Teil V: Technische Unterstützung beim Daten- und Identitätsmanagement

Die Vision eines Personal Information Management-System (PIMS) durch automatisierte Datenschutzselbstauskunft

Sebastian Wilhelm, Dietmar Jakob, Armin Gerl und Sascha Schiegg

Zusammenfassung

Mit dem Inkrafttreten der Datenschutz-Grundverordnung (DSGVO) und der Novelierung des Bundesdatenschutzgesetzes (BDSG) wurden Regelungen zum Schutz *personenbezogener Daten (pbD)* verstärkt und in einem europäischen Framework implementiert. Dies beinhaltet ein gestärktes Recht auf Auskunft über *pbD*, das jeder betroffenen Person mindestens einmal jährlich die Möglichkeit gibt, eine *Datenschutzselbstauskunft (DSA)* bei der datenverarbeitenden Stelle anzufordern (vgl. Art. 12-15 DSGVO). Die Umsetzung dieser Rechte stellt jedoch Herausforderungen für beide Seiten, Betroffene und Datenverarbeitende, dar.

Um diese Herausforderungen zu überwinden, wird in diesem Artikel ein zweiteiliges Framework eines *Personal Information Management Systems (PIMS)* vorgestellt. Dieses System soll sowohl Betroffenen als auch Datenverarbeitenden dabei helfen, *DSA-Auskünfte* anzufordern bzw. zu bearbeiten. Ein sogenanntes *Monitoring Tool for Personal Data (MoP)* unterstützt Betroffene dabei, DSA-Anfragen automatisiert zu stellen und die Datenkopien zu interpretieren. Ein Komplementärsystem namens *Tool for automated Data Self-Disclosure Request Processing (TaP)* hilft den Datenhaltenden, *DSA-Anfragen* voll-/teilautomatisch zu beantworten.

Zusammenfassend zielt das Framework auf die Wahrung der informationellen Selbstbestimmung der Bürger:innen durch eine erleichterte Anforderung einer *DSA* sowie eine ökonomischere Bearbeitung solcher Ersuchen seitens der Datenhaltenden ab.

1 Motivation und Problemstellung

Mit dem Inkrafttreten der Datenschutz-Grundverordnung (DSGVO) und der Novelierung des Bundesdatenschutzgesetzes (BDSG) wurden die Regelungen zum Schutz von *personenbezogenen Daten (pbD)* gestärkt, indem insbesondere die Betroffenenrechte weiter präzisiert wurden. Dies betrifft

unter anderem spezifizierte Regelungen zur Informations- und Transparenzpflicht (Art. 12-15 DSGVO) sowie Regelungen zur Berichtigung und Löschung, Einschränkung der Verarbeitung, Mitteilungspflicht und Datenübertragbarkeit (Art. 16-20 DSGVO). Bei der Wahrnehmung der Betroffenenrechte nach den Art. 16-20 DSGVO bzw. §§ 32-37 BDSG ergeben sich jedoch Herausforderungen für die Bürger:innen (van Ooijen und Vrabec 2018; Petrlic 2019). Im Fokus dieses Aufsatzes steht insbesondere das Recht auf Auskunft nach Art. 15 DSGVO bzw. § 34 BDSG. Nach diesen Bestimmungen haben Betroffene das Recht, eine Bestätigung darüber zu erhalten, ob *pbD* verarbeitet werden. Wenn dies der Fall ist, haben Betroffene ein Recht auf Auskunft über diese *pbD* und ergänzende Informationen darüber, mit der sog. *Datenschutz-Selbstauskunft (DSA)*. Für die betroffene Person stellen sich hier u. a. folgende Fragen: Wie und in welcher Form muss eine *DSA* angefordert werden, welche Inhalte muss diese enthalten, wie oft kann eine *DSA* angefordert werden und in welcher Form hat das datenverarbeitende Unternehmen bzw. die datenverarbeitende Organisation (im Folgenden als *Datenhaltende (DH)* bezeichnet) die Datenkopien bereitzustellen.

Im Gegensatz dazu ergeben sich für die *DH* Fragen nach der eindeutigen Identität der anfragenden Person (Petrlic 2019), in welcher Form und Frist die Auskunft zu erteilen ist, ob die Anfrage begründet ist, welche Rechte Dritter beachtet werden müssen und welche Rechtsfolgen eine Unterlassung oder unvollständige Auskunft nach sich ziehen (DSK - Datenschutzkonferenz 2017). Die wesentlichen Probleme bei der Erstellung von *DSA-Anfragen*, sowohl bei den Betroffenen als auch bei den *DH*, kann in nachfolgende drei Problemklassen zusammengefasst werden:

- *Problemklasse A: Hemmnisse bei der Erstellung von DSA-Anfragen*
  Bürger:innen müssen sich zunächst einmal daran erinnern, bei welchen *DH* potenziell Daten zur eigenen Person vorhanden sein könnten. Durch eine zunehmend datengetriebene Lebenswelt wird dies für die Bürger:innen jedoch zunehmend unüberschaubarer. Wurden die relevanten *DH* identifizieren, müssen die Bürger:innen die *DSA-Anfrage* formulieren und den *DH* mitteilen (schriftlich oder mündlich). Dabei müssen sich Bürger:innen entscheiden, ob sie konkret von Ihrem Recht gem. Art. 15 oder gem. Art. 20 DSGVO Gebrauch machen möchten (Heinemann und Straub 2019). Wenngleich es dazu zwar zahlreiche Formulierungshilfen gibt, stellt dies für die Bürger:innen eine deutliche Hemmschwelle dar, da sich zusätzlicher Aufwand in Form mit der aktiven Beschäftigung mit

den persönlichen Rechten, der Erstellung der Anfrage und dem Stellen der Anfrage manifestiert (Buchmann und Eichhorn 2019).

- *Problemklasse B: Komplexer Prozess zur Bearbeitung einer DSA-Anfrage bei den DH*
  Die Bearbeitung von *DSA-Anfragen* ist für *DH* ein komplexer Sachverhalt, da die *pbD* i. d. R. dezentral in verteilten (IT-)Systemen gespeichert werden, wodurch eine isolierte Bereitstellung einzelner Datensätze nicht ohne weiteres möglich ist (Geminn 2020). Hierbei müssen die in Informationssystemen abgelegten Daten einer Person eindeutig zugeordnet und dabei auch Verbindungen zu weiteren Personen des Datums beachtet werden. Abhängig von der Natur des digitalen Mediums birgt dies unterschiedliche Herausforderungen. So kann zum Beispiel auf einem digitalen Bild eine Personengruppe abgebildet sein, welche mehreren Personen zugeordnet werden kann bzw. auch hierbei Abgrenzungen geschaffen werden müssen. Bei Daten eines sozialen Netzwerks sind die Daten, welche eine Relation zwischen zwei Personen bilden inhärent mehreren Personen zuzuordnen. Eine Anschrift oder Adresse kann ggf. mehreren Personen zugeordnet sein. Man muss hier weiterhin unterscheiden zwischen strukturiert abgelegten Daten (z. B. in einer Datenbank) und unstrukturiert abgelegten Daten (z. B. verteilt in einem oder mehreren Text-Dokumenten in mehreren Ordnern auf mehreren PCs und Servern abgelegt), wodurch sich eine umfangreiche Komplexität ergeben kann.
  Zu berücksichtigen sind zudem Daten, die zwar strukturiert, aber nicht digital vorliegen. Diese Daten sind im Vergleich zu digitalen strukturierten Daten in Informationssystemen oder Datenbanken mit einem, um einen erheblichen Faktor größeren Aufwand zu sichten, zu erheben und an den Anfragenden zu melden. Aus Sicht der Autoren stellt dies insbesondere kleine und mittelständische Unternehmen, die häufig keine dedizierten Ressourcen hierfür aufbringen können, vor eine erhebliche Herausforderung.
  Eine weitere Herausforderung im Rahmen der Bearbeitung von *DSA-Anfragen* ergibt sich in der Identitätsprüfung der anfragenden Person (Petrlic 2019; Buchmann und Eichhorn 2019), sowie in der sicheren Übermittlung der Datenkopie.
- *Problemklasse C: Schwierigkeiten bei der Interpretation der Datenkopien*
  Bürger:innen die eine Datenkopie von einem *DH* erhalten haben, müssen diese Informationen zu interpretieren wissen. Eine erste Hürde, um die Datenkopie zu interpretieren, ist das Öffnen der Datei, wobei das

Dateiformat eine essenzielle Rolle darstellt. Der Gesetzgeber sieht zwar vor, dass die Datenkopie in einem „gängigen elektronischen Format" (Art. 15 Abs. 3, S. 3 DSGVO) zur Verfügung gestellt werden muss, dies wird jedoch nicht näher spezifiziert. *DH* können also Daten in einem (branchenüblichen) Format bereitstellen (z. B. .sql, .indd). Nicht jede betroffene Person verfügt jedoch über Mittel, um diese Dateien öffnen zu können, wodurch erhebliche Probleme verursacht werden. Die Fragestellung sollte hierbei ebenfalls beachten, dass die übermittelten Daten sowohl vom Menschen als auch von der Maschine verarbeitbar sind. Hiermit möchten wir darauf hinweisen, dass auch eine Bilddatei (.jpeg) als gängiges elektronisches Format gelten kann, jedoch ein Screenshot (im .jpeg-Format) von Datenbankeinträgen sicherlich nicht ein angemessenes Format zur Übertragung dieser Daten ist, auch wenn sie vom Menschen leicht zu öffnen, lesen und interpretieren sind. Ein besseres Format zur Übertragung wäre in Form einer, durch frei zugängliche Software, zu verarbeitendes Tabellendokument.

