Datenschutzfreundliche Vermessung und Auswertung persönlicher Daten mit Wearables zur Erhöhung der Arbeitssicherheit

Ergebnisbericht D3.2

Konzepte für Anonymisierung und Datennutzungskontrolle

Version	2.0
Datum	24.10.2024
Verfasser	Bianca Steffes (UdS) Philipp Neuschwander (IESE) Marcus-Sebastian Schröder (NMS)

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter den Förderkennzeichen 16KIS1511K, 16KIS1512 und 16KIS1665 gefördert.

Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

Ansprechperson

Marcus-Sebastian Schröder
neusta mobile solutions GmbH
Konsul-Smidt-Str. 24
28217 Bremen

E-Mail: m.schroeder@neusta.de

Inhaltsverzeichnis

Liste der Abkürzungen

API	Application Programming Interface (dt.: Anwendungsprogrammierschnittstelle)
IDS	International Data Spaces
IND²UCE	Integrated Distributed Data Usage Control Enforcement
MYDATA	MYDATA Control Technologies
ODRL	Open Digital Rights Language
PAP	Policy Administration Point
PDP	Policy Decision Point
PEP	Policy Enforcement Point
PIP	Policy Information Point
PMP	Policy Management Point
PXP	Policy Execution Point
SDK	Software Development Kit
UC	Usage Control
XACML	eXtensible Access Control Markup Language

1 Einleitung

Im Projekt WearPrivate ist das Thema Datenschutz für die Konzeption und Entwicklung ein zentrales Anliegen. Speziell die Aspekte der Anonymisierung und der Datennutzungskontrolle sind für das geplante Vorhaben relevant. Dieses Dokument soll einen Überblick über verschiedene Aspekte dieser Themenbereiche geben und mögliche Ansätze zur Umsetzung vorstellen. Zum Erstellungszeitraum ist noch nicht absehbar, welche hiervon tatsächlich in der Entwicklung zum Einsatz kommen werden. Insbesondere da in mehreren Entwicklungsiterationen gearbeitet werden soll, kann es sein, dass frühe Entwicklungsstände weniger oder andere dieser Konzepte umsetzen als spätere. Daher soll der inhaltliche Überblick über die Themen hier einerseits dokumentarischen Charakter für die im Rahmen dieses Projekts untersuchten Ansätze haben; gleichzeitig soll er aber auch als Referenz für die im Projekt tatsächlich eingesetzten Konzepte dienen.

2 Anonymisierung

In diesem Abschnitt sollen verschiedene Konzepte zur Anonymisierung der personenbezogenen Daten von Wearable-Nutzern im Kontext von WearPrivate vorgestellt werden. Wie im Ergebnisbericht D2.1 (rechtliche und ethische Bewertung) beschrieben, fallen anonymisierte Daten nicht in den Anwendungsbereich der DSGVO. Folglich können beliebige Berechnungen auf diesen Daten durchgeführt werden, ohne die Einwilligung der betroffenen Person zu benötigen. Dies bringt den Vorteil mit sich, dass auch hochsensible Daten – wie etwa Gesundheitsdaten – in anonymisierter Form verarbeitet werden können. Da der Fokus nun auf der Anonymisierung von Daten liegen soll, wird hier nur eine Teilmenge der Bausteine aus dem Ergebnisbericht D2.2 (Maßnahmenkatalog) explizit aufgegriffen. Nichtsdestotrotz sind auch weitere Bausteine wie etwa *B 4: Privacy by Design* oder *B 6: Klassisches IT-Sicherheitskonzept* in diesen Szenarien von Bedeutung.

2.1 Personenbezug von Wearable-Daten

Der Personenbezug von Wearable-Daten im Kontext von WearPrivate ist grundlegend durch zwei unabhängige Faktoren begründet. Um eine anonymisierte Verarbeitung der Sensordaten eines Wearables zu erreichen, muss der Personenbezug in diesen beiden Aspekten entfernt werden.

2.1.1 Personenbezug durch Daten

Der erste Faktor, der eine Person identifizieren kann, sind die Daten selbst. Oftmals sind Daten mit einer ID oder vielleicht sogar auch dem Namen einer natürlichen Person verbunden und können so eindeutig dieser Person zugeordnet werden. Auch das Schließen auf ein bestimmtes Gerät kann schon zur Identifizierung einer Person führen, wenn dieses Gerät ausschließlich von nur einer Person genutzt wird.

Neben einer solchen direkten Identifizierung durch einen eindeutigen Identifikator kann auch ein Quasi-Identifikator zur Identifizierung einer Person genutzt werde. Technisch gesehen ist ein Quasi-Identifikator die Teilmenge der Daten, deren Veröffentlichung zum Erhalt der Anonymität eingeschränkt werden muss [28]. Das bedeutet, dass Quasi-Identifikatoren anstelle eines direkten Identifikators zur Identifizierung einer Person genutzt werden können. Ein Beispiel eines Quasi-Identifikators könnte dabei die Kombination aus einer Postleitzahl, dem Geschlecht und dem Alter

sein. Diese Daten können bereits ausreichen, um eine Person eindeutig zu identifizieren [14]. Daher sollten auch alle Quasi-Identifikatoren aus den Daten entfernt werden.

Es gibt auch Anonymisierungsmethoden, wie das Erreichen von k-Anonymität, die darauf basieren, dass explizit diese Quasi-Identifikatoren, die eine Person eindeutig identifizieren können, generalisiert oder zusammengefasst werden, sodass mehrere Personen in einem Datensatz denselben Quasi-Identifikator besitzen und somit nicht mehr unterscheidbar sind. Im Laufe der Zeit sind jedoch immer wieder Methoden bekannt geworden, mit denen die Anonymisierung auf Basis eines Quasi-Identifikators ausgehebelt werden konnte [20], und auch aktuelle Forschungsergebnisse [5] stellen die Sicherheit der Anonymität, welche auf dem Verbergen der Quasi-Identifikatoren beruht, in Frage.

Im Verlauf der Forschung hat sich jedoch ebenfalls gezeigt, dass nicht nur einzelne Attribute der Daten und ihre Kombinationen eine Identifizierung als Quasi-Identifikator ermöglichen können. Auch kontinuierliche physiologische Daten können dazu beitragen. Unter anderem wurde gezeigt, dass Personen anhand von Gangart [30] oder auch Herzschlag [1] identifiziert werden können. Dementsprechend ist es ebenfalls notwendig, derartige Daten zu anonymisieren, etwa mittels Datenverfremdung.

2.1.2 Personenbezug durch Kommunikation

Der zweite Aspekt des Personenbezugs liegt in der Kommunikation der Daten. Auch wenn die Daten an sich keine Rückschlüsse mehr auf eine bestimmte Person zulassen, so können die Daten für eventuelle Verarbeitungen von der betroffenen Person an Dritte übersendet werden. Diese Übertragung der Daten ermöglicht eine eindeutige Zuordnung des sendenden Gerätes zu den Daten und durch das sendende Gerät eventuell auch den Rückschluss auf die betroffene Person. Sind Geräten in einem Firmennetz etwa feste IP-Adressen zugewiesen, kann von der IP-Adresse auf das Gerät und – bei der exklusiven Nutzung durch nur eine Person – auch auf die betroffene Person geschlossen werden. Andererseits können auch zeitliche oder örtliche Aspekte der Kommunikation eine Identifizierung ermöglichen. Beginnt etwa nur ein Arbeiter einer Gruppe von Arbeitern eine Stunde vor dem Rest der Gruppe zu arbeiten und Daten zu senden, so ist auch hier eine Identifizierung möglich. Ähnliches gilt für die Bestimmung des Herkunftsortes der Kommunikation (bspw. durch die IP-Adresse des Senders). Daher sollte auch die Kommunikation der Daten anonymisiert werden und keine eindeutige Identifizierung eines Gerätes oder Nutzers erlaubt werden.

2.2 Allgemeine Aspekte zu den folgenden Verfahren

In allen folgenden Szenarien wird stets davon ausgegangen, dass die Daten des Wearables zuerst verschlüsselt an das Smartphone des Nutzers versendet werden. Diese Kommunikation und die dazu anwendbaren Sicherheitsmechanismen hängen stark davon ab, welche Möglichkeiten ein Wearable bietet. Um eine möglichst hohe Datensparsamkeit zu erreichen wäre es empfehlenswert, die Daten der Mitarbeiter bereits auf dem Wearable zu anonymisieren, da Daten in der Kommunikation als auch auf dem Wearable selbst von Dritten abgefangen oder ausgelesen werden könnten. Dies hätte jedoch zur Folge, dass für den Nutzer keine lokalen Berechnungen und Analysen auf dem Smartphone mehr möglich sind. Zudem ist aufgrund der begrenzten Rechenleistung von Wearables nicht davon auszugehen, dass eine Anonymisierung der Daten effizient noch im Gerät durchgeführt werden kann. Daher soll hier davon ausgegangen werden, dass die Daten verschlüsselt vom Wearable an das Smartphone gesendet werden. Welche Art der Verschlüsselung genutzt wird, hängt dabei von den technischen Möglichkeiten des Wearables ab. In den folgenden Abbildungen wird diese Verbindung stets orange dargestellt, da sich die Sicherheitsvorkehrungen nach den Möglichkeiten des Wearables

richten. Zudem ist es für Angreifer möglich, allein aus der Tatsache, dass eine Person zu einem bestimmten Zeitpunkt Daten gesendet hat, auf die Beteiligung der Person an bestimmten Berechnungen zu schließen. Dies lässt sich durch Maßnahmen der Anonymisierung nicht lösen und wird hier daher nicht betrachtet.

2.3 Variante 1: Individueller anonymisierter Datenaustausch

Das erste Anonymisierungskonzept (Abbildung 1) basiert darauf, dass jeder Mitarbeiter seine Daten lokal auf dem eigenen Gerät anonymisiert und dann ebenfalls anonym zur Verarbeitung an die Cloud sendet.

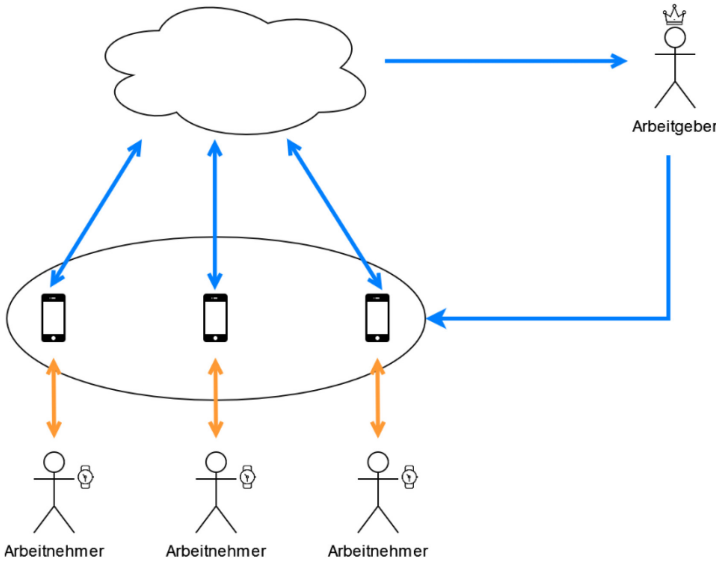

Abbildung 1 Vereinfachte Darstellung von Variante 1. Die blauen Pfeile zeigen die Kommunikation von Daten an, deren Personenbezug entfernt wurde.

Dazu werden die Daten im ersten Schritt von den Wearables der Mitarbeiter auf ihre Smartphones gesendet. Dies sollte möglichst in verschlüsselter Form stattfinden, wenn dies vom Wearable unterstützt wird. Auf dem Smartphone wird dann der Personenbezug der Daten entfernt und die so erstellten Daten auf eine anonyme Art und Weise an die Cloud gesendet. Die Cloud kann folglich zu keinem Moment bestimmen, welche Daten von welchem Mitarbeiter stammen. Daraus folgt, dass der Arbeitgeber maximal anonyme Einzelrückmeldungen der Mitarbeiter erhalten kann. Es bleibt für ihn möglich zu sehen, wenn etwa ein bestimmter Mitarbeiter besonders bedenkliche Werte hat. Dazu erhält er zwar die Information, dass es einen Mitarbeiter mit bedenklichen Werten gibt, jedoch nicht, um welchen Mitarbeiter es sich konkret handelt. Dementsprechend ist es für den Arbeitgeber nur möglich, eine Rückmeldung an die gesamte Gruppe und kein individuelles Feedback zu geben. Im Gegensatz dazu könnte aber die Cloud Antworten an die Individuen schicken, die ihr Daten zukommen lassen, ohne dabei zu wissen, welche Person es exakt ist

2.3.1 Genutzte Bausteine

B 1: Anonymisierung

B 9: Verschlüsselung

B 11: Anonyme Kommunikation

2.3.1.1 Entfernen des Personenbezugs durch Daten

Um den Personenbezug der Daten zu entfernen, sollte eine Anonymisierung der Daten durchgeführt werden. Im Folgenden sollen zwei mögliche Konzepte der Anonymisierung vorgestellt werden.

2.3.1.2 k-Anonymität

Da k-Anonymität [28] trotz der eingangs beschriebenen Schwächen ein weit verbreitetes und leicht verständliches Konzept zur Erstellung von Anonymität ist, wollen wir hier einen näheren Blick auf seine Funktionsweise werfen. Es ist dabei speziell auf die Anonymisierung von Personeninformationen ausgelegt. Dies bezieht sich vor allem auf Datenbanken und die in ihnen vorhandenen sensiblen Informationen, wie zum Beispiel Name, Alter oder Geschlecht, mit denen Rückschlüsse auf Personen gezogen werden können.

