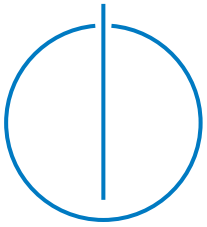
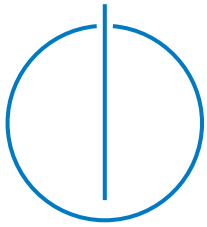
FAKULTÄT FÜR INFORMATIK
DER TECHNISCHEN UNIVERSITÄT MÜNCHEN

Umsetzung des datenschutzrechtlichen Auskunftsanspruchs auf Grundlage von Usage-Control und Data-Provenance-Technologien

Philipp Christoph Sebastian Bier





FAKULTÄT FÜR INFORMATIK
DER TECHNISCHEN UNIVERSITÄT MÜNCHEN

Umsetzung des datenschutzrechtlichen Auskunftsanspruchs auf Grundlage von Usage-Control und Data-Provenance-Technologien

Philipp Christoph Sebastian Bier

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

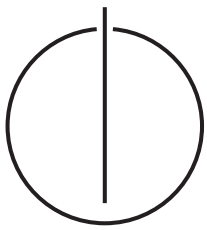
Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Prof. Dr. Uwe Baumgarten

Prüfer der Dissertation:
1. Prof. Dr. Alexander Pretschner
2. Prof. Dr. Indra Spiecker genannt Döhmann, LL.M. (Georgetown Univ.)
   Goethe-Universität Frankfurt am Main
3. Prof. Dr.-Ing. Jürgen Beyerer
   Karlsruher Institut für Technologie

Die Dissertation wurde am 11.05.2017 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 17.11.2017 angenommen.



Kurzzusammenfassung

Die wachsende Komplexität moderner Informationssysteme erschwert die Nachvollziehbarkeit der Speicherung und Verarbeitung personenbezogener Daten. Der einzelne Bürger ist den Systemen quasi ausgeliefert. Das Datenschutzrecht versucht dem entgegenzuwirken. Transparenz, das Recht zu wissen, wer was wann und bei welcher Gelegenheit über einen selbst weiß, ist ein fundamentales Grundrecht. Ein Werkzeug des Datenschutzes zur Herstellung von Transparenz ist der Auskunftsanspruch. Da bisher jedoch eine technische Unterstützung für die Erteilung von Auskünften fehlt, erhalten Betroffene meist nur statische Datenbankauszüge und keine aussagekräftigen Informationen über vergangene Datenerhebungen, Weitergaben und Informationsflüsse.

Die neuere Forschung arbeitet auf automatisierte Trackingmechanismen für personenbezogene Daten hin. Aus der entstehenden Personal-Data-Provenance kann dann eine vollständige Datenschutzauskunft abgeleitet und auf einer elektronischen Auskunftsplattform bereitgestellt werden.

Dieses Ziel wurde allerdings noch nicht vollständig erreicht. Die Anforderungen des Datenschutzes an eine elektronische Datenschutzauskunft wurden bislang nicht ausreichend systematisch berücksichtigt und umgesetzt. Die Struktur der Personal-Data-Provenance, deren konfigurierbare, verteilte Erfassung und die Zusammenführung der Personal-Data-Provenance zum Zeitpunkt eines Auskunftsersuchens sind ungeklärt. Die Datenschutzauskunft und die übrigen Betroffenenrechte sind bis dato nicht miteinander verwoben. Die durch die zusätzlichen Trackingdaten entstehenden Profilbildungsrisiken automatisierter Auskunftssysteme werden nicht adressiert.

Diese Arbeit unterzieht das Recht auf Auskunft einer kritischen Würdigung als Teil des datenschutzrechtlichen Transparenzgedankens und schafft umfassende technische Voraussetzungen für die Wahrnehmung des Rechts auf Auskunft. Die Beiträge dieser Arbeit sind wie folgt:

(1) Die Einbindung des Rechts auf Auskunft in die verfassungsrechtliche und einfachgesetzliche Struktur des Datenschutzes wird diskutiert. Die Bedeutung des Auskunftsrechts wird eingeordnet. Die Herausforderungen durch die Verabschiedung der Datenschutz-Grundverordnung, insbesondere das Recht auf Datenübertragbarkeit, werden mitberücksichtigt. Datenschutzrechtliche Anforderungen an eine automatisierte Beantwortung von Auskunftsersuchen werden unter Zuhilfenahme eines strukturierten Vorgehensmodells systematisch abgeleitet.

(2) Ein verteiltes, datenzentriertes, integriertes und betroffenenfokussiertes Datenschutzauskunftssystem wird entworfen und implementiert. Es erlaubt ein schichtenunabhängiges, semantisch konfigurierbares Provenance-Tracking über Systemgrenzen hinweg. Kombiniert mit Usage-Control-Technologien können weitergehende Datenschutzrechte wie der Löschanspruch durchgesetzt werden. Die entwickelte Auskunftsplattform erlaubt die interaktive und schrittweise Wahrnehmung des Auskunftsrechts.

(3) Entstehende Profilbildungsrisiken werden mittels einer generischen, instanziierbaren und berechenbaren Metrik für Unverkettbarkeit, der Unmöglichkeit des In-Bezug-Setzens personenbezogener Daten, sichtbar gemacht. Bestehende Konzepte für informationstheoretische Unverkettbarkeitsmetriken werden auf beliebige Verkettungsrelationen verallgemeinert. Das Schema wird für vier Relationen mit Bezug zum entwickelten Datenschutzauskunftssystem instanziiert. Das A-priori- und A-posteriori-Wissen eines Angreifers wird formalisiert und in die Metrik integriert. Die Metrik wird als Grundlage einer informierten Entscheidung des Betroffenen bezüglich des Provenance-Trackings vorgeschlagen.

Die Plausibilität des Ansatzes wird anhand eines durchgehenden Beispiels gezeigt. Das Datenschutzauskunftssystem wird darüber hinaus auf seine Skalierbarkeit hin evaluiert. Die Ergebnisse zeigen, dass eine passgenaue, datenschutzgerechte Datensammlung und eine gute Speicherskalierbarkeit miteinander Hand in Hand gehen. Für die Unverkettbarkeitsmetrik wird eine heuristische Berechnung demonstriert. Die Genauigkeit hängt vom Schwellwert des Abbruchkriteriums ab. Schließlich wird die Akzeptanz der Auskunftsplattform durch Nutzer anhand einer Studie gezeigt. Die entwickelte Plattform *PrivacyInsight* stellt sich gegenüber existierenden Formen der Visualisierung als überlegen heraus.

Abstract

The ever-growing levels of complexity of modern information systems complicate the traceability of processed and stored personal data. The individual citizen is at the mercy of the systems. Data protection law tries to counteract this. Transparency – the right to know who knows what, when, and on which occasion about oneself – is a fundamental right of the German constitution. The right to information is one instrument of data protection to establish transparency. Since there still is a lack of functionality in regards to transparency, the person concerned usually merely receives a static database snapshot. Obtaining no meaningful information about collection, transfer, and other flows of personal data.

Recent research efforts evolve around automated tracking mechanisms for personal data. All information required can be derived from the resulting personal data provenance. This information must be visualized comprehensible for the person concerned.

However, this objective has not yet been fully achieved. The data protection requirements for an electronic response to an information request have not yet been systematically considered. The data structure of personal data provenance is unclear. Its configurable, distributed collection and its aggregation at the time of an information request prove to be research gaps. It is an established fact that the right to information and other rights of the person concerned have not yet been technically interwined. The resulting tracking data raises new issues of data protection law itself. They allow extensive profiling which has not been addressed.

In the work at hand the right to information is critically assessed with a focus on the notion of transparency. Furthermore, the technical requirements to exercise the right to information are created. The contributions of this work are as follows:

(1) The implementation of the right to information in constitutional and legal structures of data protection is examined. The relevance of the right to information is evaluated. The challenges arising from the adoption of the General Data Protection Regulation, especially the right to data portability, are inspected. Data protection requirements for an automated response to information requests are systematically derived using a structured approach.

(2) A distributed, data-centric, integrated and user-focused data protection information system is designed and implemented. It allows layer-independent, semantic-configurable provenance tracking across system boundaries. Combined with usage control, further data protection rights such as the right to erasure can be enforced. The implemented privacy dashboard allows the interactive and gradual exercise of the right to information.

(3) Profiling risks are visualized by means of a generic, instantiable and calculable metric for unlinkability, the impossibility to interrelate personal data. Existing concepts for information-theoretic unlinkability metrics are generalized to arbitrary linkage relations. The schema is instantiated for four relations related to the data protection information system. A priori and a posteriori knowledge of an attacker is formalized and integrated into the metric. This metric is proposed as the basis for an informed decision of the person concerned regarding the provenance tracking.

The plausibility of the approach is demonstrated by means of a continuous example. In addition, the scalability of the data protection information system is evaluated. The results show that a precise data collection and a good memory scalability go hand in hand. A heuristic calculation is performed for the unlinkability metric. The accuracy depends on the threshold value of the termination criterion. Finally, user acceptance of the privacy dashboard is verified by a study. The developed platform platform turns out to be superior to existing visualizations.

Inhaltsverzeichnis

Abkürzungsverzeichnis

A4Cloud	Accountability For Cloud and Other Future Internet Services
a. A.	andere Auffassung
AAA	Authentication, Authorization, and Accounting
a. a. O.	am angegebenen Ort
AEUV	Vertrag über die Arbeitsweise der Europäischen Union
a. F.	alte Fassung
AG	Amtsgericht
AGB	Allgemeine Geschäftsbedingungen
AktG	Aktiengesetz
API	Application Programming Interface
A-PPL	Accountable PrimeLife Policy Language
BDSG	Bundesdatenschutzgesetz
BfD	Beauftragter für den Datenschutz
BfDI	Bundesbeauftragter für den Datenschutz und die Informationsfreiheit
BGB	Bürgerliches Gesetzbuch
BGBl.	Bundesgesetzblatt
BGH	Bundesgerichtshof
BInDSG	Gesetz zum Schutz personenbezogener Daten in der Berliner Verwaltung
BNDG	Gesetz über den Bundesnachrichtendienst
BVerfG	Bundesverfassungsgericht

BVerfSchG Gesetz über die Zusammenarbeit des Bundes und der Länder in Angelegenheiten des Verfassungsschutzes und über das Bundesamt für Verfassungsschutz

BVerwG Bundesverwaltungsgericht

Cookie-RL Richtlinie 2009/136/EG zur Änderung der Richtlinien 2002/22/EG und 2002/58/EG und der Verordnung 2006/2004

COPS Common Open Policy Service

DAX Deutscher Aktienindex

DFG Deutsche Forschungsgemeinschaft

DNS Domain Name System

DRM Digital Rights Management

DSB Datenschutzbeauftragter

DSGVO Verordnung 2016/679 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr (Datenschutz-Grundverordnung)

DSRL Richtlinie 95/46/EG zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr (Datenschutzrichtlinie)

EGMR Europäischer Gerichtshof für Menschenrechte

EMRK Europäische Menschenrechtskonvention

EPAL Enterprise Privacy Authorization Language

ePrivacy-RL Richtlinie 2002/28/EG über die Verarbeitung personenbezogener Daten und den Schutz der Privatsphäre in der elektronischen Kommunikation (Datenschutzrichtlinie für elektronische Kommunikation)

ErwGr Erwägungsgrund

EU Europäische Union

EuGH Europäischer Gerichtshof

EUV Vertrag über die Europäische Union

EVAL	Erweitertes Vorgehensmodell für Anforderungen aus dem Legalbereich
Fn.	Fußnote
GG	Grundgesetz
ggü.	gegenüber
GmbHG	Gesetz betreffend die Gesellschaften mit beschränkter Haftung
GNU	GNU's Not Unix
GRCh	Charta der Grundrechte der Europäischen Union
HTML	Hypertext Markup Language
IBAN	International Bank Account Number
ID	Identifikator
i. e. S.	im engeren Sinne
IF	Informationsfluss
IP	Internet Protocol
IRISS	Increasing Resilience in Surveillance Societies
IT	Informationstechnologie
i. V. m.	in Verbindung mit
JSON	JavaScript Object Notation
JVM	Java Virtual Machine
KG	Kammergericht
KMU	Kleine und mittlere Unternehmen
KORA	KOnkretisierung Rechtlicher Anforderungen
LAG	Landesarbeitsgericht
LDSG-SH	Schleswig-Holsteinisches Gesetz zum Schutz personenbezogener Informationen

LfD	Landesbeauftragter für den Datenschutz
LG	Landgericht
MADG	Gesetz über den militärischen Abschirmdienst
m. w. N.	mit weiteren Nachweisen
NACE	Statistische Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft, *franz. Nomenclature statistique des activités économiques dans la Communauté européenne*
n. F.	neue Fassung
NID	Namespace Identifier
o. B. d. A.	ohne Beschränkung der Allgemeinheit
ODRL	Open Digital Rights Language
OECD	Organisation for Economic Cooperation and Development
OLG	Oberlandesgericht
OPM	Open Provenance Model
OSL	Obligation Specification Language
OVG	Oberverwaltungsgericht
P3P	Platform for Privacy Preferences
PAP	Policy Administration Point
PB-Daten	personenbezogene Daten
PDF	Portable Document Format
PDP	Policy Decision Point
PEP	Policy Enforcement Point
PET	Privacy Enhancing Technology
PHP	PHP Hypertext Preprocessor
PII	Personal Identifiable Information

PIP	Policy Information Point
PMP	Policy Management Point
PRIME	Privacy and Identity Management for Europe
PrimeLife	Privacy and Identity Management in Europe for Life
ProCP	Provenance Collection Point
ProDP	Provenance Dissemination Point
ProSP	Provenance Storage Point
PRP	Policy Retrieval Point
PSM	Platform-Specific Model
PXP	Policy Execution Point
REST	Representational State Transfer
RMI	Remote Method Invocation
Rn.	Randnummer
SGB	Sozialgesetzbuch
SiG	Gesetz über Rahmenbedingungen für elektronische Signaturen
StGB	Strafgesetzbuch
SUS	System Usability Scale
TCP	Transmission Control Protocol
TDDSG	Teledienstedatenschutzgesetz
TET	Transparency Enhancing Technology
TKG	Telekommunikationsgesetz
TLS	Transport Layer Security
TMG	Telemediengesetz
UC	Usage Control

UCN	Usage Control Node
UEQ	User Experience Questionnaire
UKlaG	Gesetz über Unterlassungsklagen bei Verbraucherrechts- und anderen Verstößen
ULD	Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein
UN ISIC	United Nations International Standard Industrial Classification of all Economic Activities
URI	Uniform Resource Identifier
URN	Uniform Resource Name
UUID	Universally Unique Identifier
UWG	Gesetz gegen den unlauteren Wettbewerb
VG	Verwaltungsgericht
VGH	Verwaltungsgerichtshof
XACML	eXtensible Access Control Markup Language
XML	Extensible Markup Language
XSD	XML Schema Definition
ZPO	Zivilprozessordnung

Nomenklatur

$\mathscr{A}$	Angreifer
$\mathscr{A}^{\varsigma}$	Systemangreifer
$\mathscr{A}^{c}$	Zentraler Angreifer
$\mathscr{B}$	Betroffene ($b \in \mathscr{B}$)
$\mathcal{B}_n$	Bellsche Zahl (Anzahl der Partitionen einer n-elementigen Menge)
$\mathscr{C}$	Container ($c \in \mathscr{C}$)
$\mathscr{C}_{abstr}$	Abstrakte Container ($c_{abstr} \in \mathscr{C}_{abstr}$)
$\mathscr{C}_{\iota}$	Intermediate Container ($c_{\iota} \in \mathscr{C}_{\iota}$)
$\mathscr{C}_{real}$	Reale Container ($c_{real} \in \mathscr{C}_{real}$)
col	Kantenfärbung eines Provenance-System-Graphen ($col : \mathscr{L}^{\Sigma} \rightarrow \mathcal{P}(\mathscr{D})$)
$\mathscr{D}$	Daten ($d \in \mathscr{D}$)
Δ	Grad der Unverkettbarkeit
$\|\Delta\|$	Normierter globaler Grad der Unverkettbarkeit
Dom	Lokale Domänen
$\mathscr{E}$	Ereignisse ($e \in \mathscr{E}$)
$\mathcal{E}$	Entitäten ($\varepsilon \in \mathcal{E}$)
E	Entitätsmenge einer Entitätsklasse ($E \subset \mathcal{E}$)
H	Entropie
$\mathscr{F}$	Bezeichner ($\phi \in \mathscr{F} \subseteq Dom \times LoA \times Type \times Hndl$)

f	Benennungsfunktion ($f : \mathscr{F} \to \mathscr{C}$)
$\mathscr{G}$	Provenance-Graph $\mathscr{G} = (\mathscr{R}, \mathscr{L})$
$\mathscr{G}_t$	Provenance-Graph $\mathscr{G}_t = (\mathscr{R}_t, \mathscr{L}_t)$ zum Zeitpunkt $t \in \mathbb{N}$
$\mathscr{G}^\Sigma$	Provenance-System-Graph
Hndl	Identifikatoren von Containern
I	Beobachtungsereignis
ι	Intermediate-Container-Funktion ($\iota : \text{SCOPE} \to \mathscr{C}_\iota$)
$\mathscr{L}$	Kanten im Provenance-Graph (engl. lineage)
$\mathscr{L}^\Sigma$	Kanten im Provenance-System-Graph
l	Aliasfunktion ($l : \mathscr{C} \to \mathcal{P}(\mathscr{C})$)
LoA	Abstraktionsschichten (engl. layer of abstraction)
loc	Verortungsfunktion ($loc : \mathscr{R} \to \mathscr{S}$)
$\mathbb{N}$	Natürliche Zahlen
$\mathcal{O}$	Landau-Symbol, asymptotische obere Schranke ($\varphi \in \mathcal{O}$)
$\mathcal{P}$	Potenzmenge
$\mathbb{P}$	Wahrscheinlichkeitsmaß
π	Personenbezugsfunktion ($\pi : \mathscr{B} \to \mathcal{P}(\mathscr{D})$)
ϖ	Fortschrittsquote
$\mathscr{R}$	Repräsentationen ($\rho \in \mathscr{R} \subseteq \mathscr{D} \times \mathscr{C} \times \mathbb{N}$)
$\mathcal{R}$	Kandidatenrelationen ($R \in \mathcal{R}$)
R	Relation
R_T	Tatsächliche Relation
$R^\boxdot$	Binäre bzw. boolsche Matrix der Größe m einer Relation R mit den Einträgen r_{ij}

$R^{\triangleright}$	Datenflussrelation
$R^{\lessdot}$	Identifikationsrelation
$R^{\triangledown}$	Speicher- und Verarbeitungsrelation
$R^{\equiv}$	Verknüpfungsrelation
ϱ	Fehlerwahrscheinlichkeit
$\mathscr{S}$	Systeme ($\varsigma \in \mathscr{S}$)
$\mathcal{S}_{n,k}$	Stirling-Zahl zweiter Art (Anzahl der k-Partitionen einer n-elementigen Menge)
Σ	Systemzustände ($\sigma \in \Sigma$)
s	Speicherfunktion ($s : \mathscr{C} \rightarrow \mathcal{P}(\mathscr{D})$)
σ_0	Initialer Systemzustand
σ_t	Systemzustand $\sigma_t = (s_t, l_t, f_t)$ zum Zeitpunkt $t \in \mathbb{N}$
$\mathscr{T}$	Übergangsrelation (engl. transition relation)
$\mathscr{T}_{mod}$	Modifizierte Übergangsrelation für XLAYER- und XSYSTEM-Informationsflüsse
Θ	Datenkategorien ($\theta \in \Theta$)
τ	Schwellwert (engl. threshold)
ϑ	Datenkategorie-Zuweisungsfunktion ($\vartheta : \mathscr{D} \rightarrow \Theta$)
term	Zuweisungfunktion für den Beendigungszeitpunkt ($term : \mathscr{R} \rightarrow \mathbb{N} \cup \text{NIL}$)
Type	Containertypen
W	Matrix der bedingten Flusswahrscheinlichkeiten der Größe m mit den Einträgen w_{ij}
X	Zufallsvariable
χ	Scope-Attributspezifikationsfunktion ($\chi : \Sigma \times E \rightarrow \text{SCOPE} \times \text{DELIMITER} \times \text{BEHAVIOR} \times \text{INTER}$)
ζ	Scope-Zustandsfunktion ($\zeta : \text{SCOPE} \rightarrow \{\text{ACTIVATED}, \text{DEACTIVATED}\}$)

1 Einleitung

Moderne informationstechnische Systeme erreichen eine Komplexität,[1] die die Speicherung und Verarbeitung personenbezogener Daten nahezu undurchschaubar macht. Der einzelne Bürger ist den Systemen quasi ausgeliefert.

Das Datenschutzrecht versucht dem entgegenzuwirken. Transparenz, das Recht zu wissen, wer was wann und bei welcher Gelegenheit über mich weiß, ist ein fundamentales Grundrecht.[2] Die Transparenz des Datenumgangs[3] ist immanente Voraussetzung des Rechts auf Informationelle Selbstbestimmung. Ohne sie wird ein besonnener und rücksichtsvoller Einsatz von Informationstechnologie (IT) nicht sichtbar, eine Intervention in die Verarbeitungsvorgänge nicht möglich.[4]

Ein Werkzeug des Datenschutzes zur Herstellung von Transparenz ist der Auskunftsanspruch. Er ist Voraussetzung zur Wahrnehmung der übrigen Betroffenenrechte auf Löschung, Sperrung und Berichtigung. Trotz seiner enormen Bedeutung für einen effektiven Datenschutz wird der Auskunftsanspruch durch die Praxis vernachlässigt. Auskünfte werden zwar erteilt, jedoch nur in Form von statischen Datenbankauszügen,[5] auch wenn diese ausgesprochen umfangreich ausfallen können.[6] Dem liegt nicht unbedingt ein Unwille der beteiligten Stellen zu Grunde. Die fragmentierte Informationsverarbeitung, innerhalb und außerhalb standardisierter Prozesse, erschwert oder unterbindet das Zusammentragen von Informationen über vergangene Datenerhebungen, Weitergaben und Informationsflüsse. Eine technische Realisierung des Rechts auf Auskunft wurde bisher unzureichend angegangen.[7]

An dieser Stelle setzt Personal-Data-Provenance an. Personal-Data-Provenance ist die dokumentierte Historie eines personenbezogenen Datums. Eine Provenance-Tracking-Infrastruktur verfolgt demnach den Lebenszyklus eines personenbezogenen Datums ausgehend von der Erhebung beim Betroffenen oder einem Dritten, über einzelne Verarbeitungs- und Speicherschritte bis hin zur Übermittlung. Personenbezogene Daten

[1] Masing, NJW 2012, 2305 (2308); Weichert, DuD 2006, 694 (695).

[2] BVerfGE 65, 1 (43); BVerfGE 125, 260 (334).

[3] Die Begriffe „Datum“, „Information“, „Umgang“, „Verwendung“, „Verarbeitung“, „Nutzung“ usw. werden im Glossar am Ende der Arbeit erklärt.

[4] BVerfGE 65, 1 (43).

[5] Mehr dazu im Abschnitt 1.1.1.

[6] Beispielsweise die von Facebook bereitgestellten Datenbestände – zu finden unter http://www.europe-v-facebook.org/DE/Datenbestand/datenbestand.html, abgerufen am 9. Mai 2017.

[7] Wächter, DuD 1996, 272 (273).

werden bei der Erhebung annotiert und der Umgang mit ihnen fortlaufend und automatisiert protokolliert. Alle Schritte werden mit dem Zweck der Erhebung und Verarbeitung des personenbezogenen Datums in Bezug gesetzt. Letztendlich soll der Betroffene die Möglichkeit bekommen, über eine Auskunftsplattform jederzeit Einblick in den Umgang mit seinen personenbezogenen Daten zu nehmen.