Eine zweite Hürde für die Bürger:innen bei der Interpretation der Datenkopien stellt die Bewertung der Daten selbst dar. Die Bürger:innen müssen einschätzen, inwieweit beispielsweise die Zwecke der ursprünglichen Datenerhebung noch vorliegen, oder ob ggf. eine Löschung oder Berichtigung der Daten sinnvoll ist (Heinemann und Straub 2019). Nur dadurch können die Bürger:innen informierte Entscheidungen treffen, wie sie mit den Ergebnissen der Abfrage weiter umgehen möchten, um gegebenenfalls von weiteren ihrer persönlichen Datenschutzrechte Gebrauch zu nehmen.

Weiterhin ist zu beachten, dass durch die uneinheitliche Art der Datenkopien, auch ein Vergleich verschiedener *DH* untereinander für die Bürger:innen schwierig ist. So kann durch die unterschiedliche Bezeichnung der Datenfelder und Datenkategorien, da sie z. B. aus unterschiedlichen Informationssystemen stammen, ein Abgleich dieser nicht oder nur erschwert erfolgen. Würden die Datenfelder und Datenkategorien eine einheitliche Semantik besitzen, so könnten diese leichter verglichen werden und möglicherweise Datenflüsse zwischen verschiedenen *DH* nachvollziehbar gemacht werden.

Um diesen Problemklassen zu begegnen, stellen wir in diesem Aufsatz ein Konzept für ein *PIMS-Framework* mit zwei Hauptkomponenten vor. Mithilfe des Frameworks kann auf Betroffenenseite das Recht auf Auskunft durch Automatisierung vereinfacht werden, indem Bürger:innen dabei un-

terstützt werden, potenziell relevante *DH* zu identifizieren, bei welchen Daten zur eigenen Person potenziell vorhanden sein könnten. Bei diesen *DH* kann dann mithilfe des Frameworks eine *DSA-Anfrage*, in einem regelmäßigen Zyklus (z. B. jährlich), automatisiert elektronisch angefragt werden. Auf Seiten der *DH* kann die Anfrage mithilfe des Frameworks automatisiert entgegengenommen werden und an die Verantwortlichen weitergeleitet bzw. sogar vollautomatisiert beantwortet werden. Dies inkludiert neben der Erstellung der Datenkopie selbst, auch die Verifizierung der anfragenden Personen und die sichere/verschlüsselte Übertragung der Datenkopie. Die Datenkopie kann anschließend in standardisierter Form aufbereitet und verschlüsselt an die betroffene Person übermittelt werden. Daraufhin kann die Datenkopie bei der betroffenen Person automatisiert entschlüsselt und für den/die Bürger:in verständlich, visuell aufbereitet werden. Somit können die Hemmnisse bei der Erstellung von *DSA-Anfragen* bei den Bürger:innen und die Schwierigkeiten bei der Interpretation reduziert werden.

Mithilfe des vorgestellten Lösungsansatzes können zudem zeit- und kostenaufwändige Vorgänge zur Bearbeitung von *DSA-Anfragen* automatisiert und für beide Beteiligte (Betroffene und *DH*) effizient und auf einfache Art und Weise abgewickelt werden. Dieses übergreifende Informationssystem ermöglicht ein intuitives Datenselbstmanagement auf Betroffenenseite und stellt einen Kontrollmechanismus in Bezug zu Vollständigkeit und Rechtssicherheit auf Seiten der *DH* dar.

Zusammenfassend zielt das vorgestellte *PIMS-Framework* auf die Wahrung der informationellen Selbstbestimmung der Bürger:innen durch eine erhebliche Vereinfachung zur Anforderung einer *DSA*, sowie einer ökonomischen Bearbeitung solcher Ersuchen seitens der *DH* ab. Zudem leistet der Ansatz damit einen gewinnbringenden Beitrag zur Wahrung der Grundrechte zur Selbstbestimmung einerseits, und der Wahrung der marktwirtschaftlichen Interessen andererseits.

Der Aufsatz ist wie folgt aufgebaut: Zunächst zeigen wir in Abschn. 2 die Auswirkungen der eben genannten Problemklassen in der Praxis, indem wir eine beispielhaft durchgeführte *DSA-Anfrage* vorstellen und die Probleme bei der Bearbeitung darlegen. Anschließend gehen wir in Abschn. 3 auf die generellen Herausforderungen bei der Bearbeitung von *DSA-Anfrage* aus technischer Perspektive ein. In Abschn. 4 beleuchten wir verwandte Arbeiten und bestehende Lösungsansätze zur Umsetzung von *DSA-Anfragen*. Den von uns vorgeschlagenen Lösungsansatz zur Adressierung der genannten Problemklassen, bestehend aus einem Framework mit zwei

Hauptkomponenten, präsentieren wir in Abschn. 5 und diskutieren den Ansatz in Abschn. 6. Der Aufsatz endet mit einer Zusammenfassung und einem Ausblick in Abschn. 7.

2 Anwendungsfall aus der Praxis

Um die in Abschn. 1 aufgeführten Problemklassen, insbesondere seitens der *DH*, zu untermauern, haben wir im Rahmen dieses Aufsatzes exemplarisch eine *DSA-Anfrage* an eine Behörde gestellt. Die Anfrage erfolgte per E-Mail. Auf den Forschungshintergrund der Anfrage wurde dabei zunächst nicht hingewiesen, um eine neutrale bzw. übliche Bearbeitung der Anfrage zu garantieren.

Als erste Reaktion auf unsere *DSA-Anfrage* bat die verantwortliche Person der Behörde, die Anfrage einzuschränken, um den Aufwand bei der Bearbeitung zu reduzieren. Anschließend forderte die Behörde eine Verifikation der Identität der anfragenden Person. Dazu sollte ein Scan des Personalausweises übermittelt werden. Es wurde betont, dass nicht-wesentliche Merkmale (z. B. Bild, Personalausweisnummer, Gültigkeitsdatum) geschwärzt werden dürften. Im Forschungsfeld der Informatik, insbesondere im Fachbereich der Informationssicherheit, werden zumeist jegliche Angriffszenarien bedacht, um sie proaktiv zu mitigieren und somit das Risiko einer zukünftigen Gefährdungslage zu minimieren. Unter diesem Gesichtspunkt, ist diese Art der Identifikation als unzureichend einzustufen, da sich mit der Methodik theoretisch jeder, der einen Scan des Personalausweises besitzt oder fälschen kann, sich als eine Person identifizieren und somit eine Datenkopie anfordern könnte. Wir möchten hiermit aufzeigen, dass es sich hier um ein mögliches Problem in der Methodik handelt, jedoch davon abgrenzen, dass diese mögliche Schwäche grundsätzlich ausgenutzt wird.

Einige Tage nach der Identifizierung erfolgte die Übermittlung einer Datenkopie. Diese Datenkopie wurde passwortgeschützt/verschlüsselt per E-Mail übermittelt. Die Übermittlung des Passwortes erfolgte per SMS. Die SMS ging an diejenige Handynummer die ursprünglich beim Stellen der *DSA-Anfrage* angegeben wurde. Mit dieser Maßnahme kann sichergestellt werden, dass neben der Person, welche die *DSA-Anfrage* gestellt hat, niemand die Datenkopie einsehen kann (*Man-in-the-Middle-Angriff*). Aufgrund der unzureichenden Identitätsprüfung im Vorfeld ist jedoch nicht sichergestellt, dass es sich bei der anfragenden Person, auch um die betroffene Person handelt.

Bei einer inhaltlichen Überprüfung der Datenkopie fiel auf, dass die Daten in verschiedenen Formaten vorlagen. Teilweise wurden Screenshots einer internen Software als Bild übermittelt, teilweise CSV-Dateien. Die Semantik der Dateien, insbesondere der CSV-Dateien ist jedoch nur begrenzt interpretierbar; eine Erläuterung dazu fehlte.

Die Vollständigkeit der übermittelten Datenkopie können wir nicht bewerten. Es fällt jedoch auf, dass auch Daten über Personen enthalten sind, die nicht der anfragenden Person entsprachen (siehe Abb. 1). Es wurden also auch (personenbezogene) Daten von Dritten übermittelt.

Eine Nachfrage im Nachgang zur Anfrage bei der verantwortlichen Person der Behörde zeigte, dass die Behörde erheblichen zeitlichen Aufwand zur Bearbeitung einer entsprechenden Anfrage betreiben musste. So seien sechs Personen über mehrere Tage in die Bearbeitung der *DSA-Anfrage* involviert gewesen.

Die exemplarische Anfrage zeigt bereits deutlich die Existenz der Problemklassen in der Praxis (insb. Problemklasse B und C).