Bei k-Anonymität ist es das Hauptziel, Informationen zu einem bestimmten Grad zu erhalten, so dass sie zu sinnvollen Berechnungen genutzt werden, aber einzelne Personen nicht identifiziert werden können. Dies wird erzielt, indem der Quasi-Identifikator der Daten so verändert wird, dass mehrere Personen (insgesamt k Personen) aus dem Datensatz zu einem Quasi-Identifikator zugeordnet werden können. Dies kann erzielt werden, indem Teile des Quasi-Identifikators generalisiert oder ganz unterdrückt werden. Das Generalisieren der Daten besteht darin, diese Daten gröber zu fassen. Etwa könnte dies bei einem Alter erzielt werden, indem Altersgruppen erstellt werden (z. B. 21–25 Jahre). Wenn Daten unterdrückt werden, werden diese komplett aus dem Datensatz entfernt. Dabei sollte darauf geachtet werden, dass die unterdrückten Daten nicht von grundlegender Relevanz für die beabsichtigen Berechnungen sind. Ein Beispiel für k = 3-Anonymität ist in Tabelle 1 zu sehen.

Tabelle 1 Links: Original-Tabelle für k-Anonymität

Name	PLZ	Alter	Geschlecht	Krankheit
Patrick	66111	22	M	Kreislauf
Gerd	66123	23	M	Lungen
Hans	66127	18	M	Keine
Lena	10115	47	F	Krebs
Eva	10179	42	F	Keine
Amrei	10243	58	F	Kreislauf
George	20038	23	M	Lunge
Leon	22607	29	M	Leber
Simone	21149	18	F	Krebs

Rechts: Tabelle mit 3-Anonymität

Name	PLZ	Alter	Geschlecht	Krankheit
*	661*	22	M	Kreislauf
*	661*	23	M	Lungen
*	661*	18	M	Keine
*	10*	47	F	Krebs
*	10*	42	F	Keine
*	10*	58	F	Kreislauf
*	2*	23	*	Lunge
*	2*	29	*	Leber
*	2*	18	*	Krebs

Wie bereits einleitend beschrieben bestehen jedoch Gefahren der De-Anonymisierung bei Quasi-Identifikator-basierten Anonymisierungsverfahren.

2.3.1.3 Differential Privacy

Wie in [29] beschrieben hat sich im Laufe der Forschung herausgestellt, dass es unmöglich ist, eine (statistische) Datenbank zu veröffentlichen und gleichzeitig zu verhindern, dass jedwede Informationen über die darin enthaltenen Personen offengelegt werden [7]. Statistische Datenbanken

basieren dabei auf Aggregationen und dienen dazu, Aussagen über eine große Menge an Personen bereitzustellen (bspw. Aussagen über die gesamte Bevölkerung eines Landes). Um derartige Informationen dennoch möglichst datenschutzfreundlich nutzen zu können hat sich die Forschung darauf konzentriert, zu verhindern, dass mögliche Angreifer durch das Vorhandensein eines Datensatzes in einer statistischen Tabelle erheblich mehr Informationen über die Person erhalten können, als wenn deren Datensatz nicht in der Tabelle vorkommt. Dies soll verhindern, dass die beteiligten Personen signifikante Nachteile erhalten. Ein dafür konstruiertes Konzept ist etwa Differential Privacy [7][8]; die bekannteste Variante ist dabei ε-Differential Privacy, in der ε die garantierte Privacy festlegt [9].

In Differential Privacy wird entsprechend der oben angeführten Erkenntnis, dass keine Informationsoffenlegung unmöglich ist, bei jeder Anfrage an die Daten durch die Sensitivität bestimmt, wie viel Informationsgehalt offengelegt werden soll [6]. Der Grad des Informationsgehalts wird dabei durch ein leichtes Rauschen angepasst, wobei dieses Rauschen und die daraus folgenden Datenveränderungen von der Sensitivität, der Anzahl der Personen in der Datenbank und ε abhängt [9]. Der insgesamt offengelegte Informationsgehalt wird dabei durch ein Privacy Budget begrenzt, welches durch ε festgelegt und mit jeder Berechnung verringert wird. Ist nach einer Reihe von Berechnungen das Privacy Budget aufgebraucht, dürfen folglich keine weiteren Berechnungen durchgeführt werden [9]. Nicht jeder Anwendungsfall benötigt jedoch das komplette Privacy Budget. Die Wahl eines passenden ε für das Privacy Budget ist jedoch äußerst komplex und es muss stets für den Anwendungsfall eine individuelle Abwägung zwischen Privacy und Nutzbarkeit durchgeführt werden [16].

2.3.2 Entfernen des Personenbezugs durch Kommunikation

Werden die Daten nun individuell vom Smartphone des Mitarbeiters an die Cloud gesendet, besteht ebenfalls eine Möglichkeit zur Identifizierung durch die Kommunikation. Auch wenn die gesendeten Daten in sich anonym sind, sollte keine Verbindung der Daten mit einem konkreten Gerät erlaubt werden, welches auf einen bestimmten Nutzer schließen lässt. Dazu können verschiedene Methoden genutzt werden.

2.3.2.1 P2P-Netzwerke

Die grundlegende Basis für ein derartiges Netzwerk ist die P2P (peer-to-peer) Struktur. In P2P-Netzwerken gibt es eine Reihe an Geräten, die keine hohe Rechenleistung haben müssen. Alle Geräte des Netzwerks können miteinander kommunizieren und Informationen austauschen, was bedeutet, dass jedes Gerät sowohl empfangen als auch senden kann. Allein durch die hohe Anzahl der Geräte in diesem Netzwerk ist es für unerwünschte Personen schwer, Informationen der Teilnehmer zu finden.

Um ein P2P-Netzwerk anonym zu gestalten, wird ein zweiphasiges Model vorgeschlagen [4]: Die „Query Phase“ und die „Data Transmission Phase“. In der Query Phase werden in dem Netzwerk von allen Punkten aus an Nachbarn Initialisierungsnachrichten gesendet, um darauf basierend später inhaltliche Nachrichten versenden zu können. Es werden unter anderem kryptografische Schlüssel ausgetauscht, sodass vom Start bis zum Ziel verschlüsselte Nachrichten gesendet werden können. Wenn die Query Phase abgeschlossen ist, haben die verschiedenen Teilnehmer genügend Informationen über die anderen Kommunikationspunkte gesammelt, damit Nachrichten gesendet werden können.

In der Data Transmission Phase werden anschließend die eigentlichen Informationen gesendet. Ein Startpunkt sendet an alle seine Nachbarn eine gewünschte Nachricht. Diese untersuchen, ob sie die

Nachricht bereits gesehen haben. Falls dies der Fall ist, wird die Nachricht ignoriert. Andernfalls wird basierend auf den am Anfang ausgetauschten Informationen berechnet, mit welcher Wahrscheinlichkeit eine Nachricht weitergesendet werden soll oder nicht. Wenn die Nachricht am finalen Punkt ankommt, kann dieser die Nachricht mithilfe der in der Query Phase ausgetauschten kryptografischen Schlüssel entschlüsseln.

Im Kontext von WearPrivate könnte ein solches P2P-Netz durch die Smartphones der Arbeitnehmer erstellt werden, es könnte aber auch ein bereits existierendes Netzwerk genutzt werden. Die Anonymität in derartigen Netzen wird dabei nicht von der Vertrauenswürdigkeit der anderen Teilnehmer bedingt. Da jeder Teilnehmer des Netzes gleichzeitig sowohl Sender als auch Weiterleiter von Nachrichten ist, ist es für die Teilnehmer fast unmöglich zu bestimmen, von welchem Teilnehmer eine Nachricht ursprünglich stammt.

2.3.2.2 Crowds

Crowds [26] bestehen ebenfalls wie P2P Netzwerke aus mehreren Teilnehmern. Nutzer können diesem Netzwerk beitreten, was jedoch von den Teilnehmern des Netzwerks selbst bestätigt werden muss. Ein Benutzer wird dabei von einem lokalen Prozess namens „jondo“ (von dem englischen Namen „John Doe“) in dem Netzwerk repräsentiert. Sobald ein jondo dem Netzwerk beitritt, ist dieser in der Lage, Nachrichten von allen anderen jondos in dem Netzwerk zu empfangen und weiterzuleiten. Dies wird erzielt, indem, sobald ein jondo neu dazukommt, dies allen anderen Netzwerkbenutzer mitgeteilt wird.

Der erste jondo sendet die erste Nachricht zufällig an einen der anderen jondos. Dies kann der jondo auch selbst sein. Von dort aus entscheidet der empfangene jondo mit 50 % Wahrscheinlichkeit, ob er die Nachricht an den Empfänger sendet oder ob die Nachricht wieder zufällig an einen der übrigen jondos gesendet wird. Des Weiteren speichert der jondo sich in einer Tabelle die Ausgangsadresse und die Zieladresse, sodass im späteren Verlauf zwischen zwei jondos eine Kommunikation über das Netzwerk stattfinden kann. In allen Fällen ist die Kommunikationen zwischen den jondos verschlüsselt. Im Abbildung 2 ist ein Beispiel für eine Kommunikation gezeigt. Hier kommunizieren die jeweiligen Zahlen miteinander.

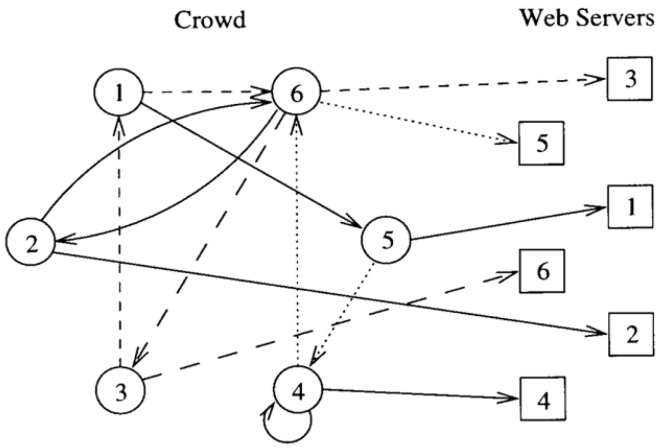

Abbildung 2 Beispielhafter Ablauf von Kommunikation in Crowds, entnommen von Reiter et al. [26]

Mit diesem System können Nachrichten anonym ausgetauscht werden. Ähnlich wie im vorherigen Absatz beschrieben könnte das Netz durch die Arbeitnehmer erstellt werden oder auch ein vorhandenes Netz genutzt werden. Das Eintreten in das Netzwerk führt dabei zu keinem Anonymitätsverlust für Nachrichten: Das Netzwerk authentifiziert zwar seine Mitglieder (kennt also die Mitglieder des Netzes), kann jedoch nicht zuordnen, welches Mitglied welche Nachricht sendet. Es

ist zudem nur schwer möglich, als lokaler Mithörer der Kommunikation zwischen jondos, also als aktiver Benutzer des Netzwerkes oder als Endserver, Informationen über die Teilnehmer oder Kommunikationsinhalte zu gewinnen.

2.3.2.3 The Onion Routing (TOR)

Eine weitere Implementierung zur Anonymisierung von Kommunikation ist die Nutzung des Onion Routing [26]. Das Onion Netzwerk ist ein großes, bereits existierendes System, das unter anderem in dem TOR Browser benutzt wird. Dieses Netzwerk ist öffentlich verfügbar.

Stark vereinfacht dargestellt werden zum Versenden von Nachrichten im Onion Routing zunächst eine Reihe verschiedener Server (onion router) aus dem Netzwerk für die Kommunikation ausgewählt. Anschließend werden aus den öffentlichen Ressourcen des Netzwerks die drei öffentlichen Schlüssel der Server, die für eine Verschlüsselung benötigt werden, eingeholt. In einem initialen Verbindungsaufbau wird dann der Schlüsselaustausch durchgeführt. Daraufhin kann eine Nachricht erstellt werden, bei der drei Schichten der Verschlüsselung über die Nachricht gelegt werden. Die jeweiligen Schichten oder auch Schalen können nur von dem dazugehörenden Server entschlüsselt werden und nur in der Reihenfolge, in der sie erstellt wurden. Daher kommt auch der Begriff Onion Router, weil hier analog wie bei einer Zwiebel Schichten nacheinander entfernt werden. In der entschlüsselten Schale ist ebenfalls festgehalten, wer der nächste Server ist. Damit weiß jeder einzelne Server lediglich, woher das Paket stammt und welcher der nächste Server ist, jedoch ist keinem einzigen Server der ganze Weg bekannt. Der letzte Server entschlüsselt dann die letzte Schale und sendet nun das Paket unverschlüsselt zu dem Empfänger. Dabei kann das Paket selbst nochmal zwischen Sender und Empfänger verschlüsselt werden, jedoch ist dies nicht mehr Teil des Protokolls des Onion Routing.

2.3.3 Zusatz: Authentizität durch Signaturen

Wird durch eines der vorhergehenden Verfahren die Kommunikation anonymisiert, so kann theoretisch der Fall eintreten, dass Personen Nachrichten an die Cloud senden, die eigentlich nicht dazu berechtigt sind. Die Cloud könnte also Probleme haben, bei einer eingehenden Nachricht zu bestimmen, ob es sich dabei um eine legitime Nachricht handelt.