Kombiniert mit Usage-Control-Technologien können weitergehende Datenschutzrechte, wie der Löschanspruch, im Nachhinein direkt über die Auskunftsplattform durchgesetzt werden. Auf der anderen Seite können die entstehenden, stark verknüpften Trackingdaten auch selbst neue datenschutzrechtlichen Probleme aufwerfen. Sie leisten einer umfangreichen und freiheitsgefährdenden Profilbildung Vorschub. In diesem Spannungsfeld und unter Berücksichtigung eines sich wandelnden und europäisierenden Datenschutzes stellt diese Arbeit ein integriertes Datenschutzauskunftssystem vor.

1.1 Motivation

In unserer Gesellschaft besteht kein Konsens mehr, was öffentlich und was privat ist.[8] Die Wertschätzung personenbezogener Daten variiert stark.[9] Die einen pflegen den digitalen Exhibitionismus, die anderen streben einen neuen Grad der Anonymität an. Nutzer digitaler Dienste werden entsprechend ihrer Einstellung zum Umgang mit personenbezogenen Daten in Datenschutzfundamentalisten, Pragmatiker, und Unbekümmerte eingeteilt.[10] In Studien wurde festgestellt, dass zusätzlich zur gesellschaftlichen Vielfalt auch das Verhalten des Einzelnen nicht konsistent ist. Behauptete Präferenzen und tatsächliches Verhalten weichen voneinander ab.[11] Das Verhalten selbst ist nicht konsistent im Hinblick auf die zu erwartenden Konsequenzen[12] und stark abhängig von kontextuellen, nicht-normativen Faktoren.[13]

Der Auskunftsanspruch ist von all dem unabhängig.[14] Das Recht auf informationelle Selbstbestimmung ist universell und kann nicht durch ein bestimmtes Verhalten des Betroffenen verwirkt werden.[15] Es wirkt grundsätzlich in alle Branchen und Anwendungsdomänen hinein.[16] Besondere Arten personenbezogener Daten, wie Gesundheitsdaten

[8] Masing, NJW 2012, 2305 (2308).
[9] Grossklags/Acquisti 2007; Beresford/Preibusch/Kübler 2010; Acquisti/John/Loewenstein 2013.
[10] Kumaraguru/Lorrie F. Cranor 2005.
[11] Berendt/Günther/Spiekermann 2005.
[12] Woodruff et al. 2014.
[13] Acquisti/John/Loewenstein 2013.
[14] Siehe Kapitel 3.4.
[15] Auch wenn Daten aus allgemein zugänglichen Quellen entnommen werden können, sieht das BDSG keine Ausnahme von der Auskunftspflicht vor – zur einschlägigen Auslegung von § 34 Abs 7 i. V. m. § 33 Abs. 2 S. 2 Nr. 7 siehe Kapitel 3.8.3.
[16] Siehe allerdings Kapitel 2.1.2 zur Problematik der Drittwirkung der Grundrechte.

und politische oder religiöse Zugehörigkeit, sind genauso geschützt wie Alltägliches und Triviales.

Insofern ergibt sich der Bedarf für ein integriertes Datenschutzauskunftssystem nicht aus einem bestimmten Geschäftsmodell, sondern aus seiner rechtlichen Relevanz. Welche offenen Forschungsfragen existieren oder ob die Anforderungen des Datenschutzes auch ohne ein Datenschutzauskunftssystem schon heute erfüllt werden, klärt der folgende Abschnitt.

1.1.1 Status Quo des Auskunftsrechts in der Praxis

Zu Beginn des Jahrtausends war es um die Situation des Auskunftsanspruchs schlecht bestellt. Noch im Jahr 2005 hat die SCHUFA Auskunftsersuchen merkblattartig beantwortet.[17] Neben einer fehlenden oder nicht rechtzeitigen Reaktion auf ein Auskunftsersuchen waren die Auskünfte häufig unvollständig oder hatten rein beschreibenden Charakter, so der ehemalige Leiter des Unabhängigen Landeszentrums für Datenschutz Schleswig-Holstein (ULD), Thilo Weichert, im Jahre 2006.[18]

Auch 2012 war das unabdingbare Recht auf Auskunft noch vielen Unternehmen unbekannt. Sein Umfang wurde unterschätzt oder schlicht ignoriert. Anfragen wurden nicht oder nur unvollständig beantwortet. Insbesondere Adresshändler waren nicht bereit, Informationen über Datenempfänger so genau mitzuteilen, dass sie von Betroffenen genutzt werden konnten. Die wesentlichen Gründe waren komplexe Unternehmensstrukturen, mangelhafte technische und organisatorische Voraussetzungen und fehlendes Wissen über die rechtlichen Anforderungen.[19]

In den aktuellen Berichten der Landesbeauftragten für den Datenschutz (LfD) erscheint die Situation nur unwesentlich besser. Unternehmen geben teilweise immer noch unzureichend oder keine Auskunft.[20] Begründet in einer extrem niedrigen Kontrolldichte,[21] ist die Berichtsdichte der LfD zum Auskunftsanspruch insgesamt gering.

Eine Erhebung im Rahmen einer DFG-Studie zum TMG ergab im Jahr 2009 eine Rücklaufquote von 54 % auf Auskunftsersuchen.[22] Ein erster unstrukturierter Versuch, Auskünfte inhaltlich zu analysieren, wurde im Rahmen des EU-Forschungsprojektes Increasing Resilience in Surveillance Societies (IRISS) unternommen.[23] Umfangreichere Untersuchungen wurden jedoch erst in den Jahren 2015/16 durchgeführt.

Die Untersuchung von Herrmann und Lindemann nahm zwei verschiedene Bereiche in

[17] LfD Hessen, 34. Tätigkeitsbericht, LT-Drs. 16/5892, 13.
[18] Weichert, DuD 2006, 694 (694).
[19] LfD Hessen, 41. Tätigkeitsbericht, 184 ff.
[20] LfD Brandenburg, 18. Tätigkeitsbericht, 164 f.
[21] Schulzki-Haddouti 2016, 114 ff.
[22] Kühling et al., DuD 2009, 335 (342).
[23] Zurawski 2014.

den Blick.[24] Im ersten Teil wurden die 100 populärsten Smartphone-Apps in Deutschland, 50 für das Googles Android-Betriebssystem und 50 für Apples iOS, installiert und genutzt. Anschließend wurde bei den Anbietern der Apps eine Auskunftsanfrage gestellt. Von den 100 Anfragen wurden 43 % im Großen und Ganzen ordnungsgemäß beantwortet. Für die deutschen Anbieter lag die Quote immerhin bei 75 %. 12 % aller Anbieter beantworteten die Anfrage nur auf abstrakte Art und Weise oder offensichtlich unvollständig. Die übrigen Anbieter antworteten entweder gar nicht, waren nicht erreichbar oder verweigerten die Auskunft.

Im zweiten Teil wurde eine Stichprobe von 120 der 500 beliebtesten Webseiten[25] in Deutschland (davon 57 aus den Top 100) ausgewählt. Auf diesen Webseiten wurde ein Benutzeraccount mit einer E-Mailadresse angelegt. Auch an die Webseitenbetreiber wurde anschließend ein Auskunftsersuchen gerichtet. Der Anteil der ordnungsgemäßen Antworten lag wieder bei 43 %, die der unvollständigen bei 16 %. Ergänzend wurden auch Anfragen von nicht registrierten E-Mailadressen gestellt. Der Umgang mit diesen Anfragen war beunruhigend. In immerhin 18 % der Fälle wurden personenbezogene Daten mitgeteilt, ohne die Berechtigung zu prüfen. Dies lässt auf unzureichende Identifikationsprozesse bei den jeweiligen Unternehmen schließen. Dritte können so leicht Zugang zu sensiblen personenbezogenen Daten erhalten.

Die eigens erstellte,[26] zeitgleiche Studie bestand aus einem quantitativen und einem sich anschließenden qualitativen Teil. Sie ist detailliert in Anhang A beschrieben.

Im qualitativen Teil wurde die Reaktion von 40 ausgewählten Unternehmen auf Auskunftsersuchen ausgewertet. Die Prüfung der Auskünfte erfolgte entlang der zwölf in Tabelle A.3 des Anhangs aufgelisteten Anforderungen. Die Anforderungen teilen sich in drei formale (# 1-3) und neun inhaltliche (# 4-12) Anforderungen auf.

Abbildung 1.1 zeigt den Erfüllungsgrad der Anforderungen durch die erteilten Auskünfte. Die Situation bei den formalen Kriterien liegt im akzeptablen Bereich. Bis auf ein Unternehmen konnten alle auf irgendeine Art und Weise kontaktiert werden. 27 Unternehmen (67,5 %) beantworteten die Auskunftsanfrage in weniger als 15 Tagen. Nur sechs Unternehmen stellten auch nach 6 Monaten noch keine Auskunft zur Verfügung. Alle Auskünfte bis auf zwei waren klar strukturiert und verständlich formuliert.

Die Lage bei den inhaltlichen Kriterien fällt deutlich schlechter aus. Insbesondere Angaben zu Speicherort, Empfängern, Datenflüssen und dem Zweck der Speicherung sind unvollständig. Der Speicherort wurde meist dann sichtbar, wenn Screenshots von Datenbankanwendungen als Teil der Auskunft mitgeliefert wurden. Dies war immerhin bei sechs Unternehmen der Fall. Die internen Flüsse personenbezogener Daten wurden in keinem Fall voll zufriedenstellend wiedergegeben. Nur wenige Auskünfte beinhalten überhaupt Angaben dazu. Positive Ausnahme war ein deutscher Versand- und Einzel-

[24] Herrmann/Lindemann 2016.
[25] Alexa top 500 sites on the web, Germany, http://www.alexa.com/topsites/countries/DE.
[26] Bier/Kömpf/Beyerer 2017.

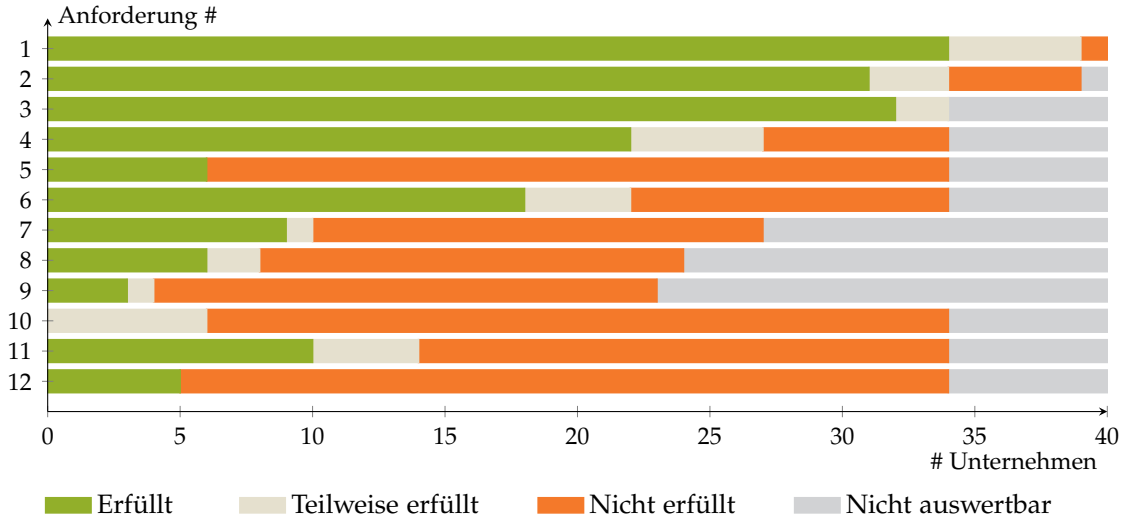


Abbildung 1.1: Erfüllungsgrad der Anforderungen durch die erteilten Auskünfte

händler, der grundlegende Informationen zu den an der Datenverarbeitung beteiligten Abteilungen mitlieferte. Er stand auch insgesamt mit 11 erfüllten oder teilweise erfüllten Anforderungen an der Spitze.

Einen Überblick über die Belastung der Unternehmen durch Auskunftsanfragen bietet die 2bAdvice-Studie zur Datenschutzpraxis.[27] Die 272 befragten Unternehmen erhalten im Durchschnitt 231 Auskunftsersuchen jährlich. Allerdings ist die Belastung durch Auskunftsersuchen stark ungleich verteilt. 100 von ihnen haben noch nie ein Auskunftsersuchen erhalten und nur 14 % erhalten über 100 Anfragen pro Jahr. Die Anzahl der Auskunftsanfragen nimmt mit der Unternehmensgröße zu.

Die Hauptlast der Auskunftsanfragen tragen wenige große oder exponierte Unternehmen wie Auskunfteien, Handelsunternehmen oder Unternehmen, die bereits wegen Datenschutzthemen im Medienfokus standen.[28] Bei einem Versandhändler mittlerer Größe sind im Kundenservice etwa 1,5 vollzeitäquivalente Stellen für die Beantwortung von Auskuftsersuchen vorgesehen. Bei einem großen Logistikkonzern mit über 100 Mitarbeitern im Datenschutz ist, zusätzlich zu einem großen Kundenservice, ein spezialisierter Mitarbeiter in Vollzeit mit der Bearbeitung von Auskunftsersuchen beschäftigt. Ein anderes exponiertes Unternehmen ist die Deutschen Telekom AG. Bei ihr arbeiten 63 Mitarbeiter in Vollzeit am Thema Datenschutz.[29]

[27] Malinka et al. 2015.
[28] Die folgenden Daten entstammen vertraulichen Expertengesprächen.
[29] Schulzki-Haddouti 2016, 117.

Insgesamt ergibt sich ein Bild des Status quo der Datenschutzauskunft, der Forschungsanstrengungen im Hinblick auf ein integriertes Datenschutzauskunftssystem in zwei Dimensionen rechtfertigt. Zum einen, um die Einhaltung datenschutzrechtlicher Anforderungen, insbesondere im Hinblick auf Informationsflüsse, zu verbessern. Zum anderen, um durch eine stärkere Automatisierung die Belastung der Unternehmen zu reduzieren.

1.1.2 Bestehende Ansätze zur elektronischen Auskunftserteilung in Literatur und Praxis

Datenschutzauskunftssysteme sind im größeren Kontext von Transparency Enhancing Technologies (TETs) zu sehen. Eine umfangreiche Bestandsaufnahme zu diesem Forschungsfeld findet sich in den Veröffentlichungen von Hedbom[30] sowie Janic et al.[31]

Eine generelle Klassifikation gegenwärtiger und zukünftiger Privacy Dashboards wurde von Zimmermann et al. entwickelt.[32]

Hilfestellungen für den Betroffenen, die allerdings das Recht auf Auskunft nicht ersetzen können, sind Browser-Plug-Ins wie Mozilla Lightbeam.[33] Lightbeam deckt die Abhängigkeit des Cookie-Trackings durch unterschiedliche Parteien auf Clientseite auf.

Das Data Disclosure Log von Kolter et al. setzt gleichermaßen auf einer Browser-Erweiterung auf.[34] Es ist eine Java-Applikation, die Datenverarbeitungsvorgänge in verschiedenen graphenbasierten Sichten darstellbar macht. Zusätzlich zu dem aus den Browserdaten entnommenen Verhalten des Betroffenen greift das Data Disclosure Log auf eine „Crowd-Sourced"-Datenbank zu, die Informationen zu den weiteren Datenverarbeitungsvorgängen bei der verantwortlichen Stelle enthält.

Standardisierte Bildsymbole (Privacy Icons) können ebenfalls als Hilfsmittel für zukünftige Auskunftsplattformen gesehen werden. Mit ihnen können der Zweck der Datenverarbeitung und Richtlinien zum Umgang mit personenbezogenen Daten visualisiert werden. Ein Überblick über Forschungsanstrengungen in dieser Richtung wurde von Hansen verfasst.[35] Standardisierte Bildsymbole werden durch ihre Verankerung in Art. 12 Abs. 7 DSGVO in Zukunft noch an Bedeutung gewinnen.[36]

Das Forschungsprojekt „Datenschutz-Auskunftsportal" konzipierte eine zentrale Plattform, die Betroffene darin unterstützt, ihr Recht auf Auskunft wahrzunehmen. Neben allgemeinen Informationen sollte die Plattform eine prozessgestützte Hilfe zur Einreichung von Auskunftsersuchen bieten.[37] Zu den Ergebnissen des Projektes auf technischer

[30] Hedbom 2009.
[31] Janic/Wijbenga/Veugen 2013.
[32] Zimmermann/Accorsi/Müller 2014.
[33] https://www.mozilla.org/de/lightbeam.
[34] Kolter/Netter/Pernul 2010.
[35] Hansen 2009.
[36] Weitere Ideen finden sich bei Holtz/Nocun/Hansen 2011 und Fischer-Hübner/Köffel et al. 2010.
[37] ULD Schleswig-Holstein, 35. Tätigkeitsbericht, LT-Drs. 18/2730, 121.

Seite ist bisher wenig bekannt. Schon seit längerer Zeit verfügbar ist die Formularhilfe selbstauskunft.net. Sie unterstützt die teilautomatisierte Anforderung von Auskünften bei 1786 Unternehmen und Behörden.[38]

Plattformen wie das Google Privacy Dashboard[39] und das Portal von acxiom[40] bieten neben der Einstellung von Datenschutzpräferenzen auch die Möglichkeit, Einblick in bestimmte personenbezogene Daten zu nehmen. Sie sind das, was heute einer Datenschutzauskunftsplattform am nächsten kommt. Allerdings erfüllen auch diese Tools nicht alle im letzten Abschnitt eingeführten Anforderungen. Insbesondere werden Herkunft und Empfänger von personenbezogenen Daten nicht vollumfänglich dargestellt. Der Zweck ist nicht entlang der stattfindenden Verarbeitungsprozesse erkennbar.

Das umfangreichste Projekt, das sich in den letzten Jahren mit technischen Lösungen für Datenschutzauskunftssysteme auseinandergesetzt hat, war das EU-Forschungsprojekt Accountability For Cloud and Other Future Internet Services (A4Cloud)[41] mit seinen beiden Vorgängerprojekten Privacy and Identity Management for Europe (PRIME)[42] und Privacy and Identity Management in Europe for Life (PrimeLife).[43] Das Projekt A4Cloud entwickelte eine Architektur für die Beschreibung und Durchsetzung von Datenschutz-Policies in der Cloud, die Protokollierung und Auditierung der Datenverarbeitung in der Cloud, sowie die Benachrichtigung des Betroffenen über die Datenverarbeitungsvorgänge. A4Cloud sieht jedoch kein, in einer durchgängigen und vollständigen Provenance resultierendes, systemübergreifendes Informationsflusstracking vor.

Die Software Data Track war eines der ersten Tools um Transparenz über die Erhebung von personenbezogenen Daten beim Betroffen zu schaffen.[44] Ursprünglich war Data Track ein clientseitiges Transaktionsprotokoll für personenbezogene Daten. Darauf aufbauend wurde das Werkzeug im Rahmen der drei Forschungsprojekte fortlaufend weiterentwickelt. In A4Cloud wurde es integraler Bestandteil der serverseitigen Transparenzplattform für die Cloud,[45] genannt GenomSynlig.[46]

GenomSynlig bietet zwei Ansichten auf vergangene Datenerhebungen, den „Trace View“ und den „Timeline View“.[47] Ersteres ist eine graphenbasierte Ansicht auf die Verknüpfungen zwischen Daten und erhebenden Stellen. Letzteres ist eine Zeitleiste, inspiriert durch Darstellungen wie die Facebook Chronik.[48] Sie zeigt Erhebungen personenbezogener Da-

[38] https://selbstauskunft.net/unternehmen, abgerufen am 9. Mai 2017.
[39] https://www.google.com/dashboard.
[40] https://aboutthedata.com/portal.
[41] http://www.a4cloud.eu.
[42] https://www.prime-project.eu.
[43] http://primelife.ercim.eu.
[44] Wästlund/Fischer-Hübner et al. 2010.
[45] Fischer-Hübner/Angulo/Pulls 2014; H. Andersson et al. 2015.
[46] http://hci.cse.kau.se:8000.
[47] Angulo et al. 2015.
[48] https://www.facebook.com/help/250714824948501, abgerufen am 9. Mai 2017.

ten in chronologischer Reihenfolge. Da GenomSynlig Cloud-basiert ist, liegt die Hauptlast für die Herstellung der Transparenz beim Cloudanbieter, also einem Auftragsdatenverarbeiter, statt bei der verantwortlichen Stelle. GenomSynlig liefert nur Informationen, wenn der Betroffene die Quelle der personenbezogenen Daten ist. Das Tool bietet keine Informationen über Datenflüsse innerhalb der verantwortlichen Stelle oder zu Übermittlungen nach außen. Empfänger personenbezogener Daten sind nicht sichtbar. Auch der Zweck kann nicht mit den unterschiedlichen Verarbeitungsschritten in Bezug gesetzt werden. Die in A4Cloud entwickelten Technologien würden zumindest konzeptionell erlauben, auch andere Schritte der Datenverarbeitung zu erfassen. Mit der Accountable PrimeLife Policy Language (A-PPL) ist eine Beschreibungssprache für Logging-Policies vorhanden.[49] Insynd, eine weitere Komponente aus A4Cloud, ist als serverseitiges Transaktionsprotokoll geeignet.[50]

Ein anderes Werkzeug, das einer Auskunftsplattform nahekommt, ist das interaktive Online-Tool, genannt Translucene Map, von Kani-Zabihi und Helmhout.[51] Es visualisiert den Fluss personenbezogener Daten in allgemeiner Form, d.h. auf Grundlage des Verfahrensverzeichnisses, nicht auf Grundlage tatsächlich stattfindender Datenflüsse. Der Betroffene ist in der Lage, den Fluss verschiedener Datenkategorien im dargestellten Graphen hervorzuheben und nachzuverfolgen.

Insgesamt ist es mit keinem der vorgestellten Ansätze möglich, eine vollständige Datenschutzauskunft automatisiert bereitzustellen. Die Anforderungen des Datenschutzes werden nicht systematisch berücksichtigt. Die Struktur der Personal-Data-Provenance, deren konfigurierbare, verteilte Erfassung und die Zusammenführung der Personal-Data-Provenance zum Zeitpunkt eines Auskunftsersuchens sind ungeklärt. Die Datenschutzauskunft und die übrigen Betroffenenrechte sind nicht miteinander verwoben. Die Profilbildungsrisiken automatisierter Auskunftssysteme werden nicht adressiert.

1.2 Zielsetzung

Die vorliegende Arbeit soll die technischen Voraussetzungen für die Wahrnehmung des Rechts auf Auskunft schaffen und das Recht auf Auskunft einer kritischen Würdigung als Teil des datenschutzrechtlichen Transparenzgedankens unterziehen. Insbesondere soll auf das Spannungsfeld zwischen Unverkettbarkeit, der Unmöglichkeit des In-Bezug-Setzens personenbezogener Daten,[52] und Transparenz eingegangen werden.

Technisch soll eine Plattform entwickelt werden, die Datenflüsse transparent macht,

[49] Butin/Chicote/Métayer 2013.
[50] Pulls/Peeters 2015.
[51] Kani-Zabihi/Helmhout 2012.
[52] ULD/TU Dresden 2007, 19 f.

Personal-Data-Provenance bereitstellt, die Interventionsrechte des Betroffenen integriert und gleichzeitig das Verbot der Verkettung und Profilbildung beachtet.

Juristisch sollen die datenschutzrechtlichen Implikationen eines Datenschutzauskunftssystems als zentrale rechtliche Problemlage beleuchtet werden. Dies umfasst sowohl die präzise Herausarbeitung von Anforderungen, als auch deren Bewertung. Grundlage ist eine nähere Auseinandersetzung mit dem datenschutzrechtlichen Auskunftsanspruch und der Einbindung der Transparenz in die datenschutzrechtliche Systematik.