Wohingegen viele große Technologie-Konzerne wie Facebook, Google oder Amazon automatisierte Methoden entwickelt haben, um *DSA-Anfragen* zu bearbeiten, haben andere *DH* – insb. KMUs oder Behörden – regelmäßig Probleme mit der Bearbeitung solcher Anfragen. Somit ergibt sich die Notwendigkeit und der Bedarf, die Prozesse zur Stellung und Beantwortung von *DSA-Anfragen* zu unterstützen.

In dieser vorgestellten Arbeit wird die Unterstützung mit Hilfe von bewährten Technologien in einem *PIMS-Framework* vorgeschlagen, wobei aber auch organisatorische Mittel zur Verbesserung der Prozesse gewählt werden können. Aus Sicht der Autoren bieten aber insbesondere technologische Ansätze erhebliche Vorteile für die Automatisierung und damit Reduktion des Aufwands in allen Prozessschritten.

previous_hospitalizations

previous	previous	previoushospitalization	previoushospitalization	previoushos	previoushospital	previoushospitalization	previoushospitalization	previoushospitalization
id	uuid	changedate	isolated	hospitalization_id	description	hospitalizationreason	otherhospitalizationreason
2533	TEZQT6-2SXPTV-WHDFB5-	2021-01-04 12:51:25.437		2464			
8068	XJWIBJ-KLNOHD-LENNAK	2021-01-19 09:19:54.009	YES	7999	[redacted] liegt auf der Station 1 in Isolation		
9308	WYPKKX-QPTVQ4-XGYZPS	2021-01-20 14:03:01.829	YES	5356			
5030	TFUMHK-KZSSA5-BLSNPZ	2021-01-20 14:22:36.827		3511	hat Lungenentzündung, keine Beatmung, nur Sauerstoff, Arzt sagt sie kann bald nach Hause		
10820	VQFAHQ-HGGOKK-UUVKFS	2021-01-22 10:53:18.866	UNKNOWN	10757			
12007	UMVB7A-O2VFNH-2XKEM7	2021-01-24 08:54:46.428		11731			
12269	UFPJ3X-Y2ET5O-4BKJLW-I	2021-01-24 13:09:56.421	UNKNOWN	12237	befindet sich in der Neurologie		
13154	TAHKJG-APNAK-G5GUX2-I	2021-01-26 10:57:30.156	YES	12775	[redacted] 26.01.2021 - Patient im Zimmer auf Station 5 - nur eine Untersuchung - Bauchsonographie - ansonsten nur Essensvergabe mit Vollschutz		
14452	WD3DTH-WRJQ6U-6HKRVA	2021-01-27 11:20:24.324	YES	12796	Wegen Atemluftbeschwerden eingeliefert		
15727	SYRAXR-CSJYJO-23DIDO-	2021-01-28 15:28:35.98	YES	8926	Wird Beatmet		
27482	QRVVF4-67BQZG-XRYCCA	2021-02-14 10:38:56.986	YES	27348	Atemnot, Fieber		
27806	UOJITX-CZ6IMF-DBUEVF-K	2021-02-15 08:51:15.288	YES	27770	Index liegt im KH [redacted]		
35220	VY3PTZ-LVVCY7-YKKEHE-J	2021-02-23 09:01:42.178	YES	21483			
35311	QTDNMV-PPH6IU-SJPOCB	2021-02-23 09:45:28.31	YES	24476			
38884	XNWB4X-CGK4?R-JQ4TXZ	2021-02-26 10:19:10.179		38687	Wurde lt. Info des Patienten gleich nach der Testung im KH [redacted] ins KH [redacted] verlegt.		
69018	QWADZB-ONKQQB-UK5DB	2021-03-23 07:36:08.84		53598			
69024	TELJ5X-WVKFV2-5PCYID-(	2021-03-23 07:39:00.208	YES	64937		OTHER	Gefäßstörung
70449	UWJKCY-5BRMPS-VWZXTA	2021-03-23 14:55:51.451	NO	70388		OTHER	Bösartiger Tumor an der Blase
71571	XJCKRJ-MVAOOU-4MPPUI	2021-03-24 13:35:53.389	YES	70368		OTHER	OP an der Prostata
77806	RX2HFJ-AHH5R7-4NDEVC-	2021-03-29 08:59:49.82	YES	67270		REPORTED_DISEASE	
116988	UUE5SM-FTMUY4-EYW2BI	2021-04-17 11:03:58.487	YES	102416		OTHER	
117104	WYTAFN-H764JQ-27CNJB-	2021-04-17 11:46:40.039	YES	108744		OTHER	Wasser im Körper,
117162	QAIQNF-7TALXU-MXGLT4-(	2021-04-17 12:06:00.098		116203	Patientin war vom [redacted] 2021 im Krankenhaus [redacted] stationär auf der Inneren Station	OTHER	rapide AZ verschlechterung
116586	7ASG21-W6AO4B-PFANZH-	2021-04-18 10:40:19.397	YES	100677		OTHER	Verdacht auf Bandscheibenvorfall
128144	VFITUR-EEU8NO-YITY4K-D	2021-04-22 11:42:14.776	UNKNOWN	125830		OTHER	Schwäche, Appetitlosigkeit
131764	UM6H1MC-WOVG5I-2NHPK(	2021-04-24 09:17:03.096	YES	81867		REPORTED_DISEASE	
132097	TSVZI5-DETPYQ-LYCJP7-L	2021-04-24 10:29:37.546	YES	65842		OTHER	Offene Füße
132522	VBXCZR-VZCS47-V2GFJE-(	2021-04-24 12:28:40.49	YES	102448		REPORTED_DISEASE	
144941	W5RWXO-USU45C-HDYPK(	2021-05-01 14:21:31.384		144632			

Abb. 1. Auszug aus einer übermittelten Datei der Datenkopie mit unklarer Semantik sowie Daten von Personen, welche nicht der anfragenden Person entsprechen. Schwärzungen durch die Autoren.

3. Herausforderungen in der praktischen IT-Umsetzung

Globalisierung, Big-Data-Ansätze und Cloud-Technologien verändern das Konzept der klassischen Datenverarbeitung in einzelnen voneinander getrennten Systemen. Der historische Ansatz einzelner Akteure, die dezentral pbD beinhalten, z. B. eine Arztpraxis, ist längst zu einem multinational verknüpften, zentralem System verschmolzen, in dem Daten wie eine Ware gehandelt, zwischen erhebenden Stellen verknüpft und darauf aufbauend analysiert werden können. Um beim Beispiel der Arztpraxis zu bleiben, wäre die Zusammenführung der dezentral vorgehaltenen Daten in eine zentrale Patientendatenbank einer bundesweit agierenden Krankenversicherung technologisch denkbar. Um die Bürger:innen vor dieser unüberblickbaren Datensammlung zu schützen und ihnen Handhabe zur Verwirklichung ihres Rechts auf informationelle Selbstbestimmung zu geben, instanziierte der Gesetzgeber Betroffenenrechte, die jedem Individuum zustehen (Hintze and El Emam 2018). Diese Individualrechte beeinflussen, welche Anforderungen an Software gestellt werden, die diese gespeicherten Daten verarbeitet.

Anhand einfacher Prozesse kann gezeigt werden, welche Problemstellungen sich in der Informatik durch die Anforderungen des Gesetzgebers entwickeln. Ein erstes Beispiel ist die Sicherung von Daten zum Schutz vor Informationsverlust bei technischen Störungen. Beruft sich ein Individuum auf sein Betroffenenrecht zur vollständigen Löschung seiner Daten, kann dies zwar im Operativsystem zeitnah erfolgen, die Datensicherung müsste jedoch ebenfalls aktualisiert werden, da sonst eine Wiederherstellung der Daten gleich gesetzt werden kann mit einer Revidierung des Löschvorgangs. Dies stört sich mit dem Ziel, eine Datensicherung so sicher wie möglich abgespalten vom Operativsystem zu betreiben. Gleichzeitig werden Sicherungen zur Speichereffizienz oft komprimiert und inkrementell aufgebaut. Da Datenpunkte somit voneinander abhängen, müssen spezielle Prozesse eingesetzt werden, um Verkettungen nicht zu zerstören, was einen allgemeineren Datenverlust zur Folge hätte.

Um die Herausgabe, Korrektur oder sonstige Verwendung von pbD eines Individuums zu autorisieren, muss sich dieses als betroffene Person zuletzt genannter pbD ausweisen. In einem Fernabwicklungsverfahren, wie dem hier beschriebenen, stellt dies den/die Sachbearbeiter:in vor ein Problem, da die betroffene Person und dessen Ausweismedium nicht direkt in Person validiert werden kann, ohne erheblichen Aufwand zu verursachen. Ein üblicherweise genutztes Verfahren ist das Übermitteln einer Personalaus-

weis-Kopie (siehe Abschn. 2). Dieses Verfahren ist jedoch hoch problematisch, da die besonders zu schützenden Daten des Personalausweises dann unkontrolliert in Umlauf gebracht werden und der Empfänger zudem nicht verifizieren kann, dass die versendende Person nicht auf sonstigem Wege an die Kopie gelangt ist. Ein vom Gesetzgeber vorgeschlagenes Verfahren stellt der digitale Personalausweis dar.