Blinde Signaturen [3] können dazu verwendet werden, um erhaltene anonymisierte Daten als legitime Nachrichten einer Gruppe zu verifizieren. Dazu wird Verschlüsselung genutzt. Als erstes kommuniziert die Person, die anonymisierte Daten senden möchte, mit einer dritten Partei. Diese dritte Partei signiert dabei die anonymisierten Daten und verifiziert so, dass die anonymisierten Daten zu einer bestimmten Person oder einer Gruppe an bestimmten Personen gehören. Daraufhin werden die anonymisierten Daten an den Adressaten gesendet. Wichtig ist hier, dass die dritte Partei, welche die Signatur durchführt, keine Information über den Inhalt der Nachricht erhält und, in Folge einer möglichen anonymisierten Kommunikation mit dem Sender der Nachricht, auch wenige Informationen über den Sender der Nachricht erhält. Hier besteht also die Problematik, dass die signierende Stelle ihre Kommunikationspartner nicht identifizieren darf, jedoch deren Zugehörigkeit zu einer legitimen Gruppe nachprüfen können muss.

Dieses Problem ist in Alternativen wie Gruppensignaturen [2] oder Ringsignaturen [27] bereits bedacht und gelöst. Vor allem in Ringsignaturen ist es so möglich, dass die Daten in der Gruppe (der Mitarbeiter) signiert wird und somit keine dritte vertrauenswürdige Partei benötigt wird.

2.3.4 Zusatz: Parametrisierung von Anonymisierungsmethoden

Die Maßnahmen der Anonymisierung erlauben es oftmals, eine Parametrisierung (*B 2: Parametrisierung*) durchzuführen, was somit eine Umsetzung von Baustein *B 3: Wahl des Schutzniveaus* ermöglicht. Die Nutzung eines schwächeren Parameters oder eines niedrigeren Schutzniveaus in einer Anonymisierung bedeutet jedoch nicht, dass nur Teile der Daten anonymisiert werden. Es erhöht sich lediglich die Wahrscheinlichkeit, dass die Anonymisierung gebrochen werden kann.

Am Beispiel von k-Anonymität etwa wäre der Parameter k variabel. Wird k auf einen hohen Wert wie bspw. 100 gesetzt, bedeutet dies, dass jeweils 100 Personen in einem Datensatz nicht zu unterscheiden sind. Dies würde ein hohes Schutzniveau darstellen. Wird k jedoch auf einen kleinen Wert wie bspw. zwei gesetzt, würde das dazu führen, dass nur jeweils zwei Personen in einem Datensatz nicht unterscheidbar sind. Hier besteht eine viel größere Gefahr, dass dennoch eine Identifizierung durch weitere Informationen ermöglicht wird, und es liegt ein niedrigeres Schutzniveau vor. Je höher jedoch das Schutzniveau gewählt ist, umso weniger Schlüsse können aus den Daten für legitime Berechnungen gezogen werden.

2.4 Variante 2: Verschlüsselte Berechnungen

In diesem Konzept (vereinfacht in Abbildung 3) werden die Daten vom Wearable des Arbeitnehmers an sein Smartphone gesendet und dort verschlüsselt.

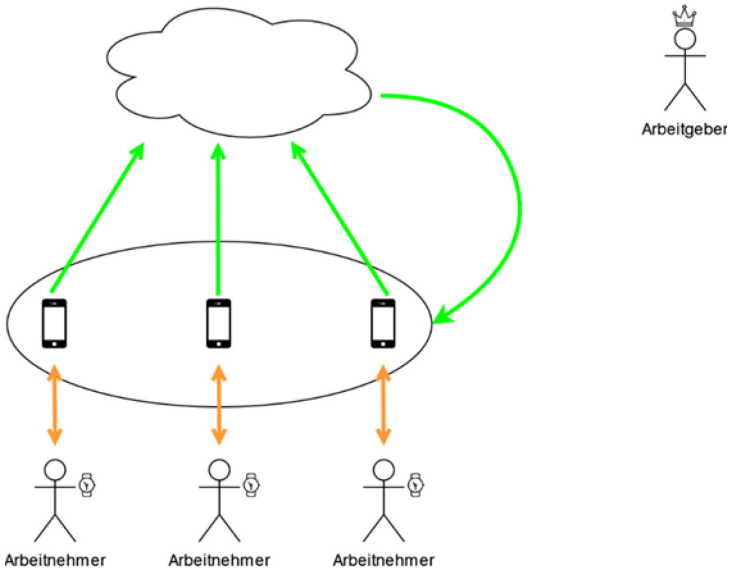

Abbildung 3 Vereinfachte Darstellung von Variante 2. Die grünen Pfeile zeigen die Kommunikation von Daten an, deren Inhalt verschlüsselt ist.

Diese (besonders) verschlüsselten Daten werden dann von jedem einzelnen Gerät an die Cloud gesendet. Die Cloud kann die Daten nicht entschlüsseln, sondern mit den verschlüsselten Daten rechnen (bspw. summieren oder den Durchschnitt berechnen). Sie weiß also in keinem Moment, welche Daten sie verarbeitet. Danach kann die Cloud nur ein gesammeltes Ergebnis für die Gruppe zur Verfügung stellen, dessen Inhalt sie aber ebenfalls nicht kennt, die Gruppenmitglieder jedoch unabhängig voneinander entschlüsseln können. Folglich ist der Arbeitgeber zu keiner Zeit über die Daten seiner Mitarbeiter informiert und kann kein Feedback erhalten oder geben. Auch die Cloud kann nicht darauf reagieren, wenn bei einem Mitarbeiter oder der Gruppe sehr extreme Werte auftreten.

2.4.1 Genutzte Bausteine

B 9: Verschlüsselung

B 10: Aggregation

2.4.2 Entfernen des Personenbezugs durch Daten und Kommunikation

Ein Verfahren, welches das Verschlüsseln und das Rechnen auf den daraufhin verschlüsselten Daten erlaubt, ist die (teil-)homomorphe Verschlüsselung [13]**.** Mit ihr ist es möglich, dass Daten verschlüsselt an Dritte versendet werden und von dieser dritten Partei dann mit anderen – ebenfalls homomorph verschlüsselten – Daten verrechnet werden können. Die dritte Partei erhält dabei zu keiner Zeit Einsicht in die zu verrechnenden Daten oder Ergebnisse und muss daher nicht vertrauenswürdig sein. Daher könnte es sich hierbei bedenkenlos um eine Cloud handeln, die Berechnungen mit hohem Rechenaufwand ausführen und die Ergebnisse dann an die Mitarbeiter zurücksenden kann. Mit der homomorphen Verschlüsselung sind jedoch nur die Rechenoperationen Addition und Multiplikation möglich.

Es gibt jedoch auch einige Unterteilungen in der homomorphen Verschlüsselung. Es gibt die drei folgenden Definitionen:

- Teil-homomorphe Verschlüsselung (PHE – Partially Homomorphic Encryption)
- Nahezu voll-homomorphe Verschlüsselung (SHE – Somewhat Homomorphic Encryption)
- Voll-homomorphe Verschlüsselung (FHE – Fully Homomorphic Encryption).

Bei der PHE ist es nur möglich, eine Rechenoperationsart auf den verschlüsselten Daten auszuführen, jedoch unbegrenzt oft. Dies wird beispielsweise in der RSA-Verschlüsselung angewandt.

Die SHE erlaubt es, beide Rechenoperationen zu benutzten, jedoch nur eine begrenzte Anzahl an Berechnungen. Nachdem eine bestimmte Grenze erreicht wurde, sind dann die berechneten Daten unbrauchbar und können nicht weiter benutzt werden.

FHE erlaubt es, die zwei Rechenoperationen beliebig oft zu benutzten. Nun wäre es naheliegend, FHE stets den anderen beiden Vorgehensweisen vorzuziehen. Jedoch hat FHE den Nachteil, dass sie im Vergleich zu anderen Verschlüsselungen nur äußerst langsam ist und für ein Smartphone im Kontext von WearPrivate zu rechenintensiv sein könnte. Denn obwohl die Verarbeitung von neuen Daten etwa alle 30 Sekunden technisch möglich sein würde, könnte dies jedoch zu einem erheblich höheren Stromverbrauch des Smartphones führen. Die Wahl der konkreten homomorphen Verschlüsselungsart für einen Anwendungsfall ist demnach stark von dessen Rahmenbedingungen abhängig.

Dadurch, dass die Daten während der gesamten Kommunikation mit der Cloud verschlüsselt sind, ist der Personenbezug durch die Daten als auch die Kommunikation entfernt. Die Cloud hat in keinem Moment eine Information darüber, welche Daten mit etwaigen Mitarbeitern zusammenhängen könnten.

2.5 Variante 3: Austausch aggregierter Daten

In diesem Konzept (vereinfacht dargestellt in Abbildung 4) werden die Daten vom Wearable der Arbeitnehmer an ihre jeweiligen Smartphones gesendet. Anschließend werden die Daten in der Gruppe auf eine anonyme oder verschlüsselte Art und Weise aggregiert und anonymisiert.

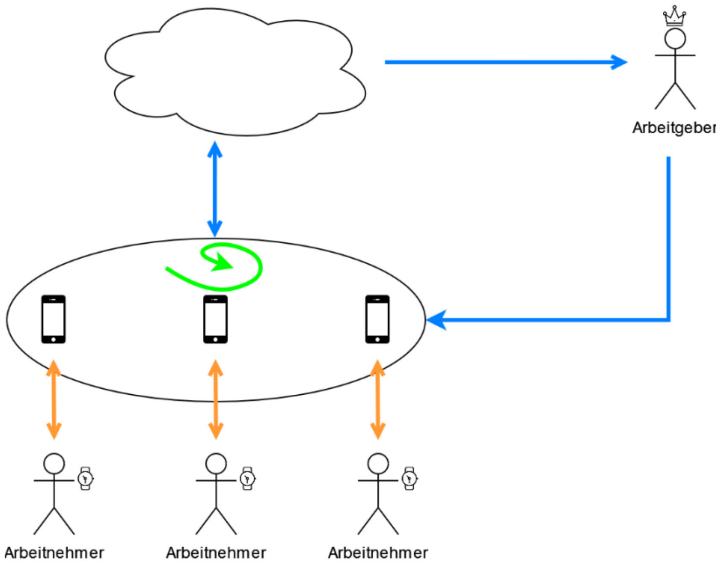

Abbildung 4 Vereinfachte Darstellung von Variante 3. Die grünen Pfeile zeigen die Kommunikation von Daten an, deren Inhalt verschlüsselt ist, die blauen Pfeile von Daten, deren Personenbezug entfernt wurde.

Danach werden die gesammelten anonymisierten Daten der Gruppe an die Cloud versendet, welche nach den durchgeführten Analysen auch ein gesammeltes Feedback an die Gruppe, aber nicht individuell für jeden Nutzer, geben kann. Der Arbeitgeber kann ebenfalls die Ergebnisse dieser anonymisierten Auswertungen erhalten und ein gesammeltes Feedback an die Gruppe geben. Hier ist es aber auch für ihn nur möglich, Aussagen über die gesamte Gruppe zu machen. Es ist nicht mehr möglich, extreme Werte einzelner Mitarbeiter zu bemerken.

2.5.1 Genutzte Bausteine

B 1: Anonymisierung

B 9: Verschlüsselung

B 10: Aggregation

2.5.2 Entfernen des Personenbezugs durch Kommunikation

In diesem Szenario bestehen zwei Möglichkeiten, in denen die Daten eines Arbeitnehmers Dritten offenbart werden können. Zum einen könnte dies im Falle der Kommunikation zwischen den Arbeitnehmern, die gemeinsam ihre Daten aggregieren, der Fall sein. Zum anderen werden die Daten final an den Arbeitgeber versendet. In der zuletzt beschriebenen Kommunikation werden jedoch aggregierte Daten versendet, die die Daten aller Arbeitnehmer beinhalten. Dementsprechend erlaubt das reine Versenden der Daten keinen Rückschluss darauf, wer die sendende Person ist. Es offenbart lediglich, welche Person Daten zu den aggregierten Daten beigesteuert haben könnte.

Die Kommunikation zwischen den Arbeitnehmern zur Aggregation der Daten kann auf verschiedene Arten durchgeführt werden, die im Folgenden kurz erläutert werden. Nicht näher erläutert wird hier die (Teil-)homomorphe Verschlüsselung, die ebenfalls zum geheimen Aggregieren genutzt werden kann, jedoch schon in Abschnitt 2.4.2 beschrieben wurde.

2.5.2.1 Slice-Mix-AggRegaTe-Algorithmus (SMART)

Der Slice-Mix-Aggregate-Algorithmus [15] ermöglicht es, Daten in drei Schritten anonym aufzuaddieren, ohne dass Beteiligten die Daten anderer Beteiligten offengelegt werden. Das Vorgehen ist in Abbildung 5 vereinfacht dargestellt.