Im Mittelpunkt steht der Betroffene als Auskunftsberechtigter. Seine Entscheidungsfreiheit soll gestärkt, Wahlmöglichkeiten auf Grundlage fundierter Informationen geschaffen werden.

Das Ziel dieser Arbeit ist die Definition, Formalisierung und Implementierung technischer Voraussetzungen für die Wahrnehmung des Rechts auf Auskunft und dessen kritische Würdigung als Teil des datenschutzrechtlichen Transparenzgedankens.

1.2.1 Forschungsfragen

„Das Auskunftsrecht ist für die Betroffenen das fundamentale Datenschutzrecht.“[53] Aus diesem Grund sollen in dieser Arbeit die verfassungsrechtlichen, europarechtlichen und einfachgesetzlichen Bestimmungen, die das Recht auf Auskunft stützen, identifiziert, analysiert und kritisch beleuchtet werden.

Wie ist das Recht auf Auskunft in die Gesamtkonzeption des Datenschutzes eingebunden und welche Anforderungen und Konsequenzen ergeben sich daraus?

- Wie ist das Recht auf Auskunft verfassungsrechtlich und europarechtlich verankert?
- Wie ist das Recht auf Auskunft in das dreistufige Schema aus datenschutzgerechter Verarbeitung, Transparenz und Intervention eingebunden?
- Welche Reichweite und welchen Umfang hat der Auskunftsanspruch?
- Wie können (datenschutz-)rechtliche Anforderungen systematisch abgeleitet und in technische Anforderungen überführt werden?
- Wie sehen die datenschutzrechtlichen Anforderungen an ein Datenschutzauskunftssystem aus?

Das Auskunftsrecht steht nicht alleine. Es ist die Voraussetzung zur Wahrnehmung der übrigen Betroffenenrechte auf Löschung, Sperrung und Berichtigung. Deshalb ergänzen

[53] Neben anderen: Dix in: Simitis, BDSG 2014, § 34 Rn. 1.

sich Provenance-Tracking und Nutzungskontrolle in der Erreichung der durch das Datenschutzrecht gestellten Anforderungen.[54] Eine Herausforderung dieser Arbeit ist der Entwurf einer integrierten Architektur für Provenance-Tracking und Nutzungskontrolle, deren Modellierung und Implementierung in einem Prototypen.

Wie sieht eine generische Architektur und ein Modell für ein integriertes Datenschutzauskunftssystem aus, das die datenschutzrechtlichen Anforderungen erfüllt und technisch umsetzbar ist?

- Wie sieht eine verteilte, gemeinsame Architektur von Provenance-Tracking und Usage-Control aus?
- Wie kann eine nahtlose Kette Erhebung – Verarbeitung – Transparenz – Intervention durch Provenance-Tracking und Usage-Control erreicht werden?
- Wie kann Personal-Data-Provenance datenminimal, dem Anwendungsfall angemessen und hinreichend präzise erfasst und gespeichert werden?
- Wie kann Personal-Data-Provenance so präsentiert und visualisiert werden, dass sie den rechtlichen Anforderungen und den Bedürfnissen der Betroffenen gerecht wird?

Personal-Data-Provenance ermöglicht Transparenz, führt aber gleichzeitig dazu, dass die Menge und Aussagekraft der erhobenen personenbezogenen Daten deutlich steigt. Dies mag zwar dem Auskunftszweck dienlich sein, stellt aber auch selbst einen Eingriff dar. Es werden Datenverknüpfungen hergestellt, die ohne Provenance-Tracking nicht bestünden. Diese Arbeit soll den genannten Konflikt beleuchten und Lösungsansätze unter Einbeziehung des Betroffenen aufzeigen.

Wie lassen sich Transparenz und Unverkettbarkeit, unter Berücksichtigung des individuellen Betroffenen, in Einklang miteinander bringen?

- Wie lässt sich Unverkettbarkeit formal beschreiben?
- Wie kann Unverkettbarkeit analytisch bestimmt und berechnet werden?
- Zu welchem Grad wird die Architektur des entworfenen Datenschutzauskunftssystems dem Ziel der Unverkettbarkeit gerecht?
- Wie kann der Betroffene in den Abwägungsprozess zwischen Transparenz und Unverkettbarkeit einbezogen werden?

Die genannten Forschungsfragen führen, unter Berücksichtigung des Ziels dieser Arbeit und unter Einbeziehung bereits existierender Forschungsansätze, zu den im nächsten Abschnitt aufgeführten Forschungshypothesen und Konstruktionszielen.

[54] Bier 2013b.

1.2.2 Forschungshypothesen und Konstruktionsziele

Die folgenden Forschungshypothesen $\mathfrak{H}$ und Konstruktionsziele $\mathfrak{K}$ strukturieren diese Arbeit. Sie werden in den einzelnen Kapiteln aufgegriffen und korrespondieren mit diesen. Die Hypothesen sind vorab getroffene Annahmen zu wissenschaftlichen Fragestellungen, die im Rahmen dieser Arbeit be- oder widerlegt werden.

$\mathfrak{H}1$ Das Recht auf Auskunft ist im Grundgesetz und in der Charta als unabdingbares Betroffenenrecht verankert.

$\mathfrak{H}2$ Der Betroffene hat das Recht, den Fluss seiner personenbezogenen Daten vollständig nachvollziehen zu können.

$\mathfrak{H}3$ Die anvisierte Implementierung von Provenance-Tracking ist datenminimal und skalierbar.

$\mathfrak{H}4$ Es ist rechtlich zulässig und technisch möglich, den Betroffenen anhand einer Metrik für Unverkettbarkeit informiert entscheiden zu lassen, ob ihm eine umfangreiche Auskunft und das Recht auf Datenübertragbarkeit oder die Unverkettbarkeit seiner personenbezogenen Daten wichtiger sind.

$\mathfrak{H}5$ Eine Datenschutzauskunftsplattform, die Flüsse personenbezogener Daten interaktiv darstellen kann, ist nutzerfreundlicher und verständlicher als bestehende Systeme zur Datenschutzauskunft.

Die Konstruktionsziele legen das gewünschte Ergebnis der systematischen Entwicklung von Modellen, Architekturen und Systemen für das Datenschutzauskunftssystem fest. Die gesetzten Ziele werden von den bereits bestehenden Ansätzen aus Literatur und Praxis nicht erreicht.

$\mathfrak{K}1$ Ein Modell zur systematischen Ableitung technischer Anforderungen an die Implementierung eines Datenschutzauskunftssystems aus den Datenschutz-Schutzzielen und den datenschutzrechtlichen Anforderungen

$\mathfrak{K}2$ Eine verteilte, generische Architektur für ein Datenschutzauskunftssystem, die die Durchsetzung von Transparenz und Intervenierbarkeit in einem gemeinsamen System erlaubt

$\mathfrak{K}3$ Die Erhebung und Speicherung von Personal-Data-Provenance in einem Datenmodell gemäß der datenschutzrechtlichen Anforderungen

$\mathfrak{K}4$ Die verteilte Speicherung und systemübergreifende Erhebung von Personal-Data-Provenance, die im Sinne einer informationellen Gewaltenteilung erst zum Anfragezeitpunkt aggregiert und an den Betroffenen kommuniziert wird

- ℜ5 Eine berechenbare Metrik zur Messung von Unverkettbarkeit in den durch die datenschutzrechtlichen Anforderungen vorgegebenen Dimensionen
- ℜ6 Der Entwurf eines Datenschutzauskunftssystems unter Berücksichtigung aller datenschutzrechtlichen Anforderungen

1.3 Lösungsstrategie

Der interdisziplinäre Charakter dieser Arbeit erfordert je nach Themen- und Fachbereich unterschiedliche Lösungsstrategien für die Bewältigung der aufgeworfenen Herausforderungen. Rechtswissenschaftliches und informationstechnisch-ingenieurwissenschaftliches Arbeiten werden miteinander verwoben.

Die Analyse der verfassungsrechtlichen, europarechtlichen und einfachgesetzlichen Verankerung des Rechts auf Auskunft wird durch die rechtswissenschaftliche Auseinandersetzung mit der maßgeblichen Fachliteratur und Rechtsprechung geprägt. Bestehende Ansätze werden eingeordnet und kritisch reflektiert, Perspektiven werden zusammengeführt und neue Schlussfolgerungen werden gezogen. Reichweite und Umfang des Rechts auf Auskunft werden diskutiert, Anforderungen an ein Datenschutzauskunftssystem werden abgeleitet. Dabei werden die aufgestellten Hypothesen im Blick behalten und kritisch beleuchtet.

Die datenschutzrechtlichen Anforderungen bilden die Brücke zwischen rechtswissenschaftlichem und informationstechnisch-ingenieurwissenschaftlichem Teil. Ein in der rechtswissenschaftlichen Literatur etabliertes Ableitungsmodell wird angepasst und so verfeinert, dass eine durchgängige Ableitungskette von den Grundrechten bis zur Implementierung möglich wird. Die rechtswissenschaftliche Anforderungsanalyse wird mit Ansätzen des Requirements-Engineering in Deckung gebracht. Die sich aus rechtlichen Kriterien ergebenden technischen Anforderungen dienen als Blaupause für die Spezifikation von Datenmodellen, Schnittstellen und den informationstechnischen Architekturentwurf des Datenschutzauskunftssystems.

Die Verbindung von Provenance-Tracking und Usage-Control wird im Modell und anhand eines Prototyps erarbeitet. Eine gemeinsame Architektur wird unter Beachtung bestehender Ansätze in der Data-Provenance- und Usage-Control-Fachliteratur entworfen. Dabei gelten die abgeleiteten Anforderungen als Leitplanke für den Entwurf. Um Usage-Control und Provenance-Tracking zu verbinden, wird das bestehende UC-Informationsflussmodell[55] um ein Provenance-Datenmodell erweitert. Bisherige generische Ansätze zur Modellierung von Data-Provenance[56] werden für den

[55] Pretschner/Lovat/Büchler 2011.
[56] Moreau/Clifford et al. 2011.

Anwendungsfall Datenschutzauskunft instanziiert und unter Einbeziehung des UC-Informationsflussmodells formalisiert.

Personal-Data-Provenance muss datenminimal, dem Anwendungsfall angemessen und hinreichend präzise vorgehalten werden. Problematisch ist insbesondere, dass der Speicherbedarf von Provenance bei naivem Vorgehen superlinear wächst (schlechte Speicherskalierbarkeit). Deshalb werden Ansätze verfolgt, die die Größe der Provenance im Verhältnis zur Menge der in jedem Schritt verarbeiteten personenbezogenen Daten maximal linear und im Verhältnis zur Menge der Operationen sublinear steigen lassen. Dies wird anhand eines Prototyps für das Betriebssystem Windows 7 evaluiert.

Personal-Data-Provenance muss über Systemgrenzen hinweg aufgezeichnet und zum Abfragezeitpunkt aggregiert werden. Deshalb wird das kombinierte Informationsfluss- und Provenancemodell zum systemübergreifenden Tracking ertüchtigt. Zu diesem Zweck wird das schichtenübergreifende Informationsflussmodel von Lovat[57] unter Berücksichtigung der Überlegungen von Kelbert[58] als generisches, systemübergreifendes, schichtenunabhängiges Tracking erweitert. Das Modell wird einer Konfiguration durch eine semantische Spezifikation zugänglich gemacht.

Der Zielkonflikt zwischen Transparenz und Unverkettbarkeit erfordert die Entwicklung einer aussagekräftigen Metrik für Unverkettbarkeit, die dem feststehenden Zuwachs an Transparenz gegenübergestellt werden kann. Bestehende Ansätze für die Bestimmung des Grads der Unverkettbarkeit werden bewertet und ihre Defizite herausgearbeitet. Eine informationstheoretische Metrik wird anhand der datenschutzrechtlichen Anforderungen in vier Dimensionen instanziiert. Für jede Dimension wird die Bestimmung des Messwertes unter den gegebenen Annahmen hergeleitet. Ein prototypischer Simulator und ein implementierter Algorithmus zeigen die Berechenbarkeit der Metrik.

Ein Datenschutzauskunftssystem ist nur so gut wie die Datenschutzauskunftsplattform, die die Informationen den Betroffenen zugänglich macht. Ein Prototyp für eine Datenschutzauskunftsplattform wird anhand von Überlegungen zur Nutzerfreundlichkeit erarbeitet und an die datenschutzrechtlichen Anforderungen angepasst. Der webbasierte Demonstrator wird gegen die neueste Datenschutzauskunftsplattform des Projektes A4Cloud[59] evaluiert. Reaktionszeiten, Erfolgsquote und Nutzerverhalten werden in einem Blickmesslabor untersucht. Eine ergänzende Nutzerbefragung bietet ein ganzheitliches Bild auf die Nutzerrezeption des Datenschutzauskunftssystems.

[57] Lovat 2015.
[58] Kelbert/Pretschner 2015.
[59] H. Andersson et al. 2015.

1.4 Wissenschaftlicher Beitrag

Diese Arbeit befasst sich erstmalig sowohl rechtlich als auch technisch mit dem Recht auf Auskunft und dessen Implikationen. Sie erweitert den bisherigen Stand der Wissenschaft und Forschung um die folgenden neuen Beiträge:

Eine Einordnung des Rechts auf Auskunft im Hinblick auf die automatisierte Beantwortung von Auskunftsersuchen. Es wird herausgearbeitet, ob und inwiefern das Recht auf Auskunft im Grundgesetz (Kapitel 2.1) sowie in den Verträgen[60] (Kapitel 2.2) verankert ist. Die verfassungsrechtliche Stellung des Rechts auf Auskunft wurde in der Literatur bisher nicht systematisch erörtert. In der bis dato oberflächlichen Betrachtung ist sie strittig.[61] Diese Arbeit positioniert sich gegen ein Recht auf Auskunft und hebt die Bedeutung des europäischen Rechts als relativierende Größe hervor.

Das Recht auf Auskunft wird in dieser Arbeit systematisch in den Kontext des Datenschutzes und die übrigen Transparenzrechte eingeordnet (Kapitel 3.1). Die Betrachtungen zum Umfang des Auskunftsrechts (Kapitel 3.7) machen deutlich, dass für eine vollständige Auskunft eine nahtlose Nachvollziehbarkeit der Verarbeitungswege personenbezogener Daten erforderlich ist. Eine solche ist letztendlich nur automatisiert zu gewährleisten. Form und Inhalt des Auskunftsersuchens (Kapitel 3.3), die Identifizierung des Auskunftsberechtigten (Kapitel 3.4) und die Art und Weise der Auskunftserteilung (Kapitel 3.5) werden deshalb im Hinblick auf eine automatisierte Auskunftserteilung analysiert.[62]

Für die Ableitung der technischen Anforderungen wird ein neues Vorgehensmodell vorgestellt (Kapitel 4.1.2). Es erweitert die Methode KORA,[63] deren Ableitung auf rechtlicher Ebene stehen bleibt. Anhand des Datenschutzauskunftssystems wird gezeigt, wie technische Anforderungen abgeleitet und in die Implementierung übertragen werden können (Kapitel 4.3).

Das erste verteilte, datenzentrierte, integrierte und betroffenenfokussierte Datenschutzauskunftssystem Die bereits vorhandenen Arbeiten zu Personal-Data-Provenance werden in Kapitel 5.1 diskutiert. Die vorliegende Arbeit geht in diesem Themenbereich insbesondere in den folgenden Punkten über den Stand der Technik hinaus:

Das vorgestellte Datenschutzauskunftssystem ist *verteilt*, da es ein schichtenunabhängiges, semantisch konfigurierbares[64] Provenance-Tracking über Systemgrenzen hinweg zulässt (Kapitel 8.1) und erst zum Anfragezeitpunkt eine Aggregation der Personal-Data-Provenance erfordert (Kapitel 8.2). Das Protokollsystem Insynd speichert die Personal-

[60] EUV und AEUV.
[61] Mallmann in: Simitis, BDSG 2014, § 19 Fn. 1.
[62] Bier, DuD 2015, 741.
[63] Vgl. Kapitel 4.1.1.
[64] Birnstill/Bier et al. 2016.

Data-Provenance dagegen auf einem zentralen Logserver und überlässt die Aggregation dem Betroffenen.[65] Kelbert stellt zwar ein verteiltes Informationsflussmodell vor,[66] leitet daraus jedoch keine Provenance ab. Auch sind die Modelle beider Ansätze nicht durch eine semantische Spezifikation konfigurierbar. Eine solche Konfigurierbarkeit ermöglicht jedoch die Integration von datenverarbeitenden Systemen ins Provenance-Tracking zur Laufzeit. Dies ist in der Anwendung unabdingbar.

Es ist *datenzentriert*, da das Informationsfluss- und Provenancemodell eine Trennung von Daten und Container vorsieht (Kapitel 6). Die Personal-Data-Provenance wird pro Datum statt pro Ereignis erfasst und lässt sich so individuell und datenminimal ablegen (Kapitel 7). Bereits das von Harvan,[67] Pretschner,[68] Lovat[69] und Kelbert[70] entwickelte und erweiterte Informationsflussmodell ist datenzentriert. Für sich alleine ist es allerdings kein Provenance-Modell. Es wird in dieser Arbeit erstmalig mit einem Provenance-Modell verknüpft und für den Zweck der Datenschutzauskunft verwendet.

Das Datenschutzauskunftssystem ist *integriert*, da es sowohl die Bereitstellung der Datenschutzauskunft als auch die Durchsetzung der übrigen Betroffenenrechte über Usage-Control in einer gemeinsamen Architektur (Kapitel 5) und einem verbundenen Modell (Kapitel 6) zulässt. Das existierende GenomSynlig sieht zwar einen Löschbutton vor, dessen Umsetzung ist jedoch offen.

Die Auskunftsplattform des Datenschutzauskunftssystems ist *betroffenenfokussiert*, da sie die interaktive und schrittweise Wahrnehmung des Auskunftsrechts und der Interventionsrechte zulässt (Kapitel 12).[71] Das Recht auf Datenübertragbarkeit und die Entscheidungsfreiheit des Betroffenen sind mitberücksichtigt (Kapitel 10). Das einzige bereits existierende System, GenomSynlig, ermöglicht es dem Betroffenen nicht, die Abfolge von Datenverarbeitungsschritten nachzuvollziehen.

Eine generische, instanziierbare und berechenbare Metrik für Unverkettbarkeit Im Konflikt der Datenschutzziele Transparenz und Unverkettbarkeit ist eine Metrik für Unverkettbarkeit eine Hilfestellung für den Betroffenen.[72] Unverkettbarkeitsmetriken wurden in der Literatur bisher auf homogene Relationen oder sogar Äquivalenzrelationen beschränkt. Personenbezogene Daten können jedoch in Relation zu mehreren anderen Entitäten, beispielsweise der Herkunft und dem Empfänger, stehen. Diese Arbeit erweitert bestehende Konzepte für informationstheoretische Unverkettbarkeitsmetriken auf belie-

[65] Pulls/Peeters 2015.
[66] Kelbert/Pretschner 2015.
[67] Harvan/Pretschner 2009.
[68] Pretschner/Lovat/Büchler 2011.
[69] Lovat/Kelbert 2014.
[70] Kelbert/Pretschner 2015.
[71] Bier/Kühne/Beyerer 2016.
[72] Bier 2016.

bige Verkettungsrelationen (Kapitel 9.3). Das Schema wird für vier Relationen mit Bezug zum entwickelten Datenschutzauskunftssystem instanziiert (Kapitel 9.7).

Des Weiteren ist in der Literatur das Hintergrundwissen eines Angreifers in den Metriken bisher vollständig unberücksichtigt. Folge ist, dass die tatsächliche Wirkung der Einführung eines Datenschutzauskunftssystems verzerrt wird. In dieser Arbeit wird deshalb das A-priori- und A-posteriori-Wissen eines modellierten Angreifers formalisiert (Kapitel 9.8) und in die Bestimmung des Grads der Unverkettbarkeit miteinbezogen.

Darüber hinaus wurde für keine der in der Literatur existierenden Metriken die tatsächliche Berechenbarkeit belegt. Die Komplexität der vollständigen Berechnung der Wahrscheinlichkeiten aller möglichen Kandidatenrelationen liegt allerdings in $\mathcal{O}(2^x)$. Deshalb wird in dieser Arbeit ein heuristischen Algorithmus zur Berechnung der Metrik hergeleitet und dessen Anwendbarkeit demonstriert (Kapitel 9.9). Die Metrik wird als Grundlage für eine informierte Entscheidung des Betroffenen bezüglich des Provenance-Trackings vorgeschlagen (Kapitel 10).

Eine vertiefte Auseinandersetzung mit verwandten Arbeiten findet sich am Ende dieser Arbeit in Kapitel 13.2.1.

1.5 Vorveröffentlichungen

Eine vollständige Liste aller Veröffentlichungen des Autors findet sich am Ende der Arbeit. Die hier referenzierten Veröffentlichungen sind direkt in diese Arbeit eingeflossen.

Die in Abschnitt 1.1.1 erwähnte und im Anhang A ausgeführte Studie zum Status Quo des Auskunftsanspruchs wurde gemeinsam mit Kömpf durchgeführt und veröffentlicht.[73] Eine Zusammenfassung des Kapitels 3 wurde bereits 2015 publiziert.[74] Die im Kapitel 5 vorgestellte Integration von Usage-Control- und Provenance-Architektur wurde bereits 2013 vorgeschlagen.[75]

Die Modelle und Verfahren der Kapitel 6.2.1 und 8.1.1 sind in Zusammenarbeit mit Birnstill entstanden.[76] Gegenüber dem veröffentlichten Stand von 2016 wurden die Modelle in dieser Arbeit um abstrakte Container erweitert. Die Verfahrensbeschreibung in Kapitel 8.1.2 berücksichtigt zusätzlich den Aufbau der Provenance.

Das Kapitel 9 zur Unverkettbarkeit basiert auf den auf der Jahrestagung Informatik 2016 vorgestellten Arbeiten.[77] Die Nutzerstudie des Kapitels 12 wurde gemeinsam mit Kühne durchgeführt und veröffentlicht.[78]

[73] Bier/Kömpf/Beyerer 2017.
[74] Bier, DuD 2015, 741.
[75] Bier 2013b.
[76] Birnstill/Bier et al. 2016.
[77] Bier 2016.
[78] Bier/Kühne/Beyerer 2016.

1.6 Fortlaufendes Beispiel

Die in dieser Arbeit vorgestellten Konzepte werden anhand eines durchgängigen Szenarios erläutert. Das Szenario ist typisch für Auskunftsersuchen und deckt die wesentlichen rechtlichen und technischen Aspekte ab, die in dieser Arbeit eingeführt werden.

Das fortlaufende Beispiel nimmt ein Unternehmen namens AdBokis Buchclub GmbH, einen fiktiven Online-Händler für Bücher und Software, in den Fokus. Alice Fox ist Kundin dieses Händlers. Sie hat im Onlineshop eine Banking-Software gekauft und sich dafür ein Benutzerkonto angelegt. Mit der Bestellbestätigung erhält Alice eine Einladung zu einem Incentive-Programm. Im Austausch gegen einen Rabatt willigt Alice darin ein, in Zukunft über interessante Produkte von AdBokis und Partnern informiert zu werden. Im weiteren Verlauf hat Alice eine Frage an den Support von AdBokis. In diesem Rahmen übermittelt sie weitere personenbezogene Daten.

Etwas später erhält Alice per E-Mail einen Newsletter von einer Bonus Card GmbH. Sie möchte jetzt wissen, wie ihre Daten an das Unternehmen gekommen sind. Da sie von der Bonus Card GmbH keine Antwort erhält erkundigt sie sich bei AdBokis. Nur dort hatte sie die für den Newsletter genutzte E-Mailadresse zur Registrierung verwendet.