Datenverbindungen bzw. deren zugrundeliegenden Protokolle sind offene Transportkanäle, wie zum Beispiel TCP oder UDP, die von allen übermittelnden Zwischenstellen im Internet mitgelesen, also abgehört und manipuliert werden können. Um die Integrität einer übermittelten Nachricht sicherzustellen, wird auf Transportverschlüsselung gesetzt, d. h. die Information wird vom Versender mit einer vorher von beiden Seiten vereinbarten Chiffre kodiert und dann vom Empfänger mit diesem dekodiert. Nutzende kennen dies vor allem aus Anwendungsbereichen wie Online-Banking oder VPN-Verbindungen. Eine Veränderung der Nachricht von Zwischenstellen würde die Chiffrierung brechen. Der Empfänger wüsste, dass der Information nicht mehr zu vertrauen ist. Dieses Verfahren nennt sich auch symmetrische Verschlüsselung, da beide Seiten dieselbe Chiffre verwenden. Um die Chiffre zwischen zwei sich nicht vorher kennenden Parteien auszutauschen, wird typischerweise asymmetrische Verschlüsselung zum Einsatz kommen. Dabei erzeugt eine Seite ein Schlüsselpaar, das besondere Eigenschaften aufweist. Ein Schlüssel, der sog. *Private Key*, darf nur dem Aussteller bekannt sein. Der andere Schlüssel, der sog. *Public Key*, kann vom Aussteller herausgegeben werden. Ein mit dem *Public Key* verschlüsseltes Objekt ist nur noch durch den *Private Key* wiederherstellbar (Simmons 1979). Um die Vertrauenswürdigkeit des Ausstellers zu verifizieren, setzt man zumeist voraus, dass eine der beiden Seiten sich von einer allgemein anerkannten Stelle zertifizieren lässt. Im Internet übernehmen diese Aufgabe Zertifizierungsstellen (Aas et al. 2019). Um nicht auf private Zertifizierungsstellen angewiesen zu sein, gäbe es auch die Möglichkeit, sich mittels des digitalen Personalausweises zu autorisieren und somit indirekt die Zertifizierung durch den Bund zu verwenden. Der *Public Key* kann dann an den *DH* gesendet werden. Der *DH* kann die herauszugebenden Informationen mit dem *Public Key* verschlüsseln und theoretisch sogar über ungeschützte Transportkanäle versenden, da nur noch der Anfragende die Informationen entschlüsseln kann. Hierdurch ist sichergestellt, dass nach erfolgter Verschlüsselung nur noch die anfragende Person selbst ihre zur Verfügung gestellten Daten lesen kann und keine etwaige Zwischenstelle

(vgl. *Man-in-the-Middle* Angriff Szenarien (Callegati, Cerroni, and Ramilli 2009)) einen Nutzen daraus ziehen kann.

Um die Sicherheit der Authentifizierung eines im Internet veröffentlichten Systems zu verstärken, wird zunehmend auf einen zweiten Faktor als Erweiterung zum herkömmlichen Passwortverfahren gesetzt (*Zwei-Faktor-Authentifizierung* bzw. *Multi-Faktor-Authentifizierung* (Dasgupta, Roy, and Nag 2017)). Dabei muss der/die Nutzer:in im Besitz eines zuvor registrierten zweiten Geräts (z. B. Smartphone), einer Empfangsmöglichkeit (z. B. Telefon) oder eines Dateischlüssels (z. B. USB-Stick) sein, um sich nach Eingabe des Passworts zu autorisieren. Da dieser zweite Faktor vorab vom Nutzenden registriert wird, ist ausgeschlossen, dass ein Angreifer aus der Ferne Zugang erhält, wenn er, entweder durch Ausprobieren von Kombinationen oder eines sonstigen Abhandenkommens, etwa durch Phishing, in Besitz des Passworts gelangt.

Auch der schon erwähnte digitale Personalausweis verwendet die oben beschriebenen Verfahren (Bundesamt für Sicherheit in der Informationstechnik 2018). Der Personalausweis selbst ist mit einem Sicherheitschip versehen, der mittels *Near Field Communication (NFC)* abgefragt werden kann. Da diese Technologie in vielen heute gängigen Smartphones integriert ist, können Nutzer:innen mit ihrem Ausweis interagieren, ohne zusätzliche Geräte beschaffen zu müssen. Der Chip ist, ähnlich dem einer Bankkarte, mit einem, hier sechs-stelligen, PIN geschützt. Nur Personen, die Kenntnis vom persönlich zu setzenden PIN haben, können die auf dem Chip hinterlegten Informationen abrufen. Dadurch ergibt sich ein Zwei-Faktor-Autorisierungssystem, da zusätzlich zum physischen Merkmal – der Karte – auch das Wissensmerkmal – die PIN – bereitgestellt werden muss. Mit beiden in Kombination kann eine Software, die mittels NFC mit dem Personalausweis kommuniziert, sich gegenüber eines vom Bund betriebenen eID-Servers ausweisen. Der eID-Server bestätigt einer anfragenden Partei daraufhin die Validität der Person. Man spricht auch von einer vertrauten Drittpartei, einer *Trusted Third Party (T3P)*, in diesem Fall dem Betreiber des eID-Servers, gleichgestellt mit der Bundesdruckerei, deren Erzeugnisse sonst anhand des Drucks einzigartiger Schutzmerkmale vertrauenswürdig erscheinen (vgl. Verordnung (EU) 2019/1157 des Europäischen Parlaments und des Rates vom 20. Juni 2019). Weitere Daten wie Name, Vorname, Adresse usw. können dann vom Personalausweis über den eID-Server an die Drittpartei übermittelt werden. Die Übermittlung der Information durch den eID-Server ist mittels Transportverschlüsselung geschützt. Der eID-Server weist sich durch das Zertifikat einer anerkannten

Zertifizierungsstelle aus. Zur Verwendung des eID-Servers muss sich eine Drittpartei zuvor registrieren lassen. Durch eine Überprüfung wird sichergestellt, dass Daten nicht an Drittparteien abfließen, die den Bürger:innen unter dem Vorspielen falscher Tatsachen zur Herausgabe der Informationen auf ihrem digitalen Ausweis drängen. Die von den Nutzer:innen des Systems verwendete Applikation kann vor Autorisierung am eID-Server darstellen, wer die Informationen mit welchem Detailgrad zugesandt bekommt.

4. Verwandte Arbeiten

Der Schutz der Privatsphäre sowie das Recht auf informationelle Selbstbestimmung durch die Regelungen der DSGVO wird in der wissenschaftlichen Literatur mehrfach diskutiert. Die Beiträge beschäftigen sich mit einem Vergleich von Datenschutzerklärungen vor und nach dem Inkrafttreten der DSGVO (Zaeem and Barber 2021), mit der Gültigkeit von Einwilligungen zu Verarbeitung von *pbD* (Sinclair and Jamal 2021) oder mit dem Datenschutz-Paradoxon (Dienlin, Masur, and Trepte 2021; Barth and de Jong 2017). Andere Arbeiten beschreiben die Vorteile und Nachteile der Einhaltung der DSGVO für die *DH*, (Këllezi 2021; Kröger, Lutz, and Ullrich 2021) oder beschäftigen sich mit der Frage, ob die DSGVO die Kontrolle der Verbraucher:innen über *pbD* aus einer Verhaltensperspektive verbessert (van Ooijen and Vrabec 2018).

Die Literatur-Recherche konnte nur wenige Beiträge identifizieren, die sich vorrangig mit der Wahrung der Betroffenenrechte durch die Einforderung einer *DSA* nach Art. 15 DSGVO befassen. Diese Bestimmung gestattet Einzelpersonen die Kontrolle über ihre *pbD* im Rahmen ihres Auskunftsrechts, zur Offenlegung aller gespeicherten *pbD* und deren Verarbeitung (di Martino et al. 2022). Buchmann und Eichhorn (2019) vertreten die Meinung, dass gerade der Art. 15 DSGVO von zentraler, datenschutzrechtlicher Bedeutung für Kund:innen von Online-Unternehmen istKlicken oder tippen Sie hier, um Text einzugeben.. Dabei können jedoch Probleme auftreten.

Nach Geminn (2020) hängt das Recht auf Auskunft, bezogen auf seine Zielerreichung wesentlich von zwei Faktoren ab: (1) Die betroffene Person muss wissen, an wen sie ein Auskunftsersuchen richten kann, und (2) die ihr gegenüber bereitgestellten Informationen müssen für sie verständlich und nützlich seinKlicken oder tippen Sie hier, um Text einzugeben.. Heine-

mann und Straub (2019) argumentieren, dass sich die Betroffenen daran erinnern müssen, welche *DH* ihre Daten (unter welchem Erlaubnistatbestand) verarbeiten. Schließlich müssen sie entscheiden, ob sie konkret von ihrem Recht gem. Art. 15 oder gem. Art. 20 DSGVO Gebrauch machenKlicken oder tippen Sie hier, um Text einzugeben.. Wie die Beiträge zeigen, ergeben sich bereits auf der Seite der Bürger:innen Probleme bei der Umsetzung des Art. 15 DSGVO. Aber auch auf der Seite der *DH* sind Schwierigkeiten in der Umsetzung dieser Vorschrift beobachtbar (Buchmann and Eichhorn 2019).