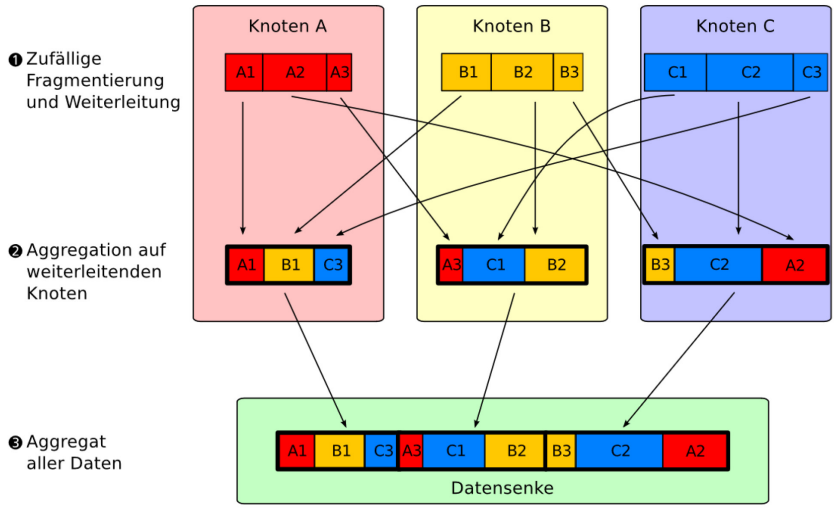

Abbildung 5 Beispiel für den Slice-Mix-Aggregate-Algorithmus für Wearable-Daten, entnommen aus Finster 2014 [12]

In einem ersten Schritt wählt jeder Beteiligte – Arbeitnehmer in unserem Fall – eine zufällige Teilmenge der anderen Beteiligten aus und zerlegt seinen zu übermittelnden Wert in zufällige Teile (Slicing-Schritt). Genauer gesagt wird der Wert in eine Summe zerlegt.

Im nächsten Schritt werden die Slices dadurch gemischt (Mix-Schritt), dass jeder Beteiligter alle Teile seiner Summe bis auf einen an die im ersten Schritt ausgewählten anderen Beteiligten sendet. Diese Kommunikation muss verschlüsselt ablaufen, um keine Informationen preiszugeben. Am Ende dieses Schrittes berechnet dann jeder Beteiligte die Summe aller Slices, die er erhalten hat mit seinem eigenen übrigen Slice.

Der letzte Schritt umfasst nun das Zusammenrechnen der Daten (Aggregate-Schritt). Hierbei werden die Daten aller Beteiligten ähnlich einer vorgegebenen Baumstruktur an den jeweiligen Vorgänger gesendet, der die neu erhaltenen Daten wiederum mit seinen bereits vorhandenen Daten aufsummiert. Dieses Verfahren wird fortgeführt, bis alle Daten aufsummiert sind. Die Kommunikation während des Aufsummierens muss nicht zwangsweise verschlüsselt sein.

Die Folge dieses Verfahrens ist es, dass es nur kommutative Aggregationsverfahren (addieren oder multiplizieren) nutzen kann. Zudem müssen die zu aggregierenden Daten dementsprechend für diese mathematischen Optionen geeignet sein. Zahlenwerte sind hier also nutzbar, wohingegen Identifikatoren oder Namen für dieses Verfahren ungeeignet sind.

2.5.2.2 Secure Multiparty Computation

Die Secure Multiparty Computation [19] geht auf Yaos Millionärsproblem [31] zurück. In diesem theoretischen Problem möchten zwei Millionäre ihr Vermögen vergleichen. Ziel ist es, dass sie erfahren, ob sie mehr oder weniger als der andere besitzen und zeitgleich jedoch nicht erfahren, um wie viel Geld es sich genau handelt. Für die Berechnung wird keine weitere Third-Party benötigt. Die Daten der beiden Parteien bleiben so anonym und sie erhalten nur das Ergebnis dieser Berechnung. Dieser Zustand ist durch den geschickten Einsatz mathematischer Berechnungen möglich.

Yao's Millionärsproblem ist in seiner ursprünglichen Version nur auf zwei Parteien ausgelegt. Im Laufe der Zeit haben sich jedoch weitere Verfahren und Protokolle herausgebildet, die mit mehr Parteien (Multiparty) arbeiten können. Dementsprechend ist Multiparty Computation nur ein Überbegriff verschiedener Verfahren, die alle dieselben Voraussetzungen oder Rahmenbedingungen erfüllen.

Die ursprüngliche Lösung von Yao war jedoch noch kein sicheres Protokoll. Erst im Laufe der Zeit ist der Fokus darauf entstanden, dass diese (mathematischen) Protokolle durch ein bösartiges Einwirken von außen, aber auch von innen, Informationen preisgeben können. Dementsprechend entstand die Secure Multiparty Computation [19], die es zum Ziel hatte, auch im Falle von Angriffen von außen wie Man-in-the-middle-Angriffen oder auch böswilligen Teilnehmern in den Berechnungen, keine Informationen über die individuellen Daten der Beteiligten preiszugeben.

Secure Multiparty Protokolle haben jedoch auch einige Einschränkungen. Eine große Problematik liegt darin, dass sie recht viel Rechenzeit und Rechenleistung benötigen. Zudem waren sie ursprünglich nicht darauf ausgelegt, Aggregationen zu berechnen. Für den Einsatz auf Wearables bleibt zu prüfen, ob es ein Protokoll zur sicheren Berechnung von Aggregationen gibt oder ein solches Protokoll überhaupt möglich ist und, sollte dies der Fall sein, ob dieses sich mit seiner erforderlichen Rechenzeit und Rechenleistung in das Anwendungsszenario von WearPrivate sinnvoll einbringen lässt.

2.5.3 Entfernen des Personenbezugs durch Daten

Da die reine Aggregation, die in dieser Variante ihre Anwendung findet, keine sichere Anonymisierung ist (unter anderem bei Dwork [7] nachzulesen) und dementsprechend nicht den Personenbezug entfernt, wird eine zusätzliche Anonymisierung benötigt. Dazu kann eines der Verfahren aus Abschnitt 2.3.2 gewählt werden.

2.6 Anwendung in WearPrivate

Die vorangegangenen Ausführungen haben verschiedene Konzepte gezeigt, die in der datenschutzfreundlichen Verarbeitung von Wearable- und Gesundheitsdaten zur Anwendung kommen könnten. Für die Umsetzung der Anwendungsfälle von WearPrivate ist jedoch nicht jedes der Konzepte nutzbar.

Verschlüsselte Berechnungen mithilfe von homomorphen Verschlüsselungen (Variante 2) sind sehr stark durch die von ihnen ermöglichten Rechenoperationen limitiert. Berechnungen, die auf maschinellem Lernen basieren, lassen sich auf diese Art und Weise nicht durchführen. Daher ist diese Variante zwar im Kontext von Wearables denkbar, jedoch nicht für den im Projekt geplanten Anwendungsfall nutzbar. Auch das Versenden von aggregierten Daten der gesamten Nutzerschaft an den Analysedienst (Variante 3) ist für die geplanten Arbeiten im Projekt unpassend: Das Zusammenfassen der Daten verschiedener Nutzer ermöglicht keine Einzelauswertungen für die jeweiligen Nutzer und gibt somit nicht die Möglichkeit, dass einzelne Personen bei kritischen Werten auf ihre Lage hingewiesen werden können. Daher haben sich die Partner im Projektkontext für die Umsetzung von Variante 1 entschieden.

3 Datennutzungskontrolle

In diesem Abschnitt sollen Konzepte zur Datennutzungskontrolle (engl.: data usage control) im Kontext von WearPrivate vorgestellt werden.

3.1 Motivation: Kontrollierte Nutzung von Wearable-Daten

Im Gesamtsystem werden Daten an verschiedenen Stellen der Systemkette verarbeitet und entlang der Systemkette weitergegeben. Als Gesundheitsdaten zählen die Wearable-Daten in unserem Anwendungsfall zu den besonderen Kategorien personenbezogener Daten, weshalb mit ihrer Verarbeitung besonders hohe datenschutzrechtliche Auflagen verbunden sind (vgl. WearPrivate-Ergebnisbericht D2.1).

Zudem sehen wir vielfältige Konfliktpotenziale zwischen den Interessen der Betroffenen einerseits und den Interessen der Datennutzer andererseits (vgl. WearPrivate-Ergebnisbericht D1.1). Die Datennutzer verwenden die Daten der Betroffenen, um aus ihnen einen Nutzen zu ziehen. Abhängig vom jeweiligen Fall kann dieser Nutzen nachteilig für die Betroffenen oder von diesen nicht gewünscht sein. Ein Beispiel aus dem Arbeitskontext für eine von den Arbeitnehmern regelmäßig unerwünschte Datennutzung ist die Verarbeitung zum Zwecke der Leistungs- und Verhaltenskontrolle. Andererseits kann jedoch eine Datennutzung, die ausschließlich dem Zweck der Gesundheit und Arbeitssicherheit der Arbeitnehmer dient, einen direkten und besonders wünschenswerten Vorteil für die Betroffenen bieten.

Allgemein kann die Nutzung von Daten vielfältige Vorteile ermöglichen, unkontrolliert jedoch auch viele Nachteile – insbesondere für die Betroffenen – mit sich bringen. Die Herausforderung besteht darin, Datennutzung in einem gewissen, gewünschten Rahmen möglich zu machen, um eine vorteilhafte Wertschöpfung aus den Daten ermöglichen zu können, und sie gleichzeitig auch genau auf diesen sinnvollen, gewünschten Rahmen zu beschränken, um Nachteile für die Betroffenen möglichst auszuschließen.

Weiterhin müssen die datenschutzrechtlichen Anforderungen und Auflagen erfüllt werden, um eine rechtmäßige Datenverarbeitung sicherzustellen. Auch dies erfordert, den Umfang der Verarbeitung und Nutzung der Daten auf ein zulässiges, geeignetes, erforderliches, angemessenes und vom Betroffenen gewünschtes Maß sowohl zu ermöglichen als auch einzuschränken.

Im Kontext von WearPrivate sollen den Arbeitnehmern als betroffene Personen möglichst umfangreiche Transparenz und Selbstbestimmung zuteilwerden. Insbesondere sollen Arbeitnehmer sich aller relevanten Datenflüsse stets bewusst sein und für jeden dieser Datenflüsse individuell nach den eigenen Vorstellungen einstellen können, ob und wem sie die Daten in welcher Form preisgeben – zum Beispiel nur aggregiert, anonymisiert oder nur an gewissen Orten oder zu gewissen Zeiten.

Datennutzungskontrolle kann in diesem Zusammenhang ein Baustein sein, um die regelkonforme Verarbeitung, Weitergabe und Nutzung der Daten sicherzustellen.

3.2 Begriffsdefinition „Datennutzungskontrolle“

Jung et al. [18]definieren Datennutzungskontrolle wie folgt:

> Datennutzungskontrolle ist ein technischer Baustein zur Umsetzung von Datensouveränität. Sie versetzt Datengebende in die Lage, frei über die Verwendung ihrer Daten zu bestimmen. Dazu werden feingranulare Auflagen für die Datennutzung spezifiziert und durch spezielle Kontrollmechanismen umgesetzt.

Sensible Daten werden häufig durch Zugriffskontrolle geschützt. Werden darüber hinaus jedoch keine Vorkehrungen getroffen, können Daten, zu denen einmal Zugriff gewährt wurde, vom Empfänger ohne weitere Einschränkungen verwendet werden [17]. Insofern stellt Datenzugriffskontrolle eine

notwendige, jedoch nicht hinreichende Maßnahme dar, um eine legitime Verwendung der Daten sicherzustellen. Datennutzungskontrolle erweitert die klassische Datenzugriffskontrolle um Mechanismen zur Kontrolle der Datennutzung, auch nachdem Zugriff auf die Daten gewährt wurde. Dies umfasst sowohl die Spezifikation als auch die Durchsetzung von Auflagen, wie mit den Daten umzugehen ist.

Datennutzungskontrolle kann also die Nutzung der Daten kontrollieren und auf ein gewünschtes Maß einschränken. Im Falle der Verarbeitung oder Nutzung personenbezogener Daten kann der betroffenen Person umfassende informationelle Selbstbestimmung ermöglicht werden, da nun nicht mehr nur die grundlegende Zugänglichmachung der Daten kontrolliert und entschieden wird, sondern auch die konkrete Nutzung der Daten nach der Zugänglichmachung.

Neben den Entscheidungen der betroffenen Person, die ihre individuellen Präferenzen für die Datennutzung wählen kann, können auch allgemeingültige Regeln vorgesehen werden, um die Datennutzung auf ein rechtlich zulässiges Maß zu beschränken. Dadurch kann bis zu einem gewissen Grad auch Gesetzeskonformität unterstützt und gewährleistet werden.

Im Kontext von WearPrivate kann durch Mechanismen der Datennutzungskontrolle die Verarbeitung und Nutzung der Daten der Wearable-Nutzer (Arbeitnehmer) kontrolliert und eingeschränkt werden. Wearable-Nutzern können Instrumente bereitgestellt werden, mit denen sie ihre individuellen Präferenzen bezüglich der Verarbeitung und Nutzung ihrer Daten festlegen können. Die dadurch zum Ausdruck gebrachten Entscheidungen (Datennutzungsregeln und Einwilligungen) sind von allen beteiligten Akteuren zu respektieren und können teilweise sogar technisch in den Komponenten der Systemkette durchgesetzt werden. Durch eine umfassende Implementierung von Mechanismen zur Datennutzungskontrolle kann die Verarbeitung und Nutzung der Daten der Arbeitnehmer gesetzeskonform und entsprechend ihren individuellen Präferenzen durchgeführt werden.

3.3 Spezifikation von Datennutzungsregeln

Datennutzungsregeln beziehungsweise Datennutzungsrichtlinien (engl.: data usage policies) beschreiben, was mit den Daten getan werden darf oder was nicht mit den Daten getan werden darf. Durch entsprechende Spezifikation von Datennutzungsregeln können also sowohl rechtliche Anforderungen als auch individuelle Wünsche oder Einwilligungen von Betroffenen ausgedrückt werden.