1.7 Inhaltsübersicht

Kapitel 1: Einleitung

Im vorangegangenen ersten Kapitel wurde der Handlungsbedarf für die Entwicklung eines integrierten Datenschutzauskunftssystems herausgearbeitet. Insbesondere wurde eine eigens erstellte Studie zum Status quo des Rechts auf Auskunft vorgestellt. Anschließend wurde die Zielsetzung der Arbeit mit den zugrundeliegenden Forschungsfragen und Forschungshypothesen erläutert. Anknüpfend wurden die in der Arbeit verwendeten Lösungsstrategien zur Erarbeitung der wissenschaftlichen Beiträge dargelegt. Zuletzt wurde das in dieser Arbeit verwendete fortlaufende Beispiel eingeführt.

Teil I: Rechtliche Grundlagen und Anforderungen

Kapitel 2: Verfassungs- und europarechtliche Grundlagen

Im zweiten Kapitel wird die Verankerung des Rechts auf Auskunft im Grundgesetz und in den Verträgen herausgearbeitet. Die dem Recht auf Auskunft entgegenstehenden Grundrechte und die aktuelle Rechtsentwicklung in Europa werden mitberücksichtigt.

Kapitel 3: Einfachgesetzliche Verankerung des Rechts auf Auskunft

Das Kapitel 3 spannt den Bogen von der Struktur des Datenschutzes über Form, Art und Inhalt eines Auskunftsbegehrens bis zum Umfang der Auskunft selbst. Dabei wird ein besonderer Schwerpunkt auf die bei der Automatisierung der Auskunft zu berücksichtigenden Aspekte gelegt.

Kapitel 4: Datenschutzrechtliche Anforderungen an ein Datenschutzauskunftssystem

Im Kapitel 4 wird ein Vorgehensmodell zur Ableitung technischer Anforderungen, ausgehend von den Grundrechten, vorgestellt. Dieses Vorgehensmodell wird anschließend auf die Anforderungen an ein Datenschutzauskunftssystem angewendet.

Teil II: Personal-Data-Provenance und Data-Usage-Control

Kapitel 5: Usage-Control & Provenance-Tracking: Einführung und Architektur

Im Kapitel 5 werden die Konzepte von Usage-Control & Provenance-Tracking eingeführt und diskutiert. Ausgehend von existierenden Forschungsansätzen wird entlang der rechtlichen Anforderungen eine integrierte, generische Architektur erarbeitet.

Kapitel 6: Ein gemeinsames Modell für IF- & Provenance-Tracking

Aufbauend auf dem vorherigen Kapitel wird im Kapitel 6 ausgehend von bestehenden Informationsfluss- und Provenancemodellen ein gemeinsames Informationsfluss- und Provenancemodell entwickelt. Aspekte der Implementierung, insbesondere der Personal-Data-Provenance-Datenhaltung werden angerissen.

Kapitel 7: Datenschutzgerechtes und skalierbares Provenance-Tracking

Im Kapitel 7 wird der Mechanismus zur Erhebung und Speicherung von Personal-Data-Provenance verfeinert, um zu einer datenminimalen und skalierbaren Umsetzung zu kommen. Ein wesentlicher Aspekt sind Abstraktionsregeln. Die vorgestellten Konzepte werden im Rahmen von Performance-Tests evaluiert.

Kapitel 8: Verteiltes Provenance-Tracking

Provenance-Tracking ist nicht auf ein lokales System beschränkt, sondern muss systemübergreifend erfolgen. Im Kapitel 8 wird ein Konzept für ein schichtunabhängiges, semantisch konfigurierbares, systemübergreifendes Provenance-Tracking vorgestellt. Ergänzend wird die Datenaggregation zum Auskunftszeitpunkt behandelt.

Teil III: Abwägung zwischen Transparenz und Unverkettbarkeit

Kapitel 9: Eine Metrik für Unverkettbarkeit

Im Kapitel 9 wird auf den Zielkonflikt zwischen Transparenz und Unverkettbarkeit eingegangen. Es wird eine informationstheoretische Metrik für Unverkettbarkeit definiert und instanziiert. Ein Verfahren zur Berechung der Metrik wird vorgestellt.

Kapitel 10: Betroffenenautonomie zwischen Transparenz und Unverkettbarkeit

Die Unverkettbarkeitsmetrik gibt dem Betroffenen ein Hilfsmittel in die Hand, um selbst abzuwägen, welches Datenschutz-Schutzziel ihm wichtiger ist. Im Kapitel 10 wird die Einbindung dieser Wahlmöglichkeit in die Auskunftsplattform diskutiert und evaluiert. Konsequenzen für das Recht auf Datenübertragbarkeit werden hervorgehoben.

Teil IV: Bewertung und Ausblick

Kapitel 11: Rechtliche Bewertung des Datenschutzauskunftssystems

Im Kapitel 11 wird das Datenschutzauskunftssystem einer abschließenden rechtlichen Bewertung auf der Grundlage der in Kapitel 2 bis 4 hergeleiteten Anforderungen unterzogen. Die Konsequenzen einer fehlerhaften Auskunft werden diskutiert.

Kapitel 12: Nutzerrezeption des Datenschutzauskunftssystems

Das Kapitel 12 stellt eine Nutzerstudie vor, in der die Angemessenheit und Nutzerfreundlichkeit der Auskunftsplattform überprüft wurde. Dazu werden konkurrierende Ansätze eingeführt und mit dem eigenen Auskunftssystem verglichen.

Kapitel 13: Zusammenfassung und Ausblick

Im letzten Kapitel werden die Ergebnisse dieser Arbeit zusammengefasst und bewertet. Die Grenzen der Arbeit werden benannt. Zuallerletzt wird ein Ausblick auf zukünftige rechtliche und technische Entwicklungen gegeben.

Teil I

Rechtliche Grundlagen und Anforderungen

2 Verfassungs- und europarechtliche Grundlagen

Das Verfassungs- und europarechtliche Gefüge aus GG, EMRK, Charta und Verträgen bildet den Rahmen für die Interpretation und Auslegung einfachen Rechts. Insbesondere in Abwägungsgrenzfällen ist es wichtig, miteinzubeziehen, welche Rechtspositionen sich auf welche Grundrechte berufen können.

Der Grundrechtscharakter des Rechts auf Auskunft ist in der Literatur bisher strittig.[1] Deshalb wird in diesem Kapitel schwerpunktmäßig behandelt, ob das Recht auf Auskunft teil des Grundrechts auf informationelle Selbstbestimmung (Abschnitt 2.1) und der europäischen Grundrechte (Abschnitt 2.2) ist. Detaillierte europarechtliche Vorgaben, wie sie sich in der DSRL und der künftig geltenden DSGVO finden, werden nicht in diesem Kapitel, sondern im jeweiligen Kontext in den Kapiteln 3 und 10 behandelt.

2.1 Verfassungsrechtliche Rahmensituation

2.1.1 Das Grundrecht auf informationelle Selbstbestimmung

Schutzgegenstand des Datenschutzrechts ist das allgemeine Persönlichkeitsrecht[2] des Einzelnen in seiner Ausprägung als Recht auf informationelle Selbstbestimmung aus Art. 2 I i. V. m. Art. 1 I GG.[3]

Das Recht auf informationelle Selbstbestimmung steht allen lebenden natürlichen Personen unabhängig von ihrer Staatsangehörigkeit zu, soweit sie unmittelbar betroffen sind.[4] Es ist Teil der Informations- und Kommunikationsverfassung des Grundgesetzes[5] und Auffanggrundrecht neben den spezielleren Freiheitsrechten der Art. 10, 13 GG.[6] Es steht

[1] Mallmann in: Simitis, BDSG 2014, § 19 Fn. 1; für ein Auskunftsrecht als Bestandteil des Rechts auf informationelle Selbstbestimmung Di Fabio in: Maunz/Düring, GG 2016, Art. 2 Abs. 1 Rn. 178; Rudolf in: Merten/Papier, HGR IV 2011, § 90 Rn. 49; Gola/Schomerus, BDSG 2015, § 19 Rn. 2; Dix in: Simitis, BDSG 2014, § 34 Rn. 2; Weichert, NVwZ 2007, 1004 (1005).

[2] BVerfGE 54, 148 (153).

[3] BVerfGE 65, 1 (41).

[4] Rudolf in: Merten/Papier, HGR IV 2011, § 90 Rn. 36.

[5] Gallwas, NJW 1992, 2785 (2785).

[6] BVerfGE 67, 157 (171); BVerfGE 100, 313 (358); BVerfGE 118, 168 (183). Umstände der Telekommunikation, beispielsweise Verkehrs- und Verbindungsdaten, befinden sich im Schutzbereich des Art. 10 GG, soweit

neben anderen Schutzrechten wie dem Schutz der Privatsphäre, dem Recht am eigenen Bild und dem Recht am eigenen Wort.[7] In seinem Bezug auf die Menschenwürde[8] geht es über den Schutzbereich des Art. 2 Abs. 1 GG hinaus.[9]

In seiner subjektiven Ausprägung gewährleistet das Grundrecht „die Befugnis des Einzelnen, grundsätzlich selbst über die Preisgabe und Verwendung seiner persönlichen Daten zu bestimmen."[10] Objektivrechtlich schützt es die freiheitlich demokratische Kommunikationsverfassung. Eine Gesellschaftsordnung, „in der Bürger nicht mehr wissen können, wer was wann und bei welcher Gelegenheit über sie weiß" ist mit dem Recht auf informationelle Selbstbestimmung unvereinbar.[11] Bürger, die dauerhaft versuchen, nicht durch abweichende Verhaltensweisen aufzufallen, können einen demokratischen pluralistischen Rechtsstaat wie die Bundesrepublik Deutschland nicht stützen. „Selbstbestimmung [ist] eine elementare Funktionsbedingung eines auf Handlungsfähigkeit und Mitwirkungsfähigkeit seiner Bürger begründeten freiheitlich-demokratischen Gemeinwesens."[12]

Der Einzelne hat jedoch kein absolutes, eigentumsähnliches Herrschaftsrecht über seine personenbezogenen Daten. Das Grundgesetz folgt einem Menschenbild, das den Menschen nicht allein als isoliertes Individuum betrachtet, sondern ihn in seiner ganzen Gemeinschaftsbezogenheit als soziales Wesen erfasst.[13] Kommunikationsbeziehungen, veröffentlichte und ausgetauschte Informationen sind ein Abbild der sozialen Realität und bilden mannigfache Bezüge.[14]

2.1.2 Drittwirkung der Grundrechte

Die Grundrechte binden den Staat unmittelbar. Sie sind „Abwehrrechte des Bürgers gegen den Staat".[15] Aufgrund der „Asymmetrie des Rechtsstaats"[16] gilt diese Bindung für den Bürger oder privatwirtschaftliche Unternehmen nicht. Art. 1 Abs. 3 GG trifft dazu eine elementare Unterscheidung.[17] „Private stehen sich in Freiheit gegenüber."[18]

sie nicht im Herrschaftsbereich eines Kommunikationsteilnehmers gespeichert sind. – BVerfGE 115, 166 (166); BVerfGE 124, 43 (54).

[7] BVerfGE 34, 238 (246).
[8] BVerfGE 30, 1 (25 f.).
[9] BVerfGE 35, 202 (220); BVerfGE 109, 279 (313).
[10] BVerfGE 65, 1 (43).
[11] Ebd.
[12] Ebd.
[13] BVerfGE 4, 7 (15 f.); BVerfGE 35, 202 (220).
[14] BVerfGE 65, 1 (43 f.).
[15] BVerfGE 7, 198 (204).
[16] Masing, NJW 2012, 2305 (2306).
[17] BVerfGE 128, 226 (244).
[18] Masing, NJW 2012, 2305 (2307).

Dennoch ist das Recht auf informationelle Selbstbestimmung Teil der Wertordnung des Grundgesetzes und damit objektives Prinzip für die Gestaltung des Privatrechts.[19] Daraus ergeben sich Schutzpflichten für die staatlichen Organe gegenüber dem Bürger in seinem Verhältnis zu privaten Dritten.[20] Das Recht auf informationelle Selbstbestimmung entfaltet damit mittelbar auch seine Wirkung für das Verhältnis Privater untereinander.[21]

Dies bedeutet jedoch nicht, dass für das Verhältnis Privater die selben Standards gelten wie im öffentlichen Bereich.[22] Kollidierende Grundrechte sind in ihrer Wechselwirkung so zu begrenzen, dass sie in praktischer Konkordanz nebeneinander bestehen können.[23] Allerdings können je nach Intensität des Machtungleichgewichts die an den Gesetzgeber gestellten Anforderungen für die Regelung einer Datenverarbeitung durch Private genauso streng sein wie die Anforderungen, die durch die Grundrechte an staatliche Akteure gerichtet sind.[24]

2.1.3 Verankerung des Rechts auf Auskunft im Grundgesetz

Ein Recht auf Auskunft kann sich nicht auf auf die Informationsfreiheit nach Art. 5 Abs. 1 S. 1, 2. Hs. GG beziehen, sondern nur aus dem Recht auf informationelle Selbstbestimmung speisen.

Ob das Auskunftsrecht Bestandteil des Rechts auf informationelle Selbstbestimmung ist, ist in der Literatur strittig.[25]

Das Bundesverfassungsgericht (BVerfG) hat die Frage nach einem Leistungsrecht auf Auskunft selbst lange, insbesondere im Bezug auf das Volkszählungsurteil, explizit offen gelassen. Die dortigen Nennungen von Auskunfts-, Aufklärungs- und Löschpflichten[26] stehen exemplarisch für mögliche verfahrensrechtliche Sicherungen.[27]

Unstrittig war schon immer, dass ein fehlender Zugang zu gespeicherten personenbezogenen Daten das Verhältnismäßigkeitsprinzip als Teil des Rechts auf informationelle Selbstbestimmung berühren kann, unabhängig davon, ob es ein eigenes Zugangsrecht gibt oder nicht.[28] Verfahrensgarantien wird ein hoher Stellenwert eingeräumt.[29]

[19] BVerfGE 7, 198 (198); BVerfGE 81, 242 (254).
[20] BVerfGE 117, 202 (227 ff.
[21] BVerfGE 7, 198 (205); BVerfGE 84, 192 (194 f.).
[22] Rudolf in: Merten/Papier, HGR IV 2011, § 90 Rn. 29.
[23] BVerfGE 89, 214 (232).
[24] BVerfGE 81, 242 (255); Masing, NJW 2012, 2305 (2308).
[25] Mallmann in: Simitis, BDSG 2014, § 19 Fn. 1; für ein Auskunftsrecht als Bestandteil des Rechts auf informationelle Selbstbestimmung Di Fabio in: Maunz/Düring, GG 2016, Art. 2 Abs. 1 Rn. 178; Rudolf in: Merten/Papier, HGR IV 2011, § 90 Rn. 49; Gola/Schomerus, BDSG 2015, § 19 Rn. 2; Dix in: Simitis, BDSG 2014, § 34 Rn. 2; Weichert, NVwZ 2007, 1004 (1005).
[26] BVerfGE 65, 1 (46); BVerfGE 113, 29 (58).
[27] BVerfG, NVwZ 2001, 185 (185); BVerfG, NJW 2006, 1116 (1117) und dem folgend BVerwGE 130, 29 (36).
[28] BVerfG, NJW 2006, 1116 (1117).
[29] BVerfGE 113, 29 (57 f.).

Darüber hinaus ist der Anspruch auf Kenntnis ein in sich eigenes und spezifisches Datenschutzrecht.[30]

Erst 2008 hat das Bundesverfassungsgericht klargestellt, dass das Grundgesetz dem Gesetzgeber nicht vorgibt, wie die Möglichkeit der Kenntnisnahme auszugestalten ist. „Das Grundrecht auf informationelle Selbstbestimmung gewährt [..] keinen Anspruch auf eine bestimmte Art der Informationserlangung."[31]

Die Prüfung der Verhältnismäßigkeit eines Eingriffs kommt bei heimlichen Datenerhebungen eher zu dem Ergebnis, dass eine Benachrichtigung geboten ist.[32] Besteht eine solche Verpflichtung nicht und sind Eingriffe nicht klar abzuschätzen, kommt dem, durch den Betroffenen initiativ wahrzunehmenden, Auskunftsrecht eine zentrale Bedeutung zu.[33] Umgekehrt ist eine Benachrichtigung immer dann erforderlich, wenn Daten heimlich erhoben werden und Auskunftsansprüche nicht eingeräumt werden.[34]

Ein wirksamer Anspruch muss sich auf alle Erhebungs- und Verarbeitungsprozesse und jede Nutzung erstrecken.[35] Eine reine Benachrichtigung zum Erhebungszeitpunkt ist deshalb verfassungsrechtlich nicht ausreichend. Wird auf ein Auskunftsrecht verzichtet, ist der Betroffene wiederholt zu benachrichtigen.

In summa kann festgestellt werden, dass es kein verfassungsrechtlich verankertes Recht auf Auskunft gibt. Verfassungsrechtlich geschützt ist dagegen die Kenntnisnahme in einer offenen Ausprägung.

Der Gesetzgeber hat jedoch u. a. im § 19 Abs. 1 und § 34 Abs. 1 BDSG entschieden, Transparenz durch einen Auskunftsanspruch herzustellen. Der Gesetzgeber überträgt durch seine Rechtsetzung die an ihn verfassungsmäßig gestellten Anforderungen auf das Verhältnis zwischen Privaten bzw. Bürgern. Die Prüfung und Interessenabwägung der verantwortlichen Stelle ist dementsprechend entlang ähnlicher Maßstäbe durchzuführen, wie sie auch für die exekutive Gewalt gelten.

Der Auskunftsanspruch ist „Ausdruck demokratischer und rechtsstaatlicher Grundvorstellungen." Er „wirkt der Gefahr entgegen, dass sich der Bürger unkontrollierbarem Wissen und damit unkontrollierbarer Macht ausgeliefert sieht und deshalb [...] auf die Ausübung von Grundrechten verzichtet."[36]

Im Verhältnis zwischen den Bürgern soll durch das Datenschutzrecht gleichfalls ein Machtungleichgewicht reduziert werden. Aus diesem Grunde gilt das BDSG nicht für die

[30] BVerfGE 100, 313 (361).

[31] BVerfGE 120, 351 (363); bereits zu Art. 10 GG in BVerfGE 100, 313 (361) und zu Art. 13 GG in BVerfGE 109, 279 (363); abweichende Interpretation Gola/Schomerus, BDSG 2015, § 19 Rn. 2, die das Recht auf Kenntnisnahme mit dem Recht auf Auskunft gleichsetzen.

[32] BVerfGE 120, 351 (363).

[33] BVerfGE 120, 351 (364); BVerfG, Urteil des Ersten Senats vom 20. April 2016 – 1 BvR 966/09 Rn. 137, http://www.bverfg.de/e/rs20160420_1bvr096609.html, abgerufen am 9. Mai 2017.

[34] BVerfGE 109, 279 (363 f.).

[35] BVerfGE 100, 313 (361).

[36] Mallmann in: Simitis, BDSG 2014, § 19 Rn. 3.

Erhebung, Verarbeitung und Nutzung personenbezogener Daten im privaten, höchstpersönlichen Bereich. Für diesen Bereich sind der Unterlassungsanspruch, das Kunsturhebergesetz und strafrechtliche Sanktionen im Rahmen der üblen Nachrede relevant.

Das Auskunftsrecht trägt der Gewährung effektiven Rechtsschutzes aus Art. 19 Abs. 4 GG Rechnung.[37] Es schafft die Voraussetzung dafür, dass der Betroffene Kenntnis von unrechtmäßiger Datenerhebung und -verwendung erlangen kann, und ist damit die Grundlage für die Beschwerde bei der Aufsichtsbehörde und die Anrufung der zuständigen Gerichte. Die Möglichkeit, von Eingriffen in das Recht auf informationelle Selbstbestimmung zu erfahren, bietet Orientierung und Erwartungssicherheit. Der Anspruch ist jedoch nicht auf den gerichtlichen Rechtsschutz nach Art. 19 Abs. 4 GG verengt.[38] Informationsmöglichkeiten für den Betroffenen sind Voraussetzung dafür, dass er seine übrigen Betroffenenrechte geltend machen kann.[39]

Das durch Art. 2 Abs. 1 i. V. m. Art. 1 Abs. 1 GG gewährleistete allgemeine Persönlichkeitsrecht schützt den selbst definierten sozialen Geltungsanspruch eines jeden einzelnen. Es berechtigt zur Korrektur von Äußerungen, die dem einzelnen unrechtmäßig in den Mund gelegt werden.[40] Damit ist das allgemeine Persönlichkeitsrecht auch ein möglicher Ausgangspunkt für den datenschutzrechtlichen Löschanspruch.

Adäquate technische und organisatorische Vorkehrungen, um im Fall einer Betroffenenanfrage Auskunft geben zu können, sind Erfordernisse eines effektiven Grundrechtsschutzes.[41] Daraus lässt sich ein verfassungsrechtlicher Auftrag herleiten, wirksame technische Systeme zur Auskunftserteilung zu entwickeln.

2.1.4 Schranken

Beschränkungen des Rechts auf Kenntnisnahme müssen gegenläufigen Interessen dienen, die von größerem Gewicht sind als das Recht selbst.[42] Sie unterliegen den gleichen Voraussetzungen wie Beschränkungen des Rechts auf informationelle Selbstbestimmung.[43]

2.1.5 Entgegenstehende Grundrechte anderer Personen

In Frage kommende Grundrechte

Das Sammeln und die Verwendung personenbezogener Daten ist für Private (nichtöffentliche Stellen) durch Art. 2 Abs. 1 GG (Allgemeine Handlungsfreiheit), Art. 5 Abs.

[37] BVerfGE 65, 1 (70); BVerfGE 120, 351 (362).
[38] BVerfGE 100, 313 (361).
[39] BVerfGE 120, 351 (362 f.).
[40] BVerfG, NJW 1980, 2070.
[41] BVerfGE 35, 79 (120); BVerfGE 56, 216 (238).
[42] BVerfGE 120, 351 (365); BVerfGE 133, 277 (368).
[43] OVG Bremen, NJW 1987, 2393 (2394); analog zu Art. 13 GG: BVerfGE 109, 279 (364).

1 S. 1, 2. Hs. (Informationsfreiheit), aber auch durch Art. 4 (Religions- und Gewissensfreiheit), Art. 12 (Berufsfreiheit) und Art. 14 (Eigentumsfreiheit) geschützt.[44] Natürliche Personen haben als Teil des Rechts auf informationelle Selbstbestimmung ein Recht zur Verwendung von Daten mit mehrfachem Personenbezug. Juristische Personen können sich nicht auf das Recht auf informationelle Selbstbestimmung berufen.[45] Aus dem Recht auf Verwendung ergibt sich jedoch noch kein Recht auf Zugang zu fremden personenbezogenen Daten (Erhebung).[46] Ein Recht auf Übermittlung wird ergänzend durch die Meinungsfreiheit gestützt. Sie hat einschränkenden Charakter gegenüber dem Recht auf informationelle Selbstbestimmung.[47]

In den genannten Grundrechtspositionen Dritter beziehungsweise der verantwortlichen Stelle, soweit sie grundrechtsberechtigt sind, findet das Recht auf Kenntnisnahme seine Grenzen.[48]

Dem allgemeinen Persönlichkeitsrecht entspringt ein personenbezogenes Abwägungsgebot zwischen Abwehrinteresse und Beeinträchtigungsinteresse.[49] Durch die Abwägung kann es in jedem Schritt wieder zu neuen Abwehrrechten und Ansprüchen des Grundrechtsberechtigten gegenüber dem Grundrechtsverpflichteten kommen.[50] Der Anspruch auf Auskunft ist solch ein nachgelagerter Gegenstand.

