Modelle

6.1.1.1 Modell P1 – Die Datenquellen pseudonymisieren



Bei den Datenquellen wird für jede reale Person ein eindeutiges Pseudonym anhand einheitlicher Personenkriterien erzeugt und zusammen mit den unverschlüsselten Nutzdaten an die Datensammelstelle übermittelt. Alle Datenquellen verwenden das selbe Pseudonymisierungsverfahren und denselben Schlüssel.

Das einstufige Modell "Datenquellen pseudonymisieren" ist für die Praxis ungeeignet.

Unter diesen Voraussetzungen ist die Datenzusammenführung bei der Datensammelstelle gewährleistet, ohne dass der Bezug zu den Individualdaten hergestellt werden kann. Jedoch können einzelne Datenquellen durch Auswertung der zusammengeführten Daten auf die personenbezogenen Inhalte schließen, da das Pseudonymisierungsverfahren bekannt ist. Damit kann das Verfahren vor diesem Hintergrund schon als korrumpiert angesehen werden.

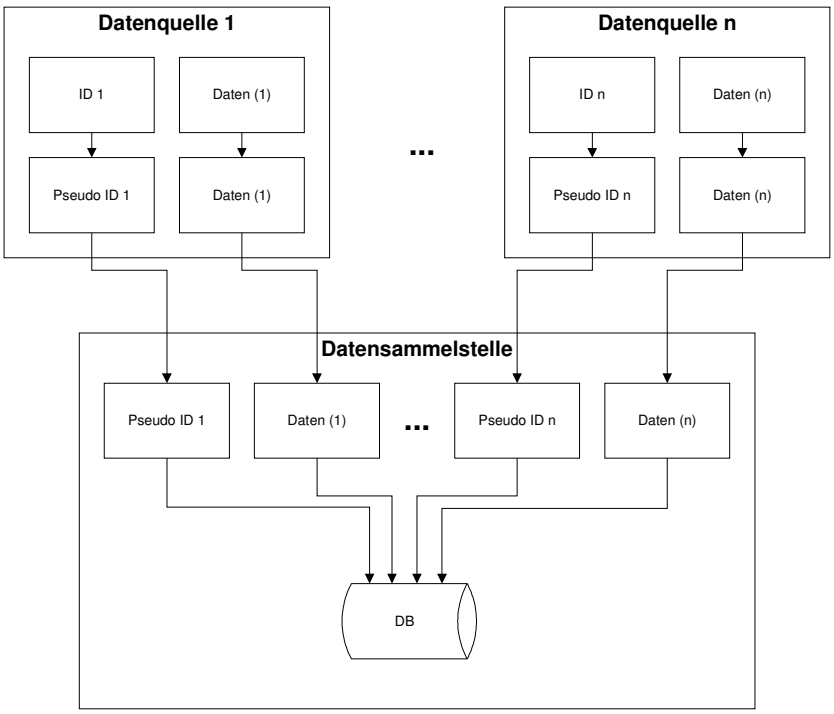


Abbildung 2 Modell P1 – Datenquelle pseudonymisiert

Bewertung:

- Sicherheit

  Vorteil:
  - Die Personendaten sind nicht außerhalb der Datenquelle bekannt.
  - Die Datensammelstelle erhält keine personenbezogenen Daten.

  Nachteil:
  - Der Datensammelstelle ist die Datenquelle bekannt (evtl. Rückschlüsse auf Personendaten möglich).
  - Nur einstufiges Verfahren.
  - Alle Datenquellen pseudonymisieren mit identischem Schlüssel (Gefahr des Bekanntwerdens des Schlüssels).

- Aufwand
  - Organisatorisch / Verfahrensaufbau

    Vorteil:
    - Kein Aufbau einer zusätzlichen Organisation.
    - Keine Online-Fähigkeit und einheitliche Datenbereitstellung durch die Datenquelle erforderlich.

- Administrativ / Betrieb des Verfahrens

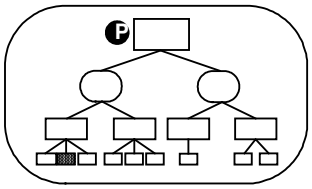
  Vorteil: - Nur einstufiger Datenfluss.

  - Kein Online-Zugang und keine ständige Datenbereithaltung durch die Datenquelle notwendig.

  Nachteil: - Viele Stellen (alle Datenquellen) pseudonymisieren.

Modell	Sicherheit	Aufwand	
		Organisatorisch	Administrativ
P1	☹☹	☺☺	☺

6.1.1.2 Modell P2 – Die Datensammelstelle pseudonymisiert



Die Datenquellen übermitteln die Nutzdaten zusammen mit den einheitlichen Personenkriterien an die Datensammelstelle. Die Nutzdaten liegen in der Datensammelstelle unverschlüsselt vor. Bei der Datensammelstelle wird für jede reale Person ein eindeutiges Pseudonym anhand der einheitlichen Personenkriterien erzeugt.

Die Sicherheit des Modells ist nicht ausreichend, da die Datensammelstelle Kenntnis über die personenbezogenen Daten hat.

Unter diesen Voraussetzungen ist die Datenzusammenführung bei der Datensammelstelle gewährleistet. Den einzelnen Datenquellen ist bei diesem Verfahren ein Rückschluss auf die personenbezogenen Inhalte nicht möglich.

Die Datensammelstelle verfügt damit über alle personenbezogenen Klartextdaten aller Datenquellen.

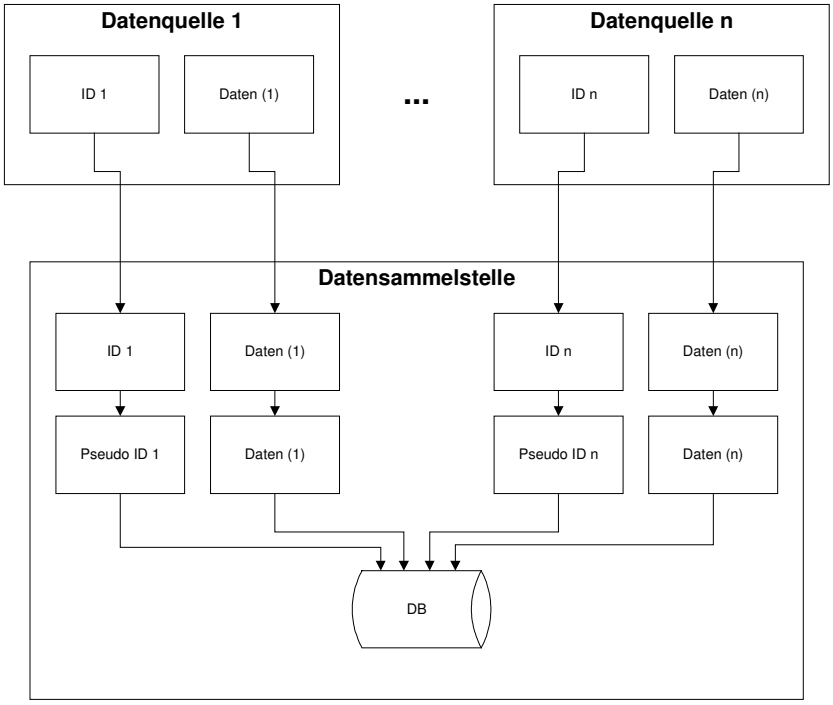


Abbildung 3 Modell P2 – Die Datensammelstelle pseudonymisiert

Bewertung:

- Sicherheit

  Nachteil: - Die Personendaten sind außerhalb der Datenquellen bekannt.
  - Die Datensammelstelle erhält die personenbezogenen Daten.
  - Der Datensammelstelle sind die Datenquellen bekannt.

- Aufwand
  - Organisatorisch / Verfahrensaufbau

    Vorteil: - Kein Aufbau einer zusätzlichen Organisation.
    - Keine Online-Fähigkeit und einheitliche Datenbereitstellung durch die Datenquellen erforderlich.

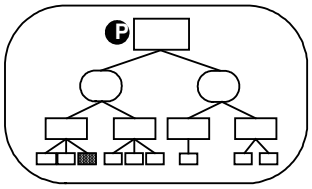
  - Administrativ / Betrieb des Verfahrens

    Vorteil: - Nur einstufiger Datenfluss.

- Zentrale Pseudonymisierung
- Kein Online-Zugang und keine ständige Datenbereithaltung durch die Datenquellen notwendig.

Modell	Sicherheit	Aufwand	
		Organisatorisch	Administrativ
P2	☹☹	☺☺	☺☺

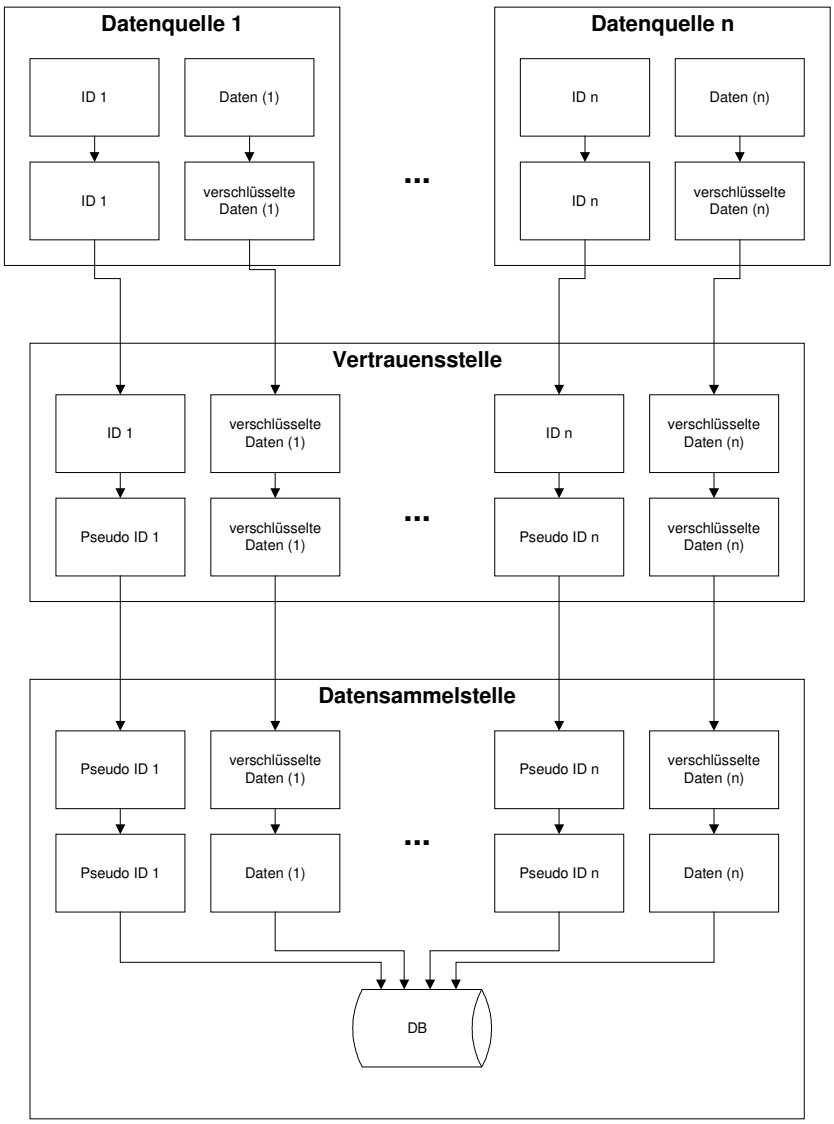
6.1.1.3 Modell P3 – Eine Vertrauensstelle pseudonymisiert



Die Datenquellen übermitteln die verschlüsselten Nutzdaten zusammen mit den unverschlüsselten einheitlichen Personenkriterien an eine unabhängige Vertrauensstelle. Die Vertrauensstelle erzeugt aus eindeutigen Personenkriterien für jede reale Person ein eindeutiges Pseudonym, welches zusammen mit den verschlüsselten Nutzdaten an die Datensammelstelle weitergeleitet wird

In der Datensammelstelle werden die Nutzdaten zu den Pseudonymen entschlüsselt, zusammengeführt und damit auswertbar gemacht. Unter diesen Voraussetzungen ist die Datenzusammenführung bei der Datensammelstelle gewährleistet, ohne dass der Bezug zu den Individualdaten hergestellt werden kann. Auch die einzelnen Datenquellen sind nicht in der Lage auf die personenbezogenen Inhalte zu schließen. Nur die Vertrauensstelle kann eine Depseudonymisierung durchführen.