Die Spezifikation von Datennutzungsregeln kann grundsätzlich auf verschiedene Arten und in verschiedenen Formen erfolgen. In der einfachsten Form passiert dies verbal. Damit einher geht jedoch die Schwäche, dass sich Parteien falsch an die Übereinkunft erinnern können, dass es zu Missverständnissen kommt, dass Unwahrheiten behauptet werden oder dass die vereinbarten Regeln schlichtweg nicht beachtet werden. Eine formalere und stärker bindende Form sind schriftliche Vereinbarungen, wie zum Beispiel Verträge. Diese Formen von Datennutzungsregeln eigenen sich zur Aushandlung und Befolgung durch Menschen; es erfolgt keine technische Durchsetzung der Regeln. Bei Verstößen streiten sich die Beteiligten üblicherweise nachträglich vor Gericht um die Interpretation von Formulierungen im Vertragswerk.

Im Rahmen von WearPrivate sind wir insbesondere an technischen Möglichkeiten der Durchsetzung interessiert. Für eine technische Durchsetzung bedarf es einer formalen Spezifikation, die durch Computer interpretiert und verstanden werden kann. Hier können Sprachen wie etwa die Open Digital Rights Language (ODRL) oder die eXtensible Access Control Markup Language (XACML) helfen, die jeweiligen Regeln auszudrücken.

Im Kontext von WearPrivate können wir den betroffenen Personen (Arbeitnehmern) jedoch nicht zumuten, Datennutzungsregeln oder ihre Präferenzen selbst in einer solchen formalen Sprache auszudrücken. Für sie müssen einfach nutzbare und verständliche Bedienelemente vorgesehen werden, mit denen sie ihre Selbstbestimmung souverän ausüben können. Der WearPrivate-Ergebnisbericht D4.2 beschreibt in diesem Zusammenhang relevante Interaktionskonzepte für informationelle Selbstbestimmung und transparente Datennutzung. Der Dienstanbieter muss eine adäquate Umsetzung gewährleisten und den Betroffenen zugänglich machen.

Konzeptionell kann zwischen allgemeingültigen Datennutzungsregeln ohne individuelle Konfiguration, allgemeingültigen Datennutzungsregeln mit individueller Konfiguration und individuellen Datennutzungsregeln unterschieden werden. Dies soll durch nachfolgendes Beispiel verdeutlicht werden:

- Allgemeingültige Datennutzungsregeln ohne individuelle Konfiguration:
    - *Daten zur Herzratenvariabilität dürfen für die individuelle Belastungsmessung verwendet werden. Sie dürfen nicht für Marktforschung verwendet werden.*
- Allgemeingültige Datennutzungsregeln mit individueller Konfiguration
    - *Die Daten zur Herzratenvariabilität dürfen für die individuelle Belastungsmessung verwendet werden, wenn die jeweilige betroffene Person dem zugestimmt hat.*
    - *Die Daten zur Herzratenvariabilität dürfen für Marktforschung verwendet werden, wenn die jeweilige betroffene Person dem zugestimmt hat.*

      Zusätzliche Individuelle Konfigurationen der Betroffenen:

      Konfiguration Person A:

        - *Herzratenvariabilität für individuelle Belastungsmessung: JA*
        - *Herzratenvariabilität für Marktforschung: NEIN*

      Konfiguration Person B:

        - *Herzratenvariabilität für individuelle Belastungsmessung: JA*
        - *Herzratenvariabilität für Marktforschung: JA*
- Individuelle Datennutzungsregeln
    - *Die Daten zur Herzratenvariabilität von Person A dürfen für die individuelle Belastungsmessung verwendet werden, sofern sie auf dem Werksgelände erfasst wurden. Sie dürfen nicht für Marktforschung verwendet werden, es sei denn, dass das aktuelle Jahr ein Schaltjahr ist.*
    - *Die Daten zur Herzratenvariabilität von Person B dürfen für die individuelle Belastungsmessung und für Marktforschung verwendet werden, sofern sie während der Arbeitszeit erfasst wurden.*

Das Konzept allgemeingültiger Datennutzungsregeln ohne individuelle Konfiguration eignet sich besonders zur Formulierung von grundsätzlichen Rahmenbedingungen, die zum Beispiel gesetzlich vorgegeben sein könnten und für alle Betroffenen einheitlich anzuwenden sind. Individuelle Datennutzungsregeln bieten hingegen das höchste Maß an Individualisierung, da sie die Nutzung der Daten eines individuellen Betroffenen im Detail beschreiben. Sind bei der Spezifikation individueller

Datennutzungsregeln keine Grenzen gesetzt, können Betroffene möglicherweise sehr ausgefallene Datennutzungsregeln und Abhängigkeiten im Rahmen ihrer Selbstbestimmung zum Ausdruck bringen (wie im obigen Beispiel exemplarisch durch die Bedingung „Schaltjahr“ dargestellt). Die Ausdruckskraft und -freiheit der Betroffenen bezüglich der Datennutzungsregeln hängt hier wesentlich von der Ausgestaltung der für sie zugänglichen Bedienelemente ab. Dem gegenüber stehen allgemeingültige Datennutzungsregeln mit individueller Konfiguration, die eine Brücke zwischen diesen beiden Extremen schlagen können. Allgemeingültige Datennutzungsregeln mit individueller Konfiguration können allgemeingültig formuliert werden und zugleich die individuellen Präferenzen der jeweiligen Betroffenen berücksichtigen. Die Betroffenen sind jedoch in ihrer Selbstbestimmung insofern eingeschränkt, als sie nur aus den für sie vorgesehenen Einstellungsmöglichkeiten wählen können und ihre Präferenzen nur so weit berücksichtigt werden, wie es die jeweiligen Datennutzungsregeln vorsehen.

Neben dieser Art der Kategorisierung von Datennutzungsregeln kann allgemein auch zwischen erlaubenden und verbietenden Datennutzungsregeln unterschieden werden. Erlaubende Datennutzungsregeln erlauben eine gewisse Nutzung der Daten (ggf. unter Bedingungen oder mit Auflagen). Ein einfaches Beispiel: „Die Daten zur Herzratenvariabilität von Person A dürfen für die individuelle Belastungsmessung verwendet werden, sofern sie auf dem Werksgelände erfasst wurden.“ Dem gegenüber untersagen verbietende Datennutzungsregeln gewisse Datennutzungen. Zum Beispiel: „Die Daten zur Herzratenvariabilität von Person A dürfen NICHT für die individuelle Belastungsmessung verwendet werden, sofern sie außerhalb des Werksgeländes erfasst wurden.“ Ein Beispiel für Auflagen als Teil einer Datennutzungsregel wäre: „Der Belastungsverlauf von Person A darf an medizinische Forschungseinrichtungen in der EU weitergegeben werden. Die Daten müssen jedoch vor der Weitergabe anonymisiert werden und Person A muss informiert werden, an welche Forschungseinrichtung seine Daten weitergegeben wurden.“ Bezüglich der Erlaubnis einer Datenweitergabe könnte der Betroffene möglicherweise auch Datennutzungsregeln für den Empfänger der Daten spezifizieren, die von diesem einzuhalten wären.

Unter Umständen ist für manche Datennutzungen keine Regel vorhanden, die eine Aussage darüber trifft, ob die jeweilige Datennutzung erlaubt oder verboten sein soll. Für diesen Fall ist festzulegen, welche Datennutzungsentscheidung getroffen werden soll, wenn keine Regel für eine spezifische Datennutzung vorliegt. Hier wäre denkbar, dass entweder (A) alle Datennutzungen grundsätzlich erlaubt sind, die nicht ausdrücklich verboten wurden, oder, dass (B) alle Datennutzungen grundsätzlich verboten sind, die nicht ausdrücklich erlaubt wurden.

In Fall A besteht die Gefahr, dass unerwünschte Datennutzungen vergessen wurden zu verbieten und daher ungewollt möglich sind. Darüber hinaus sind die erlaubten Datennutzungen nicht aus den Datennutzungsregeln erkennbar.

In Fall B besteht die Gefahr, dass gewünschte Datennutzungen vergessen wurden zu erlauben und daher ungewollt nicht möglich sind; jedoch sind alle erlaubten Datennutzungen aus den Datennutzungsregeln erkennbar.

Den Betroffenen sollte bekannt und verständlich sein, wie sich Spezifikationslücken in den Datennutzungsregeln auf die Nutzung ihrer Daten auswirken. Um Spezifikationslücken jedoch erkennen und schließen zu können, ist ein detailliertes Wissen bezüglich der möglichen Datennutzungen notwendig, welches in den meisten Fällen den Betroffenen jedoch nicht unterstellt werden kann. Entsprechend sollte das System die Betroffenen unterstützen, ihre Datennutzungsregeln und Präferenzen möglichst lückenlos zu spezifizieren, so dass nur die gewünschten Datennutzungen ermöglicht werden.

Für den Fall, dass es keine Aussage darüber gibt, ob eine spezifische Datennutzung erlaubt oder verboten werden soll, sollte nach dem Prinzip Privacy-by-Default eine Datennutzung im Zweifel eher untersagt werden (Fall B); der Betroffene sollte Datennutzungen ausdrücklich zulassen müssen. Abhängig vom konkreten Anwendungsfall könnte – nach einer Abwägung – aber auch eine andere Umsetzung für manche Datennutzungen sinnvoll sein (Opt-out statt Opt-in). Beispiele hierfür könnten etwa Situationen oder Umgebungen mit akuter Gefahr für Leib und Leben sein.

Darüber hinaus sind widersprüchliche Datennutzungsregeln denkbar. Wie dieser Widerspruch bei der Durchsetzung der Datennutzungsregeln aufgelöst wird, ist abhängig von der Implementierung und könnte zu für den Betroffenen unerwarteten Datennutzungskontrollentscheidungen führen.

So könnte beispielsweise eine Strategie verwendet werden, die eine Datennutzung verbietet, sobald mindestens eine Regel die Datennutzung verbietet (sog. Blacklisting). Alternativ könnte eine Datennutzung erlaubt werden, sobald mindestens eine Datennutzungsregel die Datennutzung erlaubt (sog. Whitelisting).

Eine weitere Option wäre, den Datennutzungsregeln Prioritäten zuzuordnen. Beispielsweise könnten Regeln, die auf gesetzlichen Anforderungen basieren, die individuellen Regeln des Betroffenen bei der Entscheidungsfindung überstimmen.

Um die Spezifikation der Datennutzungsregeln und Präferenzen für die Betroffenen möglichst einfach und die daraus folgenden Datennutzungskontrollentscheidungen für sie möglichst nachvollziehbar und erwartbar zu halten, sollte das System widersprüchliche Datennutzungsregeln möglichst verhindern. Zudem sollten Betroffene auch nur solche Datennutzungen zulassen können, die nicht in Konflikt mit gesetzlichen Anforderungen stehen.

3.4 Verwaltung und Austausch von Datennutzungsregeln

Für die Verwaltung und den Austausch der Datennutzungsregeln gibt es verschiedene konzeptionelle Ansätze, die nachfolgend näher beschrieben werden.

Datennutzungsregeln können an die Daten, die sie betreffen, angeheftet werden und mit diesen eine Einheit bilden; man spricht dann von sogenannten Sticky Policies [21]. Alternativ können die Datennutzungsregeln auch unabhängig von den Daten verwaltet werden und die Daten, für die sie gelten sollen, lediglich referenzieren.

Wenn Daten weitergegeben werden, sind auch die zugehörigen Datennutzungsregeln zu übermitteln, damit der Datenempfänger weiß, was er mit den Daten tun darf. Sticky Policies haben hier den Vorteil, dass sie bei der Weitergabe der Daten ebenfalls weitergegeben werden und sodann dem Empfänger bekannt sind. Ein Nachteil könnte sein, dass dadurch eine spätere Veränderung der Nutzungsregeln erschwert wird, da diese Änderungen an alle Stellen propagiert und in den Daten aktualisiert werden müssten. Zudem kann ein Datenempfänger möglicherweise erst nachdem er die Daten bereits empfangen hat prüfen und sicherstellen, dass er die Daten korrekt handhaben kann und wird.

Bei einer Regelverwaltung unabhängig von den Daten könnten die Regeln zum Beispiel an zentraler Stelle verwaltet und verändert werden. Dies hat jedoch den Nachteil, dass dann alle Datennutzer oder Datenempfänger an diese zentrale Stelle angebunden sein müssen, um die jeweils gültigen Regeln erhalten und berücksichtigen zu können.

Auch ist denkbar, dass die Datennutzungsregeln vor dem Austausch der eigentlichen Daten übermittelt und im Zielsystem konfiguriert werden. Das hat den Vorteil, dass der Empfänger prüfen und

sicherstellen kann, dass die jeweiligen Datennutzungsregeln von ihm umgesetzt werden können und auch werden, noch bevor er die Daten erhält.

Anstelle einer einseitigen Vorgabe der Datennutzungsregeln durch den Dateneigentümer (im vorliegenden Fall der Arbeitnehmer mit seinen persönlichen Vitaldaten) können Datennutzungsregeln zwischen den Beteiligten auch ausgehandelt werden. Der Datenkonsument kann zum Beispiel bestimmte Mindestforderungen an seine Nutzungsrechte stellen und dafür im Gegenzug dem Datenanbieter bestimmte Vorteile einräumen, wie etwa verbesserte Service-Qualität oder höhere Vergütungen. Ein Konzept für die Aushandlung von Datennutzungsrichtlinien zwischen Datenanbietern und Datenkonsumenten bieten etwa die International Data Spaces (IDS) [10].