Die einfachgesetzlichen Regelungen zur Datenverarbeitung Privater (z. B. § 28 ff. BDSG) haben zum Ziel, die mehrseitigen Freiheitsrechte in Ausgleich zu bringen.

Betriebs- und Geschäftsgeheimnisse

Die Erteilung einer Datenschutzauskunft durch Unternehmen berührt insbesondere deren geschäftliche Interessen. Der Schutz von Betriebs- und Geschäftsgeheimnissen hat einen ähnlichen Stellenwert für eine freie, selbstbestimmte Gesellschaft, wie der Schutz von personenbezogenen Daten.[51] Eine Abwägung zwischen diesen Schutzgütern ist deshalb geboten. Gemäß der Rechtsprechung des Bundesverwaltungsgerichts (BVerwG) sind Betriebs- und Geschäftsgeheimnisse gemeinsam durch Art. 12 und Art. 14 GG geschützt.[52] Das Grundrecht auf Berufsfreiheit ist auf juristische Personen[53] sowie auf Handelsgesellschaften, die keine juristischen Personen sind, aber für eine Personenmehrheit natürlicher

[44] Gallwas, NJW 1992, 2785 (2787); Masing, NJW 2012, 2305 (2307).
[45] Gurlit, NJW 2010, 1035 (1036).
[46] Gallwas, NJW 1992, 2785 (2788).
[47] Gallwas, NJW 1992, 2785 (2789).
[48] BVerfG, NJW 1999, 1777 (1777).
[49] Gallwas, NJW 1992, 2785 (2785).
[50] Gallwas, NJW 1992, 2785 (2786).
[51] Spiecker genannt Döhmann 2009, 30.
[52] BVerwG, NVwZ 2009, 1114 (1116).
[53] BVerfGE 50, 290 (363).

Personen auftreten, anwendbar.[54] Aus Sicht der Eigentumsfreiheit bilden Betriebs- und Geschäftsgeheimnisse einen eigenen Vermögenswert.[55] Das BVerfG hat bisher offen gelassen, ob Betriebs- und Geschäftsgeheimnisse von der Eigentumsgarantie geschützt sind oder nur durch die Berufsfreiheit.[56] Jedenfalls würde ein Schutz durch Art. 14 GG nicht weiter gehen als der durch Art. 12 GG.[57]

„Als Betriebs- und Geschäftsgeheimnisse werden alle auf ein Unternehmen bezogene Tatsachen, Umstände und Vorgänge verstanden, die nicht offenkundig, sondern nur einem begrenzten Personenkreis zugänglich sind und an deren Nichtverbreitung der Rechtsträger ein berechtigtes Interesse hat. Betriebsgeheimnisse umfassen im Wesentlichen technisches Wissen im weitesten Sinne; Geschäftsgeheimnisse betreffen vornehmlich kaufmännisches Wissen.“[58]

Eine Veröffentlichung von Betriebs- und Geschäftsgeheimnissen wäre eine Einschränkung der erfolgreichen Berufsausübung, da mögliche Marktvorsprünge im Wettbewerb nivelliert würden.[59]

Damit eine Information vom Schutzbereich des Betriebs- und Geschäftsgeheimnisses erfasst wird, müssen vier Voraussetzungen erfüllt sein:[60] Erstens muss die Information einen Unternehmensbezug besitzen. Rein private oder wissenschaftliche Informationen sind ausgeschlossen. Zweitens muss die Nichtoffenkundigkeit der Information gegeben sein. Ist die Information unter Zuhilfenahme erlaubter Mittel für Dritte zugänglich, ist sie nicht schützenswert. Drittens muss das Unternehmen seinen Geheimhaltungswillen erklärt oder konkludent erkennbar gemacht haben. Zuletzt muss ein berechtigtes Interesse an der Geheimhaltung der Information vorhanden sein. Das Interesse muss wirtschaftlicher Natur sein. Seine Berechtigung ist abhängig von der Stellung des Unternehmens im Wettbewerb. Für einen Monopolisten ist es schwieriger zu argumentieren, dass eine Veröffentlichung unternehmensbezogener Informationen seine Marktstellung gefährdet.

2.2 Europarechtliches Fundament

Das Datenschutzrecht ist seit der DSRL von 1995 stark durch das Europarecht geprägt. Der Einfluss des Europarechts wird durch die DSGVO in den nächsten Jahren noch steigen.

Die Europäisierung des Datenschutzrechts bewegt sich in einem komplexen Rechtsschutzgebilde aus EMRK, Unionsrecht und nationalem Verfassungsrecht. Daher ist zunächst zu klären, wie die unterschiedlichen rechtlichen Ordnungen zueinander in Bezie-

[54] BVerfGE 10, 89 (99).
[55] Kloepfer/Greve 2011, 579.
[56] BVerfGE 115, 205 (229).
[57] BVerfGE 115, 205 (248).
[58] BVerfGE 115, 205 (230 f.).
[59] Kloepfer/Greve 2011, 578.
[60] Kloepfer/Greve 2011, 580 ff.; Spiecker genannt Döhmann 2009, 32.

hung stehen. Darauf aufbauend ist zu ergründen, inwiefern das Europarecht ein Recht auf Auskunft vorsieht und ob sich die Ausgestaltung des Rechts auf Auskunft durch die DSGVO wesentlich ändert.

2.2.1 Verhältnis der rechtlichen Ordnungen

Das Verhältnis der europäischen und nationalen Rechtsordnungen ist äußerst komplex und vielfältig. Es kann an dieser Stelle nur angerissen, aber nicht abschließend beleuchtet werden. Entscheidend sind die Auswirkungen auf das Datenschutzrecht, insbesondere die Transparenzrechte.

EMRK und nationales Recht Die im Rahmen des Europarates entstandene EMRK gilt in der Bundesrepublik Deutschland im Range des Zustimmungsgesetzes nach Art. 59 Abs. 2 S. 1 Alt. 2 GG, eines einfachen Bundesgesetzes.[61] Damit könnte die EMRK nach dem Grundsatz „lex posterior derogat legi priori" durch neuere Bundesgesetzgebung verdrängt werden.[62]

Allerdings sind nach der Rechtsprechung des BVerfG der Inhalt und der Entwicklungsstand der EMRK bei der Auslegung der korrespondierenden Grundrechte in Betracht zu ziehen.[63] Die Rechtsprechung des Europäischen Gerichtshofs für Menschenrechte (EGMR) dient als Auslegungshilfe, erlangt jedoch keine unmittelbare Bindungswirkung für die nationalen Gerichte. Gerichte müssen sich nichtsdestotrotz mit Entscheidungen des EGMR, die für den konkreten Einzelfall getroffen wurden, erkennbar und nachvollziehbar auseinandersetzen.[64] Wegen des völkerrechtsfreundlichen Charakters des Grundgesetzes sind Konventionsverstöße, auch über den konkreten Einzelfall hinaus, zu vermeiden.[65]

Art. 8 Abs. 1 EMRK schützt in allgemeiner Form die elektronische Kommunikation[66] und die Verarbeitung personenbezogener Daten.[67] Der verwendete Begriff der „Privatheit" ist entsprechend weit.[68]

Ergänzt wird die EMRK durch das Übereinkommen 108 des Europarates zum Schutz des Menschen bei der automatischen Verarbeitung personenbezogener Daten vom 28.1.1981. In diesem Übereinkommen wird im Art. 8 Lit. a und b ein Recht auf Kenntnisnahme statuiert. Das Übereinkommen 108 wurde von allen Mitgliedsstaaten der Europäische Union unterzeichnet und ratifiziert.[69]

[61] BGBl. 1954 II S. 14.
[62] Herdegen 2016, § 3 Rn. 53.
[63] BVerfGE 75, 358 (370); BVerfGE 111, 307 (317).
[64] BVerfGE 111, 307 (324).
[65] BVerfGE 128, 326 (368 f.).
[66] EGMR, MMR 2007, 431 (432).
[67] EGMR, NJW 2011, 1333 (1334 f.).
[68] EGMR, NJW 2014, 1645 (1646).
[69] http://www.coe.int/de/web/conventions/full-list/-/conventions/treaty/108/signatures, abgerufen am 9. Mai 2017.

Unionsrecht und Grundgesetz Das Unionsrecht ist in den Mitgliedsstaaten unmittelbar wirksam.[70] Das nationale Recht ist primär- und sekundärrechtskonform auszulegen. Die Mitgliedsstaaten sind verpflichtet ihre Rechtsordnung dem Unionsrecht anzupassen.[71] Dies gilt nach Inkrafttreten der DSGVO insbesondere für das BDSG.

Die allgemeinen Rechtsgrundsätze der Mitgliedsstaaten und die Charta der Grundrechte der Europäischen Union stehen nach Art. 6 Abs. 1 EUV auf der Ebene des primären Unionsrechts.[72] Sie sind den Verträgen rechtlich gleichrangig.

Das Unionsrecht enthält Grundfreiheiten und Grundrechte. Während die Grundfreiheiten nur bei grenzüberschreitendem Bezug ihre Wirksamkeit entfalten, gelten die Grundrechte auch bei Inlandssachverhalten.[73] Die Adressaten der Charta sind nach Art. 51 GRCh die Union selbst, ihre Organe, Einrichtungen und Stellen sowie die Mitgliedsstaaten bei der Durchführung von Unionsrecht. Bezüglich des Anwendungsbereichs der Charta ist noch ungeklärt, ob der notwendige Unionsbezug restriktiv oder weit auszulegen ist.[74]

Der Europäische Gerichtshof (EuGH) geht von einem strikten Vorrang des Unionsrechts gegenüber nationalem Recht aus.[75] Er begründet dies unter anderem mit der Stellung des Unionsrechts als eigene autonome Rechtsordnung und mit der Sicherung der Funktionsfähigkeit der Union.[76] Innerhalb der Charta statuiert Art. 53 GRCh, dass die Charta keinen geringeren Grundrechtsschutz bietet als die Verfassungen der Mitgliedsstaaten. Nach Auffassung des EuGH ist Art. 53 eng auszulegen.[77] Er bedeutet keine Einschränkung des Vorrangs des Unionsrechts.

Beim Vorrang des Unionsrechts handelt sich um einen Anwendungsvorrang, nicht um einen Geltungsvorrang.[78] Deutsche Regelungen zum Datenschutz gelten deshalb auch nach Inkrafttreten der DSGVO unverändert weiter,[79] dürfen jedoch dort nicht angewendet werden, wo Konflikte auftreten.[80] Dies schließt indirekte Kollisionen mit ein.[81] Ergänzende Regelungen bleiben bestehen, soweit das Unionsrecht einen Sachverhalt nicht abschließend regelt.[82] Insbesondere kommen dafür die Öffnungsklauseln der DSGVO in Frage.

Der Anwendungsvorrang des Unionsrechts ist im Prinzip unangefochten.[83] Allerdings

[70] Ehlers in: Schulze/Zuleeg/Kadelbach 2015, § 11 Rn. 9.
[71] Zuleeg/Kadelbach in: Schulze/Zuleeg/Kadelbach 2015, § 8 Rn. 6.
[72] Ebd.
[73] Streinz/Michel in: Streinz, EUV/AEUV 2012, Art. 6 EUV Rn. 34.
[74] Herdegen 2016, § 68 Rn. 33 f.
[75] EuGH, NJW 1964, 2371.
[76] Ehlers in: Schulze/Zuleeg/Kadelbach 2015, § 11 Rn. 13.
[77] EuGH, Rs. C-399/11, ECLI:EU:C:2013:107, Rn. 58 ff.
[78] Roßnagel in: Roßnagel 2017, § 2 Rn. 5 ff. Herdegen 2016, § 10 Rn. 3; BVerfGE 126, 286 (301).
[79] Roßnagel in: Roßnagel 2017, § 2 Rn. 4.
[80] Roßnagel in: Roßnagel 2017, § 2 Rn. 5.
[81] Roßnagel in: Roßnagel 2017, § 2 Rn. 10.
[82] Roßnagel in: Roßnagel 2017, § 2 Rn. 14.
[83] Zuleeg/Kadelbach in: Schulze/Zuleeg/Kadelbach 2015, § 8 Rn. 15; BVerfGE 123, 267 (402).

bestehen Differenzen mit den Verfassungsgerichten der Mitgliedsstaaten bezüglich des Verhältnisses Unionsrecht – Verfassungsrecht. Insbesondere das Bundesverfassungsgericht zieht Grenzen in zweierlei Hinsicht:[84] Die Grenze des Nichtübertragbaren und die Grenze des Nichtübertragenen.[85]

Die Grenze des Nichtübertragbaren ist die Identität der geltenden Verfassungsordnung. Darunter fallen insbesondere die konstituierenden Strukturen der Bundesrepublik Deutschland und die völkerrechtliche Souveränität.[86] Daneben steht, dass bei der Übertragung von Hoheitsrechten nach Art. 23 Abs. 1 S. 1 GG ein im wesentlichen vergleichbarer Grundrechtsschutz gewährleistet sein muss. Das BVerfG hat diesen bis auf weiteres für die Union festgestellt und übt in diesem Rahmen keine Grundrechtsprüfung mehr aus.[87] Es beschränkt sich auf eine generelle Gewährleistung der unabdingbaren Grundrechtsstandards.[88] Für die Datenschutzgrundrechte der Charta ist ein den deutschen Grundrechten vergleichbares Schutzniveau gegeben.[89] Deshalb ist im europäischen Bezugsrahmen die Charta vorrangig. Eine Auslegung der DSGVO am Maßstab des Grundgesetzes ist nicht vorgesehen.[90] Ein möglicher Grenzfall werden die aus den Öffnungsklauseln der DSGVO resultierenden nationalen Regelungen sein.[91]

Die Grenze des Nichtübertragenen ergibt sich aus der fehlenden Staatlichkeit der Union.[92] Der Europäischen Union fehlt als wesentliches Merkmal der Staatlichkeit ein europäisches Staatsvolk.[93] Sie ist eine Union der Völker Europas.[94] Des Weiteren fehlt der Union die Kompetenz zur selbstständigen Erweiterung ihrer Befugnisse (Kompetenz-Kompetenz).[95] Die Mitgliedsstaaten sind die „Herren der Verträge“.[96] Ultra-Vires-Akte sind unzulässig.[97]

Dennoch ist das Grundgesetz im Grundsatz europarechtsfreundlich.[98] Das BVerfG sieht sich in einem Kooperationsverhältnis zum EuGH.[99] Es erlegt sich selbst auf, eine Vorabentscheidung des EuGH anzustreben, bevor es eine Verletzung der unabdingbaren

[84] BVerfGE 123, 267 (400 ff.).
[85] Ehlers in: Schulze/Zuleeg/Kadelbach 2015, § 11 Rn. 21 ff.
[86] BVerfGE 37, 271 (279); BVerfGE 123, 267 (400).
[87] BVerfGE 73, 339 (387).
[88] BVerfGE 89, 155 (175).
[89] Johannes in: Roßnagel 2017, § 2 Rn. 64.
[90] Johannes in: Roßnagel 2017, § 2 Rn. 65.
[91] Johannes in: Roßnagel 2017, § 2 Rn. 108.
[92] Ehlers in: Schulze/Zuleeg/Kadelbach 2015, § 11 Rn. 25.
[93] Herdegen 2016, § 5 Rn. 16; BVerfGE 89, 155 (184).
[94] Präambel des EUV.
[95] BVerfGE 89, 155 (192 ff.).
[96] Herdegen 2016, § 6 Rn. 1.
[97] BVerfGE 126, 286 (302 ff.).
[98] BVerfGE 123, 267 (347).
[99] BVerfGE 89, 155 (175).

Grundrechtsstandards oder einen Ultra-Vires-Akt annimmt.[100]

Neben der Bundesrepublik Deutschland wird auch in anderen Mitgliedsstaaten der Vorrang des Unionsrechts nicht uneingeschränkt akzeptiert.[101]

Wie im nationalen Verfassungsrecht ist der Schutz personenbezogener Daten in der Charta substantiell verankert. Das Recht auf Auskunft ist in Art. 8 Abs. 2 Satz 2 GRCh explizit kodifiziert. Art. 16 Abs. 2 AEUV enthält die Kompetenz zum Erlass von Vorschriften zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten. Darauf beruht die Verabschiedung der DSGVO.

Daneben finden sich, wie in der nationalen Verfassung, in der Charta in Art. 15 die Berufsfreiheit und in Art. 16 GRCh die unternehmerische Freiheit als Gegeninteressen, die in die gesetzgeberische Abwägung mit einzubeziehen sind.[102]

EMRK und Unionsrecht Die Charta und die EMRK widersprechen sich nicht, sondern ergänzen sich.[103] Soweit die Charta Rechte enthält, die jenen in der EMRK entsprechen, gleichen Bedeutung und Tragweite der Rechte in der Charta nach Art. 52 Abs. 3 S. 1 GRCh der Bedeutung und Tragweite der Rechte der EMRK. Nach Art. 52 Abs. 3 S. 2 GRCh schließt dies nicht aus, dass „das Recht der Union einen weiter gehenden Schutz gewährt." Bezogen auf die Datenschutzgrundrechte vertritt die Artikel-29-Datenschutzgruppe die Auffassung, dass Art. 8 Abs. 1 EMRK und Art. 7 und 8 GRCh die gleiche Bedeutung und die gleiche Tragweite haben.[104] Der sachliche Schutzbereich des Art. 8 GRCh geht allerdings über die EMRK hinaus.[105]

Darüber hinaus sind nach Art. 6 Abs. 3 EUV die Grundrechte der EMRK und der gemeinsamen Verfassungsüberlieferung als allgemeine Grundsätze Teil des Rechts der Union.

Die EMRK ist Rechtserkenntnisquelle statt Rechtsquelle des EuGH, aus der rechtsvergleichend die Unionsgrundrechte hergeleitet werden.[106] Dies gilt so lange die Union der EMRK noch nicht nach Art. 6 Abs. 2 S. 1 EUV beigetreten ist. Der EuGH sperrt sich bisher gegen einen Beitritt.[107] Der Beitritt der Union zur EMRK würde der EMRK im

[100] BVerfGE 126, 286 (304).

[101] Ehlers in: Schulze/Zuleeg/Kadelbach 2015, § 11 Rn. 36.

[102] Im Sekundärrecht positioniert sich der Erwägungsgrund (ErwGr) 41 DSGVO entsprechend. Er klassifiziert zwar das Recht auf Auskunft als ein grundlegendes Recht eines jeden Unionsbürgers, fordert aber zugleich die Berücksichtigung unternehmerischer Freiheiten. Insbesondere in Bezug auf den logischen Aufbau von IT-Systemen wird auch das Urheberrecht als Schranke aufgeführt. Inwieweit das Urheberrecht eine über Betriebs- und Geschäftsgeheimnisse hinausgehende Schutzwirkung für den logischen Aufbau von IT-Systemen bieten soll, bleibt unklar.

[103] EuGH, EuZW 2010, 939 (941).

[104] Artikel-29-Datenschutzgruppe 2014, 4 f.

[105] Holznagel/Dietze in: Schulze/Zuleeg/Kadelbach 2015, § 37 Rn. 3.

[106] Streinz/Michel in: Streinz, EUV/AEUV 2012, Art. 6 EUV Rn. 25.

[107] EuGH, Gutachten 2/13, ECLI:EU:C:2014:2454.

gleichen Maße Anwendungsvorrang gegenüber nationalem Recht geben, wie dies für das Primärrecht der Union der Fall ist.

Aus Sicht des EGMR ist das Grundrechtsniveau des Unionsrechts mit der EMRK vergleichbar.[108] Der EGMR folgt der Solange-II-Rechtsprechung[109] des BVerfG.[110]

2.2.2 Das Recht auf Auskunft im Unionsrecht

Nach Auffassung des EuGH ist das Auskunftsrecht essentiell notwendig, um dem Betroffenen die Wahrnehmung seiner weiteren Rechte zu ermöglichen.[111] Aus der Verankerung des Rechts auf Auskunft im Primärrecht der Union folgt, dass ein Fehlen des Rechts im Grundgesetz in den Bereichen, in denen die Union materiell zuständig ist, keine praktischen Auswirkungen hat.

Die DSGVO ist gemäß Art. 99 am 20. Mai 2016 in Kraft getreten und gilt ab dem 25. Mai 2018 verbindlich und unmittelbar in jedem Mitgliedsstaat.[112] Sie wird das europäische Datenschutzrecht der Zukunft entscheidend prägen und gibt in den Artikeln 12 bis 15 und 20 neue Impulse für die datenschutzrechtliche Transparenzverfassung. Das eigentliche Auskunftsrecht bleibt im Grundsatz gleich. Seine Ausgestaltung weicht nur in wenigen Aspekten von der Umsetzung der DSRL durch das BDSG ab. Diese Aspekte werden in Kapitel 3 themenbezogen mitbehandelt. Die in dieser Arbeit auf Grundlage des BDSG getroffenen, fundamentalen Aussagen zum Auskunftsrecht dürften wegen der vergleichbaren Rechtslage unter der DSGVO auch europaweit Geltung haben.

Das neuartige Recht auf Datenübertragbarkeit in Art. 20 DSGVO wird in Kapitel 10.2 gesondert beleuchtet. Sein innovativer Charakter macht, abweichend vom übrigen Vorgehen, eine herausgehobene Erörterung zukünftigen Rechts notwendig.

Gegenwärtig ist die DSRL noch der sekundärrechtliche Rahmen des EU-Datenschutzrechts für die einfachgesetzliche Umsetzung im nationalen Recht. In Art. 12 wird das Recht auf Auskunft kodifiziert. In Art. 13 werden mögliche Ausnahmeregelungen aufgezeigt, die der nationale Gesetzgeber wahrnehmen kann. Ob das BDSG europäisches Recht richtlinienkonform umsetzt, wird in den nachfolgenden Kapiteln diskutiert.

[108] Herdegen 2016, § 3 Rn. 62.
[109] BVerfGE 73, 339 (387).
[110] EGMR, NJW 2006, 197 (203).
[111] EuGH, EuZW 2009, 546 (548).
[112] Vgl. Art. 288 AEUV.

2.3 Zwischenfazit

Das Recht auf Auskunft ist entgegen der Annahme aus Hypothese $\mathfrak{H}_1$ kein generell verfassungsrechtlich gebotenes Betroffenenrecht. Das vom Bundesverfassungsgericht aus dem Recht auf informationelle Selbstbestimmung abgeleitete Recht auf Kenntnisnahme führt allerdings zum gleichen Schutzniveau ohne den Gesetzgeber in der Realisierung des Transparenzgedankens für die Zukunft einzuschränken.

Fraglich ist, welchen Stellenwert das Recht auf Kenntnisnahme in Zukunft noch haben wird. Mit dem in Art. 8 Abs. 2 Satz 2 der Charta explizit kodierten Recht auf Auskunft wurde auf europäischer Ebene eine starke Festlegung getroffen. Vermittels der DSGVO wird die Bedeutung der nationalen Grundrechte weiter zurückgehen. Der EuGH wird das Datenschutzrecht maßgeblich mitprägen.

Wie sich das Bundesverfassungsgericht nach Solange I,[113] Solange II,[114] dem Maastricht-Urteil[115] und der Lissabon-Entscheidung[116] positioniert, bleibt offen. Einer Konkretisierung des Rechts auf Kenntnisnahme durch die Verträge, die Charta und die Rechtsprechung des EuGH wird es wohl nicht entgegentreten.

[113] BVerfGE 37, 271.
[114] BVerfGE 73, 339.
[115] BVerfGE 89, 155.
[116] BVerfGE 123, 267.