Durch Einschaltung einer Vertrauensstelle ist in diesem Modell zwar grundsätzlich ein geeigneter Schutz der Personendaten möglich, jedoch sind personenbezogene Daten außerhalb der Datenquelle bekannt.



Die Vertrauensstelle erhält vom Dateninhalt keine Kenntnis.

Abbildung 4 Modell P3 – Eine Vertrauensstelle pseudonymisiert

Bewertung:

- Sicherheit

  Vorteil:
  - Die Datensammelstelle erhält keine personenbezogenen Daten.
  - Der Datensammelstelle sind die Datenquellen nicht bekannt (keine Rückschlüsse auf Personendaten möglich).
  - Zentrale Pseudonymisierung durch eine Vertrauensstelle (Schlüssel nur an einer Stelle bekannt)

Nachteil: - Die Personendaten sind außerhalb der Datenquelle bekannt.

- Nur einstufiges Verfahren.

- Aufwand
  - Organisatorisch / Verfahrensaufbau

    Vorteil: - Keine Online-Fähigkeit und einheitliche Datenbereitstellung durch die Datenquelle erforderlich.

    Nachteil: - Aufbau einer zusätzlichen Organisation.

  - Administrativ / Betrieb des Verfahrens

    Vorteil: - Zentrale Pseudonymisierung.

    - Kein Online-Zugang und keine ständige Datenbereithaltung durch die Datenquelle notwendig.

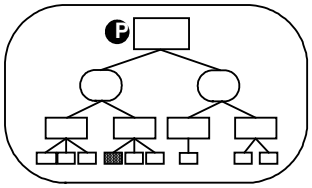
    Nachteil: - Zweistufiger Datenfluss

Modell	Sicherheit	Aufwand	
		Organisatorisch	Administrativ
P3	😐	☺	☺

6.1.2 Zweistufige Modelle

Nachfolgend wird davon ausgegangen, dass pro Stelle nur ein Pseudonymisierungsprozess durchlaufen wird, da eine zweistufige Pseudonymisierung innerhalb einer Stelle keine Änderung in der Bewertung bedeutet.

6.1.2.1 Modell P4 – Zweistufiger Pseudonymisierungsprozess ohne Vertrauensstelle



Bei den Datenquellen wird für jede reale Person ein eindeutiges Pseudonym anhand einheitlicher Personenkriterien erzeugt und zusammen mit den unverschlüsselten

Der zweistufige Pseudonymisierungsprozess ohne Vertrauensstelle ist bzgl. der Sicherheit angreifbar.

Nutzdaten an die Datensammelstelle übermittelt. Alle Datenquellen verwenden dasselbe Pseudonymisierungsverfahren.

Die Datensammelstelle pseudonymisiert die Pseudonyme und führt die Nutzdaten zusammen. Weder die Datensammelstelle noch die Datenquellen können jeweils für sich die zweistufige Pseudonymisierung aufheben, da das jeweils andere Pseudonymisierungsverfahren nicht bekannt ist (vgl. auch das im Abschnitt 4.1.1 beschriebene DAPI-Verfahren).

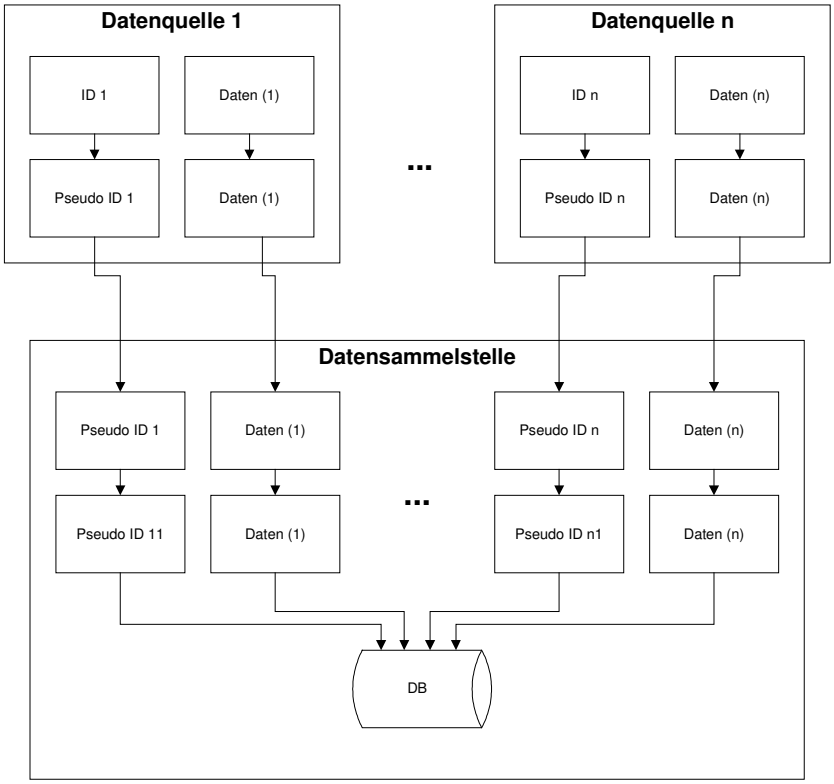


Abbildung 5 Modell P4 – Zweistufige Pseudonymisierung ohne Vertrauensstelle

Bewertung:

- Sicherheit

  Vorteil: - Die Personendaten sind nicht außerhalb der Datenquelle bekannt.
  - Die Datensammelstelle erhält keine personenbezogenen Daten.
  - Zweistufiges Verfahren

Nachteil: - Der Datensammelstelle ist die Datenquelle bekannt (evtl. Rückschlüsse auf Personendaten möglich).

- Alle Datenquellen pseudonymisieren mit identischem Schlüssel (Gefahr des Bekanntwerdens des Schlüssels).

- Aufwand
  - Organisatorisch / Verfahrensaufbau

    Vorteil: - Kein Aufbau einer zusätzlichen Organisation.

    - Keine Online-Fähigkeit und einheitliche Datenbereitstellung durch die Datenquelle erforderlich).

  - Administrativ / Betrieb des Verfahrens

    Vorteil: - Nur einstufiger Datenfluss.

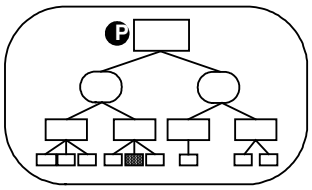
    - Kein Online-Zugang und keine ständige Datenbereithaltung durch die Datenquelle notwendig.

    Nachteil: - Viele Stellen (alle Datenquellen) pseudonymisieren.

Modell	Sicherheit	Aufwand	
		Organisatorisch	Administrativ
P4	😐	☺☺	☺

Bei Verwendung von Einweg-Pseudonymisierungsverfahren siehe Ausführungen unter Kapitel 6.6.1

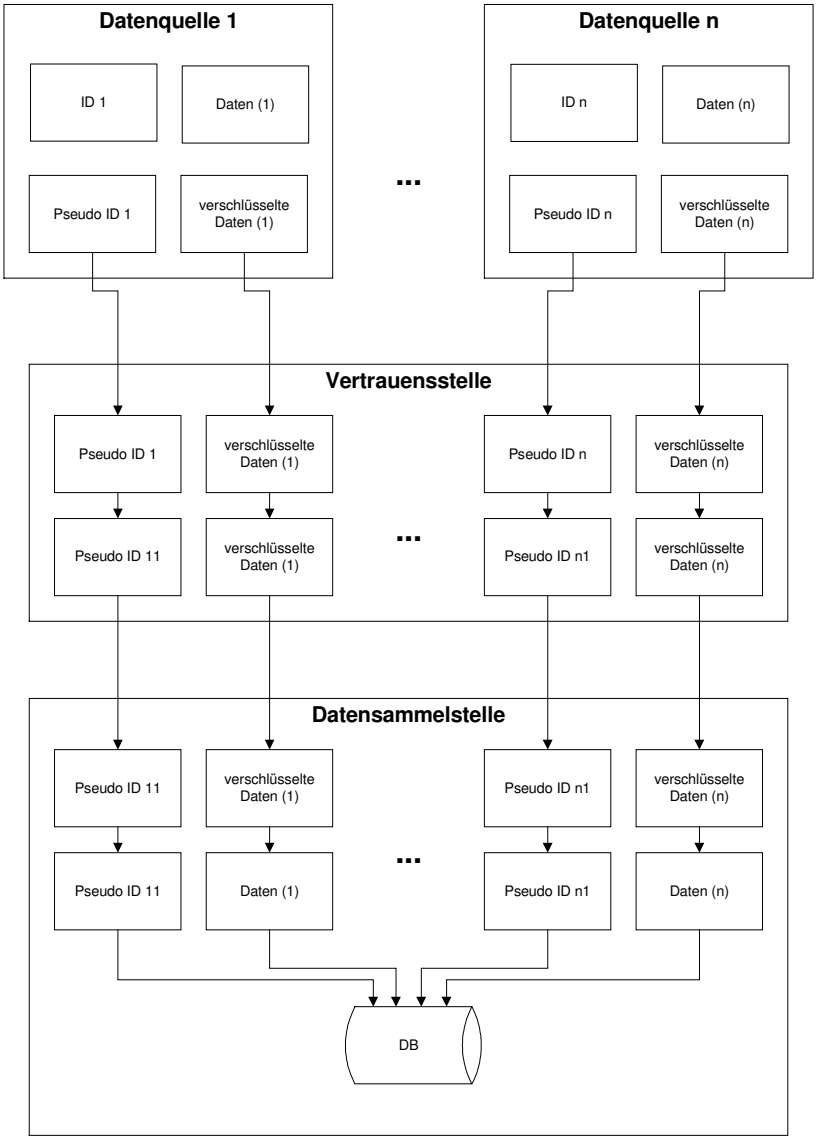
6.1.2.2 Modell P5 – Zweistufiger Pseudonymisierungsprozess mit einer Vertrauensstelle



a) Pseudonymisierung durch Datenquelle und Vertrauensstelle

Bei den Datenquellen wird für jede reale Person ein eindeutiges Pseudonym anhand einheitlicher Personenkriterien erzeugt und zusammen mit den verschlüsselten Nutzdaten an die Vertrauensstelle übermittelt. Die Vertrauensstelle erzeugt ein weiteres eindeutiges Pseudonym, welches zusammen mit den verschlüsselten Nutzdaten an die Datensammelstelle weitergeleitet wird. Dort werden die Nutzdaten entschlüsselt, und anhand der neuen Pseudonyme zusammengeführt und damit auswertbar gemacht.
Damit entspricht dieses Modell in seinen Konsequenzen Modell P4

P5a stellt ein sicheres Modell mit höherem organisatorischen und administrativen Aufwand dar.