Solche Verhandlungsstrategien sind überwiegend im kommerziellen Umfeld sinnvoll, um Marktinteressen abzugleichen oder das günstigste Preis-Leistungs-Verhältnis für die Beteiligten zu bestimmen. Im Kontext von Vitaldatenanalysen im Arbeitnehmerkontext erwarten wir jedoch eher nicht, dass die Betroffenen in technisch unterstützte Datennutzungsverhandlungen eintreten, sondern eher, dass die betroffenen Arbeitnehmer im Rahmen der verfügbaren technischen Optionen selbstbestimmt ihre Datennutzungsregeln festlegen. Verhandlungslösungen wären dem gegenüber wenig nutzerfreundlich und relativ zeitraubend für alle Beteiligten.

3.5 Durchsetzung von Datennutzungsregeln

Datennutzungsregeln können auf verschiedene Weise durchgesetzt werden. Im einfachsten Fall wird Datennutzungskontrolle durch organisatorische Maßnahmen betrieben. Beispielsweise können Verträge und Arbeitsanweisungen einen Rahmen bieten, die Datennutzung rechtlich auf den gewünschten und erlaubten Umfang zu beschränken. Diese Form der Durchsetzung basiert im Wesentlichen auf Vertragstreue, Vertrauen und Strafandrohung. Weitergehende Datennutzungen werden jedoch technisch nicht verhindert.

Datennutzungskontrolle kann auch durch technische Maßnahmen betrieben werden, sofern die Datennutzungsregeln formalisiert vorliegen. Statt allein auf Verträge und Arbeitsanweisungen zu vertrauen, werden hierbei technische Kontroll- und Durchsetzungsmechanismen in die jeweiligen Systeme integriert, die eine Nutzung der Daten zur Laufzeit ausschließlich entsprechend den aktuell gültigen Nutzungsrichtlinien zulassen. Interne Datenflüsse können hierzu von den integrierten Kontrollmechanismen kontrolliert und zugelassen oder unterbunden werden. Während klassische Zugriffskontrolle vorwiegend Ressourcenzugriffe durch externe Entitäten kontrolliert und über deren Datenzugang befindet, stehen bei der Datennutzungskontrolle insbesondere auch die systeminternen Ressourcenzugriffe zum Zwecke einer Verarbeitung im Fokus.

Effektive Datennutzungskontrolle erfordert ein geschlossenes, kontrollierbares System. Nur in einem solchen Umfeld können sämtliche Datenflüsse kontrolliert werden. Sind weitere Systeme involviert, so sind auch in diesen entsprechende Kontrollmechanismen vorzusehen, um die Einhaltung der Datennutzungsregeln sicherzustellen. Weiterhin ist eine Kontrolle der Datenflüsse zwischen den Systemen denkbar, um die Datennutzungsregeln auch hier anzuwenden.

Fortschrittliche Mechanismen für Datennutzungskontrolle können einen Datenfluss nicht nur unverändert zulassen und unterbinden, sondern diesen zudem mit veränderten Daten zulassen. Beispielsweise könnten sensitive Informationen aus den Daten entfernt oder verändert werden, bevor diese weiterfließen dürfen. Das ist zum Beispiel sinnvoll, wenn personenbezogene Daten das geschlossene, kontrollierbare System verlassen oder an einen Dritten übermittelt werden sollen. In einem solchen Fall könnten Datennutzungsregeln diesen Datenfluss nicht nur im Ganzen unterbinden

und zulassen, sondern diesen auch unter gewissen Auflagen zulassen, beispielsweise unter der Auflage, dass die Daten vor der Übermittlung zunächst anonymisiert werden müssen.

Im Rahmen von WearPrivate möchten wir – soweit möglich – insbesondere technische Möglichkeiten für Datennutzungskontrolle untersuchen und erproben. Ein technisches System kann die individuellen Datennutzungspräferenzen der Arbeitnehmer nur dann korrekt umsetzen, wenn es diese technisch berücksichtigt und die Verarbeitung zur Laufzeit entsprechend gesteuert wird. Insofern ist die technische Durchsetzung von Datennutzungskontrolle gewissermaßen auch notwendig, um den Arbeitnehmern eine Selbstbestimmung zu ermöglichen, die über eine allgemeine, pauschale Teilnahmeentscheidung hinausgeht. Weiterhin erhoffen wir uns, durch technische Durchsetzungsmaßnahmen, das Risiko eines Datenmissbrauchs reduzieren zu können, was im Hinblick auf die Sensibilität der Daten sinnvoll scheint.

3.6 Kontextsensitive Datennutzungskontrolle

Kontextsensitive Datennutzungsregeln können Kontextinformationen berücksichtigen und die Nutzbarkeit der Daten davon abhängig machen. So kann die Nutzbarkeit der Daten beispielsweise vom Wochentag, von der Uhrzeit, von der aktuellen Position, vom aktuellen Systemzustand oder von Geräteparametern abhängen. Auch ist denkbar, dass Daten nur für ausgewählte Zwecke genutzt werden dürfen.

Im Kontext von WearPrivate ist beispielsweise denkbar, dass die Datennutzungsregeln berücksichtigen, ob sich der Arbeitnehmer gerade am Arbeitsplatz befindet oder nicht. Möglicherweise werden Daten außerhalb des Arbeitsplatzes gar nicht erhoben oder zumindest besonders gekennzeichnet, um sie später von manchen Verarbeitungen auszunehmen oder anders zu behandeln.

Im Projektkontext von WearPrivate vermuten wir einen Bedarf der Arbeitnehmer an kontextsensitiver Datennutzungskontrolle und entsprechenden kontextsensitiven Datennutzungsregeln, Dies ist jedoch noch mit Vertretern der Zielgruppe und weiteren Stakeholdern zu überprüfen.

3.7 MYDATA Control Technologies

MYDATA Control Technologies (kurz MYDATA) ist eine technische Lösung für Datennutzungskontrolle des Fraunhofer-Institut für Experimentelles Software Engineering IESE [22]. MYDATA kann in Software integriert werden, um Datenflüsse entsprechend den konfigurierten Datennutzungsregeln zu kontrollieren. MYDATA basiert auf dem Integrated Distributed Data Usage Control Enforcement (IND²UCE) Framework, welches das Fraunhofer IESE im Rahmen seiner Forschung im Bereich der Datensouveränität entwickelt hat.

In Anlehnung an den XACML-Standard [24] unterscheidet MYDATA zwischen den nachfolgenden Systemkomponenten:

- Policy Enforcement Point (PEP)
  Für jedes MYDATA-fähige System muss ein sogenannter Policy Enforcement Point (PEP) in das System integriert werden. Seine Hauptaufgabe besteht darin, Systemereignisse und Datenflüsse innerhalb des Systems abzufangen bzw. zu überwachen, eine Datennutzungskontrollentscheidung vom Policy Decision Point (PDP) zu erfragen und den Datenfluss entsprechend dieser Entscheidung zu ändern.

- Policy Decision Point (PDP)
  Der Policy Decision Point (PDP) verarbeitet die Entscheidungsanforderungen der PEPs. Dazu wertet er die zutreffenden Datennutzungsregeln anhand des übermittelten Systemereignisses aus und leitet daraus eine Datennutzungskontrollentscheidung ab. Diese wird dem PEP zur technischen Durchsetzung mitgeteilt.
- Policy Information Point (PIP)
  Policy Information Points (PIPs) stellen zusätzliche Informationen für einen PDP bereit, um den Entscheidungsprozess zu unterstützen. PIPs erweitern MYDATA um die Möglichkeit, Informationen abzurufen, die unabhängig vom aktuellen Systemereignis sind. Standardmäßig hat der PDP keinen Zugriff auf andere Informationen als die, die von einem eingehenden Ereignis geliefert oder in der jeweiligen Datennutzungsregel selbst angegeben wurden. Es gibt jedoch viele Beispiele, in denen dies nicht ausreicht und zusätzliche Informationen für eine Datennutzungskontrollentscheidung erforderlich sind. Beispiele hierfür könnten sein:
    - Kontextinformationen: Befindet sich die Person gerade an ihrem Arbeitsplatz, im Außeneinsatz oder in einer privaten Situation?
    - Verzeichnisinformationen: Welche Rolle hat die Person in ihrer Organisation?
    - Einstellungen: Hat die Person bestimmte Datenschutzeinstellungen aktiviert?
- Policy Execution Point (PXP)
  Eine Datennutzungsregel kann in bestimmten Fällen auch das Auslösen einer Aktion beinhalten. Diese Aktionen werden von sogenannten Policy Execution Points (PXPs) ausgeführt. Die Ausführung der Aktion wird im MYDATA-Modell bei Bedarf durch den PDP ausgelöst. Beispiele für Aktionen sind: Benachrichtigungsversand, Datenlöschung oder Kontosperrung.
- Policy Management Point (PMP)
  Der sogenannte Policy Management Point (PMP) verwaltet die Komponenten und Datennutzungsregeln technisch. Er stellt dem PDP die aktuell gültigen Datennutzungsregeln zur Entscheidungsfindung zur Verfügung. Der PMP kommuniziert alle Änderungen an der Menge der aktiven Datennutzungsregeln an den PDP.
- Policy Administration Point (PAP)
  Der Policy Administration Point (PAP) bietet eine Benutzeroberfläche zur Spezifikation von Datennutzungsregeln. Die damit spezifizierten Datennutzungsregeln werden an den PMP übertragen, der diese in seinen Regelbestand aufnimmt und verwaltet.

Abbildung 6 gibt einen Überblick über diese Komponenten und ihre Beziehungen zueinander.

Zur Laufzeit arbeiten diese Komponenten zusammen, um Datennutzungskontrolle im System technisch umzusetzen. Die entsprechenden Kommunikationsflüsse sind in Abbildung 7 dargestellt. Grundsätzlich lassen sich die Aktivitäten „Erzeugung von Datennutzungsregeln“ und „Technische Durchsetzung der Datennutzungsregeln“ unterscheiden.

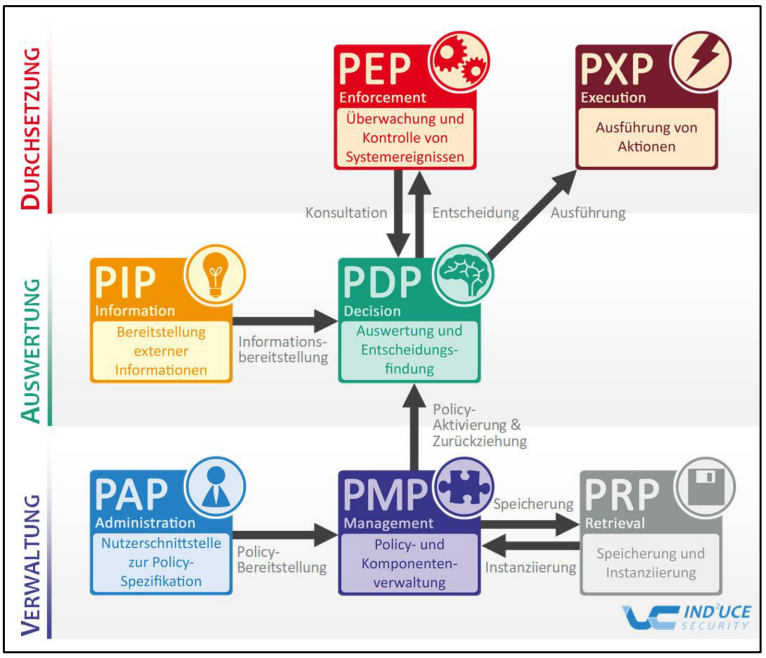

Abbildung 6 Komponenten von MYDATA gemäß IND²UCE-Framework [11]

Erzeugung von Datennutzungsregeln:

1. Der Nutzer spezifiziert eine Datennutzungsregel (Policy) mit dem PAP und der PAP überträgt diese an den PMP
2. Der PMP nimmt die Policy in seinen Bestand auf und installiert sie durch Übermittlung an den PDP im System

➔ Die Datennutzungsregel ist dem PDP nun bekannt und wird künftig bei der Entscheidungsfindung berücksichtigt

Technische Durchsetzung der Datennutzungsregeln:

1. Der PEP fängt den Datenfluss ab
2. Der PEP fordert eine Datennutzungskontrollentscheidung beim PDP an.
3. Der PDP wertet die aktiven und passenden Datennutzungsregeln aus.
   Optional:
   a. Der PDP konsultiert einen PIP, um weitere Informationen zu berücksichtigen
   b. Der PIP liefert die gewünschte Information als Antwort an den PDP

   Optional:
   c. Der PDP konsultiert einen PXP um eine Aktion auszulösen
   d. Der PXP antwortet dem PDP, ob die gewünschte Aktion erfolgreich durchgeführt werden konnte
4. Der PDP ermittelt die finale Datennutzungskontrollentscheidung
5. Der PDP telt dem PEP seine Datennutzungskontrollentscheidung mit

6. Der PEP setzt die Datennutzungskontrollentscheidung des PDP durch (Datenfluss zulassen, verändern oder verhindern)

➔ Die Datennutzung wurde entsprechend den geltenden Datennutzungsregeln zugelassen, verhindert oder mit veränderten Daten zugelassen.

Damit Datennutzungskontrolle mit MYDATA funktioniert, müssen PEPs an den richtigen Stellen im System integriert und mit einem PDP verbunden werden. Anschließend können geeignete Datennutzungsregeln (Policies) spezifiziert und installiert werden.