Für die *DH* stellt ein Auskunftsersuchen einen erheblichen wirtschaftlichen Aufwand dar. Buchmann und Eichhorn (2019) vermuten, dass *DH* durch komplizierte Prozesse oder aufwändige Identitätsprüfungen deshalb versuchen, ihre Kund:innen von Auskunftsersuchen abzubringen. Speziell die Form der Darstellung der gespeicherten *pbD* scheint, nach einigen Autor:innen, den *DH* Schwierigkeiten zu bereiten.

Erhebliche Unsicherheiten für die Bürger:innen bestehen, insbesondere bezogen auf die Reichweite des Rechts auf Erhalt einer Kopie aus Art. 15 Abs. 3 DSGVO. Umstritten sind sowohl die Stellung des Rechts auf Erhalt einer Kopie als auch Inhalte und Reichweite, nicht zuletzt, weil die Verordnung selbst zu all diesen Aspekten schweigt – von der Möglichkeit der Erhebung eines angemessenen Entgelts für weitere Kopien und der Pflicht zur Bereitstellung in einem gängigen elektronischen Format bei elektronischer Antragstellung abgesehen (Geminn 2020). In einer Studie von Bowyer u.a. (2022) mit zehn Teilnehmenden, in der jede Person vier bis fünf *DSA-Anfragen* stellte, wurde beobachtet, dass die erhaltenen Daten in frustrierenden Formaten, darunter Screenshots, Ausdrucke oder Dateien, die mit Akronymen übersät waren, an Betroffene übermittelt wurden. Die Daten waren zu technisch, um sie zu verstehen und die Informationen waren nicht verwendbar. Des Weiteren kamen die Autor:innen zu dem Ergebnis, dass die Qualität der erhaltenen Informationen unvollständig, ungenau, unbrauchbar und als nicht nützlich von den Studienteilnehmenden beurteilt wurdenKlicken oder tippen Sie hier, um Text einzugeben..

Zu ähnlichen Ergebnissen kommen Kroeger, Lutz und Ullrich (2021). in ihrer Studie, in der sie *DSA-Anfragen* an Anbieter von 225 beliebten mobilen Apps versandten. Die von den Anbietern übermittelten Informationen enthielten Formatierungsfehler, in einigen Fällen bestanden die Daten sogar aus einem kontinuierlichen Block alphanumerischer Zeichen ohne Überschriften, Leerzeichen und Zeilenumbrüche und waren damit unbrauchbar. In den meisten Fällen wurden *pbD* in Anhängen in ver-

schiedenen Dateiformaten (nämlich .pdf, .html, .json, .csv, .jpeg, .png, .docx und .txt) und als Klartext im E-Mail-Text bereitgestellt. Buchmann und Eichhorn (2019) stellten in ihrer Studie fest, dass von insgesamt 14 angefragten *DH*, nur sieben vollständige Auskünfte erteilten. Die Autor:innen berichten zudem von wenig datenschutzfreundlichen Identitätsprüfungen. Des Weiteren stellen sie fest, dass auf Nachfragen bei unvollständigen Informationen von den *DH* keine Rückmeldungen mehr erfolgten

Die dargestellten Probleme im Zusammenhang mit der Anforderung einer *DSA* durch die betroffene Person einerseits, sowie die Bearbeitung durch die *DH* andererseits, verlangen nach Lösungen. Bowyer u.a. (2022) schlagen diesbezüglich vor, den Betroffenen Zusammenfassungen über die gespeicherten *pbD* zur Verfügung zu stellen, damit diese einen Überblick über vorhandene Daten bekommenKlicken oder tippen Sie hier, um Text einzugeben.. Nach Geminn (2020) fehlen standardisierte Formatvorgaben seitens des Gesetzgebers und eine speziell für die Betroffenen erstellte digitale Akte", in der alle gespeicherten Informationen ersichtlich" sindKlicken oder tippen Sie hier, um Text einzugeben.. Dashboards könnten nach Heinemann und Straub (2019) auch ein Mittel sein, um den Betroffenen alle relevanten Informationen übersichtlich und gebündelt zu präsentierenKlicken oder tippen Sie hier, um Text einzugeben.. Buchmann und Eichhorn (2019) verweisen auf die Internetseite *selbstauskunft.net*, auf der *DSA* auch für Laien auf einfache Art und Weise angefordert werden könnenKlicken oder tippen Sie hier, um Text einzugeben..

Zusammenfassend ist festzustellen, dass speziell die Ausübung des Betroffenenrechts nach Art. 15 DSGVO in der wissenschaftlichen Literatur noch weitestgehend unerforscht ist, und deshalb Handlungsbedarf besteht.

5. Lösungsansatz MoP und TaP

Um Bürger:innen die Möglichkeit zu geben, eine automatisierte *DSA-Anfrage* an Unternehmen, Behörden oder sonstige *DH* zu stellen, schlagen wir ein Framework mit zwei Hauptkomponenten vor. Auf der Seite der Betroffenen ist es erforderlich, Unterstützung bei der Erstellung und Übermittlung von *DSA-Anfragen* zu bieten, hierfür schlagen wir das Modul *Monitoring Tool for Personal Data (MoP)* für Betroffene vor (siehe Abschn. 5.1). Um die Anfragen auf Seite des *DH* aufzunehmen und (teil-)automatisiert zu verarbeiten, schlagen wir das Modul für die Datenverantwortlichen *Tool for automated Data Self-Disclosure Request Processing (TaP)* vor (siehe

Abschn. 5.2). Durch die Zusammenarbeit beider Komponenten können die beschriebenen Problemklassen systematisch und holistisch gelöst werden. So unterstützt das *MoP* dabei die Hemmnisse bei der Erstellung von DSA-Anfragen bei Bürger:innen abzubauen (Problemklasse A), da eine zentrale Schnittstelle für die Erstellung ebendieser sowie für die Rückübermittlung der pbD geschaffen wird. Das *TaP* fokussiert sich hierbei insbesondere auf die Lösung der Problemklasse B, wobei die Beantwortung der *DSA-Anfragen* unterstützt wird. Um einen Lösungsansatz für die Problemklasse C zu bieten, sind beide Komponenten *MoP* und *TaP* notwendig, bzw. die Schaffung der Schnittstellen dieser. So kann die Interpretation von Datenkopien für die Bürger:innen durch technische Systeme am besten unterstützt werden, falls einheitliche Austauschformate im „gängigen elektronischen Format" (Art. 15 Abs. 3 S.3 DSGVO) genutzt werden. Diese müssen von *MoP* bereitgestellt werden und von *TaP* verarbeitet werden. Weiterhin kann basierend auf einheitlichen Formaten mit *TaP* eine Aufbereitung und Visualisierung vorgenommen werden, damit die Bürger:innen die übermittelte Antwort auf ihre *DSA-Anfrage* transparent und verständlich präsentiert bekommen. Dadurch können nur durch das Zusammenspiel, der im Folgenden weiter präzisierten Komponenten des vorgeschlagenen Frameworks, die Problemklassen adressiert werden.

Schematisch ist das Gesamtsystem in Abb. 2 dargestellt.

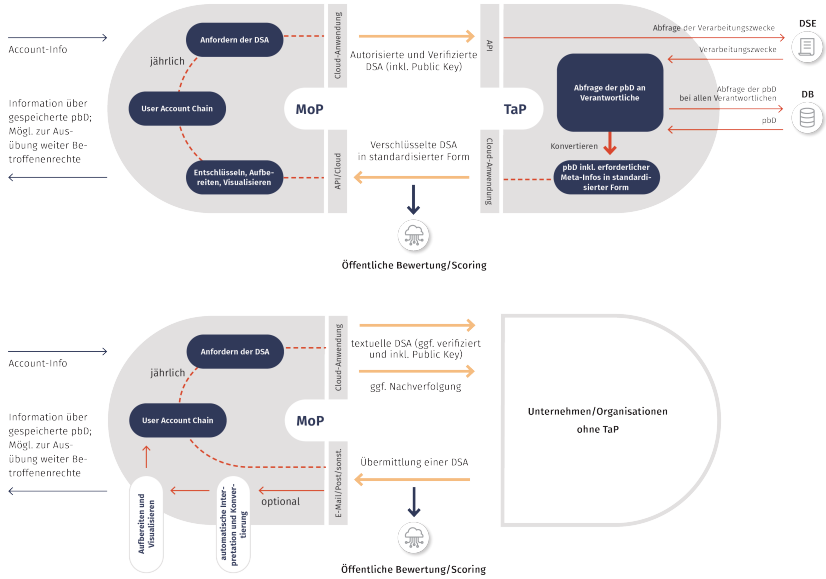

Abb. 2. Schematische Darstellung des PIMS bestehend aus der Bürger:innen-Einheit Monitoring Tool for Personal Data (MoP) und dem Komplementärsystem für DH: Tool for automated Data Self-Disclosure Request Processing (TaP).

5.1 Monitoring Tool for Personal Data (MoP)

Mithilfe des sog. *Monitoring Tool for Personal Data (MoP)* werden Bürger:innen dabei unterstützt, automatisiert *DSA-Anfragen* zu erstellen und diese an die *DH* zu übermitteln. Ferner werden die Bürger:innen unterstützt, die erhaltenen Rückantworten zu interpretieren.