Die Vertrauensstelle erhält vom Dateninhalt keine Kenntnis.

Abbildung 6 Modell P5a – Zweistufige Pseudonymisierung mit einer Vertrauensstelle (Pseudonymisierung durch Datenquelle und Vertrauensstelle)

Bewertung:

- Sicherheit

  Vorteil:
  - Die Personendaten sind nicht außerhalb der Datenquelle bekannt.
  - Die Datensammelstelle erhält keine personenbezogenen Daten.
  - Der Datensammelstelle ist die Datenquelle nicht bekannt (Keine Rückschlüsse auf Personendaten möglich).
  - Zweistufiges Verfahren

Nachteil: - Alle Datenquellen pseudonymisieren mit identischem Schlüssel (Gefahr des Bekanntwerdens des Schlüssels).

- Aufwand
  - Organisatorisch / Verfahrensaufbau

    Vorteil: - Keine Online-Fähigkeit und einheitliche Datenbereitstellung durch die Datenquellen erforderlich.

    Nachteil: - Aufbau einer zusätzlichen Organisation.
  - Administrativ / Betrieb des Verfahrens

    Vorteil: - Kein Online-Zugang und keine ständige Datenbereithaltung durch die Datenquellen notwendig.

    Nachteil: - Zweistufiger Datenfluss

    - Viele Stellen (alle Datenquellen) pseudonymisieren.

Modell	Sicherheit	Aufwand	
		Organisatorisch	Administrativ
P5a	☺	☺	😐

Bei Verwendung von Einweg-Pseudonymisierungsverfahren siehe Ausführungen unter Kapitel 6.6.1

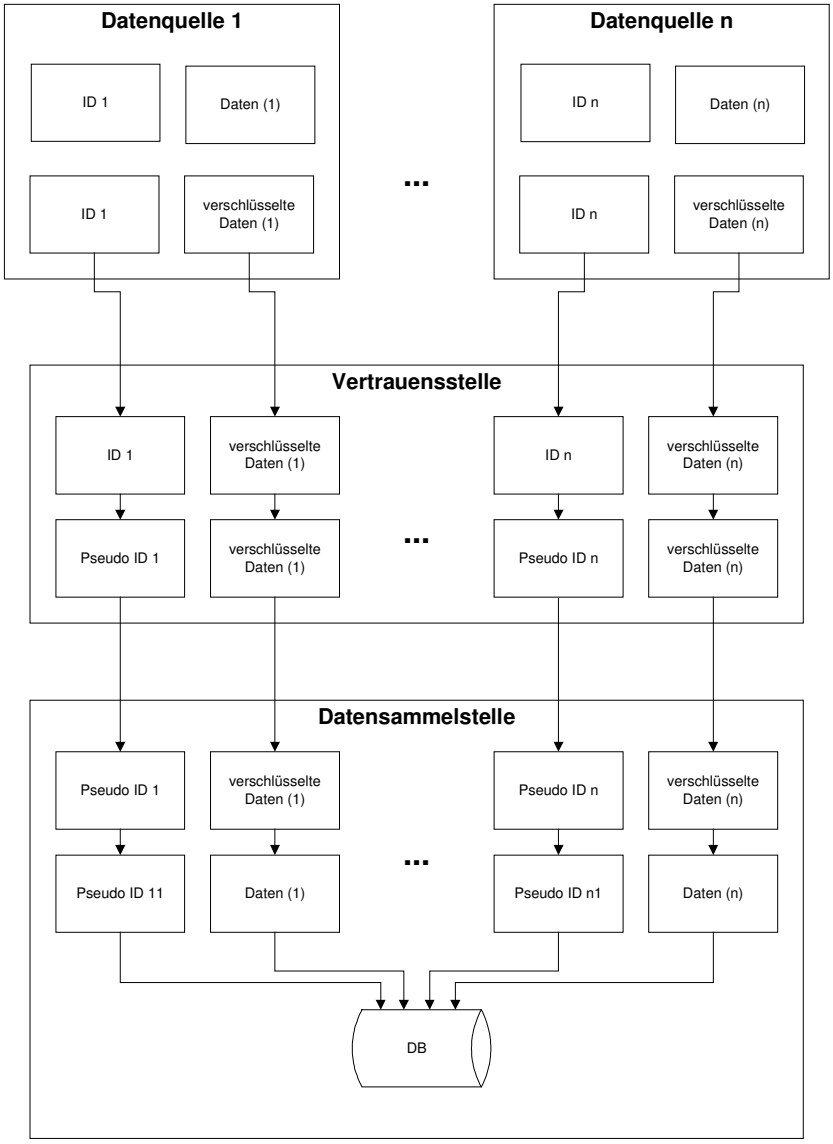
b) Pseudonymisierung durch Vertrauensstelle und Datensammelstelle

P5b ist eine sichere Modellvariante mit dem Nachteil, dass die Vertrauensstelle Kenntnis der Personendaten hat.

Die Datenquellen übermitteln die verschlüsselten Nutzdaten zusammen mit den unverschlüsselten einheitlichen Personenkriterien an eine unabhängige Vertrauensstelle. Die Vertrauensstelle erzeugt aus den eindeutigen Personenkriterien für jede reale Person ein eindeutiges Pseudonym, welches zusammen mit den verschlüsselten Nutzdaten an die Datensammelstelle weitergeleitet wird. Diese pseudonymisiert die Pseudonyme erneut, entschlüsselt die Nutzdaten, führt sie zusammen und macht sie damit auswertbar.

Bei diesem Verfahren bestehen im Falle einer Depseudonymisierung zwei definierte Ansprechpartner. Die Vertrauensstelle verfügt über alle personenidentifizierenden

Daten aller Datenquellen. Die Nutzdaten liegen der Vertrauensstelle in verschlüsselter Form vor.



Die Vertrauensstelle erhält vom Dateninhalt keine Kenntnis.

Abbildung 7 Modell P5b – Zweistufige Pseudonymisierung mit einer Vertrauensstelle (Pseudonymisierung durch Vertrauensstelle und Datensammelstelle)

Bewertung:

- Sicherheit

  Vorteil: - Datensammelstelle erhält keine personenbezogenen Daten.

- Der Datensammelstelle ist die Datenquelle nicht bekannt (keine Rückschlüsse auf Personendaten möglich).
- Zweistufiges Verfahren
- Die Pseudonymisierung erfolgt zentral durch zwei Vertrauensstellen (geringe Gefahr, dass Schlüssel bekannt wird).

Nachteil: - Personendaten sind außerhalb der Datenquellen bekannt.

- Aufwand
  - Organisatorisch / Verfahrensaufbau

    Vorteil: - Keine Online-Fähigkeit und einheitliche Datenbereitstellung durch die Datenquellen erforderlich.

    Nachteil: - Aufbau einer zusätzlichen Organisationen
  - Administrativ / Betrieb des Verfahrens

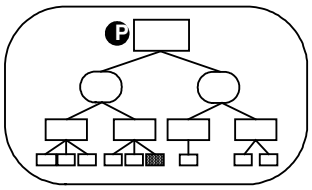
    Vorteil: - Zentrale Pseudonymisierung
    - Kein Online-Zugang und keine ständige Datenbereithaltung durch die Datenquellen notwendig.

    Nachteil: - Zweistufiger Datenfluss.

Modell	Sicherheit	Aufwand	
		Organisatorisch	Administrativ
P5b	☺	☺	☺

Siehe auch Ausführungen unter Kapitel 7.6.1.

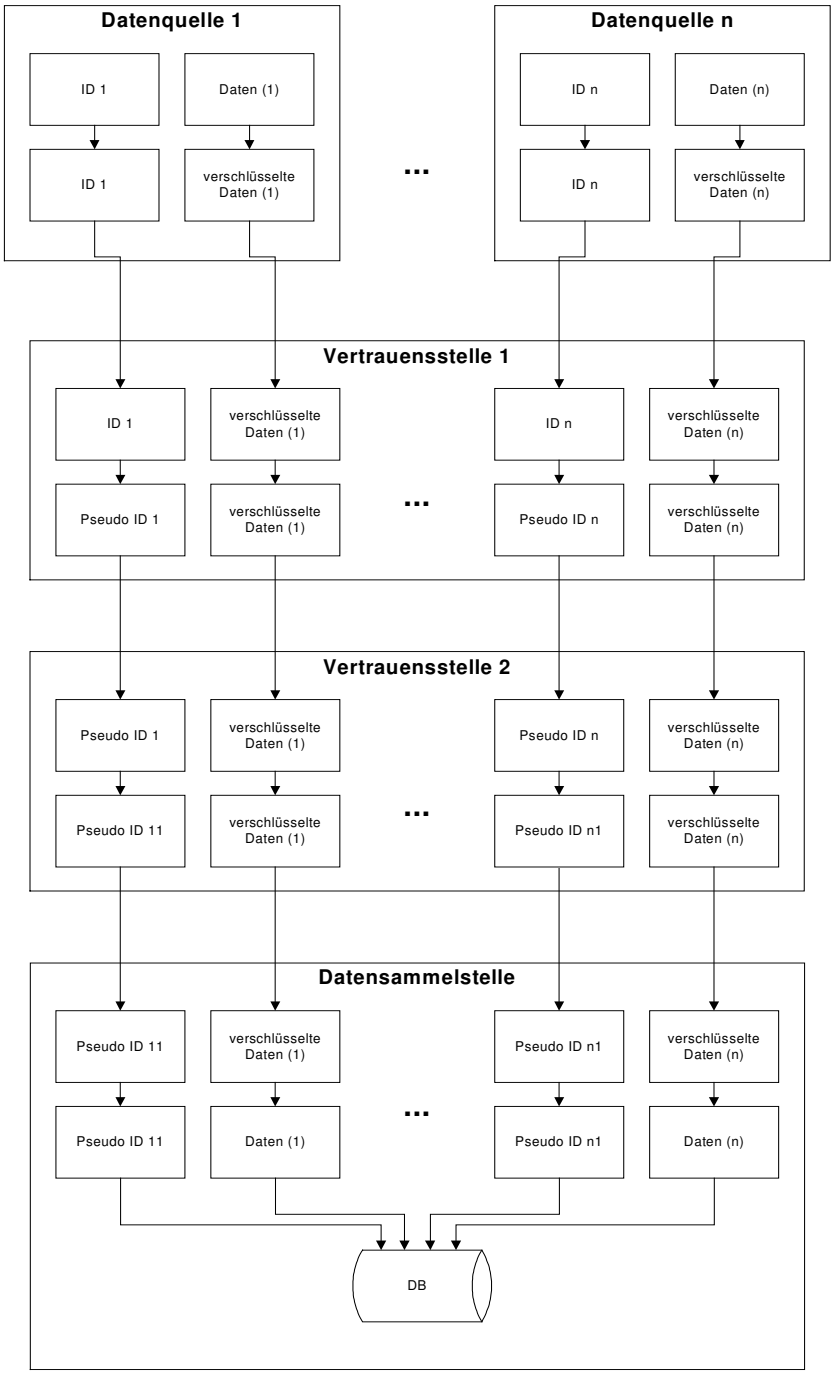
6.1.2.3 Modell P6 – Zweistufiger Pseudonymisierungsprozess mit zwei Vertrauensstellen



Eine sichere Modellvariante mit höherem Aufwand und dem Nachteil, dass eine Vertrauensstelle Kenntnis der Personendaten hat.