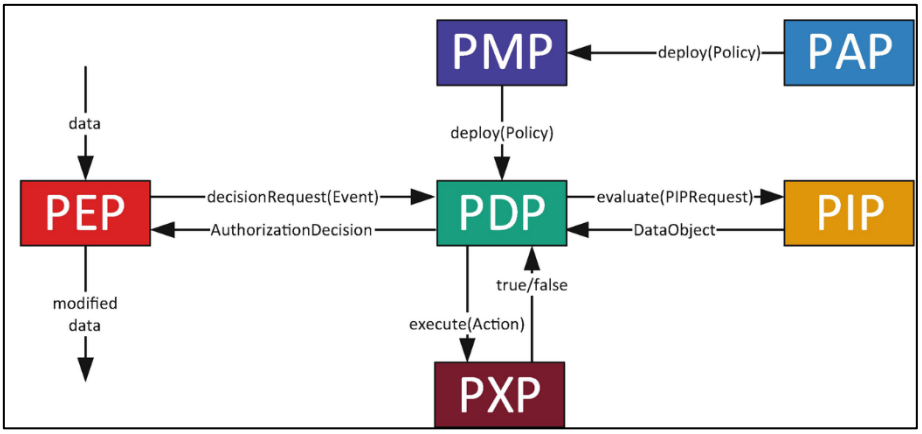

Abbildung 7 Kommunikationsflüsse zur technischen Durchsetzung der Datennutzungskontrolle [17]

Der von MYDATA angebotene PAP richtet sich an Experten, die die Datennutzungsregeln in der MYDATA-eigenen Sprache für Datennutzungsregeln spezifizieren. Sie werden dabei durch Funktionen zur automatischen Vervollständigung und einer Prüfung der Regel auf syntaktische Korrektheit unterstützt. Benutzerfreundliche PAPs für Endnutzer sind, abhängig vom konkreten Anwendungsfall und der konkreten Zielgruppe, durch das MYDATA-nutzende Unternehmen selbst für ihre Lösung zu implementieren und können per API an bereits vorhandene Komponenten (z. B. den PMP) andocken.

Im Kontext von WearPrivate müsste entweder ein PAP für die Arbeitnehmer entwickelt werden, damit diese individuelle Datennutzungsregeln erzeugen und konfigurieren können, oder – sofern allgemeingültige Datennutzungsregeln mit individueller Konfiguration verwendet werden – zumindest eine Benutzerschnittstelle zur Verwaltung der individuellen Präferenzen, die dann bei der Auswertung der Datennutzungsregeln via PIP berücksichtigt werden. Im letztgenannten Fall würden die allgemeingültigen Datennutzungsregeln mit individueller Konfiguration von einem Experten spezifiziert, der sich des von MYDATA bereitgestellten PAP für Experten bedienen könnte. Bei allgemeingültigen Datennutzungsregeln mit individueller Konfiguration ergibt sich jedoch die Herausforderung, dass die Datennutzungsregeln ausreichend umfassend und ausreichend individuell konfigurierbar gestaltet werden müssen, so dass die Selbstbestimmungsbedarfe aller teilnehmenden Arbeitnehmer befriedigt werden können. Auch eine Kombination aus allgemeingültigen Datennutzungsregeln ohne individuelle Konfiguration, allgemeinen Datennutzungsregeln mit individueller Konfiguration und individuellen Datennutzungsregeln ist denkbar.

MYDATA bietet mit seinem Open Source SDK eine Implementierung für die Programmiersprache Java an. Sollten im Rahmen von WearPrivate andere Programmiersprachen für die Komponenten der Systemkette verwendet werden, sind verschiedene Optionen denkbar, wie MYDATA dennoch integriert und verwendet werden kann:

1. Entwicklung von PEPs in der jeweiligen Programmiersprache und Verwendung des bestehenden PDP über programmiersprachenunabhängige Schnittstellen
2. Java-PEP externalisieren und als Dienst bzw. „Sidecar“ neben der Komponente betreiben; Integration eines leichtgewichtigen PEP in der jeweiligen Programmiersprache, welcher den Java-PEP über eine programmiersprachenunabhängige Schnittstelle einbindet und Teile der Durchsetzung an diesen delegiert.

Im Kontext der International Data Spaces (IDS) wurde eine Integration von MYDATA in die Konnektoren nach Option 2 durchgeführt [10]. Die Java-Komponente, welche MYDATA beinhaltet, wird dort als „Usage Control App“ bzw. „Usage Control Container“ bezeichnet und bei der technischen Durchsetzung der Datennutzungsregeln eingesetzt.

Weitere technische Details zu MYDATA werden in der MYDATA Entwicklerdokumentation [22] beschrieben.

3.8 Datennutzungskontrolle in den WearPrivate-Systemkomponenten

Für eine umfassende Kontrolle der Datennutzung im Gesamtsystem empfiehlt sich eine Integration der Kontrollmechanismen in alle Komponenten der Systemkette. Gemäß den jeweiligen Datennutzungsregeln können dann entsprechende Maßnahmen in den Komponenten durchgeführt und der regelkonforme Umgang mit den Daten entlang der gesamten Systemkette sichergestellt werden. Darüber hinaus bietet eine Integration in allen Systemkomponenten den Vorteil, dass Maßnahmen – abhängig von den technischen Möglichkeiten – möglichst früh in der Systemkette greifen können, also sensible Daten beispielsweise erst gar nicht oder zumindest nicht ohne vorherige Anonymisierung übermittelt werden. Das zahlt unter anderem auf den Verarbeitungsgrundsatz der Datensparsamkeit ein und folgt dem Need-to-know-Prinzip.

Die effektive Durchsetzung der Datennutzungskontrolle in den Komponenten der Systemkette erfordert eine korrekte Integration der Kontrollmechanismen an allen relevanten Stellen. Hierbei können Fehler gemacht werden. Wiederkehrende Sicherheitskontrollen der Software und externe Reviews können dazu beitragen, die korrekte Integration der Kontrollmechanismen und deren Effektivität sicherzustellen.

Nachfolgend wird beschrieben, welche Wirkung die Integration von Datennutzungskontrolle in die Komponenten der Systemkette jeweils haben kann.

3.8.1 Datennutzungskontrolle auf dem Wearable

Datennutzungskontrolle auf dem Wearable kann dafür sorgen, dass

- Daten nur dann an das Smartphone und die App übertragen werden, wenn der Nutzer dies wünscht (Unterbinden der Datenbereitstellung);
- nur die Daten übertragen werden, die der Nutzer wünscht (nicht alle Metriken bereitstellen);
- die Daten vor einer Übertragung verändert (z. B. anonymisiert) werden (veränderte Werte bereitstellen);
- jegliche Datennutzung, -verarbeitung, und -weitergabe protokolliert wird

Eine Durchsetzung von Datennutzungskontrolle im Wearable erfordert entsprechende Anpassungen an der Firmware. Im Rahmen des Projekts haben wir – aufgrund des Wechsels im Konsortium – keine

Möglichkeit mehr, Änderungen an der Firmware des Wearables vorzunehmen. Folglich werden sich die Arbeiten diesbezüglich auf Konzepte beschränken müssen.

3.8.2 Datennutzungskontrolle auf dem Smartphone / in der App

Datennutzungskontrolle auf dem Smartphone oder in der App kann dafür sorgen, dass

- Daten nur dann von der App verarbeitet werden, wenn der Nutzer dies wünscht (Unterbindung der Datenannahme);
- nur die Daten von der App verarbeitet werden, die der Nutzer wünscht (nicht alle Daten annehmen);
- nur jene Verarbeitungen in der App durchgeführt werden, die der Nutzer wünscht;
- die Daten nur so lange in der App gespeichert werden, wie es der Nutzer wünscht;
- Daten nur an die vom Nutzer gewünschten Empfänger / Systeme / Anbieter übertragen werden („An wen?“);
- Daten nur dann an den Analyseanbieter übertragen werden, wenn der Nutzer dies wünscht („Wann?“ / „In welchen Fällen?“);
- nur die Daten an den Analyseanbieter übertragen werden, die der Nutzer wünscht („Welche Daten?“)
- die Daten vor einer Übertragung an den Analyseanbieter verändert werden („In welcher Form?“)
- jegliche Datennutzung, -verarbeitung, und -weitergabe protokolliert wird

Die gewünschten Maßnahmen können auch kontextspezifisch angewendet werden. Möglicherweise soll die Datenübertragung an die Cloud zum Beispiel nur in den Mittagspausen (Zeit) oder außerhalb des Betriebsgeländes (Ort) unterbunden werden.

Denkbare Kontextinformationen sind zum Beispiel:

- Zeit
    - Uhrzeit
    - Wochentag
    - Datum
- Ort/Position
    - GPS (longitude, latitude)
    - Höhe (height)
    - Geo-Fences (z. B. nur in Reichweite eines lokalen WLANs)
- Situation / Kontext
    - Arbeit / Freizeit
    - In Gefahr / nicht in Gefahr
- Geräteparameter
    - Verbindung zu einem bestimmten Netzwerk

    - Modus
- Einstellungen des Nutzers
    - Einwilligungen, Opt-In, Opt-Out
    - Ausgewählte Option

Die Durchsetzung der Datennutzungskontrolle in der App erfordert eine korrekte Integration der Kontrollmechanismen im Code der Anwendung. Zudem müssen die vom Anwender konfigurierten Einstellungen und Datennutzungsregeln von der Implementierung korrekt berücksichtigt und durchgesetzt werden. Im Rahmen des Projekts wird die App im Konsortium neu entwickelt. Insofern besteht hier die Möglichkeit, Mechanismen der Datennutzungskontrolle (z. B. MYDATA) zu integrieren und zu erproben.

3.8.3 Datennutzungskontrolle im System des Analyseanbieters

Datennutzungskontrolle im System des Analyseanbieters kann dafür sorgen, dass

- Daten nur dann im System des Analyseanbieters verarbeitet werden, wenn der Nutzer dies wünscht (Unterbindung der Datenannahme) – idealerweise sollten die Daten gar nicht erst an den Analyseanbieter übertragen werden, wenn der Betroffene keine Verarbeitung durch diesen wünscht.
- nur die Daten im System des Analyseanbieters verarbeitet werden, die der Nutzer wünscht (nicht alle Daten annehmen) – idealerweise sollten die Daten gar nicht erst an den Analyseanbieter übertragen werden.
- Daten zunächst anonymisiert werden, bevor diese gespeichert oder anderweitig verarbeitet werden – idealerweise sollte das bereits erfolgt sein, bevor die Daten an den Analysedienst übertragen werden.
- nur jene Verarbeitungen im System des Analyseanbieters durchgeführt werden, die der Nutzer wünscht
    - Daten nur dann für die Erzeugung eines Gruppenberichts verwendet werden, wenn der Nutzer dies wünscht
    - Nur die Daten für die Erzeugung eines Gruppenberichts verwendet werden, die der Nutzer wünscht
    - Nur dann Alerts erzeugt werden, wenn der Nutzer dies wünscht
    - Nur solche Alerts erzeugt werden, die der Nutzer wünscht
- jegliche Datennutzung, -verarbeitung, und -weitergabe protokolliert wird
- Daten aus dem System des Analyseanbieters gelöscht werden, sobald diese nicht mehr benötigt werden (z. B. Sensor-Rohdaten)
- Daten nur so lange im System des Analyseanbieters gespeichert werden, wie es der Nutzer wünscht (z. B. Belastungsverläufe)
- Daten nur dann weitergegeben werden (z. B. an den Arbeitgeber), wenn es der Nutzer wünscht
- Daten nur in dem Umfang weitergegeben werden, wie es der Nutzer wünscht

- Daten vor einer Weitergabe verändert (z. B. anonymisiert) werden
- Daten des Nutzers gelöscht werden, wenn dieser nicht mehr teilnehmen möchte
- Daten, die in der Freizeit erhoben wurden, anders verarbeitet werden als Daten, die im Arbeitskontext erhoben wurden.

Die Durchsetzung der Datennutzungskontrolle im System des Analyseanbieters erfordert eine korrekte Integration der Kontrollmechanismen an allen relevanten Stellen im System des Analyseanbieters. Zudem müssen die vom Nutzer konfigurierten Einstellungen und Datennutzungsregeln von der Implementierung korrekt berücksichtigt und durchgesetzt werden. Im Rahmen des Projekts entwickelt ein Partner die Cloud-Komponenten des Analyseanbieters und hat die Kontrolle über deren Quelltext. Insofern besteht hier die Möglichkeit, Mechanismen der Datennutzungskontrolle (z. B. MYDATA) zu integrieren und zu erproben.

3.8.4 Datennutzungskontrolle beim Arbeitgeber

Datennutzungskontrolle beim Arbeitgeber kann dafür sorgen, dass

- Daten nur von spezifischen Mitarbeitern eingesehen werden können
- Daten nur betrachtet, aber nicht exportiert oder weitergegeben werden können
- jegliche Datennutzung, -verarbeitung, und -weitergabe protokolliert wird
- Zugriff auf die Daten nur an ausgewählten Orten und Geräten möglich ist
- Zugriff auf die Daten nur zu bestimmten Uhrzeiten möglich ist
- Daten automatisch gelöscht werden
- Daten vor der Anzeige verändert werden (Filtern, Maskieren, Wasserzeichen, …)

Abhängig von der Art und Weise, in der der Arbeitgeber Zugriff auf Berichte oder Daten erhält, kann Datennutzungskontrolle unterschiedlich durchgeführt werden. Erhält der Arbeitgeber die Daten analog per Post (z. B. als Brief), so ist die Datennutzungskontrolle durch organisatorische Maßnahmen sicherzustellen. Erhält er die Daten per E-Mail (z. B. als PDF-Datei im Anhang), so kann durch technische und organisatorische Maßnahmen eine Datennutzungskontrolle durchgeführt werden. Erhält er Zugriff auf die Daten durch ein technisches System, so können in diesem System Datennutzungskontrollmechanismen technisch implementiert werden.