Für die Bürger:innen ist die zentrale Schnittstelle mit *MoP* eine sog. *User Account Chain*. Diese *User Account Chain* enthält, angelehnt an einen digitalen Schlüsselbund, Informationen darüber, bei welchen *DH* potenziell Daten zur Person vorhanden sein könnten. Der Aufbau dieser *User Account Chain* ist zentraler Baustein für die Funktionalität. Es muss hierbei ein geeignetes Datenschema entwickelt werden, welches definiert, welche Informationen gespeichert werden. Unter anderem sollten hierbei in Anlehnung an die DSGVO, insbesondere die benötigten transparenten Informationen für die Informationspflicht – Datengruppen, Zwecke, Löschfristen, etc. –

(vgl. Art. 12 - 14 DSGVO) und ggf. weitere notwendige Daten und Informationen zur Inanspruchnahme der Betroffenenrechte, insbesondere für die *Datenschutz-Selbstauskunft (DSA)* in diesem Kontext. Als Grundlage für eine derartige *User Account Chain* können *Domain Specific Languages*, insbesondere *Privacy Languages*, dienen. Beispiele hierfür sind die *Layered Privacy Language* mit Framework, welches sich darauf fokussiert, Datenschutzerklärungen für Maschinen und Menschen lesbar strukturiert abzubilden und mit Hilfe des Frameworks die Möglichkeit bietet, automatisiert Techniken zur Anonymisierung und Pseudonymisierung auf die Rohdaten zweckgebunden anzuwenden, oder das *SPECIAL* Projekt, welches ebenfalls einen Vorschlag für die Abbildung von Datenschutzerklärungen erstellt hat und Forschung zur Visualisierung dieser durchgeführt hat. Weiterhin bietet das im *World Wide Web Consortium (W3C)* vorgeschlagene *Data Privacy Vocabulary (DPV)* Ansatzpunkte zur semantischen Vereinheitlichung der Terminologie im Datenschutzbereich. Somit können *Privacy Languages* als Basis für ein Austauschformat dienen und semantische Standards wie das *Data Privacy Vocabulary (DPV)* verwendet werden, um die Inhalte mehrheitlich einheitlich zu standardisieren. Auf Basis dieser Technologien kann die *User Account Chain* aufgebaut werden, um möglichst viele Informationen zu speichern, welche *DH* potenziell Daten zu einer Person besitzen könnten; u. a. auch Daten von *DH*, bei denen die Person keinen direkten User Account besitzt (z. B. Akte in der Arztpraxis; Kund:innenkartei in der Kfz-Werkstatt). Die Informationen zu den potenziellen *DH* aus der *User Account Chain* kann als Grundlage dienen, um periodisch (i. d. R. jährlich) eine *DSA-Anfrage* an den *DH* zu stellen. Dabei können folgende zwei Fälle betrachtet werden:

- *DSA-Anfrage an DH, die TaP verwenden:* Die *DH* erhalten eine voll-elektronische und standardisierte *DSA-Anfrage* über die in *TaP* dafür vorgesehene API. Die anfragende Person wird durch *MoP* direkt verifiziert (z. B. mithilfe des digitalen Personalausweises). Ferner wird durch *MoP* ein *Public-Key* der *DSA-Anfrage* hinzugefügt, welcher später zur sicheren Übertragung der *pbD* dient.
- *DSA-Anfrage an DH, die TaP nicht verwenden:* Die *DH* erhalten über *MoP* automatisiert eine textuelle *DSA* per E-Mail. *MoP* überwacht anschließend auf Seiten der Bürger:innen den Bearbeitungsstand der *DSA-Anfrage*. Sollten *DH* innerhalb der (gesetzlichen) Frist die *DSA-Anfrage* nicht beantworten, so wird dies durch *MoP* moniert.

Unabhängig von der Art der *DSA-Anfrage* werden die Rückantworten der *DH* anschließend wieder in *MoP* gesammelt. Bei *DH*, die *TaP* verwenden, werden die Daten durch die *DH* direkt in das *MoP* System mittels einer Schnittstelle übertragen. Bei *DH* ohne *TaP* kann weiterhin eine manuelle Import-Funktion für die Bürger:innen bereitgestellt werden, bei denen die relevanten Informationen mittels einer Benutzeroberfläche eingegeben werden können und damit manuell übertragen werden. Die automatisierte Übertragung mittels einer Schnittstelle ist hier jedoch zu präferieren, da sie weniger fehleranfällig ist und auch keinen zusätzlichen Aufwand verursacht. Die *pbD* werden anschließend durch *MoP* intelligent aufbereitet, visualisiert und in der *User Account Chain* abgespeichert. Bei der Visualisierung sind hierbei unterschiedliche Ansichten für den Nutzenden denkbar, z. B. können die Informationen nach unterschiedlichen Prioritäten dargestellt werden, wie eine Ansicht, welche basierend auf den Verarbeitungszwecken die Daten gruppiert, während eine andere Ansicht basierend auf den Datengruppen gruppiert. Weitere Filter- und Sortierfunktionen sind durch eine derartige elektronische Aufbereitung und Visualisierung ebenfalls realisierbar, um Mehrwerte zu generieren.

Das Konzept könnte weiterhin mit den persönlichen Datenschutz-Präferenzen der Nutzenden erweitert werden, bei denen diese Präferenzen in einem strukturierten Format persistent gespeichert werden und mit den im *MoP* vorhandenen Daten verglichen werden. Dadurch könnten die Nutzenden, mit Hilfe einer geeigneten Darstellung, eigene datenschutzbezogene Verhalten, anhand zuvor definierter Präferenzen, reflektieren und ggf. anpassen.

5.2 Tool for automated Data Self-Disclosure Request Processing (TaP)

Mithilfe des sog. *Tool for automated Data Self-Disclosure Request Processing (TaP)* werden *DH* dabei unterstützt, voll-/teil-automatisiert *DSA-Anfragen* systematisch zu bearbeiten. *TaP* ist als Komplementärsystem zu *MoP* zu betrachten und setzt voraus, dass die *DSA-Anfragen* auch über ein *MoP* bzw. einer Schnittstelle gestellt werden.

Eine Schlüsselinnovation von *TaP* ist, dass die durch *MoP* erzeugten *DSA-Anfragen* direkt über eine API entgegengenommen werden. Durch eine Standardisierung des Anfrageformats und einer direkten, bereits in der Anfrage integrierten Autorisierung der Anfrage, ermöglicht *TaP* eine voll-automatisierte Bearbeitung der *DSA-Anfrage*.

Zur Umsetzung der Autorisierung sollte hierbei auf bestehende und bewährte Standards zurückgegriffen werden; auch die Nutzung des elektronischen Personalausweises zur Authentifizierung wäre als geeignetes Verfahren möglich.

Zur inhaltlichen Bearbeitung der *DSA-Anfrage* und zur Erstellung der Datenkopie sind mehrere Prozessschritte notwendig. Zunächst identifiziert *TaP* aus *elektronischen Datenschutzerklärungen (DSE)* alle für die Bearbeitung der *DSA-Anfrage* relevanten Verarbeitungszwecke. Anschließend werden die *pbD* aus allen Verarbeitungszwecken abgefragt. Dies kann entweder voll-automatisiert durchgeführt werden, z. B. durch eine Abfrage aus bestehenden Datenbanken, oder manuell durch eine Abfrage bei den jeweiligen Verfahrensverantwortlichen. Eine Herausforderung bei der voll-automatisierten Abfrage aus Datenquellen besteht darin, dass eine Verknüpfung bzw. ein Mapping zwischen den abgefragten Daten und Datengruppen zu den einzelnen Daten (in der Terminologie von Datenbanken: *Attributen*) der Datenquellen hergestellt werden muss. Dadurch ergibt sich ebenfalls eine Klassifizierung der Daten in den Datenquellen in persönliche Daten und nicht persönliche Daten, wobei man aus Sicht der Informatik persönliche Daten noch weiter differenzieren würde in (Sweeney 2002; Venkataramanan and Shriram 2016):

- *Explizite Identifikatoren (EI)* definiert Attribute, die Bürger:innen eindeutig identifizieren. Beispiele für *EI* sind die Reisepass-ID, der Name oder die Sozialversicherungsnummer, wobei der Name auch für die folgende Datenkategorie zugeordnet werden könnte, da Namen bei großen Datensammlungen möglicherweise nicht eindeutig sind und zusätzliche Attribute zur eindeutigen Identifizierung eines/einer Benutzer:in erfordern.
- *Quasi-Identifikatoren (QI)* definiert Attribute, die in Kombination mit anderen *QI* die Identifizierung eines Benutzers ermöglichen. Beispiele für *QI* sind IP-Adresse, Postleitzahl, Geburtstag, Alter, Geschlecht und andere demografische Informationen. *QI* sind oft öffentlich zugänglich, zum Beispiel in Telefonbüchern, Wählerdatenbanken oder anderen Quellen.
- *Sensitive Data (SD)* definiert Attribute, die für Benutzer:innen vertraulich sind. Beispiele für *SD*-Attribute sind Gesundheitsdaten, Finanzdaten oder andere Informationen, die je nach Zweck nicht mit dem/der Nutzer:in in Verbindung gebracht werden sollten.