Die Datenquellen übermitteln die verschlüsselten Nutzdaten zusammen mit den unverschlüsselten einheitlichen Personenkriterien an eine unabhängige Vertrauensstelle. Die Vertrauensstelle erzeugt aus den eindeutigen Perso-

nenkriterien für jede reale Person ein eindeutiges Pseudonym, welches zusammen mit den verschlüsselten Nutzdaten an eine zweite Pseudonymisierungsstelle weitergeleitet wird. Diese pseudonymisiert die Pseudonyme erneut und leitet die Daten an die Datensammelstelle weiter. Diese entschlüsselt die Nutzdaten, führt sie zusammen und macht sie damit auswertbar.



Die Vertrauensstelle erhält vom Dateninhalt keine Kenntnis.

Eine Depseudonymisierung ist nur unter Einschaltung beider Vertrauensstellen möglich (vier-Augen-Prinzip).

Abbildung 8 Modell P6 – Zweistufige Pseudonymisierung mit zwei Vertrauensstellen

Bewertung:

- Sicherheit

  Vorteil:
  - Die Datensammelstelle erhält keine personenbezogenen Daten.
  - Der Datensammelstelle sind die Datenquellen nicht bekannt (keine Rückschlüsse auf Personendaten möglich).
  - Zweistufiges Verfahren.
  - Die Pseudonymisierung erfolgt zentral durch zwei Vertrauensstellen (geringe Gefahr, dass Schlüssel bekannt wird).

  Nachteil:
  - Personendaten sind außerhalb der Datenquellen bekannt.

- Aufwand
  - Organisatorisch / Verfahrensaufbau

    Vorteil:
    - Keine Online-Fähigkeit und einheitliche Datenbereitstellung durch die Datenquelle erforderlich .

    Nachteil:
    - Aufbau zweier zusätzlicher Organisationen.

- Administrativ / Betrieb des Verfahrens

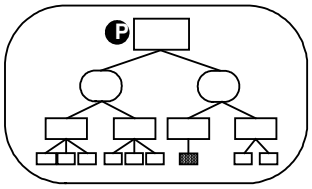
  Vorteil:
  - Zentrale Pseudonymisierung
  - Kein Online-Zugang und keine ständige Datenbereitstellung durch die Datenquellen notwendig.

  Nachteil:
  - Dreistufiger Datenfluss

Modell	Sicherheit	Aufwand	
		Organisatorisch	Administrativ
P6	☺	😐	😐

6.2 Pseudonymisierung bei dezentraler Datenhaltung

6.2.1 Modell P7 – Einstufige Pseudonymisierung an der Datenquelle



Die datenanfordernde Stelle richtet in diesem Modell ihre Anfrage an mehrere datenhaltende Stellen und erhält von diesen jeweils die entsprechenden Ergebnismengen.

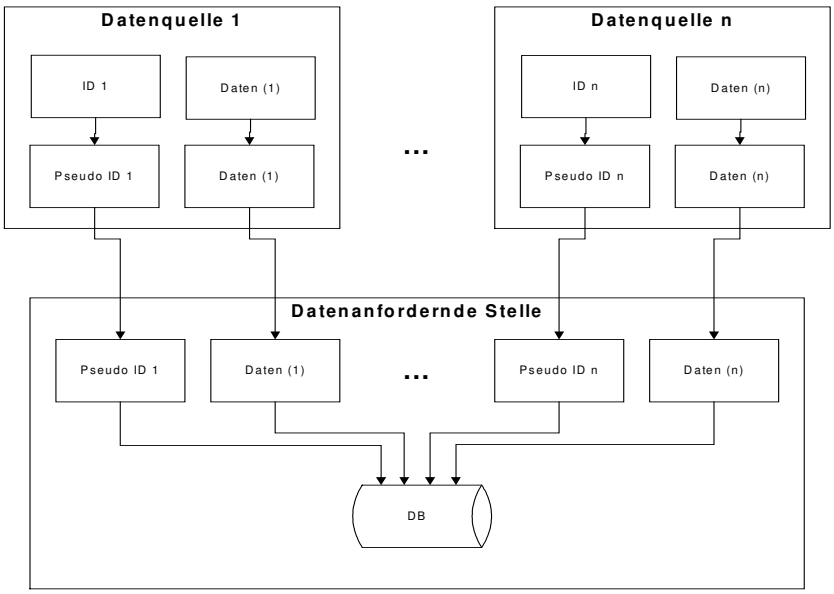
Die einstufige Pseudonymisierung an der Datenquelle ist kein sicheres Modell.

Die Synchronisation der Ergebnismengen und damit die Zusammenführung der Daten obliegt der datenanfordernden Stelle.

Die Pseudonymisierung des Personenbezuges wird in diesem Modell in den jeweiligen Datenquellen vorgenommen.

Alle Datenquellen verwenden das selbe Pseudonymisierungsverfahren und denselben Schlüssel.

Jedoch können einzelne Datenquellen durch Auswertung der zusammengeführten Daten auf die personenbezogenen Inhalte schließen, da das Pseudonymisierungsverfahren bekannt ist. Damit kann das Verfahren vor diesem Hintergrund schon als korrumpiert angesehen werden.



Die datenanfordernde Stelle kann auch eine der Datenquellen sein.

Abbildung 9 Modell P7 – Einstufige Pseudonymisierung an den Datenquellen

Bewertung:

- Sicherheit

  Vorteil: - Die Personendaten sind nicht außerhalb der Datenquelle bekannt.

  - Die datenanfordernde Stelle erhält keine personenbezogenen Daten.

  Nachteil: - Der datenanfordernden Stelle ist die Datenquelle bekannt (evtl. Rückschlüsse auf Personendaten möglich).

  - Nur einstufiges Verfahren.

  - Alle Datenquellen pseudonymisieren mit identischem Schlüssel (Gefahr des Bekanntwerdens des Schlüssels).

- Aufwand
  - Organisatorisch / Verfahrensaufbau

    Vorteil: - Kein Aufbau einer zusätzlichen Organisation.

    Nachteil: - Alle Datenquellen müssen "onlinefähig" gemacht werden und die relevanten Daten per einheitlichem Modell zum Abruf verfügbar halten.

- Administrativ / Betrieb des Verfahrens

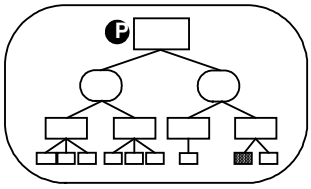
  Vorteil: - Nur einstufiger Datenfluss.

  Nachteil: - Viele Stellen (alle Datenquellen) pseudonymisieren.

  - Alle Datenquellen müssen online sein und die Daten zum Abruf bereit halten.

Modell	Sicherheit	Aufwand	
		Organisatorisch	Administrativ
P7	☹☹	☹	😐

6.2.2 Modell P8 – Zweistufige Pseudonymisierung



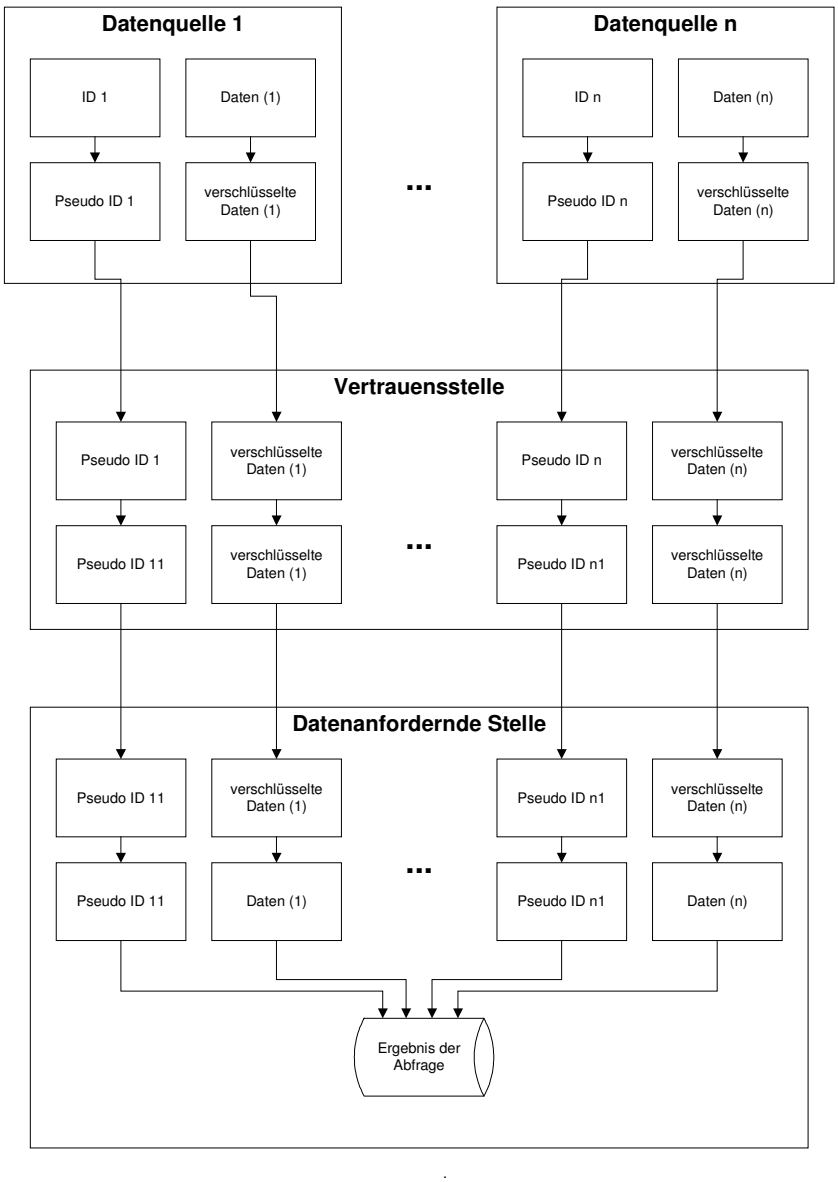
In diesem Modell richtet die datenanfordernde Stelle ihre Anfrage nicht an mehrere datenhaltende Stellen sondern an eine vorgeschaltete Vertrauensstelle. Diese leitet die Anfrage an die einzelnen Datenquellen weiter und erhält pseudonymisierte Ergebnismengen zurück. Diese Ergebnisse werden in der Folge mit einem anderen Pseudonymisierungsverfahren erneut pseudonymisiert, bevor sie an die datenanfordernde Stelle gesandt werden. Durch die Vertrauensstelle wird für eine Pseudonymisierung jedes Mal ein anderer Schlüssel verwendet.

Die zweistufige Pseudonymisierung (durch Datenquellen und Vertrauensstelle) bei dezentraler Datenhaltung ist ein sicheres Verfahren mit erhöhtem Aufwand, da es permanente einheitliche Datenbereitstellung durch die einzelnen Datenquellen erfordert.

Das Zusammenführen der Ergebnismengen kann nach wie vor durch die datenanfordernde Stelle erfolgen.