3 Einfachgesetzliche Verankerung des Rechts auf Auskunft

„Das Auskunftsrecht ist für die Betroffenen das fundamentale Datenschutzrecht.“[1] Es wird sogar als die „Magna Charta des Datenschutzes“ bezeichnet.[2] Es ist gemäß § 6 Abs. 1 BDSG unabdingbares Recht des Betroffenen und kann nicht ausgeschlossen oder beschränkt werden. Das Auskunftsrecht ist die Voraussetzung zur Wahrnehmung der übrigen Betroffenenrechte auf Löschung, Sperrung und Berichtigung. Es ist die wirksamste Ausprägung des datenschutzrechtlichen Transparenzgebots.[3]

Das Auskunftsrecht kombiniert Selbstdatenschutz[4] (Kontrollrecht des Einzelnen) und externe Kontrolle. Es ist ein Element des ersteren und ermöglicht damit letzteres.

Dieses Kapitel diskutiert und beleuchtet die Einbindung des Rechts auf Auskunft in die Systematik des Datenschutzrechts (Abschnitt 3.1), Aspekte der Auskunftserteilung und der Reichweite des Auskunftsrechts (Abschnitt 3.2 bis 3.4) sowie den Umfang des Auskunftsanspruchs (Abschnitt 3.6 bis 3.8).[5] Ergänzend zu den Vorschriften des BDSG wird, wo erforderlich, themenbezogen auf grundrechtliche Aspekte sowie die bestehende (DSRL) und kommende (DSGVO) Rechtslage der Europäischen Union eingegangen.

3.1 Einbindung des Rechts auf Auskunft in die Systematik des Datenschutzrechts

Die datenschutzrechtlichen Vorschriften können in drei wesentliche Bausteine untergliedert werden: Den datenschutzgerechten Umgang mit personenbezogenen Daten, die Transparenz darüber und die darauf aufbauenden Interventionsmöglichkeiten (Abbildung 3.1). Transparenz und Intervenierbarkeit haben jeweils eine interne und eine externe Komponente. Interne Transparenz wird über Verfahrensverzeichnisse und Datenschutzaudits (§ 9a BDSG) hergestellt. Externe Transparenz entsteht vor allem durch ein umfangreiches Informationsangebot nach außen hin. Intervenierbarkeit manifestiert sich intern im

[1] Dix in: Simitis, BDSG 2014, § 34 Rn. 1.

[2] Neben anderen Wedde in: Roßnagel 2003, Kap. 4.4 Rn. 2; Dix in: Simitis, BDSG 2014, § 34 Rn. 1; Weichert, DuD 2006, 694 (694).

[3] Rudolf in: Merten/Papier, HGR IV 2011, § 90 Rn. 48.

[4] Roßnagel in: Roßnagel 2003, Kap. 3.4 Rn. 1 ff.

[5] Eine Zusammenfassung dieses Kapitels wurde bereits in Bier, DuD 2015, 741 veröffentlicht.

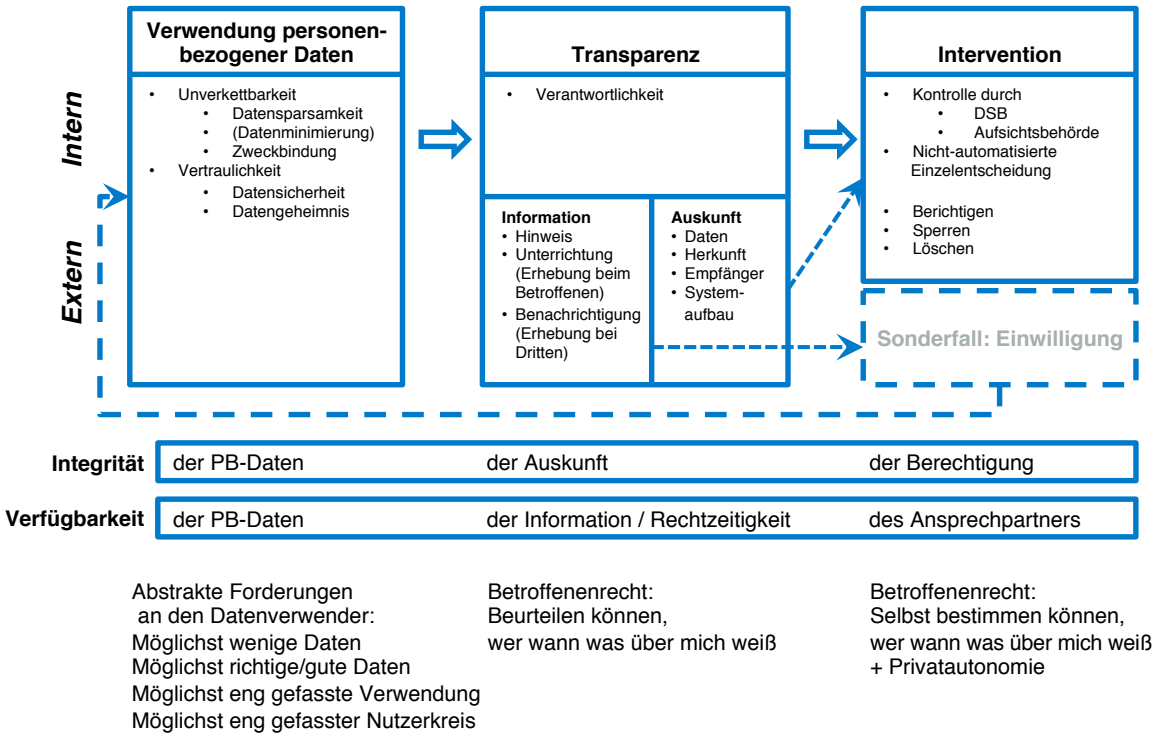


Abbildung 3.1: Struktogramm Datenschutz (mit angedeuteter Abfolge der Bausteine)

Verbot der automatischen Einzelentscheidung (§ 6a BDSG), extern durch umfangreiche Gestaltungsrechte des Betroffenen (§ 35 BDSG).

Die datenschutzgerechte Erhebung, Verarbeitung und Nutzung personenbezogener Daten ist die Grundlage auf der alle Betroffenenrechte aufsetzen. Sie ist Schwerpunkt der Datenschutzziele[6] Integrität, Verfügbarkeit und Vertraulichkeit (siehe dazu auch Kapitel 4.1.2).[7]

Transparenz ist die Voraussetzung für die Wahrnehmung weiterer Kontroll-, Abwehr- und Gestaltungsrechte.[8] Der Transparenzgedanke der EU-Datenschutzrichtlinie, die auch das Bundesdatenschutzgesetz prägt, verlangt schon bei der Erhebung eine umfangreiche Information des Betroffenen.[9]

An erster Stelle stehen deshalb *Hinweis- und Kennzeichnungspflichten*. Geübte Praxis ist die Kennzeichnung durch Hinweisschilder im Umfeld der Videoüberwachung (§ 6b Abs. 2

[6] In § 5 Abs. 1 LDSG-SH verankert.
[7] Rost/Bock, DuD 2011, 30 (32).
[8] Mallmann in: Simitis, BDSG 2014, § 19 Fn. 1.
[9] Dix in: Simitis, BDSG 2014, § 34 Rn. 4.

BDSG) und durch ein digitales Äquivalent in der Umsetzung der EU-Cookie-Richtlinie.[10] Ebenso zählt die bei der Einwilligung erforderliche Belehrung (§ 4a Abs. 1 S. 2 BDSG) zu den einer Erhebung vorausgehenden Hinweispflichten. Hinweis- und Kennzeichnungspflichten sind unabhängig von einer tatsächlichen Erhebung personenbezogener Daten bei jedem Einzelnen, der den Hinweis wahrnimmt. Sie sollen vor der potentiellen Erhebung warnen, ihren Zweck darlegen und über die Folgen der Erhebung und anschließenden Verwendung aufklären. Der Betroffene hat dann die Möglichkeit schon auf den Erhebungsvorgang selbst einzuwirken. Wie beispielsweise bei Videoüberwachungsanlagen finden sie meistens dort Anwendung, wo Daten über eine unspezifische und nicht ohne weiteres zu benennende Menge an Personen erhoben werden. Teil eines Hinweises vor der Erhebung personenbezogener Daten ist auch die Erläuterung einer eventuellen gesetzlichen Auskunftsverpflichtung des Betroffenen (§ 4a Abs. 1 S. 2 BDSG) und eine Aufklärung über mögliche Folgen einer Verweigerung der Datenerhebung (§ 4a Abs. 1 S. 2 BDSG).

Die *Unterrichtung* ist das geeignete Instrument zur Information des Betroffenen nach der Erhebung personenbezogener Daten bei ihm selbst (§ 4 Abs. 3 S. 1 BDSG). Sie ist spezifisch auf einen tatsächlich vorgenommenen Erhebungsvorgang hin gemünzt und soll den Betroffenen dann über die Erhebung und ihre Umstände informieren, wenn er von ihr nicht selbst hat Kenntnis nehmen können und sie nicht zumindest nach dem im Geschäftsverkehr üblichen zu erwarten hatte.[11] Digital findet man sie als Datenschutzerklärungen auf Webseiten (§ 13 Abs. 1 TMG).

Die *Benachrichtigung* (§§ 19a, 33 BDSG) ist das Gegenstück bei der Erhebung personenbezogener Daten bei Dritten. Sie gibt dem unbeteiligten Betroffenen die Möglichkeit der Kenntnisnahme. Sie flankiert damit den Hinweis und die Unterrichtung in der Ermächtigung (Empowerment) des Betroffenen zu weiteren Schritten. Benachrichtigung und Unterrichtung sind Spielarten der aktiven Transparenz und gehen von der verantwortlichen Stelle aus. Sie sind Bringschuld der verantwortlichen Stelle. Gemeinsam schaffen sie die Voraussetzung dafür, dass der Betroffene im zweiten Schritt selbst aktiv wird und ein Auskunftsersuchen stellt.[12] Die Brücke zwischen beiden Transparenzarten kann beispielsweise die Vergabe von Schlüsselkennungen (Credentials) bilden.

Die *Auskunft* (§§ 19, 34 BDSG) ist als passive Transparenzmaßnahme Holschuld des Betroffenen.[13] Sie ist dort Auffangrecht für den Betroffenen, wo die übrigen Transparenzregelungen lückenhaft sind.[14]

Die Übersicht für Jedermann nach § 4g Abs. 2 S. 2 i. V. m. § 4e S. 1 Nr. 1 bis 8 BDSG steht strukturell zwischen den Hinweispflichten und der Auskunft. Sie ist einerseits wie die

[10] Art. 5 Abs. 3 ePrivacy-RL in der durch die Cookie-RL geänderten Fassung.

[11] Im Besonderen ist sie deshalb als Maßnahme bei mobilen personenbezogenen Speicher- und Verarbeitungsmedien in § 6c Abs. 1 BDSG zu finden.

[12] Meents/Hinzpeter in: Taeger/Gabel, BDSG 2013, BDSG § 34 Rn. 1.

[13] Dix in: Simitis, BDSG 2014, § 34 Rn. 6.

[14] Weichert, DuD 2006, 694 (694).

Hinweispflichten unabhängig von der Betroffeneneigenschaft und macht ausschließlich allgemeine Informationen zugänglich. Sie ist andererseits wie die Auskunft Holschuld der interessierten Person, wogegen Hinweise und Kennzeichen so angebracht werden müssen, dass sie ein potentiell Betroffener mit großer Sicherheit wahrnehmen kann.[15] Die Übersicht für Jedermann ist nicht nur potentiell Betroffenen, natürlichen Personen, sondern auch juristischen Personen zugänglich.[16] Die Übersicht für Jedermann führt weder zu einer Erleichterung des Zugangs zur Auskunft, noch kann sie Auskunftsersuchen Betroffener in irgendwie gearteten Fällen erledigen.[17] Sie erscheint nicht nur deshalb als ein bürokratisches Relikt.[18]

Stellt der Betroffene Ungereimtheiten in den über ihn gespeicherten Daten oder die Unrechtmäßigkeit der Verarbeitung fest, kann er im Folgeschritt seine *Interventionsrechte* wahrnehmen. Bei einer Integration in ein interaktives Auskunftssystem kann er in einem direkten Folgeschritt die Berichtigung (§ 20 Abs. 1, § 35 Abs. 1 BDSG), Löschung (S 20 Abs. 2, § 35 Abs. 2 BDSG) oder Sperrung (§ 20 Abs. 3, 4 u. 6, § 35 Abs. 3 u. 4 BDSG) der Daten anordnen und ihrer weiteren Verwendung widersprechen (§ 20 Abs. 5, § 35 Abs. 5 BDSG).

Die Systematik der Transparenzregelungen bleibt im Rahmen der zukünftigen DSGVO grundsätzlich gleich.[19] Allerdings kommt das Recht auf Datenübertragbarkeit in Art. 20 DSGVO als Neuschöpfung hinzu.[20]

3.2 Reichweite des Rechts auf Auskunft

Auskunft meint im Rahmen dieser Arbeit immer den datenschutzrechtlichen Auskunftsanspruch als Recht des Betroffenen aus § 19, 34 BDSG und verwandten Spezialgesetzen.[21] Die Auskunftspflicht der verantwortlichen Stelle gegenüber der Aufsichtsbehörde (§ 38 Abs. 3 BDSG) oder anderen Behörden[22] ist nicht Teil dieser Diskussion.[23] Außen vor ist auch die Verpflichtung des Betroffenen zur Auskunft aufgrund einer Rechtsvorschrift entsprechend § 4 Abs. 3 S. 2 BDSG.

[15] Scholz in: Simitis, BDSG 2014, § 6b Rn. 107.
[16] Scheja in: Taeger/Gabel, BDSG 2013, BDSG § 4g Rn. 33.
[17] A.A. Gola/Schomerus, BDSG 2015, § 4g Rn. 25
[18] Scheja in: Taeger/Gabel, BDSG 2013, BDSG § 4g Rn. 34.
[19] Unterrichtung in Art. 13 DSGVO, Benachrichtigung in Art. 14 DSGVO, Auskunft in Art. 15 DSGVO, Berichtigung in Art. 16 DSGVO, Löschung in Art. 17 DSGVO, Sperrung in Art. 18 DSGVO, Widerspruch in Art. 21 DSGVO.
[20] Siehe Kapitel 10.2.
[21] U. a. § 13 TMG, § 93 TKG, § 83 SGB X.
[22] Beispielsweise gemäß § 110 ff. TKG gegenüber Sicherheitsbehörden.
[23] Die Übermittlung an Dritte nach § 28 Abs. 2 Nr. 2 BDSG wird ebenfalls nicht einbezogen. Siehe zur entgegenstehenden Zweckbindung auch BGH, DuD 2014, 715 (716).

Auch zivil-[24] sowie gesellschaftsrechtliche[25] Ansprüche des Betroffenen bleiben unberücksichtigt. Zivilrechtliche Ansprüche können im Gegensatz zu datenschutzrechtlichen Ansprüchen gegenüber jedem Privaten durchgesetzt werden. Sie sind Rechte unter gleichen. Der datenschutzrechtliche Auskunftsanspruch gleicht dagegen ein informationelles Machtgefälle aus. Der Umgang mit personenbezogenen Daten aus einer rein persönlichen oder privaten Tätigkeit heraus ist deshalb nach § 1 Abs. 2 Nr. 3 BDSG vom datenschutzrechtlichen Auskunftsanspruch ausgenommen.

Auskunftsverpflichtete ist die verantwortliche Stelle. Der Anwendungsbereich des § 34 BDSG bestimmt sich grundsätzlich nach § 27 BDSG und erstreckt sich damit auf nichtöffentliche Stellen und öffentlich-rechtliche Wettbewerbsunternehmen des Bundes und (mit Einschränkungen) der Länder (§ 27 Abs. 1 S. 1 Nr. 1 u. 2 BDSG) soweit mit personenbezogenen Daten umgegangen wird (§ 27 Abs. 1 S. 1 BDSG).[26] Der Anwendungsbereich des § 19 BDSG folgt dem des § 12 BDSG. Er bezieht sich auf den Umgang mit personenbezogenen Daten durch öffentliche Stellen des Bundes (§ 12 Abs. 1 BDSG). Weil mittlerweile alle Länder eigene Datenschutzgesetze besitzen, spielt § 12 Abs. 2 BDSG keine Rolle mehr.

Der Gesetzgeber hat den Auskunftsanspruch des Betroffenen im BDSG einschließlich etwaiger Spezialgesetze abschließend geregelt. Deshalb kann ein Auskunftsersuchen grundsätzlich nicht mit dem allgemeinen Persönlichkeitsrecht begründet werden.[27]

Die einzige Voraussetzung der Auskunftserteilung ist ein Auskunftsverlangen durch den Betroffenen. Das Auskunftsverlangen unterliegt selbst keinerlei Voraussetzung. Insbesondere muss es nicht begründet werden und bedarf keines Zweckes oder Anlasses. Auch ein berechtigtes Interesse muss nicht dargelegt werden.[28]

3.3 Form und Inhalt des Auskunftsersuchens

Es besteht keinerlei Formerfordernis an das Auskunftsersuchen selbst.[29] Das Verlangen kann schriftlich, in Textform oder mündlich an die verantwortliche Stelle gerichtet werden. Es steht dem Betroffenen auch frei, ob er sich elektronisch per E-Mail oder Webformular oder analog an die verantwortliche Stelle wendet. Die verantwortliche Stelle hat zumindest die Kommunikationskanäle offen zu halten, auf denen sie auch selbst mit Dritten kommuniziert.[30] Verwendet die verantwortliche Stelle E-Mails zur werbenden Ansprache,

[24] §§ 259, 260, 402, 666, 1004, 1379, 1580, 1605, 1799, 1839, 2027, 2057, 2127, 2314 BGB.
[25] § 131 AktG; § 51a GmbHG.
[26] Erheben, Verarbeiten und Nutzen von Daten unter Einsatz von Datenverarbeitungsanlagen oder in/aus nicht automatisierten Dateien.
[27] BGH, NJW 1984, 1886 (1886).
[28] Dix in: Simitis, BDSG 2014, § 34 Rn. 12.
[29] Dix in: Simitis, BDSG 2014, § 34 Rn. 13; Gola/Schomerus, BDSG 2015, § 34 Rn. 1.
[30] So auch die DSGVO: Werden personenbezogene Daten elektronisch verarbeitet, sollte gemäß ErwGr 59 eine elektronische Antragstellung möglich sein.

muss sie auch sicherstellen, dass eine E-Mailadresse, unter der sie erreichbar ist, öffentlich bekannt ist. Die internen Prozesse der verantwortlichen Stelle müssen so organisiert sein, dass auf allen Kanälen eingehende Auskunftsverlangen bearbeitet werden können. Diesbezüglich obliegt es der verantwortlichen Stelle, die eigenen Mitarbeiter, insbesondere Kundenservice, Vertrieb, und Empfang, angemessen zu schulen. Ein intern einheitlicher Ansprechpartner, wie der betriebliche oder behördliche Datenschutzbeauftragte, ist nicht von außen her verpflichtend zu kontaktieren.

Gemäß § 6 Abs. 2 S. 2 BDSG kann sich der Betroffene sogar an eine andere als die speichernde Stelle wenden, sofern bei einer verteilten automatisierten Verarbeitung nicht ersichtlich ist, welche Stelle speichert. Die kontaktierte Stelle ist dann verpflichtet, das Auskunftsersuchen an die speichernde Stelle weiterzuleiten.

Die nach § 34 Abs. 1 S. 2 BDSG zu erfolgende nähere Bezeichnung der Art personenbezogener Daten, die beauskunftet werden sollen, ist eine Sollvorschrift. Die nähere Bezeichnung ist nicht Voraussetzung für die Auskunftserteilung. Das Auskunftsersuchen darf nicht aufgrund fehlender Präzision verweigert werden.[31]

Da nur Betroffene ein Recht auf Auskunft haben, ist für eine gerichtliche Geltendmachung des Rechtes ausreichend darzulegen, dass personenbezogene Daten von einer verantwortlichen Stelle verwendet wurden. Es ist kein Erwirken der gerichtlichen Durchsetzung „ins Blaue hinein“ möglich.[32] Die Auskunftsansprüche und ihre Begründung müssen im Sinne des § 253 Abs 2 Nr. 2 ZPO hinreichend präzisiert werden.[33]

Die Ausführungen zum BDSG haben unter der DSGVO weiterhin Geltung. Die Rechtslage ändert sich im Wesentlichen nicht. Allerdings führt Art. 26 DSGVO den Begriff des *gemeinsam für die Verarbeitung Verantwortlichen* als neue Rechtsfigur ein. Verarbeiten mehrere Stellen personenbezogene Daten im Verbund und legen gemeinsam die Zwecke der Verarbeitung fest, können sie intern, aber in für den Betroffenen transparenter Form, regeln, wer welchen Teil der Informationspflichten nach Art. 13 und 14 DSGVO erfüllt. In der Vereinbarung kann ein einheitlicher Ansprechpartner für den Betroffenen festgelegt werden. Eine Vereinbarung nach Art. 26 DSGVO hat wie gehabt keine Außenwirkung (Abs. 3). Der Betroffene ist nicht verpflichtet die benannte Stelle zu kontaktieren, sondern kann sich weiterhin an jede verantwortliche Stelle wenden. Die Bestimmungen des Art. 26 DSGVO sollen jedoch zukünftig dazu führen, dass Auskunftsanfragen effizienter und für den Betroffenen im Regelfall schneller beantwortet werden können.

Einschränkend kann eine verantwortliche Stelle in Zukunft gemäß ErwGr 63 DSGVO verlangen, dass bei einer Vielzahl gespeicherter personenbezogener Daten die gewünschten Daten präzisiert werden. Der Betroffene kann jedoch weiterhin antworten, dass er

[31] Dix in: Simitis, BDSG 2014, § 34 Rn. 41; Gola/Schomerus, BDSG 2015, § 34 Rn. 5; Meents/Hinzpeter in: Taeger/Gabel, BDSG 2013, BDSG § 34 Rn. 15.

[32] Gola/Schomerus, BDSG 2015, § 34 Rn. 5a; Landgericht München II, Urteil vom 20.09.2005, MMR Heft 12, XXIII (Kurzinformation).

[33] LAG Hessen, Urteil vom 29.01.2013, BeckRS 2013, 67364.

gerne alle Daten hätte.[34] Die nähere Bezeichnung bleibt optional.

3.4 Der Auskunftsberechtigte und dessen Identifizierung

Das Recht auf Auskunft steht allen Betroffenen zu – unabhängig von Alter, Nationalität und Wohnsitz. Es gibt jedoch kein Auskunftsrecht für Dritte. Eine Weitergabe personenbezogener Daten an Dritte ist eine Übermittlung.[35] Ihrem Regelungsgehalt nach, und da die Auskunft ein Auskunftsverlangen des Betroffenen voraussetzt, kann die verantwortliche Stelle diese Übermittlung personenbezogener Daten an Dritte niemals auf das Auskunftsrecht stützen. Der Antrag auf Auskunft kann allerdings durch einen Bevollmächtigten des Betroffenen gestellt werden.[36] In diesem Fall ist durch die verantwortliche Stelle die Vollmacht gesondert zu prüfen.

Anders sieht es für die Rechtsnachfolge der Erben aus. Sie können grundsätzlich keinen Auskunftsanspruch geltend machen, da das Auskunftsrecht ein höchstpersönliches Recht ist.[37] Nichtsdestoweniger wird vertreten, dass das Auskunftsrecht auf den Erben übergeht, da die Daten nach Tod des Erblassers ansonsten komplett schutzlos wären.[38] Diese Ansicht verkennt indes, dass nicht Daten, sondern die Persönlichkeitsrechte des Betroffenen das Schutzgut des Datenschutzes sind. Das postmortale Persönlichkeitsrecht ist jedoch auf den eng gefassten Schutzbereich der Menschenwürde aus Art. 1 Abs. 1 GG reduziert.[39] Mit dem postmortalen Würdeschutz wird der Achtungsanspruch des Verstorbenen gegen Verobjektivierung und Herabwürdigungen nach dem Tode geschützt.[40] Das Andenken des Verstorbenen ist geschützt, ein Missbrauch seiner Daten zur Verunglimpfung ist unzulässig.[41] Ein Datenschutzrecht auf Auskunft der Erben ergibt sich jedoch nicht.