- *Non-Sensitive Data (NSD)* definiert Attribute, die weder Benutzer:innen identifizieren noch für Benutzer:innen sensibel sind. Daher sind *NSD*-Attribute alle Attribute, die nicht einer der anderen Datenkategorien *EI*, *QI* oder *SD* zugeordnet werden können und entsprechen nicht persönlichen Daten.

Die Zuordnung der Attribute der Datenquellen zu den Datengruppen, kann in den Datenquellen „*by Design*" durchgeführt werden, oder bei bestehenden Systemen durch den Einsatz von zusätzlicher Middleware umgesetzt werden.

Informationen aus den *DSE*, wie Zweck und Rechtsgrundlage sowie Löschfristen, werden gemeinsam mit den *pbD* aus den unterschiedlichen Verarbeitungszwecken in ein standardisiertes Schema konvertiert, zusammengefasst und mit relevanten Zusatzinformationen, bspw. Hinweisen auf das Beschwerderecht, ergänzt. Als Grundlage für das standardisierte Schema können *Privacy Languages* genutzt werden. Abschließend stellt *TaP* der anfragenden Person über eine Cloud-Schnittstelle von *MoP* verschlüsselt die Daten aus der *DSA* im vereinheitlichten Austauschformat zur Verfügung. Die Bearbeitung der *DSA-Anfrage* ist damit für den *DH* vollständig abgeschlossen.

6. Diskussion

Der vorgeschlagene Lösungsansatz mit *Monitoring Tool for Personal Data (MoP)* und *Tool for automated Data Self-Disclosure Request Processing (TaP),* als zwei Komponenten eines *PIMS-Frameworks*, ist ein technischer Ansatz, um die beschriebenen Problemklassen zu bewältigen.

Um die Problemklasse A, die Hemmnisse der Bürger:innen bei der Erstellung von *DSA-Anfragen*, entgegenzuwirken wird insbesondere das *MoP* eingesetzt, das Informationen über mögliche *DH* mit vorliegenden *pbD* verwaltet und es vereinfacht, Anfragen zu stellen. Das vorgestellte *MoP* fokussiert sich hierbei nur auf einen Aspekt der Problemklasse, dem Stellen von *DH-Anfragen*, jedoch können diese Hemmnisse auch mit weiteren Mitteln abgebaut werden. So könnten die Bürger:innen bereits bei der Registrierung bei Web-Anwendungen durch eine geeignete Benutzeroberfläche bzw. Visualisierung darüber informiert werden, welche ihrer Daten für welche Zwecke und Verarbeiter genutzt werden (Tran-Van, Anciaux, and Pucheral 2017). Weiterhin kann den Bürger:innen dargestellt werden, welche Risiken mit der Verwendung einer Web-Anwendung einhergehen

(Yee 2007). Damit können vor der Nutzung der Daten Hemmnisse der Bürger:innen abgebaut werden, bzw. diese ausreichend informiert werden, wobei das vorgestellte *MoP* komplementär als Werkzeug nach der Verarbeitung der Daten genutzt werden kann.

Um der Problemklasse B, die Komplexität des Prozesses zur Bearbeitung von DSA-Anfragen, entgegenzuwirken wird insbesondere das *TaP* eingesetzt, wobei Informationen über gespeicherte *pbD* gehalten werden, und diese strukturiert abgerufen und übermittelt werden können. In diesem Konzept wird die grundlegende Funktionsweise des *TaP* erläutert, jedoch muss auch beachtet werden, dass die technischen Systeme und organisatorischen Maßnahmen der *DH* angemessen gestaltet werden. Die zugrundeliegenden technischen Systeme müssen dahingehend gestaltet sein, dass sie es dem *TaP* ermöglichen, die angemessenen Daten für den Zweck der *DSA-Anfrage* zu erheben, jedoch sollten diese technischen Systeme diese Daten auch grundsätzlich nur an Personen oder Drittsysteme für die Zwecke weitergeben, die in der Datenschutzerklärung festgelegt sind. Hierbei gibt es den Ansatz des *Purpose-based Access Control* (Ji-Won Byun, Bertino, and Li 2005), wobei dieser bereits in verschiedene Datenbanksysteme experimentell integriert wurde (Ji-Wonand Byun and Li 2008; Colombo and Ferrari 2017). Das *TaP* sollte nicht nur als Werkzeug zur Verbesserung des Datenschutzes für Bürger:innen dienen, sondern auch selbst unter Aspekten der Privatheit (*Privacy by Design*) und Informationssicherheit gestaltet werden. Hierfür wurden außerhalb Europas bereits Richtlinien und Standards geschaffen, wie der *AS 27701 PIMS-Standard*, welcher auf der *ISO/IEC 27001* und *ISO/IEC 27002* aufbaut und den Zweck verfolgt, die Compliance von Unternehmen bezüglich des komplexen Themas Datenschutz zu steigern (Christie 2022).

Neben technischen Standards müssen auch die organisatorischen Maßnahmen des Unternehmens zur Verbesserung des Datenschutzes beachtet werden. Neben der Sensibilisierung des Führungspersonals als auch der Mitarbeitenden, welche mit den Daten und Informationen in ihrer täglichen Arbeit umgehen, müssen insbesondere die Prozesse angemessen gestaltet werden, um Datenschutz und Informationssicherheit zu garantieren, wobei ggf. auch mit unerwarteten Konsequenzen umgegangen werden muss (Parks et al. 2017).

Um der Problemklasse C, den Schwierigkeiten bei der Interpretation der Datenkopien, entgegenzuwirken werden Schnittstellen und strukturierte Datenformate zwischen *MoP* und *TaP* definiert, sowie in *MoP* die übermittelten Daten visuell aufbereitet und präsentiert. Dies umfasst damit einen

technischen Lösungsansatz, wobei auch ethische und juristische Fragestellungen betrachtet werden müssen (Grout 2019; Balthasar and Gerl 2019). So könnten verbindliche Richtlinien für die Übertragung von verschiedene Datenformaten (Daten, welche Texte, Bilder, Audio, etc. repräsentieren) geschaffen werden, welche zum einen eine technische Umsetzung schaffen aber zum anderen auch den Aufwand zur Umsetzung für *DH* minimieren. Hierbei müssen unterschiedliche technologische, ethische, juristische und ökonomische Gesichtspunkte miteinander abgewogen werden.

Zusammenfassend bietet der vorgestellte Lösungsansatz mit *MoP* und *TaP* einen technologischen Lösungsansatz, um *DSA-Anfragen* strukturiert umzusetzen. Weiterhin bietet dieser Lösungsansatz eine Grundlage auf welcher weitere technologische, juristische, ethische und ökonomische Lösungsansätze angewendet werden können, um diesen *PIMS-Ansatz* zu erweitern und zu erforschen.

7. Zusammenfassung und Ausblick

In diesem Aufsatz haben wir uns mit dem Recht auf Auskunft gemäß Art. 15 DSGVO bzw. § 34 BDSG auseinandergesetzt und die Umsetzungsperspektive näher beleuchtet. Dabei haben wir aus der Literatur drei wesentliche Problemklassen identifiziert, die im Zusammenhang mit dem Recht auf Auskunft auftreten können (siehe Abschn. 1 und Abschn. 4). Auf Seite der Bürger:innen sind diese Probleme zum einen, dass Hemmnisse bei der Erstellung von *DSA-Anfragen* existieren (Problemklasse A) und dass die von den *DH* erhaltenen Datenkopien für die Bürger:innen teilweise schwierig zu interpretieren sind (Problemklasse C). Auf der Seite der *DH* ist der Prozess zur Bearbeitung einer *DSA-Anfrage* komplex und zeitaufwändig (Problemklasse B). Die *DH* müssen aus oft historisch gewachsenen IT-Systemen sämtliche *pbD* extrahieren, was eine Herausforderung darstellen kann, da viele IT-Systeme darauf nicht ausgelegt sind (siehe Abschn. 3). Zudem gestaltet sich die eindeutige Identifizierung der anfragenden Person und die sichere Übermittlung der Datenkopien häufig schwierig.

Um die genannten Probleme zu adressieren, stellt dieser Aufsatz ein Framework für ein zweiteiliges *PIMS* vor, welches zum einen Bürger:innen bei der Erstellung von *DSA-Anfragen* sowie bei der Interpretation der Datenkopien unterstützt. Zum anderen erlaubt das vorgestellte Framework, eine (voll-)automatisierte Bearbeitung von *DSA-Anfragen* bei *DH*.

Auf Seiten der Bürger:innen soll ein sogenanntes *Monitoring Tool for Personal Data (MoP)* eingesetzt werden, das Informationen über mögliche *DH* enthält, bei denen *pbD* vorliegen könnten. Das *MoP* fordert periodisch eine *DSA* bei allen potenziell relevanten *DH* ein. Die Antworten/Datenkopien werden dann im *MoP* erfasst und die gespeicherten *pbD* in Bürger:innen-Interesse aufbereitet und visualisiert. Dadurch kann die betroffene Person ein besser informiertes Datenselbstmanagement betreiben und das Grundrecht auf informationelle Selbstbestimmung wird gestärkt.