Zwar verwenden beim hier skizzierten Verfahren alle Datenquellen dasselbe Pseudonymisierungsverfahren und denselben Schlüssel, durch das wechselnde Pseudonymisierungsverfahren bei der Vertrauensstelle ist die Wiederherstellung des Personenbezuges jedoch nicht mehr möglich.

Dieses Vorgehen hätte allerdings zur Folge, dass eine Datenquelle, wollte sie ihre eigenen Daten mit den Daten anderer Datenquellen zusammenführen, bei einer entsprechenden Abfrage auch ihre eigenen Daten mit erfragen müsste.



Die Vertrauensstelle erhält vom Dateninhalt keine Kenntnis.

Die datenanfordernde Stelle kann auch eine der Datenquellen sein.

Abbildung 10 Modell P8 – Zweistufige Pseudonymisierung bei dezentraler Datenhaltung

Obige Abbildung verdeutlicht den Datenfluss im Falle einer Antwort von zwei Datenquellen an eine datenanfordernde Stelle. Im obigen Fall liefern die Quellen jeweils unterschiedliche Nutzdaten (Daten 1 und Daten 2) zur selben Person, die mit Pseudo-ID 1 gekennzeichnet ist. Die vertrauenswürdige Stelle pseudonymisiert das jeweiligen Pseudonym (Pseudo-ID 1) jeweils für jede Sitzung neu, so dass die datenanfordernde Stelle ein sitzungsgebundenes Pseudonym in der Ergebnismenge (hier: Pseudo-ID 11) erhält.

Bewertung:

- Sicherheit

  Vorteil:
  - Die Personendaten sind nicht außerhalb der Datenquelle bekannt.
  - Die datenanfordernde Stelle erhält keine personenbezogenen Daten.
  - Der datenanfordernden Stelle ist die Datenquelle nicht bekannt (keine Rückschlüsse auf Personendaten möglich).
  - Zweistufiges Verfahren

  Nachteil:
  - Alle Datenquellen pseudonymisieren mit identischem Schlüssel (Gefahr des Bekanntwerdens des Schlüssel).

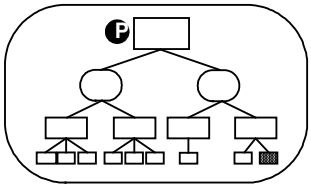
- Aufwand
  - Organisatorisch / Verfahrensaufbau

    Nachteil:
    - Aufbau einer zusätzlichen Organisation.
    - Alle Datenquellen müssen "onlinefähig" gemacht werden und die relevanten Daten per einheitlichem Modell zum Abruf verfügbar halten.
  - Administrativ / Betrieb des Verfahrens

    Nachteil:
    - Zweistufiger Datenfluss
    - Viele Stellen (alle Datenquellen) pseudonymisieren.
    - Alle Datenquellen müssen online sein und die Daten zum Abruf bereit halten.

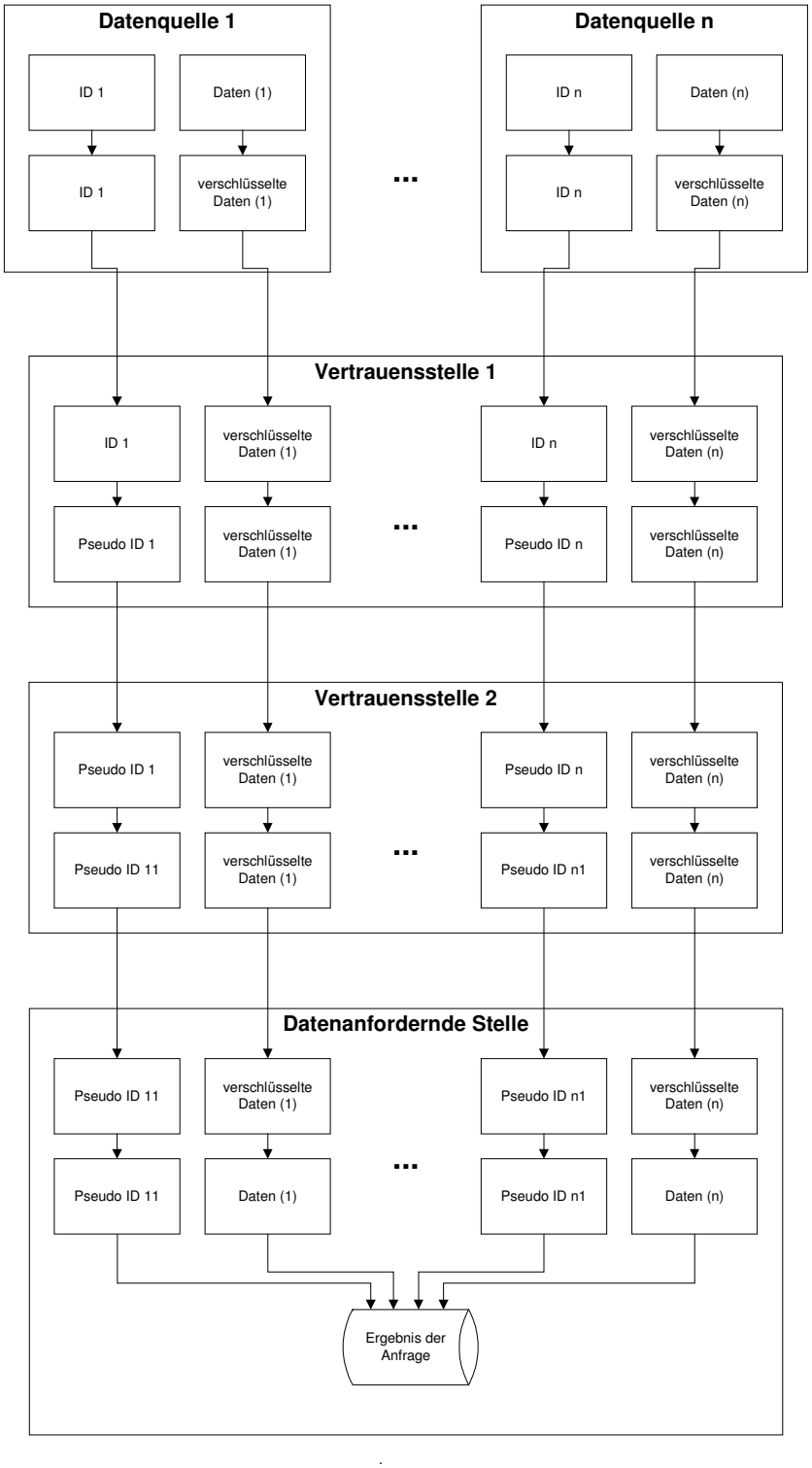
Modell	Sicherheit	Aufwand	
		Organisatorisch	Administrativ
P8	☺	☹☹	☹

6.2.3 Modell P9 – Zweistufige Pseudonymisierung durch zwei Vertrauensstellen



Die Datenquellen übermitteln die verschlüsselten Nutzdaten zusammen mit den unverschlüsselten einheitlichen Personenkriterien an eine unabhängige Vertrauensstelle. Die Vertrauensstelle erzeugt aus den eindeutigen Personenkriterien für jede reale Person ein eindeutiges Pseudonym, welches zusammen mit den verschlüsselten Nutzdaten an eine zweite Pseudonymisierungsstelle weitergeleitet wird. Diese pseudonymisiert die Pseudonyme erneut und leitet die Daten an die datenanfordernde Stelle weiter. Diese entschlüsselt die Nutzdaten, führt sie zusammen und macht sie damit auswertbar.

Die zweistufige Pseudonymisierung (durch zwei Vertrauensstellen) bei dezentraler Datenhaltung ist ein sicheres Verfahren mit stark erhöhtem Aufwand.



Die Vertrauensstellen erhalten vom Dateninhalt keine Kenntnis.

Abbildung 11 Modell P9 – Zweistufige Pseudonymisierung, zwei Vertrauensstellen, dezentrale Datenhaltung

Bewertung:

- Sicherheit

  Vorteil: - Die datenanfordernde Stelle erhält keine personenbezogenen Daten.

  - Der datenanfordernden Stelle ist die Datenquelle nicht bekannt (keine Rückschlüsse auf Personendaten möglich).

  - Zweistufiges Verfahren

  - Die Pseudonymisierung erfolgt mit unterschiedlichen Schlüsseln (geringe Gefahr, dass Schlüssel bekannt wird).

  Nachteil: - Die Personendaten sind außerhalb der Datenquelle bekannt.

- Aufwand
  - Organisatorisch / Verfahrensaufbau

    Nachteil: - Aufbau zweier zusätzlicher Organisationen.

    - Alle Datenquellen müssen "online-fähig" gemacht werden und die relevanten Daten per einheitlichem Modell zum Abruf verfügbar halten.

  - Administrativ / Betrieb des Verfahrens

    Vorteil: -Zentrale Pseudonymisierung

    Nachteil: - Dreistufiger Datenfluss

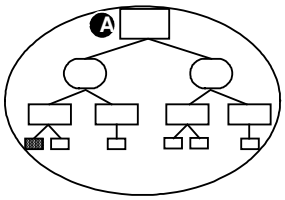
    - Alle Datenquellen müssen online sein und die Daten zum Abruf bereit halten.

Modell	Sicherheit	Aufwand	
		Organisatorisch	Administrativ
P9	☺	☹☹	☹

6.3 Anonymisierung bei zentraler Datenhaltung

6.3.1 Einstufige Modelle

6.3.1.1 Modell A1 – Die Datenquellen anonymisieren



Bei den Datenquellen wird für jede reale Person ein Anonym erzeugt und zusammen mit den unverschlüsselten Nutzdaten an die Datensammelstelle übermittelt. Die Anwendung einheitlicher Verfahren bei der Bildung des Anonyms ist dabei nicht erforderlich.

Modell ist durch die Kenntnis des Absenders der Datensammelstelle als nur eingeschränkt sicher zu bezeichnen.

Bei diesem Modell hat die Datensammelstelle zu keiner Zeit Zugang zum von der Anonymisierung betroffenen Personenbezug, kennt jedoch den Absender der Daten.

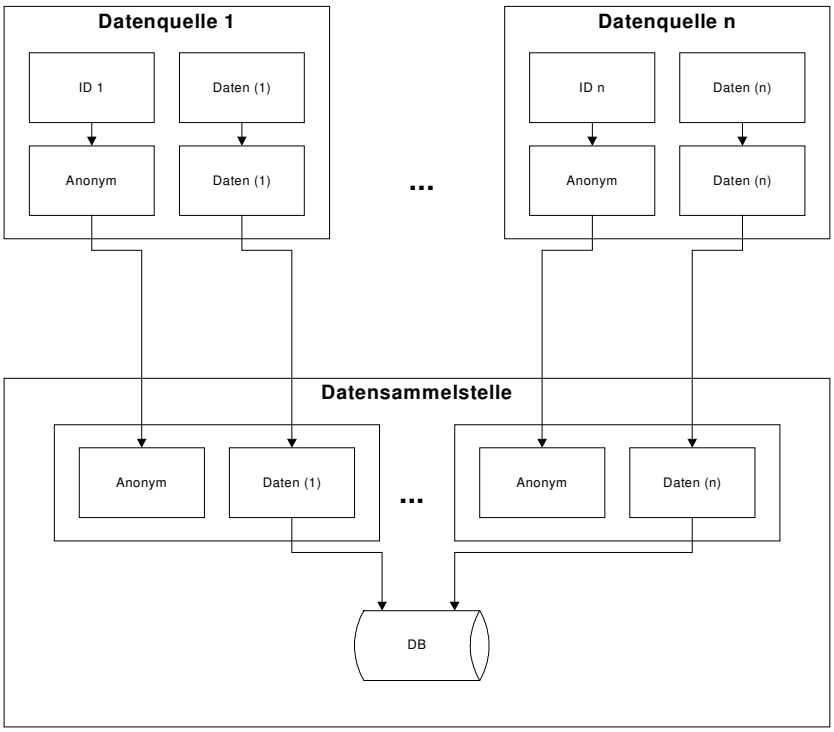


Abbildung 12 Modell A1 – Zentrale Datenhaltung, Datenquellen anonymisieren

Bewertung:

- Sicherheit

  Vorteil: - Die Personendaten sind nicht außerhalb der Datenquelle bekannt.