3.9 Datennutzungsregeln für Selbstbestimmung

Im Kontext von WearPrivate können bezüglich der Selbstbestimmung der Arbeitnehmer die nachfolgenden, grundlegenden Bereiche unterschieden werden:

1. Selbstbestimmung über die grundsätzliche Teilnahme am Programm
2. Selbstbestimmung darüber, welche Daten der Analysedienstleister erhält
   ➔ Datenübertragung an den Analysedienstleister kontrollieren und einschränken
3. Selbstbestimmung darüber, was der Analysedienstleister mit den Daten tun darf
   ➔ Datennutzung durch den Analysedienstleister kontrollieren und einschränken

Bezüglich der Selbstbestimmung der Arbeitnehmer in den Bereichen 2 und 3 können Datennutzungsregeln spezifiziert und durchgesetzt werden. Datennutzungsregeln im Bereich 2 sind

vor einer Datenübertragung an den Analysedienstleister durchzusetzen (z. B. in der App), so dass der Analysedienstleister nur genau die Daten erhält, die der Arbeitnehmer diesem bereitstellen möchte. Datennutzungsregeln im Bereich 3 sind beim Analysedienstleister durchzusetzen (z. B. in dessen Systemen), so dass nur genau die Verarbeitungen und Nutzungen der Daten möglich sind, mit denen der Arbeitnehmer einverstanden ist.

Abhängig davon, wie weitreichend die Selbstbestimmung der Arbeitnehmer im jeweiligen Anwendungsfall sein soll und welche Entscheidungen die Arbeitnehmer treffen können sollen, sind geeignete Dialoge mit Einstellungsmöglichkeiten vorzusehen, mit denen die Arbeitnehmer ihre Präferenzen festlegen können.

Beispiele für mögliche Einstellungen im Kontext von WearPrivate sind:

- Teilnahme am Gruppenbericht: Ja/Nein
    - Durchsetzung beim Analysedienstleister
- Daten verrauschen: nein / beste Datenqualität vs. ja / bessere Anonymität; ggf. auch mehrere Anonymisierungsstufen (0, 1, 2, …)
    - Durchsetzung in der App
- Datennutzung zur Verbesserung des Dienstes: Ja/Nein
    - Durchsetzung beim Analysedienstleister
- Datennutzung für Forschung: Ja/Nein
    - Durchsetzung beim Analysedienstleister
- Weitergabe anonymer Daten an Dritte: Ja/Nein
    - Durchsetzung beim Analysedienstleister
- Datenerfassung während der Arbeitszeiten: Ja/Nein
    - Durchsetzung in der App
- Datenerfassung außerhalb der Arbeitszeiten: Ja/Nein
    - Durchsetzung in der App
- Datenerfassung innerhalb des Betriebsgeländes: Ja/Nein
    - Durchsetzung in der App
- Datenerfassung außerhalb des Betriebsgeländes: Ja/Nein
    - Durchsetzung in der App
- Erzeugung von Mitteilungen: Nein / Ja, alle / manche
    - Durchsetzung beim Analysedienstleister
- Speicherdauer des persönlichen Belastungsverlaufs: Unbegrenzt, 5 Jahre, 2 Jahre, 1 Jahr, 6 Monate, 2 Monate
    - Durchsetzung beim Analysedienstleister
- Auswahl der Sensordaten, die übermittelt werden: Herzratenvariabilität, Beschleunigung, Standort, Lautstärke, …
    - Durchsetzung in der App

3.10 Datennutzungsregeln für Transparenz

Datennutzungsregeln können die Protokollierung von bestimmten Datennutzungen vorschreiben. Ausgehend von den so entstehenden Datennutzungsprotokollen können die tatsächlich erfolgten Datennutzungen identifiziert und in einer geeigneten Form dargestellt werden, um Transparenz über die tatsächlich erfolgten Datennutzungen zu ermöglichen.

Weiterhin kann die Menge der Datennutzungsregeln selbst Aufschluss geben, welche Datennutzungen im Gesamtsystem zulässig und deshalb möglich sind. Um sicherzustellen, dass keine darüberhinausgehenden, verborgenen Datennutzungen möglich sind, kann der Grundsatz „was nicht explizit erlaubt ist, ist verboten“ verwendet werden. Dadurch würde sichergestellt, dass für jede mögliche Datennutzung eine entsprechende datennutzungsgewährende Datennutzungsregel existiert.

3.11 Grenzen der Selbstbestimmung

Im Kontext von WearPrivate stehen sich Selbstbestimmungs- und Datennutzungsinteressen gegenüber. Selbstbestimmungsinteressen sehen wir vor allem auf der Seite der Arbeitnehmer, die die Kontrolle über die Verarbeitung ihrer sensiblen Vitaldaten behalten möchten und vor negativen Folgen der Datenverarbeitung zu schützen sind. Datennutzungsinteressen sehen wir bei allen Beteiligten (also Arbeitnehmer, Arbeitgeber und Analyseanbieter).

Eine Herausforderung in diesem Zusammenhang ist, Datennutzung zu ermöglichen und gleichzeitig noch größtmöglichen Schutz und Selbstbestimmung für die Arbeitnehmer zu bieten. Arbeitnehmer können ihre Selbstbestimmung jedoch nicht grenzenlos ohne Folgen ausüben. Ab einem gewissen Punkt kann die vom Analyseanbieter angebotene Dienstleistung schlichtweg nicht mehr erbracht werden. Ist ein Arbeitnehmer mit einer notwendigen Verarbeitung nicht einverstanden, dann kann die jeweilige Dienstleistung eben nicht durchgeführt werden. Der Arbeitnehmer muss zumindest mit den minimal erforderlichen Datenverarbeitungen einverstanden sein, damit zumindest eine grundlegende Diensterbringung möglich ist; ansonsten steht es ihm natürlich frei – im Rahmen seiner Selbstbestimmung – nicht am Programm teilzunehmen und ganz auf die Dienstleistung zu verzichten.

Ebenso ist eine obere Schranke denkbar, die die Freiheit der Arbeitnehmer begrenzt, um sie vor zu großen Risiken und Datenschutzfolgen zu schützen und um Konformität mit anderen gesetzlichen Auflagen sicherzustellen.

Im Rahmen dieser Leitplanken kann der Arbeitnehmer dann seine Dienstnutzung und die Nutzung seiner Daten durch den Dienst selbstbestimmt gestalten.

Nach unserer Einschätzung ergibt es jedoch wenig Sinn, Arbeitnehmer beliebige Datennutzungsregeln formulieren zu lassen. Vielmehr sehen wir den Analyseanbieter in der Pflicht zu definieren, welche Daten und Datenverarbeitungen für die von ihm angebotenen Dienstleistungen minimal erforderlich sind und welche Daten und Datenverarbeitungen optional darüber hinaus noch wünschenswert sind, um zum Beispiel mehr Funktionen bereitzustellen oder um bessere Ergebnisse zu erzielen. Es gilt zu erarbeiten, welche sinnvollen Möglichkeiten zur Selbstbestimmung dem Arbeitnehmer im jeweiligen Anwendungsfall angeboten werden können und sollen.

Effektive Anonymisierung und eine datenschutzfreundliche, verständliche und nachvollziehbare Ausgestaltung des Gesamtsystems kann in diesem Zusammenhang hilfreich sein, um Arbeitnehmer mit großen Datenschutzbedenken zu überzeugen und auch diesen einen Nutzen aus der Verarbeitung ihrer Daten zu ermöglichen.

Quellenverzeichnis

[1] Alduwaile D. et al. (2020): Single Heartbeat ECG Biometric Recognition using Convolutional Neural Network. In *2020 International Conference on Advanced Science and Engineering (ICOASE).*

[2] Camenisch, J. et al. (1997): Efficient group signature schemes for large groups. In: Advances in Cryptology — CRYPTO '97. Lecture Notes in Computer Science, vol 1294. Springer, Berlin, Heidelberg.

[3] Chaum, D. (1983): Blind signature system. In Advances in Cryptology—CRYPTO'83, Santa Barbara, CA, USA.

[4] Chou, C. -c. et al. (2007): An efficient anonymous communication protocol for peer-to-peer applications over mobile ad-hoc networks. In *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 1.

[5] Cohen, A. (2022): Attacks on Deidentification's Defenses. To appear in USENIX Security 2022.

[6] Dwork C. et al. (2006): Calibrating Noise to Sensitivity. In Private Data Analysis, Proceedings of the 3rd Theory of Cryptography Conference.

[7] Dwork, C. (2006): Differential privacy. In: Automata, languages and programming. Springer Berlin Heidelberg.

[8] Dwork, C. (2008): Differential privacy: a survey of results. In Proceedings of the 5th international conference on Theory and applications of models of computation (TAMC'08). Springer Berlin Heidelberg.

[9] Dwork C. et al. (2014): The Algorithmic Foundations of Differential Privacy. In: Foundations and Trends in Theoretical Computer Science, Vol. 9, No. 3–4, pp. 211–407, https://doi.org/10.1561/0400000042

[10] Eitel, A., Jung, C., Brandstädter, R., Hosseinzadeh, A., Bader, S., Kühnle, C., Birnstill, P., Brost, G., Gall, M., Bruckner, F., Weißenberg, N., & Korth, B. (2021): Usage control in the international data spaces. International Data Spaces Association. https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-Position-Paper-Usage-Control-in-the-IDS-V3.pdf 8zuletzt besucht: 2023-03-290

[11] Feth, D., Jung, C. (2019): 10 Jahre Forschung zu Datennutzungskontrolle am Fraunhofer IESE. https://www.iese.fraunhofer.de/blog/10-jahre-datennutzungskontrolle-am-fraunhofer-iese/ [zuletzt besucht: 2023-03-27}

[12] Finster, S. (2014): Protokolle für privatsphärengerechtes Smart Metering. Dissertation, Fakultät für Informatik, Karlsruher Institut für Technologie, DOI: 10.5445/IR/1000042824, https://publikationen.bibliothek.kit.edu/1000042824/3188654

[13] Gentry, C. et al. (2009): A fully homomorphic encryption scheme. PhD Thesis, Stanford: Stanford University, https://dl.acm.org/doi/10.5555/1834954

[14] Golle, P. (2006): Revisiting the uniqueness of simple demographics in the US population. In Proceedings of the 5th ACM workshop on Privacy in electronic society (WPES '06). Association for Computing Machinery, New York, NY, USA.

[15] He, W. et al. (2007): PDA: Privacy-preserving data aggregation in wireless sensor networks. In: IEEE INFOCOM 2007 - 26th IEEE International Conf. Comput. Commun., Anchorage.

[16] Hsu, J. et al. (2014): Differential Privacy: An Economic Method for Choosing Epsilon. In: IEEE 27th Computer Security Foundations Symposium.

[17] Jung, C., Dörr, J. (2022): Data Usage Control. In: Otto, B., ten Hompel, M., Wrobel, S. (eds) Designing Data Spaces. Springer, Cham. https://doi.org/10.1007/978-3-030-93975-5_8

[18] Jung, C., Eitel, A., Feth, D. (2022): Datensouveränität in Digitalen Ökosystemen: Daten nutzbar machen, Kontrolle behalten. In: Rohde, M., Bürger, M., Peneva, K., Mock, J. (eds) Datenwirtschaft und Datentechnologie. Springer Vieweg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-65232-9_15

[19] Lindell, Y. (2005): Secure multiparty computation for privacy preserving data mining. In Encyclopedia of Data Warehousing and Mining. IGI Global.

[20] Machanavajjhala, A. et al. (2007): L-diversity: Privacy beyond k-anonymity. In ACM Trans. Knowl. Discov. Data 1, 1.

[21] Mont, M. C., & Pearson, S. (2011): Sticky policies: An approach for managing privacy across multiple parties. *Computer, 44*(9), 60–68.

[22] MYDATA Control Technologies. https://www.mydata-control.de/ [zuletzt besucht: 2023-03-27]

[23] MYDATA Control Technologies – Developer Documentation. https://developer.mydata-control.de/, zuletzt besucht: 2023-03-27

[24] Parducci, B., Lockhart, H., & Rissanen, E. (2013): eXtensible Access Control Markup Language (XACML) Version 3.0. OASIS Standard [Online]. Available: https://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.html

[25] Reed, M. et al. (1998): Anonymous connections and onion routing. In: IEEE Journal on Selected areas in Communications 16.4: 482-494.

[26] Reiter, M. et al. (1998): Crowds: Anonymity for web transactions. In: ACM transactions on information and system security (TISSEC) 1.1.

[27] Rivest R. et al. (2006): How to leak a secret: theory and applications of ring signatures. In: Theoretical Computer Science: essays in Memory of Shimon Even. Springer-Verlag, Berlin, Heidelberg.

[28] Samarati, P. et al. (1998): Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. In: Proceedings of the IEEE Symposium on Research in Security and Privacy, Oakland, CA.

[29] Steffes, B. et al. (2022): Einwilligung oder Anonymisierung? Rechtliche Implikationen der Datenverarbeitung im Beschäftigungskontext. In: Recht DIGITAL – 25 Jahre IRIS: Tagungsband des 25. Internationalen Rechtsinformatik Symposions IRIS 2022, Österreichische Computer Gesellschaft.

[30] Xu et al. (2019): KEH-Gait: Using Kinetic Energy Harvesting for Gait-based User Authentication Systems. In IEEE Transactions on Mobile Computing, vol. 18, no. 1.

[31] Yao, A. C. (1982): Protocols for secure computations. In: Proceedings of 23th Annual Symposium on Foundations of Computer Science (FOCS ’82).