Auf Seiten der *DH* schlagen wir ein Komplementärsystem namens *Tool for automated Data Self-Disclosure Request Processing (TaP)* vor, dass die *DH* bei der Bearbeitung von *DSA-Anfragen* unterstützt. Mit *TaP* soll es möglich sein, *DSA-Anfragen*, die über *MoP* gestellt werden, (voll-)automatisiert zu bearbeiten. Dabei ist in *TaP* (in Verbindung mit *MoP*) die Identifikation der anfragenden Person und ein Kanal zur sicheren Übermittlung der Datenkopien bereits enthalten. Wird *TaP* in die IT-Systeme des *DH* integriert, kann auch die Datenkopie vollautomatisiert erstellt werden. Ansonsten unterstützt *TaP* den *DH* auf Grundlage von Informationen aus den elektronischen *DSE* bei der Erstellung einer vollständigen und rechtssicheren *DSA*. Mithilfe von *TaP* kann die Erteilung von *DSA* ökonomischer erfolgen.

In künftigen Arbeiten gilt es, ein Teilfunktionsmuster des skizzierten Frameworks zu entwickeln und die Funktionsweise beider Komponenten (*TaP* und *MoP*) praktisch zu evaluieren sowie zu prüfen, welche organisatorischen und personellen Rahmenbedingungen bei der Einführung und Umsetzung dieses Systems zu berücksichtigen sind. Insbesondere die Integration des *TaP* in die bestehenden Systeme der *DH* gilt es dabei genauer zu untersuchen. Ferner soll ein Standard zum Datenaustausch zwischen *MoP* und *TaP* definiert werden.

Für den Einsatz von *MoP* ohne das Komplementärsystem *TaP* , sollte durch den Gesetzgeber zudem das Format der Datenkopien (siehe Art. 15 Abs. 3 S. 3 DSGVO) näher spezifiziert werden.

Literatur

Aas, Josh; Barnes, Richard; Case u. a. (2019): Let's Encrypt: An Automated Certificate Authority to Encrypt the Entire Web. In: *CSS'19: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*. New York, NY, USA: ACM, S. 2473-2487. doi: 10.1145/3319535.3363192.

Balthasar, Mandy und Gerl, Armin (2019): Privacy in the toolbox of freedom. In: *2019 12th CMI Conference on Cybersecurity and Privacy (CMI)*. Kopenhagen: IEEE, S. 1-4. doi: 10.1109/CMI48017.2019.8962146.

Barth, Susanne und de Jong, Menno D.T. (2017): The privacy paradox -Investigating discrepancies between expressed privacy concerns and actual online behavior - A systematic literature review. *Telematics and Informatics*, 34 (7), S. 1038–1058. doi: 10.1016/j.tele.2017.04.013.

Bowyer, Alex; Holt, Jack u.a. (2022): Human-GDPR Interaction: Practical Experiences of Accessing Personal Data. New Orleans, LA, USA: *CHI'22*. doi: 10.48550/ARXIV.2203.05037.

Buchmann, Erik und Eichhorn, Susanne (2019): Auskunftsersuchen nach Art. 15 DSGVO. *Datenschutz und Datensicherheit – DuD*, 43 (2), S. 65–70. doi: 10.1007/s11623-019-1065-y.

Bundesamt für Sicherheit in der Informationstechnik (2018): Technische Richtlinie TR-03127: eID-Karten mit eID- und eSign-Anwendung basierend auf Extended Access Control. Bonn: Bundesamt für Sicherheit in der Informationstechnik. https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/TechnischeRichtlinien/TR03127/BSI-TR-03127.pdf.

Byun, Ji-Won; Bertino, Elisa und Li, Ninghui (2005): Purpose Based Access Control of Complex Data for Privacy Protection. In: *SACMAT'05: Proceedings of the Tenth ACM Symposium on Access Control Models and Technologies*. New York, NY, USA: Association for Computing Machinery, S. 102-110. doi: 10.1145/1063979.1063998.

Byun, Ji-Won und Li, Ninghui (2008): Purpose based access control for privacy protection in relational database systems. *The VLDB Journal*, 17 (4), S. 603-619. doi: 10.1007/s00778-006-0023-0.

Callegati, Franco; Cerroni, Walter und Ramilli, Marco (2009): Man-in-the-Middle Attack to the HTTPS Protocol. *IEEE Security & Privacy Magazine*, 7 (1), S. 78–81. doi:10.1109/MSP.2009.12.

Christie, Alec (2022): AS 27701: the PIMS standard you can't afford to ignore. *Privacy Law Bulletin*, 19 (5), S. 92–95. doi: 10.3316/agispt.20220830073173.

Colombo, Pietro und Ferrari, Elena (2017): Enhancing MongoDB with Purpose-Based Access Control. *IEEE Transactions on Dependable and Secure Computing*, 14 (6), S. 591–604. doi: 10.1109/TDSC.2015.2497680.

Dasgupta, Dipankar; Roy, Arunava und Nag, Abhijit (2017): Multi-Factor Authentication. In: *Advances in User Authentication*. Cham: Springer, S. 185-233. doi: 10.1007/978-3-319-58808-7_5.

Dienlin, Tobias; Masur, Philipp K. und Trepte, Sabine (2021): A longitudinal analysis of the privacy paradox. *New Media & Society*, 15 (5), S. 1043-1064. doi: 10.1177/14614448211016316.

DSK - Datenschutzkonferenz (2017): Auskunftsrecht der betroffenen Person, Art. 15 DS-GVO. Kurzpapier Nr. 6. https://www.datenschutzkonferenz-online.de/media/kp/dsk_kpnr_6.pdf.

Geminn, Christian L. (2020): Betroffenenrechte verbessern. *Datenschutz und Datensicherheit – DuD*, 44 (5), S. 307–11. doi: 10.1007/s11623-020-1273-5.

Grout, Vic (2019): No More Privacy Any More? *Information*, 10 (1), doi: 10.3390/info10010019.

Heinemann, Andreas und Straub, Tobias (2019): Datenschutz muss benutzbar sein. *Datenschutz und Datensicherheit – DuD*, 43 (1), S. 7–12. doi: 10.1007/s11623-019-1052-3.

Hintze, Mike und El Emam, Khaled (2018): Comparing the benefits of pseudonymisation and anonymisation under the GDPR. *Journal of Data Protection & Privacy*, 2 (2), S. 145–58.

Këllezi, Pranvera (2021): Consumer Choice and Consent in Data Protection. Antitrust Chronicle. https://www.competitionpolicyinternational.com/category/antitrust-chronicle.

Kröger, Jacob Leon; Lutz, Otto Hans-Martin und Ullrich, Stefan (2021): The myth of individual control: Mapping the limitations of privacy self-management. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3881776.

Di Martino, Mariano; Meers, Isaac u.a. (2022): Revisiting Identification Issues in GDPR `Right Of Access' Policies: A Technical and Longitudinal Analysis. *Proceedings on Privacy Enhancing Technologies*, 2022 (2), S. 95–113. doi: 10.2478/popets-2022-0037.

Ooijen, I. van und Vrabec, Helena U. (2018): Does the GDPR Enhance Consumers' Control over Personal Data? An Analysis from a Behavioural Perspective. *Journal of Consumer Policy*, 42 (1), S. 91–107. doi: 10.1007/s10603-018-9399-7.

Parks, Rachida; Xu, Heng; Chu, Chao-Hsien und Lowry, Paul Benjamin (2017): Examining the intended and unintended consequences of organisational privacy safeguards. *European Journal of Information Systems*, 26 (1), S. 37–65. doi: 10.1057/s41303-016-0001-6.

Petrlic, Ronald (2019): Identitätsprüfung bei elektronischen Auskunftsersuchen nach Art. 15 DSGVO. *Datenschutz und Datensicherheit – DuD*, 43 (2), S. 71–75. doi: 10.1007/s11623-019-1066-x.

Simmons, Gustavus J. (1979): Symmetric and Asymmetric Encryption. *ACM Computing Surveys*, 11 (4), S. 305–330. doi: 10.1145/356789.356793.

Sinclair, David und Jamal, Arshad (2021): Does the GDPR Protect UK Consumers from Third Parties Processing Their Personal Data for Secondary Purposes? A Systematic Literature Review. In: *Cybersecurity, Privacy and Freedom Protection in the Connected World*. Cham: Springer. S. 379-394. doi: 10.1007/978-3-030-68534-8_24.

Sweeney, Latanya (2002): k-Anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10 (05), S. 557–570. doi: 10.1142/S0218488502001648.

Tran-Van, Paul; Anciaux, Nicolas und Pucheral, Philippe (2017): SWYSWYK: A Privacy-by-Design Paradigm for Personal Information Management Systems. In: *International Conference on Information Systems Development (ISD)*. Cyprus. https://hal.inria.fr/hal-01675090.

Venkataramanan, Nataraj und Shriram, Ashwin (2016): *Data privacy: principles and practice*. Chapman and Hall/CRC.

Yee, George (2007): Visual Analysis of Privacy Risks in Web Services. In: *IEEE International Conference on Web Services (ICWS 2007)*. S. 671–78. Doi: 10.1109/ICWS.2007.189.

Zaeem, Razieh Nokhbeh und Barber, Suzanne K. (2021): The Effect of the GDPR on Privacy Policies. *ACM Transactions on Management Information Systems*, 12 (1), S. 1–20. doi: 10.1145/3389685.