  - Die Datensammelstelle erhält keine personenbezogenen Daten.

  Nachteil: - Der Datensammelstelle ist die Datenquelle bekannt (evtl. Rückschlüsse auf Personendaten möglich).

- Aufwand
  - Organisatorisch / Verfahrensaufbau

    Vorteil: - Kein Aufbau einer zusätzlichen Organisation.

    - Keine Online-Fähigkeit und einheitliche Datenbereitstellung durch die Datenquelle erforderlich.

  - Administrativ / Betrieb des Verfahrens

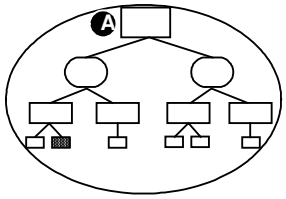
    Vorteil: - Nur einstufiger Datenfluss

    - Kein Online-Zugang und keine ständige Datenbereithaltung durch die Datenquellen notwendig.

    Nachteil: - Viele Stellen (alle Datenquellen) anonymisieren.

Modell	Sicherheit	Aufwand	
		Organisatorisch	Administrativ
A1	☺	☺☺	☺

6.3.1.2 Modell A2 – Die Datensammelstelle anonymisiert



Die Datenquellen übermitteln die unverschlüsselten Nutzdaten personenbezogen an die Datensammelstelle. Bei der Datensammelstelle wird für jede reale Person ein Anonym erzeugt. Für weitergehende Auswertungen steht hierdurch der Personenbezug nicht zur Verfügung.

Modell A2 stellt ein unsicheres Verfahren dar, da die Datensammelstelle auf alle Daten inklusive der Personendaten Zugriff hat.

Die Datensammelstelle verfügt damit – zumindest temporär – über alle personenbezogenen Klartextdaten aller Datenquellen.

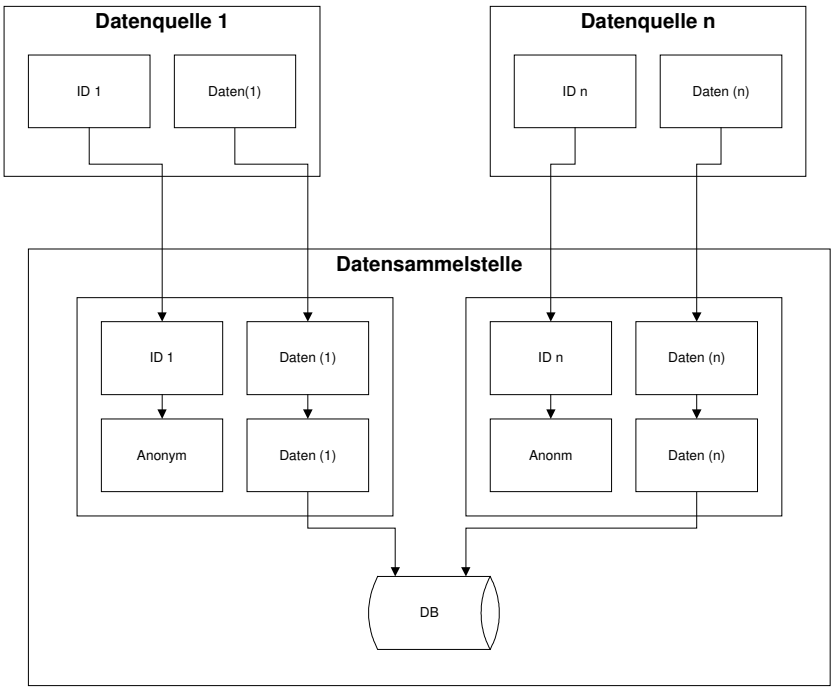


Abbildung 13 Modell A2 – Zentrale Datenhaltung, Datensammelstelle anonymisiert

Bewertung:

- Sicherheit

  Nachteil:
  - Die Personendaten sind außerhalb der Datenquelle bekannt.
  - Die Datensammelstelle erhält die personenbezogenen Daten.
  - Der Datensammelstelle ist die Datenquelle bekannt.

- Aufwand
  - Organisatorisch / Verfahrensaufbau

    Vorteil: - Kein Aufbau einer zusätzlichen Organisation

    - Keine Online-Fähigkeit und einheitliche Datenbereitstellung durch die Datenquelle erforderlich.
  - Administrativ / Betrieb des Verfahrens

    Vorteil: - Nur einstufiger Datenfluss.

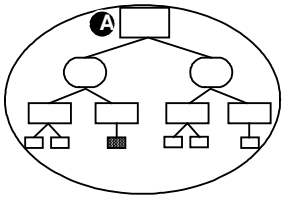
    - Zentrale Anonymisierung.

    - Kein Online-Zugang und keine ständige Datenbereithaltung durch die Datenquelle notwendig.

Modell	Sicherheit	Aufwand	
		Organisatorisch	Administrativ
A2	☹☹	☺☺	☺☺

6.3.2 Zweistufige Modelle

6.3.2.1 Modell A3 – Datenquelle(n) und Vertrauensstelle anonymisieren



Modell A3 stellt ein uneingeschränkt sicheres Verfahren dar.

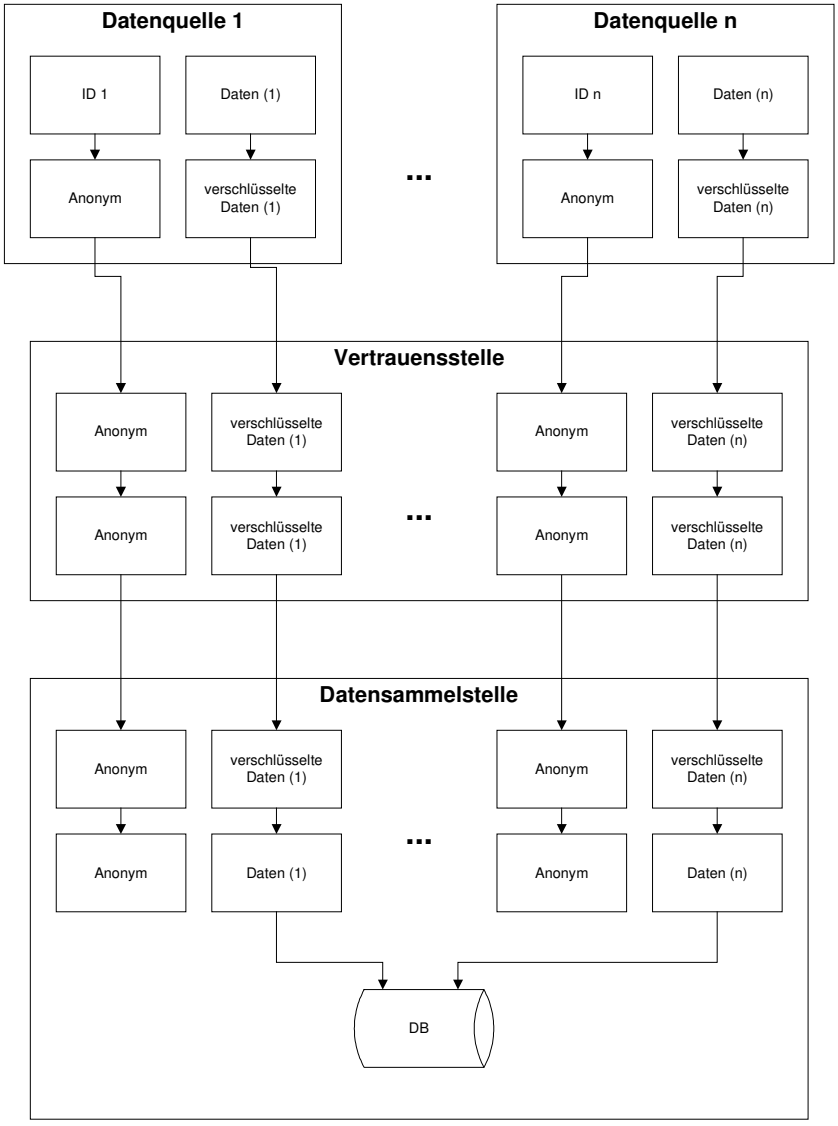
Bei diesem Modell nimmt die Datenquelle eine Anonymisierung vereinbarter Inhalte vor und leitet diese mit den verschlüsselten Nutzdaten an eine Vertrauensstelle weiter.

Die Vertrauensstelle ihrerseits anonymisiert den Absenderbezug und leitet die Daten an die Datensammelstelle weiter.

Sofern die Anonymisierung der Absenderangaben nicht ausreicht (z.B. weil das Volumen der Daten Rückschlüsse auf den Absender zulässt), können in der Vertrauensstelle

weitergehende Maßnahmen zur Herstellung einer vollständigen Anonymisierung getroffen werden.

Bei diesem Modell hat die Datensammelstelle zu keiner Zeit Zugang zum von der Anonymisierung betroffenen Personenbezug und hat auch keine Kenntnis vom Absender der Daten (vgl. auch das im Abschnitt 4.1.2 beschriebene GAmSi-Verfahren).



Die Vertrauensstelle anonymisiert den Absenderbezug.

Abbildung 14 Modell A3 – Zentrale Datenhaltung, Datenquelle und Vertrauensstelle anonymisieren

Bewertung:

- Sicherheit

  Vorteil: - Die Personendaten sind nicht außerhalb der Datenquelle bekannt.
  - Die Datensammelstelle erhält keine personenbezogenen Daten.
  - Der Datensammelstelle ist die Datenquelle nicht bekannt (keine Rückschlüsse auf Personendaten möglich.

- Aufwand
  - Organisatorisch / Verfahrensaufbau

    Vorteil: - Keine Online-Fähigkeit und einheitliche Datenbereitstellung durch die Datenquelle erforderlich.

    Nachteil: - Aufbau einer zusätzlichen Organisation
  - Administrativ / Betrieb des Verfahrens

    Vorteil: - Kein Online-Zugang und keine ständige Datenbereithaltung durch die Datenquelle notwendig.

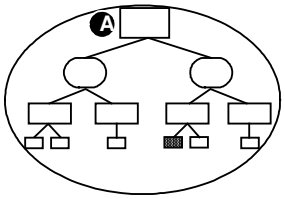
    Nachteil: - Zweistufiger Datenfluss
    - Viele Stellen (alle Datenquellen) anonymisieren.