Vermögensrechtliche Ansprüche auf Auskunft aus Vertrag i. V. m. § 1922 Abs. 1 BGB stehen unabhängig daneben.[42] Sie erstrecken sich auch auf den nicht-vermögensrechtlichen Teil des Erbes.[43] Der Auskunftsanspruch aus Vertrag findet seine Grenzen im postmortalen Persönlichkeitsrecht des Erblassers. Soweit es sich beim Erblasser um einen Minderjährigen handelt, steht es dem Auskunftsanspruch der Erziehungsberechtigten als mögliche Erben und Sachverwalter des Persönlichkeitsrechts nicht entgegen.[44]

[34] Laue/Nink/Kremer 2016, § 4 Rn. 27.
[35] Dix in: Simitis, BDSG 2014, § 34 Rn. 14.
[36] Mallmann in: Simitis, BDSG 2014, § 19 Rn. 34.
[37] Dix in: Simitis, BDSG 2014, § 34 Rn. 43; Klas/Möhrike-Sobolewski, NJW 2015, 3473 (3474)
[38] Solmecke/Köbrich/Schmitt, MMR 2015, 291 (293).
[39] BVerfGE 30, 173 (194).
[40] Dreier in: Dreier, GG 2013, Art. 1 Abs. 1 Rn. 76.
[41] Strafrechtlicher Schutz in § 189 StGB; darüber hinausgehend § 4 Abs. 1 S. 2 BlnDSG.
[42] Klas/Möhrike-Sobolewski, NJW 2015, 3473 (3478).
[43] LG Berlin, ZD 2016, 182 (183) = MDR 2016, 165 (165).
[44] LG Berlin, ZD 2016, 182 (184).

Ob eine grundsätzliche Pflicht zur Feststellung der Identität des Auskunftsersuchenden besteht, ist in der Literatur umstritten.[45] Aufgrund des bußgeldbewehrten Verbots der unbefugten Übermittlung personenbezogenen Daten an Dritte,[46] liegt es in jedem Fall im Interesse der verantwortlichen Stelle, die Identität des Auskunftsersuchenden gegen den eigenen Datenbestand zu prüfen. Die Intensität der Prüfung hängt von der Sensitivität der gespeicherten Daten ab und ist mit dieser in ein sinnvolles Verhältnis zu setzen.[47]

Für die Zuordnung der gespeicherten Daten zu der auskunftsersuchenden Person ist gegebenenfalls die Einholung ergänzender Informationen erforderlich (§ 34 Abs. 1 S. 2 BDSG). Die Auskunft darf nicht unter dem Vorwand abgelehnt werden, dass kein Name und keine Kontaktadresse des Betroffenen gespeichert ist.[48] Nur wenn nach Einholung weiterer Informationen die Zuordnung zu einem bestimmten Betroffenen nicht möglich ist entfällt die Auskunftspflicht. In diesem Falle handelt es sich zumindest subjektiv nicht um personenbezogene Daten (relativer Personenbezug).

Müsste der Betroffene die über ihn gespeicherten Daten jedoch voll umfänglich nennen, um zu seinem Recht auf Auskunft zu kommen, würde die Intention des Anspruchs pervertiert. Denn die Auskunft soll ja gerade erst dem Betroffenen jene Einblicke in die Speicherung und Weitergabe gewähren, die er bisher nicht hatte. Sie soll auch bei bisher nicht bekanntem Umgang mit personenbezogenen Daten für Transparenz sorgen. Daraus entsteht eine Fürsorgepflicht der verantwortlichen Stelle gegenüber dem Betroffenen, die Daten so zu erheben, zu speichern und zu verwenden, dass eine vollständige Beauskunftung im Nachhinein noch möglich bleibt.

Diese Fürsorgepflicht entspricht den Interessen des Betroffenen und bildet i. V. m. § 31 BDSG die Grundlage für eine Speicherung von Herkunft, Empfänger und Identifikatoren über den eigentlichen Zweck der Basisdaten hinaus (siehe auch Abschnitt 3.6). Eine ergänzende Erhebung personenbezogener Daten ist jedoch ausgeschlossen.[49] Es geht darum, die Struktur der Ablage personenbezogener Daten so zu gestalten, dass der Personenbezug nicht nur bei der Verwendung durch die verantwortliche Stelle, sondern auch im Zuge der Auskunft herstellbar ist. Dies gilt insbesondere bei pseudonym gespeicherten Daten.

Die spezialgesetzliche Regelung des § 13 Abs. 7 TMG verpflichtet explizit zu einer Auskunftserteilung über zu einem Pseudonym gehörende personenbezogene Daten. Ein entsprechendes Auskunftsersuchen kann im Geltungsbereich des TMG allein über das Pseudonym stattfinden, ohne die eigentliche Identität preiszugeben. Die Authentizität des Pseudonymverwenders kann am besten mit Hilfe eines gemeinsamen Geheimnis-

[45] Dix in: Simitis, BDSG 2014, § 34 Rn. 43 m. w. N. Gola/Schomerus, BDSG 2015, § 34 Rn. 6 für eine Prüfungspflicht mit der im Verkehr erforderlichen Sorgfalt.
[46] § 43 Abs. 2 Nr. 1 i. V. m. Abs. 3 S. 1 BDSG.
[47] Heinemann/Wäßle, MMR 2010, 600 (603).
[48] Dix in: Simitis, BDSG 2014, § 34 Rn. 14.
[49] BVerfGE 109, 279 (365).

ses überprüft werden.[50] Diese vorausschauende Regelung für die digitalen Telemedien lässt erkennen, wie eine zukünftige automatisierte Auskunft generell ausgestaltet werden könnte.

In Situationen, in denen personenbezogene Daten nicht mit einem authentifizierbaren Pseudonym, einem Berechtigungsnachweis (Credential) oder einem anderen eindeutigen Identifikator verknüpft werden können oder solch eine Verknüpfung den Interessen des Betroffenen widersprechen würde, kann kein vollständiger Auskunftsanspruch gewährleistet und gerechtfertigt werden. Wurden Daten vom Betroffenen oder einem Dritten erhoben oder an den Betroffenen oder einen Dritten weitergegeben, und sind diese Vorgänge nur einem eingeschränkten Personenkreis bekannt, kann ein Bezug auf diese Vorgänge als ergänzende Information für ein Auskunftsersuchen ausreichen. Damit lässt sich jedoch keine vollständige Auskunft erreichen, sondern nur solch eine, die die mit dem Vorgang verknüpften Daten umfasst.

Die Identifizierung des Auskunftsberechtigen wird in der DSGVO den selben Stellenwert haben, wie im BDSG. In Art 12 Abs. 6 bestimmt die DSGVO, dass bei begründeten Zweifeln über die Identität des Betroffenen zusätzliche Informationen zur Identitätsfeststellung eingeholt werden können. Kann die verantwortliche Stelle den Betroffenen mit bereits gespeicherten Daten nicht identifizieren, so ist sie nach ErwGr 57 allerdings nicht verpflichtet, weitere Informationen einzuholen. Die verantwortliche Stelle darf sich jedoch nicht weigern, solche Informationen anzunehmen, soweit sie ihr vom Betroffenen übermittelt werden. Vor einer Auskunft hat die verantwortliche Stelle nach ErwGr 64 alle vertretbaren Mittel, insbesondere Online-Kennungen, zu nutzen, um die Identität des Auskunftsersuchenden zu überprüfen und sicherzustellen. Eine präventive Speicherung von Identifikationsmerkmalen, allein für den Zweck der Auskunft, sollte sie indessen nicht vornehmen.

3.5 Art und Weise der Auskunftserteilung

Die Auskunft hat umfassend zu erfolgen. Eine unvollständige oder unrichtige Auskunft ist nach § 43 Abs. 1 Nr. 8a i. V. m. Abs. 3 S. 1 BDSG bußgeldbewehrt. Eine Differenzauskunft[51] ist nur bei Einwilligung des Betroffenen zulässig.[52] Sie würde den Betroffenen sonst überproportional belasten. Unter Berücksichtigung möglicher Assistenzsysteme für Betroffene, die Differenzauskünfte auf Seiten des Betroffenen zu einer einheitlichen Auskunft zusammenfügen, kann unter Einwilligung des Betroffenen eine Differenzauskunft

[50] Dix in: Simitis, BDSG 2014, § 34 Rn. 45.

[51] Eine Differenzauskunft ist eine Auskunft über die Änderungen im Bestand der gespeicherten Daten und über neu hinzugekommene Datenweitergaben seit der letzten Auskunft. Sie stellt das Delta zwischen einer vollständigen Auskunft und den bereits erfolgten Auskünften dar. Sie ist einem inkrementellen Backup von der Funktionsweise her ähnlich.

[52] Dix in: Simitis, BDSG 2014, § 34 Rn. 21; a. A. siehe Fn. 53, 54 a. a. O.

dennoch für beide Seiten die ökonomischere Lösung sein. Der Auskunftsanspruch umfasst alle Daten, also auch die Daten, von denen der Betroffene bereits weiß, dass sie bei der verantwortlichen Stelle gespeichert sind.[53]

Bei Zweifel bezüglich der Vollständigkeit besteht vor Gericht die Möglichkeit der Verurteilung auf Erklärung der Vollständigkeit der Auskunft an Eides statt.[54]

Die Auskunft hat unverzüglich, ohne schuldhaftes Zögern (§ 121 Abs. 1 S. 1 BGB) zu erfolgen. Die im Geschäftsverkehr üblichen Fristen müssen allerdings eingeräumt werden.[55] Lediglich besondere Umstände lassen längere Fristen zu.[56]

Eine Fristbestimmung, wie sie beispielsweise im irischen Datenschutzrecht[57] oder in der DSGVO[58] zu finden ist, findet sich im deutschen Datenschutzrecht bisher nicht. Die Auskunft ist jedenfalls innerhalb solch einer Zeitspanne zu erteilen, die es ermöglicht, den Zweck des Auskunftsersuchens noch zu erreichen. Falls eine besondere Dringlichkeit vorliegt, hat der Auskunftsberechtigte darauf hinzuweisen. Die verantwortliche Stelle hat dann im Rahmen der Verhältnismäßigkeit zu prüfen, ob ihr eine entsprechende Beschleunigung der Bearbeitungsvorgänge zumutbar ist.

Auch wenn der Betroffene keine terminlichen Zwänge vorbringen kann, hat er in jedem Fall das Recht auf eine zügige Bearbeitung nach Treu und Glauben. Der vorgeschlagene Rahmen bewegt sich von wenigen Tagen bis zu 3 Monaten.

Die Auskunft ist nach § 34 Abs. 8 und § 19 Abs. 7 BDSG unentgeltlich. Ausnahmeregelungen gibt es allein für Auskunfteien (§ 34 Abs. 8 S. 3 ff. BDSG).

Die Auskunft muss nicht durch die verantwortliche Stelle selbst erfolgen. Sie kann auch einen Auftragsdatenverarbeiter anweisen, die Auskunft in ihrem Namen und Auftrag zu erteilen.[59]

Das Verfahren zur Auskunftserteilung ist in § 34 Abs. 1 S. 2 BDSG nur rudimentär geregelt. Damit besteht für die verantwortliche Stelle die nötige Flexibilität das Auskunftsverfahren an die eigenen Geschäftsprozesse anzupassen. Eine elektronische Plattform, auf der permanent eine aktuelle Auskunft erhalten werden kann, ist nicht vorgeschrieben. Für eine Schaffung einer permanent verfügbaren Online-Auskunft haben sich jedoch schon in den Anfangstagen der Internetwirtschaft Stimmen erhoben.[60]

Seit der Novellierung des BDSG im Jahr 2009 ermöglicht § 34 Abs. 6 BDSG die Auskunft

[53] Gola/Schomerus, BDSG 2015, § 19 Rn. 4; a. A. LAG Hessen, Urteil vom 29.01.2013, BeckRS 2013, 67364.
[54] AG Geislingen, RDV 2004, 178.
[55] Gola/Schomerus, BDSG 2015, § 34 Rn. 16.
[56] Dix in: Simitis, BDSG 2014, § 34 Rn. 42.
[57] 40 Tage – Abschnitt 4 (1) (a) Irish Data Protection Act, 1988.
[58] 1 Monat – Art. 12 Abs. 2 S. 1 DSGVO; mit einer Verlängerungsmöglichkeit um weitere 2 Monate bei komplexen Anträgen oder einer Vielzahl weiterer Anträge.
[59] Gola/Schomerus, BDSG 2015, § 34 Rn. 13.
[60] Roßnagel/Pfitzmann/Garstka 2001, 174 f.

in Textform[61] nach § 126b BGB (vormals: Schriftform), soweit nicht aufgrund besonderer Umstände eine andere Form angemessen ist. Dadurch ist grundsätzlich die Wahl einer elektronischen Beauskunftung gewährleistet.[62] Erforderlich ist jedoch die Sicherstellung der dauerhaften Wiedergabe. Die reine Bereitstellung der Auskunft auf einer Webseite ist deshalb nicht ausreichend.[63] Es liegt außerhalb der Einflusssphäre des Auskunftsverpflichteten, ob der Betroffenen die Auskunft herunterlädt und abspeichert oder nicht.[64]

Sobald jedoch eine integritätssichernde Download-Möglichkeit zur Verfügung gestellt wird, wird dem Empfänger die dauerhafte Wiedergabe ermöglicht. Entscheidend ist, ob die Auskunft als Zugegangen zu werten ist. Es kann nicht darauf abgestellt werden, dass der Auskunftsverpflichtete keinerlei Einfluss auf die Akzeptanz des Download-Vorschlags hat. Wird die Webseite als regelmäßiges Kommunikationsmedium zwischen den Beteiligten genutzt oder wird der Download in direktem Zusammenhang mit einem elektronischen Auskunftsersuchen auf der Webseite angeboten, kann davon ausgegangen werden, dass schon ein passwortgeschützter Bereich auf der Webseite als Herrschaftsbereich des Auskunftsersuchenden zu werten ist.[65]

Gibt es keinen entsprechenden Bereich, wäre ein weiterer möglicher Lösungsweg zumindest die Initialisierung des Downloads der aktuellen Auskunft vor der Nutzung einer interaktiven Plattform zu erzwingen. Eine solche Maßnahme ist jedoch weder im Interesse des Betroffenen noch des Auskunftsverpflichteten. Die Textform soll den Betroffenen schützen und ihn nicht in der Art der Wahrnehmung seiner Rechte einschränken. Deshalb scheint eine weite Auslegung des Begriffs der besonderen Umstände aus § 34 Abs. 6 BDSG angebracht. Ist es den Umständen nach im Interesse aller Beteiligten von der Erzwingung der Textform abzusehen, ist dementsprechend zu verfahren. Ein normaler Download-Link ist deshalb ausreichend.

Die DSGVO erlaubt bzw. fordert zukünftig eine elektronische Antwort, soweit die Anfrage elektronische gestellt wurde (Art. 12 Abs. 3 S. 4 DSGVO). Insofern werden die Anforderungen an die verantwortliche Stelle eher entschärft.

Besondere Vorsicht bei der Wahl der Form ist bei besonders sensiblen personenbezogenen Daten (z. B. medizinischen Daten) angebracht. Eine mündliche Auskunft erlaubt es dem Betroffenen eher, seine personenbezogenen Daten vor Dritten zu verheimlichen. Besteht diese Gefahr nicht, sind besondere Vorsichtsmaßnahmen bei der Übermittlung der besonders sensiblen Daten zu treffen. Eine Übermittlung im verschlossenen Umschlag

[61] Abweichend obliegt die Form der Auskunft öffentlicher Stellen nach § 19 Abs. 1 S. 4 BDSG pflichtgemäßem Ermessen.

[62] Nach § 13 Abs. 7 S. 2 TMG ist auf Verlangen des Nutzers immer die elektronische Form zu wählen.

[63] KG Berlin, CR 2006, 680; Meents/Hinzpeter in: Taeger/Gabel, BDSG 2013, BDSG § 34 Rn. 41; a. A. für den Anwendungsbereich des TMG Moos in: Taeger/Gabel, BDSG 2013, TMG § 13 Rn. 59.

[64] Dix in: Simitis, BDSG 2014, § 34 Rn. 49.

[65] Robrecht 2015, 20 hält bereits die Zusendung eines zugangsgeschützten Links für ausreichend.

oder in digital verschlüsselter Form erhöht die Vertraulichkeit und ist für solche Datenarten unumgänglich.

Je nach Charakter der Daten ist namentlich für medizinische Daten eine interaktive Einsichtsgewährung vorzuziehen.[66] Medizinische Informationen sind erklärungsbedürftig und verlangen eine Vermittlung durch fachkundiges Personal. Eine rein faktenbasierte Auskunft kann zu Fehlschlüssen und einer Gefährdung von Patient und Behandlungserfolg führen.[67]

Die Art der Auskunft ist an den gesundheitlichen Bedürfnissen des Patienten auszurichten. Es ist die Ansprache derjenigen Sinne erforderlich, die noch am besten funktionieren (Sehen, Hören, Fühlen).[68]

3.6 Speicherfrist

Eine generelle Festlegung der erforderlichen Speicherfristen für die vom Auskunftsanspruch umfassten Informationen ist durch den Gesetzgeber bisher nicht erfolgt. Es obliegt deshalb der verantwortlichen Stelle eine angemessene Frist festzulegen.[69] Eine Orientierung bietet die Speicherfrist der Basisdaten. Die Speicherfrist der Basisdaten ergibt sich aus den gewählten Löschfristen.[70] Nach ErwGr 39 DSGVO ist die verantwortliche Stelle künftig angehalten, solche Löschfristen oder regelmäßige Überprüfungen für personenbezogene Daten vorzusehen. Die Löschfristen sind gemäß Art. 25 Abs. 2 S. 2 DSGVO automatisiert durchzusetzen oder organisatorisch zu gewährleisten.

Werden Daten länger gespeichert, als von einer gesetzlichen Aufbewahrungsfrist vorgeschrieben, sind sie für den längeren Zeitraum zu beauskunften.[71]

Für einzelne Sonderfälle finden sich Anhaltspunkte für die Speicherfrist im Gesetz. Bei der Übermittlung listenmäßiger personenbezogener Daten zum Zwecke der Werbung gemäß § 28 Abs. 3 S. 4 BDSG gilt nach § 34 Abs. 1a BDSG eine Speicherfrist für Herkunft und Empfänger von zwei Jahren nach der Übermittlung. Scoring-Daten unterliegen einer Speicherfrist von sechs Monaten gemäß § 34 Abs. 2 S. 1 Nr. 1 BDSG. Wer personenbezogene Daten zum Zweck der Übermittlung erhebt, speichert oder verändert, hat nach § 34 Abs. 4 S. 1 Nr. 1 BDSG die innerhalb der letzten zwölf Monate vor Zugang des Auskunftsverlangens übermittelten Wahrscheinlichkeitswerte (Scores) und deren Empfänger zu beauskunften und dementsprechend auch zu speichern. Eine Vereinbarkeit dieser

[66] Scheiwe 1998, 319.
[67] BGH, NJW 1983, 328 (329 f.).
[68] BVerfG, NJW 2005, 1103 (1104).
[69] Dix in: Simitis, BDSG 2014, § 34 Rn. 23.
[70] Für solche Löschfristen wurde die Norm DIN 66398:2016-05 „Leitlinie zur Entwicklung eines Löschkonzepts mit Ableitung von Löschfristen für personenbezogene Daten“ entwickelt, die sich an den durch Hammer, DuD 2011, 890 dargelegten Verfahrensweisen orientiert.
[71] BGH, BeckRS 2001, 30158545.

einfach-gesetzlichen Regelungen mit der Rechtsprechung des EuGH, insbesondere bei längerer Speicherung der Basisdaten, ist zweifelhaft.[72]

Das Recht auf Auskunft gilt auch für die Vergangenheit und die gewählte Speicherfrist muss die Ausübung dieses Rechts wirksam ermöglichen.[73] Zwischen den Rechten des Betroffenen und der Belastung für die verantwortliche Stelle muss ein angemessener Ausgleich stattfinden.[74]

Wenn sich die Speicherfrist der für die Auskunft relevanten Informationen allein an der entsprechenden Frist der Basisdaten orientiert, wären Datenpfade nach Löschung der Basisdaten nicht mehr reproduzierbar. Damit wird dem Betroffenen die Möglichkeit genommen, Schritt für Schritt nachzuvollziehen, welchen Weg seine personenbezogenen Daten genommen haben und wer sie aktuell vorhält. Dementsprechend darf die Angabe der Empfänger erst dann gelöscht werden, wenn auch die Basisdaten bei Empfängern und Folgeempfängern gelöscht sind.

Für die solcherart gespeicherten Daten gilt eine besonders strenge Zweckbindung an die Auskunftserteilung. Für alle anderen Zwecke sind sie zu sperren (§ 34 Abs. 5 BDSG).

3.7 Umfang des Rechts auf Auskunft

Die grundlegenden Regelungen zum Recht auf Auskunft sind für öffentliche und nicht-öffentliche Stellen identisch. § 19 Abs. 1 S. 1 sowie § 34 Abs. 1 S. 1 BDSG entsprechen sich und sind in Teilen sogar wortgleich. Einzig das Abstellen auf den Antrag als Willenserklärung ist eine Besonderheit des öffentlichen Rechts.

Die vorgenannten Rechtsnormen geben den Umfang des Auskunftsanspruchs vor. Der Auskunftsanspruch umfasst nach § 19 Abs. 1 S. 1 bzw. § 34 Abs. 1 S. 1 BDSG[75]

- die gespeicherten personenbezogenen Daten,
- ihre Herkunft,
- die Empfänger der Daten und
- den Zweck der Speicherung.

Art. 15 Abs. 1 DSGVO ergänzt die Auskunft zukünftig um die geplante Speicherdauer, die Kategorien der personenbezogenen Daten, alle Verarbeitungszwecke sowie um eine

[72] Dix in: Simitis, BDSG 2014, § 34 Rn. 32.
[73] EuGH, EuZW 2009, 546 (549).
[74] EuGH, EuZW 2009, 546 (550).
[75] § 83 SGB X unterscheidet sich inhaltlich nicht wesentlich von den Vorgaben des BDSG; § 13 TMG und § 93 TKG verweisen jeweils auf § 34 BDSG.

Unterrichtung über die Rechte des Betroffenen und die Garantien bei Drittlandübermittlungen.[76]

Herkunft- und Empfängerangaben unterschiedlicher verantwortlicher Stellen ergänzen sich zu einer durchgängigen Speicher- und Verarbeitungshistorie eines jeden personenbezogenen Datums. Jeder Übermittlung, ausgehend von der einen verantwortlichen Stelle, steht eine Erhebung bei einer anderen verantwortlichen Stelle gegenüber.[77]

3.7.1 Gespeicherte personenbezogene Daten

Legaldefinition und Reichweite des Begriffs

Anknüpfungspunkt des Schutzbereichs des Datenschutzrechts ist das personenbezogene Datum.[78] Denn das Datenschutzrecht soll die Beeinträchtigung der Persönlichkeitsrechte des Einzelnen durch die Erhebung und Verwendung seiner personenbezogenen Daten verhindern.[79] Ist der Umgang mit personenbezogenen Daten nicht transparent, ist es dem betroffenen Bürger nicht möglich, selbst über den Umgang zu verfügen oder ihn zu beeinflussen.