Modell	Sicherheit	Aufwand	
		Organisatorisch	Administrativ
A3	☺☺	☺	😐

6.4 Anonymisierung bei dezentraler Datenhaltung

6.4.1 Einstufige Modelle

6.4.1.1 Modell A4 – Die Datenquellen anonymisieren



Bei den Datenquellen wird für jede reale Person ein Anonym erzeugt und zusammen mit den unverschlüsselten Nutzdaten an die datenanfordernden Stelle übermittelt. Die Anwendung einheitlicher Verfahren bei der Bildung des Anonyms ist dabei nicht erforderlich.

Die Sicherheit des Modells ist nicht so hoch wie im Modell A3, da die datenanfordernde Stelle Kenntnis über die Datenabsender hat.

Bei diesem Modell hat die datenanfordernde Stelle zu keiner Zeit Zugang zum von der Anonymisierung betroffenen Personenbezug, kennt jedoch den Absender der Daten.

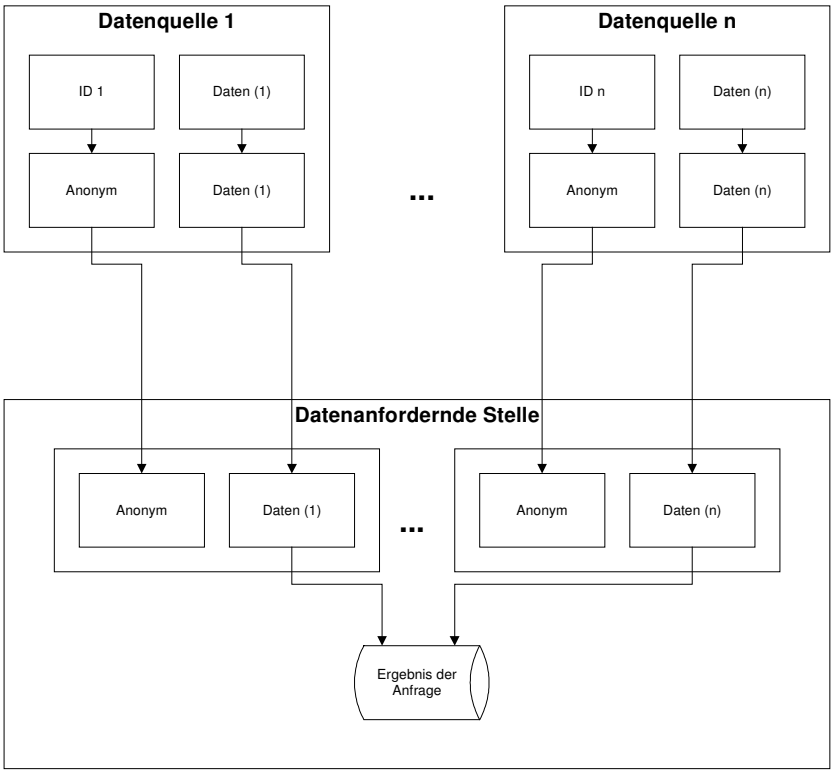


Abbildung 15 Modell A4 – Dezentrale Datenhaltung, Datenquelle anonymisiert

Bewertung:

- Sicherheit

  Vorteil: - Die Personendaten sind nicht außerhalb der Datenquelle bekannt.

  - Die datenanfordernde Stelle erhält keine personenbezogenen Daten.

  Nachteil: - Der datenanfordernden Stelle ist die Datenquelle bekannt (evtl. Rückschlüsse auf Personendaten möglich).

- Aufwand
  - Organisatorisch / Verfahrensaufbau

    Vorteil: - Kein Aufbau einer zusätzlichen Organisation.

    Nachteil: - Alle Datenquellen müssen "onlinefähig" gemacht werden und die relevanten Daten per einheitlichem Modell zum Abruf verfügbar halten.
  - Administrativ / Betrieb des Verfahrens

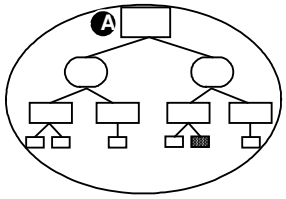
    Vorteil: - Nur einstufiger Datenfluss.

    Nachteil: - Viele Stellen (alle Datenquellen) anonymisieren.

    - Alle Datenquellen müssen online sein und die Daten zum Abruf bereit halten.

Modell	Sicherheit	Aufwand	
		Organisatorisch	Administrativ
A4	☺	☹	😐

6.4.1.2 Modell A5 – Die Vertrauensstelle anonymisiert



Modell A5 stellt ein sicheres Modell dar, ist jedoch mit vergleichsweise hohem organisatorischen Aufwand verbunden.

Die Datenquellen übermitteln die verschlüsselten Nutzdaten zusammen mit den unverschlüsselten einheitlichen Personenkriterien an eine unabhängige Vertrauensstelle.

Die Vertrauensstelle anonymisiert den Absenderbezug und erzeugt Anonyme, welche zusammen mit den verschlüsselten Nutzdaten an die datenanfordernde Stelle weitergeleitet werden.

Sofern die Anonymisierung der Absenderangaben nicht ausreicht (z.B. weil das Volumen der Daten Rückschlüsse auf den Absender zulässt), können in der Vertrauensstelle weitergehende Maßnahmen zur Herstellung einer vollständigen Anonymisierung getroffen werden.

Bei diesem Modell hat die datenanfordernde Stelle zu keiner Zeit Zugang zum von der Anonymisierung betroffenen Personenbezug und hat auch keine Kenntnis vom Absender der Daten.

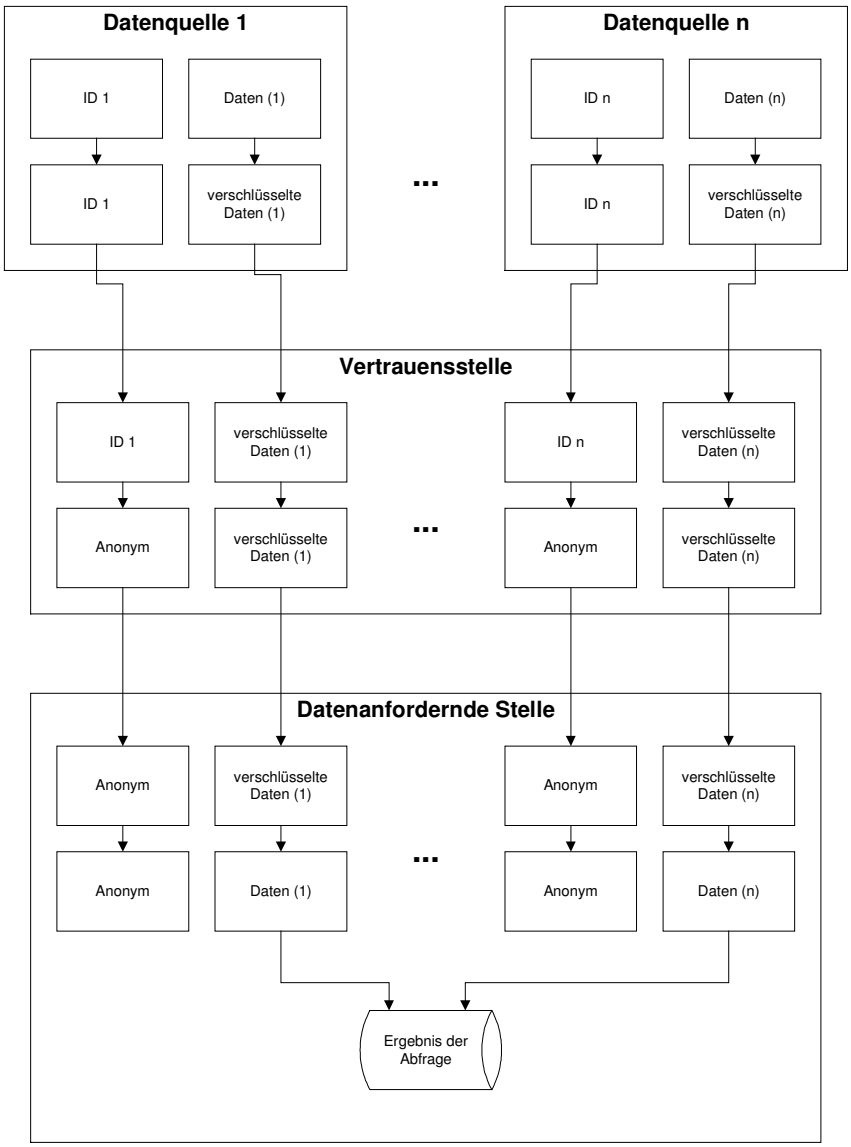


Abbildung 16 Modell A5 – Dezentrale Datenhaltung, Vertrauensstelle anonymisiert

Bewertung:

- Sicherheit

  Vorteil: - Die datenanfordernde Stelle erhält keine personenbezogenen Daten.

  - Der datenanfordernden Stelle ist die Datenquelle nicht bekannt (keine Rückschlüsse auf Personendaten möglich.

  Nachteil: - Die Personendaten sind außerhalb der Datenquelle bekannt.

- Aufwand
  - Organisatorisch / Verfahrensaufbau

    Nachteil: - Aufbau einer zusätzlichen Organisation.

    - Alle Datenquellen müssen "online-fähig" gemacht werden und die relevanten Daten per einheitlichem Modell zum Abruf verfügbar halten.
  - Administrativ / Betrieb des Verfahrens

    Vorteil: - Zentrale Anonymisierung

    Nachteil: - Zweistufiger Datenfluss

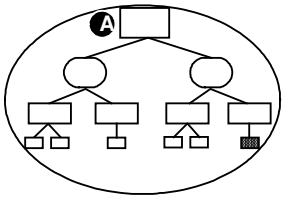
    - Alle Datenquellen müssen online sein und die Daten zum Abruf bereit halten.

Modell	Sicherheit	Aufwand	
		Organisatorisch	Administrativ
A5	☺	☹☹	😐

6.4.2 Zweistufiges Modell

6.4.2.1 Modell A6 – Datenquelle(n) und Vertrauensstelle anonymisieren

Modell A6 stellt ein organisatorisch und administrativ aufwändiges aber sehr sicheres Modell dar.

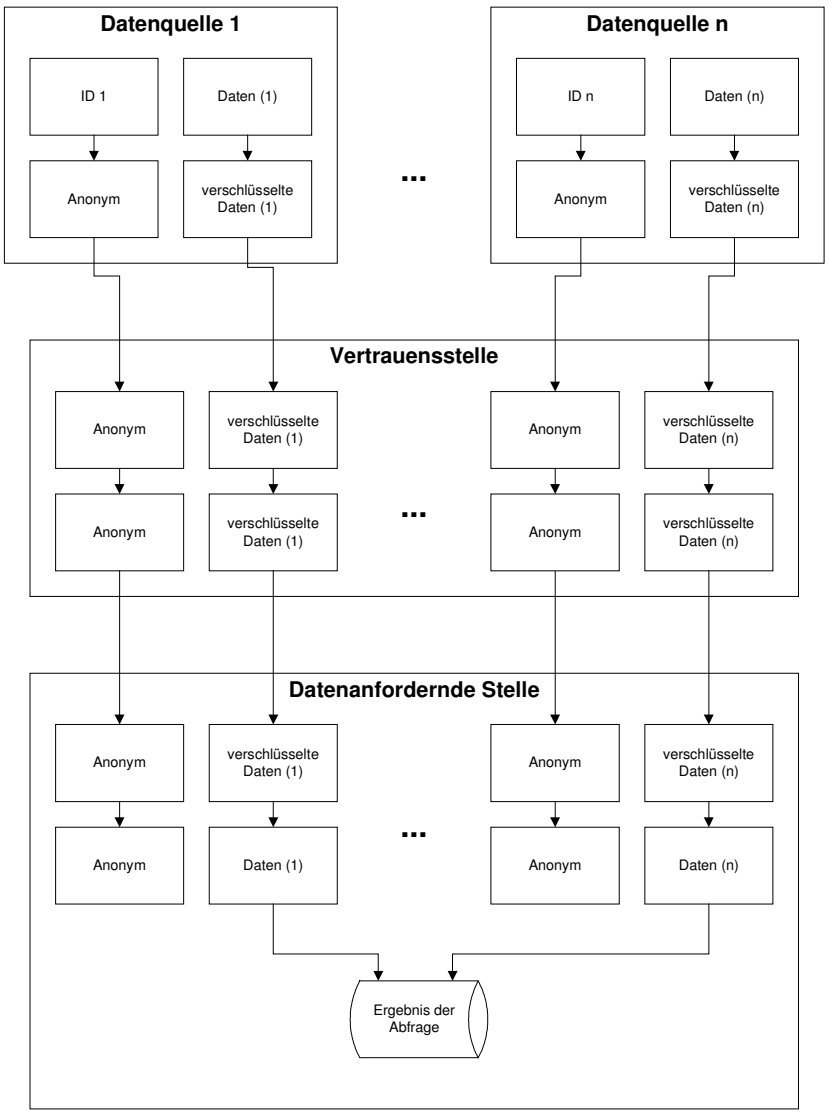


Bei diesem Modell nimmt die Datenquelle eine Anonymisierung vereinbarter Inhalte vor und leitet diese mit den verschlüsselten Nutzdaten an eine Vertrauensstelle weiter.

Die Vertrauensstelle ihrerseits anonymisiert den Absenderbezug und leitet die Daten an die datenanfordernde Stelle weiter.

Sofern die Anonymisierung der Absenderangaben nicht ausreicht (z. B. weil das Volumen der Daten Rückschlüsse auf den Absender zulässt), können in der Vertrauensstelle weitergehende Maßnahmen zur Herstellung einer vollständigen Anonymisierung getroffen werden.

Bei diesem Modell hat die datenanfordernde Stelle zu keiner Zeit Zugang zum von der Anonymisierung betroffenen Personenbezug und hat auch keine Kenntnis vom Absender der Daten.



Die Vertrauensstelle anonymisiert den Absenderbezug.

Abbildung 17 Modell A6 – Dezentrale Datenhaltung, Datenquelle und Vertrauensstelle anonymisiert

Bewertung:

- Sicherheit

  Vorteil:
  - Die Personendaten sind nicht außerhalb der Datenquelle bekannt.
  - Die datenanfordernde Stelle erhält keine personenbezogenen Daten.
  - Der datenanfordernden Stelle ist die Datenquelle nicht bekannt (keine Rückschlüsse auf Personendaten möglich).

- Aufwand
  - Organisatorisch / Verfahrensaufbau

    Nachteil: - Aufbau einer zusätzlichen Organisation.

    - Alle Datenquellen müssen "online-fähig" gemacht werden und die relevanten Daten per einheitlichem Modell zum Abruf verfügbar halten.
  - Administrativ / Betrieb des Verfahrens

    Nachteil: - Zweistufiger Datenfluss

    - Viele Stellen (alle Datenquellen) anonymisieren.

    - Alle Datenquellen müssen Online sein und die Daten zum Abruf bereit halten.

Modell	Sicherheit	Aufwand	
		Organisatorisch	Administrativ
A6	☺☺	☹☹	☹

6.5 Modelle mit "parallelen" Vertrauensstellen

Es sind auch Modelle denkbar, bei denen parallele Vertrauensstellen eingesetzt werden.

Die vorab beschriebenen Anonymisierungs-/Pseudonymisierungsmodelle gelten analog auch bei Betrieb mehrerer Vertrauensstelle auf einer horizontalen Ebene. Die Datenquellen haben hierbei die Möglichkeit, ihre Daten (Anonym, Pseudonym, Nutzdaten) an eine der Vertrauensstellen abzugeben.

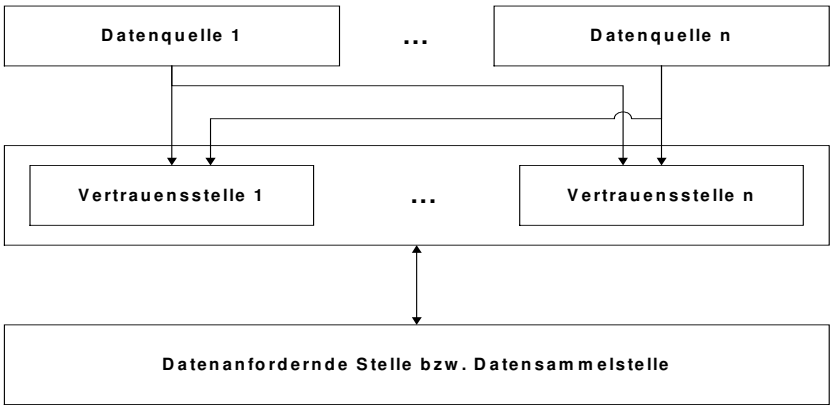


Abbildung 18 Einsatz paralleler Vertrauensstellen

6.6 Zusammenfassende Bewertungsmatrix

Die Bewertung der Verfahren untereinander erfolgt getrennt nach Pseudonymisierungs- und Anonymisierungsverfahren, da eine Anonymisierung grundsätzlich einer Pseudonymisierung vorzuziehen ist.

Beim Vergleich der Modelle untereinander wurde bei der Priorisierung die Sicherheit vorrangig vor dem organisatorischen Aufwand und dieser wiederum vor dem administrativen Aufwand gewertet.

6.6.1 Pseudonymisierungsverfahren

Modell	Sicherheit	Aufwand		Priorität
		Organisatorisch	Administrativ	
P1 (S.21)	☹☹	☺☺	☺	9
P2 (S. 23)	☹☹	☺☺	☺☺	8
P3 (S. 25)	😐	☺	☺	5
P4 (S. 27)	😐	☺☺	☺	3
P5a (S. 30)	☺	☺	😐	2
P5b (S. 32)	☺	☺	☺	1
P6 (S. 34)	☺	😐	😐	4
P7 (S. 37)	☹☹	☹	😐	10
P8 (S. 39)	☺	☹☹	☹	6
P9 (S. 42)	☺	☹☹	☹	7

Grundsätzlich ist die Auswahl des Verfahrens abhängig vom Auswertungsinteresse der Daten.

Betrachtet man ausschließlich die Relation zwischen Sicherheit und organisatorischem/administrativem Aufwand wäre zur Pseudonymisierung aus Sicht des ATG das Mo-

dell P5b zu empfehlen (Pseudonymisierung durch Vertrauensstelle und Datensammelstelle, vgl. Seite 32).

Die Datensammelstelle sollte grundsätzlich alle Datenmengen, die sie auf Anfragen nach außen gibt, mit einem Session Key versehen. Hierdurch wird das gesamte Verfahren weniger anfällig gegen Korrumpierung.

Durch das zweistufige Verfahren erhält weder die Datensammelstelle, noch die Vertrauensstelle Einblick in personenbezogene Daten. Allenfalls wären die Datenquellen in der Lage, aufgrund von Alleinstellungsmerkmalen in den von ihnen gelieferten Nutzdaten Rückschlüsse auf Personenidentitäten der Daten anderer Datenquellen zu ziehen.

Es ist daher vertraglich zu regeln, dass durch Alleinstellungsmerkmale nicht auf reale Personen geschlossen werden darf. Ferner sollten nach Möglichkeit nur Auswertungen auf aggregierten Datenbeständen definiert werden.

Unter bestimmten Voraussetzungen kann es erforderlich sein, eine erneute Pseudonymisierung der in der Datensammelstelle befindlichen Daten vorzunehmen (z. B. Korrumpierung des Verfahrens). Sofern kein Einweg-Pseudonymisierungsverfahren gewählt wurde, ist dieses im vorgeschlagenen Modell durch eine Depseudonymisierung und anschließende Pseudonymisierung zwischen der Datensammelstelle und der Vertrauensstelle recht einfach zu realisieren, da hierbei ausschließlich die IDs betroffen sind.

Es ist aber zu beachten, dass – sofern keine Depseudonymisierung im Einzelfall benötigt wird – die Anwendung von Einweg-Pseudonymisierungsverfahren zu bevorzugen ist. In diesem Fall wird bezüglich der Abschätzung von Sicherheit und organisatorischem/administrativem Aufwand das Modell P4 (siehe Kapitel 6.1.2.1) empfohlen. Die im Vergleich zu P5b scheinbar geringere Sicherheit durch die dezentrale Pseudonymisierung durch die Datenquellen (keine Vertrauensstelle) wird durch das nicht rückführbare Einweg-Pseudonymisierungsverfahren aufgehoben. Modell P5a (siehe Kapitel 6.1.2.27.1.2.2) beinhaltet zwar als zusätzlichen Sicherheitsaspekt die Anonymisierung der Absenderbezuges; dies allein rechtfertigt jedoch nicht den organisatorischen/administrtiven Mehraufwand von Aufbau und Betrieb einer Vertrauensstelle.

6.6.2 Anonymisierungsverfahren

Modell	Sicherheit	Aufwand		Priorität
		Organisatorisch	Administrativ	
A1 (S. 45)	☺	☺☺	☺	2
A2 (S. 47)	☹☹	☺☺	☺☺	6
A3 (S. 48)	☺☺	☺	😐	1
A4 (S. 51)	☺	☹	😐	4
A5 (S. 52)	☺	☹☹	😐	5
A6 (S. 55)	☺☺	☹☹	☹	3

Grundsätzlich ist die Auswahl des Verfahrens abhängig vom Auswertungsinteresse der Daten.

Betrachtet man ausschließlich die Relation zwischen Sicherheit und organisatorischem/administrativem Aufwand wäre aus den in der Bewertungsmatrix gegenüber gestellten Modellen aus Sicht des ATG das Modell A3 zu empfehlen (Datenquelle(n) und Vertrauensstelle anonymisieren, vgl. Kapitel 6.3.2.1).

Die das ATG tragenden Organisationen in alphabetischer Reihenfolge:

- **Bundesärztekammer**
- **Bundesknappschaft**
- **Bundesverband der Allgemeinen Ortskrankenkassen**
- **Bundesverband der Betriebskrankenkassen**
- **Bundesverband der Innungskrankenkassen**
- **Bundesverband der landwirtschaftlichen Berufsgenossenschaften**
- **Bundesverband der landwirtschaftlichen Krankenkassen**
- **Bundesvereinigung Deutscher Apothekerverbände**
- **Bundesversicherungsanstalt für Angestellte**
- **Bundeszahnärztekammer**
- **Deutsche Krankenhausgesellschaft**
- **Hauptverband der gewerblichen Berufsgenossenschaften e.V.**
- **Kassenärztliche Bundesvereinigung**
- **Kassenzahnärztliche Bundesvereinigung**
- **Verband der Angestelltenkrankenkassen**
- **Verband der privaten Krankenversicherung e.V.**
- **Zentralverband der Krankengymnasten und Physiotherapeuten e.V.**