Nach der Legaldefinition des § 3 Abs. 1 BDSG sind personenbezogene Daten alle „Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person [...].“ Die Person ist bestimmt, wenn sich der Bezug zwischen der Person und den Einzelangaben direkt ergibt und eindeutig feststeht.[80] Der Bezug ergibt sich direkt, wenn die Einzelangabe durch die auskunftspflichtige Stelle mit einem eindeutigen Identifikator verknüpft ist.[81]

Ein Identifikator kann ein expliziter Identifikator oder ein aus mehreren Attributen zusammengesetzter Quasi-Identifikator sein. Ein expliziter Identifikator ist selbst ein personenbezogenes Datum. Beispiele sind Kundennummern, Sozialversicherungsnummern, oder Ausweisnummern. Solche Nummern werden im Regelfall selbst erst dadurch zum Identifikator, dass sie, beispielsweise in einer Kundendatenbank, auf Quasi-Identifikatoren (Name, Geburtsdatum, Geburtsort, Adresse, ...) verweisen. Ihr Personenbezug entsteht also durch den Kontext.[82] Einzig charakteristische biometrische Merkmale und genetische Marker sind originäre explizite Identifikatoren.[83]

[76] Hohmann in: Roßnagel 2017, § 3 Rn. 135. Eine tabellarische Übersicht zum Vergleich des inhaltlichen Umfangs des Auskunftsrechts in DSGVO und BDSG findet sich in Laue/Nink/Kremer 2016, § 4 Rn. 24.

[77] Übermittlung und Erhebung bedürfen jeweils einer eigenen Rechtsgrundlage – BVerfGE 130, 151 (184).

[78] Bei personenbezogenen Daten handelt es sich eigentlich um Informationen; vgl. Spiecker genannt Döhmann 2016, 137.

[79] BVefGE 65, 1.

[80] Dammann in: Simitis, BDSG 2014, § 3 Rn. 22.

[81] Art. 4 Nr. 1 DSGVO spricht statt von „Bestimmbarkeit“ von „Identifizierbarkeit“, meint jedoch dasselbe.

[82] Dammann in: Simitis, BDSG 2014, § 3 Rn. 61.

[83] Dammann in: Simitis, BDSG 2014, § 3 Rn. 73.

Die Bestimmbarkeit einer natürlichen Person ergibt sich im Falle des Auskunftsanspruchs aus der Möglichkeit der Zuordnung der Daten zum Betroffenen durch die verantwortliche Stelle. Der Personenbezug ist relativ.[84] Es sind die rechtlichen und tatsächlichen Mittel zu berücksichtigen, die von der verantwortlichen Stelle eingesetzt werden können.[85] Dabei müssen nicht alle zur Identifizierung des Betroffenen notwendigen Informationen in der Hand der verantwortlichen Stelle liegen. Es muss ihr nur möglich sein, sie in rechtlich zulässiger Weise zu einem späteren Zeitpunkt zu erlangen.[86]

Nicht jedes potentiell zuordenbare Datum ist zum Zwecke der Auskunft dem Betroffenen zuzuordnen. Es ist nicht im Sinne des Datenschutzes, dort zwanghaft Bezüge zu natürlichen Personen herzustellen, wo sie für eine verantwortliche Stelle nicht mit vertretbarem Aufwand herstellbar sind und ohne das Auskunftsverlangen auch nie hergestellt würden. Auf der anderen Seite umfasst der Auskunftsanspruch auch Daten, die gegenwärtig noch keinen Personenbezug aufweisen, deren Personenbezug aber im Rahmen bestimmter Prozesse im Regelfall hergestellt wird.[87] Entsprechende Regelungen finden sich insbesondere für das Scoring (§ 34 Abs. 2 S. 2 Nr. 1 BDSG) und für die geschäftsmäßige Speicherung personenbezogener Daten zur Übermittlung (Auskunfteien, Adresshandel – § 34 Abs. 3 S. 2 Nr. 1 BDSG).

Persönliche und sachliche Verhältnisse eines Betroffenen sind in Summe alle Angaben zu einer Person, unabhängig von ihrer jeweiligen Semantik (Bedeutung), Sigmatik (Hinweisfunktion und Zweckbezug) oder Pragmatik (Verwendung und Kontext).[88] Seinen datenschutzrelevanten Gehalt bekommt das Datum schon durch seinen Bezug zur Person. Die Relevanz ergibt sich aus dem Zusammenhang. Kein Datum kann abstrakt als belanglos qualifiziert werden.[89] Sachliche Verhältnisse, die einer Person zugeordnet werden können, aber keine persönlichkeitsrechtliche Relevanz in sich tragen, unterliegen nicht dem Datenschutzrecht.[90] Nach der „3-Elemente-Theorie“ ist eine Relevanz immer dann anzunehmen, wenn das Datum ein Inhaltselement (Information *über* den Betroffenen), ein Ergebniselement (Auswirkung *auf* den Betroffenen) oder ein Zweckelement (Vorhaben, den Betroffenen auf Grundlage des Datums zu beurteilen) in sich trägt.[91]

Eine abstrakte Entscheidung, ob eine Angabe ein personenbezogenes Datum oder ein Teil eines Datums ist, oder ob mehrere Daten vorliegen, ist nicht möglich.[92] Beispielsweise kann man einen Namen als Ganzes als personenbezogenes Datum auffassen. Man kann

[84] Gola/Schomerus, BDSG 2015, § 3 Rn. 10; Dammann in: Simitis, BDSG 2014, § 3 Rn. 32 m. w. N. in Fn. 109; a. A. z. B. Weichert, DuD 2007, 113 (115); Pahlen-Brandt, DuD 2008, 34 (37 ff.).
[85] ErwGr 26 DSRL; ErwGr 26 DSGVO; EuGH, EuZW 2016, 909.
[86] EuGH, EuZW 2016, 909 (911).
[87] Gola/Schomerus, BDSG 2015, § 34 Rn. 8b.
[88] Dammann in: Simitis, BDSG 2014, § 3 Rn. 6.
[89] BVerfGE 65, 1 (45).
[90] Weichert, VuR 2009, 323 (325).
[91] Artikel-29-Datenschutzgruppe 2007, 11 ff.
[92] Dammann in: Simitis, BDSG 2014, § 3 Rn. 133.

jedoch auch seine Bestandteile Titel, Familienname und Vornamen als kleinste Einheit heranziehen. Es ist sogar noch möglich in einzelne Vornamen oder Namensbestandteile zu differenzieren. Entscheidend für die Granularität personenbezogener Daten sind der Verwendungszusammenhang und die allgemeine Verkehrsanschauung.[93] Die technische Realisierung kann nur mittelbar als Ergebnis einer beim Entwurf einer Datenverarbeitung erfolgten Wertung herangezogen werden. Von der Festlegung hängt ab, ob eine Ergänzung im Datenbestand eine Änderung eines personenbezogenen Datums oder eine zusätzliche Speicherung ist.[94]

Einzelne Arten personenbezogener Daten

Im Folgenden werden personenbezogene Daten detailliert beleuchtet, die für das Recht auf Auskunft von besonderer Wichtigkeit sind.

Von Belang ist die *Verortung* eines Datums. Der Ort der Speicherung (IT-System, Akte, Datenbank,...) und Verarbeitung i. e. S. (Prozess) hat selbst Aussagekraft über die betroffene Person.[95] Ist ein Datum im IT-System für die Inkassoabwicklung gespeichert, hat bereits diese Information Einfluss auf die Bonität. Befindet sich der Name einer natürlichen Person in einer Liste wertvoller Stammkunden, hat auch dieser Zusammenhang Aussagekraft, ohne dass es einer weiteren Kennzeichnung bedarf. Besonders zu beachten sind auch Dateinamen. Sie weisen dem personenbezogenen Datum Attribute zu (Erstellungszeitpunkt, Kategorie, o.ä.) und dienen als Selektionskriterien.[96] Ähnliche Schlussfolgerungen lassen sich auch aus der Struktur der Datenablage ziehen. Analog zu § 6a BDSG ist deshalb auch der logische Aufbau der gespeicherten personenbezogenen Daten Teil des Auskunftsanspruchs.[97]

Der *Zeitpunkt* einer Erhebung, Verarbeitung und Nutzung personenbezogener Daten kann auch personenbezogenes Datum sein. Im Falle der Erhebung beim Betroffenen gibt er Auskunft über eine Interaktion zwischen Betroffenem und verantwortlicher Stelle. So kann der Zeitpunkt einer Erhebung beispielsweise einen Kaufzeitpunkt für ein bestimmtes Produkt oder den Zugriffszeitpunkt auf eine Webseite widerspiegeln. Aus solchen Informationen lassen sich umfangreiche Persönlichkeitsprofile bilden. Auch der Zeitpunkt einer Kommunikation mit Dritten (Erhebung und Übermittlung) ist für den Betroffenen von Relevanz. Wann wer was über ihn weiß beeinflusst direkt seine Handlungsmöglichkeiten.

Schwieriger ist der Zeitpunkt der internen Verarbeitung und der Nutzung zu fassen. Da mit personenbezogenen Daten des Betroffenen umgegangen wird, ergibt sich zwar

[93] Dammann in: Simitis, BDSG 2014, § 3 Rn. 133.
[94] Dammann in: Simitis, BDSG 2014, § 3 Rn. 132.
[95] Betrifft auch den geografischen Ort, z. B. beim Cloud Computing, so Dix in: Simitis, BDSG 2014, § 34 Rn. 23.
[96] HessVGH, RDV 1991, 187 (188); Dix in: Simitis, BDSG 2014, § 34 Rn. 17; Gola/Schomerus, BDSG 2015, § 34 Rn. 9 und Mallmann in: Simitis, BDSG 2014, § 19 Rn. 21.
[97] Meents/Hinzpeter in: Taeger/Gabel, BDSG 2013, BDSG § 34 Rn. 17.

zwangsläufig ein Personenbezug, fraglich ist jedoch die Auskunftsrelevanz, vor allem im Vergleich mit den Interessen der verantwortlichen Stelle. Genaue Verarbeitungszeitpunkte legen Betriebsgeheimnisse offen und setzten bei nicht-vollautomatisierter Verarbeitung außerdem die Mitarbeiter der verantwortlichen Stelle unter einen Überwachungsdruck. Wie diese Rechtsgüter abzuwägen sind, kann nur im Einzelfall beantwortet werden.[98]

Innerhalb einer *Kommunikationsbeziehung* sind Herkunft[99] und Empfänger[100] nicht nur auskunftspflichtige Informationen, weil sie explizit aufgeführt sind, sondern auch, weil es sich bei diesen Angaben selbst um personenbezogene Daten des Betroffenen handelt.[101] Dadurch, dass Sender und Empfänger jeweils personenbezogene Daten des Betroffenen verarbeiten, stehen sie in einer Beziehung zu diesem. Sind Herkunft und Empfänger natürliche Personen, ergibt sich ein doppelter Personenbezug[102] und dadurch das Erfordernis einer Abwägung zwischen den Rechten des einen Betroffenen auf Auskunft und den Rechten des anderen Betroffenen auf das Datengeheimnis.

Auch andere Kommunikationskonstellationen und Sozialbeziehungen zwischen natürlichen Personen führen im Regelfall zu einem doppelten Personenbezug. Gleiches gilt für Bilddaten (Fotografien, Videoaufnahmen): Sind auf ihnen mehrere Personen abgebildet, entsteht zu jeder einzelnen Person ein Personenbezug. Ein doppelter oder mehrfacher Personenbezug gespeicherter Daten führt zu einem Auskunftsanspruch jeder betroffenen Person.[103]

*Gesperrte personenbezogene Date*n sind weiterhin gespeicherte Daten und unterliegen damit der Auskunftspflicht.[104] Häufig handelt es sich bei gesperrten Daten jedoch um Daten, die gesetzlichen, satzungsmäßigen oder vertraglichen Aufbewahrungsvorschriften unterliegen. Solche Daten sind nach § 34 Abs. 7 i. V. m. § 33 Abs. 2 S. 1 Nr. 2 nicht verpflichtend zu beauskunften.[105]

Der Auskunftsanspruch umfasst nicht nur die gespeicherten Daten, sondern auch *übermittelte Daten*.[106] Dies erfordert eine durchgängige und feingranulare Historie von Übermittlungen mit Bezug auf gespeicherte Daten.[107] Eine eigene Speicherung der Inhalte übermittelter Daten für sonst bereits gelöschte Daten ist jedoch unverhältnismäßig.

Für *medizinische Daten* hat die Rechtsprechung weitergehende Vorgaben gemacht. Neben dem datenschutzrechtlichen Auskunftsrecht wurde ein zivilrechtliches Einsichtsrecht

[98] Mehr dazu in Abschnitt 3.8.
[99] Mehr dazu in Abschnitt 3.7.3.
[100] Mehr dazu in Abschnitt 3.7.2.
[101] Dammann in: Simitis, BDSG 2014, § 3 Rn. 47; Dix in: Simitis, BDSG 2014, § 34 Rn. 22; a. A. BGH, NJW 1981, 1738 (1739) für das BDSG 1977.
[102] Dammann in: Simitis, BDSG 2014, § 3 Rn. 43, 45.
[103] Zur Abwägungen im Einzelfall siehe Abschnitt 3.8.
[104] Dix in: Simitis, BDSG 2014, § 34 Rn. 19.
[105] Dix in: Simitis, BDSG 2014, § 34 Rn. 57.
[106] Dix in: Simitis, BDSG 2014, § 34 Rn. 23.
[107] Sydow, NVwZ 2013, 467 (470).

entwickelt, welches auch für nicht in Dateien abgelegte Aufzeichnungen gilt.[108] Diese ergänzenden Rechte liegen auch darin begründet, dass medizinische Daten besonders sensible Daten des Betroffenen im Sinne des § 3 Abs. 9 BDSG sind. Speziell vor Abschluss der Behandlung besteht ein gesteigertes Interesse des Betroffenen, die weitere Verwendung seiner personenbezogenen Daten zu kontrollieren und ihren Charakter zu analysieren.[109] Auf der anderen Seite sind berechtigte Interessen des behandelnden Arztes zu berücksichtigen (siehe Abschnitt 3.8.2).

Aufgezeichnete Daten, die den bisherigen Umgang mit einem personenbezogenen Datum beschreiben (*Data-Provenance*), sind auch selbst personenbezogene Daten. Eine Auskunft nach § 34 BDSG ist keine Übermittlung, da sie nicht gegenüber einem Dritten erfolgt. Als Nutzung personenbezogener Daten ist sie dennoch selbst zu beauskunften. So hat der Betroffene auch die Möglichkeit zu überprüfen, ob sich unberechtigte Dritte im Wege der Auskunft Zugang zu seinen personenbezogenen Daten verschafft haben.

3.7.2 Empfänger

Gemäß § 3 Abs. 8 S. 1 BDSG ist Empfänger „jede Person oder Stelle, die Daten erhält." Mit Personen sind sowohl natürliche als auch juristische Personen gemeint. Der Betroffene selbst kann auch Empfänger sein. Der Begriff der Stelle ist komplexer und führt zu einem differenzierten Empfängerbegriff.

Stelle

Der Stellenbegriff des Datenschutzrechtes geht über den verwaltungsrechtlichen Stellenbegriff nach § 1 Abs. 4 VwVfG hinaus. Er bezeichnet auch nicht-öffentliche, private Gebilde. Der Stellenbegriff kann sowohl funktional als auch organisatorisch interpretiert werden.[110]

Der funktionale Stellenbegriff ist zweckbestimmt. Er macht den Übergang eines Datums von einer Stelle auf eine andere am Wechsel der Aufgabe bezüglich der Verarbeitung und Nutzung der Daten fest. Ein IT-System oder eine IT-Applikation ist dann einer dezidiert anderen Stelle zuzuordnen, wenn ihm eine inhärent andere Rolle und Funktionalität in der Datenverarbeitung zukommt.

Der organisatorische Stellenbegriff ist von der juristischen Person bzw. dem Rechtsträger verschieden.[111] Er orientiert sich an der aufbauorganisatorischen Struktur einer Organisation. In hierarchischen Organisationen geht ein Stellenwechsel meist mit einer Änderung der Weisungsbefugnis, zumindest auf der niedrigsten Verantwortungsebene, einher. Filialen oder Betriebe sind keine Stellen.[112] Die räumliche Struktur einer Organisation ist

[108] BGH, NJW 1983, 328; BGH, NJW 1983, 330.
[109] BGH, NJW 1985, 674 (675).
[110] Dammann in: Simitis, BDSG 2014, § 2 Rn. 15.
[111] Dammann in: Simitis, BDSG 2014, § 3 Rn. 231.
[112] Dammann in: Simitis, BDSG 2014, § 3 Rn. 233.

für den organisatorischen Stellenbegriff unerheblich. Eine beschäftigte natürliche Person ist, soweit sie in ihrer dienstlichen Funktion handelt, dem Arbeitgeber bzw. Dienstherren zuzuordnen und damit keine eigene Stelle.[113]

Indem der Gesetzgeber die Zweckbindung lückenlos gewährleistet[114] folgt er im Ergebnis dort dem funktionalen Stellenbegriff,[115] wo umfassende Nutzungsregelungen fehlen.[116]

Zusammenfassend kann festgestellt werden, dass der Empfängerbegriff alle Organisationseinheiten,[117] denen die Daten zur Nutzung zur Verfügung gestellt werden,[118] und alle dem Zweck nach vom Sender verschiedene IT-Systeme und IT-Applikationen innerhalb der verantwortlichen Stelle umfasst.[119] „Offenzulegen ist somit auch der interne Datenfluss und zwar auch bei nicht-automatisierter Verarbeitung."[120]

Die DSGVO führt eine rechtsdefinitorische Klärung des Empfängerbegriffs herbei. Die in den obigen Ausführungen zum BDSG vertretene Interpretation wird bestätigt. In Art. 4 Nr. 9 DSGVO ist festgehalten, dass jede Person oder Stelle Empfänger ist, der gegenüber personenbezogene Daten offengelegt werden, „unabhängig davon, ob es sich bei ihr um einen Dritten handelt, oder nicht".[121] Eine in der Auslegung offene Frage ist, ob alle Organisationseinheiten innerhalb eines Unternehmens Empfänger sind oder ob eine gewisse Eigenständigkeit verlangt wird.[122] Im Ergebnis sollte die Eigenständigkeitserwägung zu den selben Resultaten kommen, wie die organisatorische und funktionale Interpretation des Stellenbegriffs.

Dritte

Nach § 3 Abs. 8 S. 2 BDSG „ist jede Person oder Stelle außerhalb der verantwortlichen Stelle" Dritter.[123] Ausgenommen sind der Betroffene (§ 3 Abs. 8 S. 3 Alt. 1 BDSG) und der Auftragsdatenverarbeiter (§ 3 Abs. 8 S. 3 Alt. 2 BDSG). Beispielhaft sind zwei Behörden, zwei Ämter oder zwei natürliche bzw. juristische Personen Dritte füreinander. Organe

[113] Dammann in: Simitis, BDSG 2014, § 3 Rn. 234.
[114] § 4 Abs. 1 BDSG i.V.m § 14 Abs. 1 für öffentliche Stellen bzw. § 28 Abs. 1 S. 2 für eigene Geschäftszwecke nicht-öffentlicher Stellen.
[115] Dammann in: Simitis, BDSG 2014, § 2 Rn. 16.
[116] Dammann in: Simitis, BDSG 2014, § 2 Rn. 18.
[117] Unabhängig davon, ob sie Dritte sind oder nicht. – EuGH, C-553/07, ECLI:EU:C:2009:293, Rn. 11.
[118] Gola/Schomerus, BDSG 2015, § 3 Rn. 51.
[119] Weitere Ausführungen zur verantwortlichen Stelle finden sich in Abschnitt 3.7.4.
[120] Gola/Schomerus, BDSG 2015, § 34 Rn. 2.
[121] Schreiber in: Plath, BDSG/DSGVO 2016, Art. 4 DSGVO, Rn. 29 geht im Widerspruch zu Art. 4 Nr. 10 DSGVO davon aus, dass Auftragsverarbeiter Dritte sind und begrenzt irrigerweise den Empfängerbegriff auf solche Stellen, die nicht Teil der verantwortlichen Stelle sind.
[122] Ernst in: Paal/Pauly, DSGVO 2017, Art. 4 DSGVO, Rn. 57.
[123] In Zukunft sinngemäß gleich: Art. 4 Nr. 10 DSGVO.

einer verantwortlichen Stelle sind keine Dritten.[124]

Eine organisationsinterne Weitergabe von personenbezogenen Daten zwecks Arbeitsteilung ist keine Übermittlung, da sie innerhalb der verantwortlichen Stelle stattfindet.[125] Handelt es sich bei dem Empfänger personenbezogener Daten um einen Dritten, ist der Tatbestand der Übermittlung erfüllt. Die Weitergabe personenbezogener Daten an einen Empfänger, der nicht Dritter ist, ist als Nutzung zu klassifizieren.[126] Die rechtlichen Voraussetzungen für eine Übermittlung sind strenger als für eine Nutzung. Beide unterliegen jedoch der Zweckbindung. Für den nachgelagerten Auskunftsanspruch ist es unerheblich, ob der Empfänger Dritter ist oder nicht. Sowohl die einfache Weitergabe als auch die Übermittlung personenbezogener Daten sind zu beauskunften. Allerdings ist das Erfordernis einer Auskunft über Empfänger, die Dritte sind, zu personenbezogenen Daten, die aus öffentlichen Quellen entnommen wurden, fraglich.[127]

Während die Auskunft bezüglich der Nutzung und Weitergabe personenbezogener Daten durch die verantwortliche Stelle abschließend erfolgen kann, eröffnet sich durch die Auskunft über eine Übermittlung an Dritte ein neuer Auskunftsbedarf, der an den Empfänger zu richten ist. Ein Empfänger personenbezogener Daten, der Dritter ist, ist wiederum selbst verantwortliche Stelle für die Erhebung, Verarbeitung und Nutzung personenbezogener Daten. Die Relevanz einer Auskunft über Übermittlungen ist für den Betroffenen also ungleich höher. Die in der Auskunft gemachten Angaben müssen eine weitere Inanspruchnahme des Auskunftsrechts beim Empfänger ermöglichen.[128] Es muss deshalb nicht nur eine Adresse oder Kontaktmöglichkeit angegeben werden,[129] sondern es müssen auch all die Informationen, die eine Übermittlung für den Empfänger referenzierbar machen (z. B. Zeitpunkt, Inhalt der Übermittlung, Gegenstelle beim Empfänger), enthalten sein.[130]

Die Auskunft legt auf der anderen Seite die Geschäftsbeziehungen der verantwortlichen Stelle offen und berührt damit im Besonderen Geschäftsgeheimnisse der verantwortlichen Stelle.

Kategorien von Empfängern

Die Beauskunftung gemäß § 34 Abs. 1 S. Nr. 2 Alt. 2 von Kategorien von Empfängern statt der Empfänger selbst erhält ihren Sinn in der Bevorzugung der Weitergabe von personenbezogenen Daten an gleichartige Empfänger zu gleichen Zwecken. Dahinter steht die

[124] Dammann in: Simitis, BDSG 2014, § 3 Rn. 238 ff.
[125] Dammann in: Simitis, BDSG 2014, § 3 Rn. 234.
[126] Gola/Schomerus, BDSG 2015, § 3 Rn. 49; Dammann in: Simitis, BDSG 2011, § 3 Rn. 193
[127] AG Leipzig, BeckRS 2014, 14721.
[128] Gebot des effektiven Rechtsschutzes, BVerfGE 65, 1 (70).
[129] Für das Scoring festgelegt in § 34 Abs. 4 S. 1 Nr. 1 BDSG.
[130] Für die Herkunft von Score-Daten entsprechend § 34 Abs. 2 S. 4 BDSG.

Annahme, dass diese gleichartige Weitergabe auch eine gleichartige, geringfügige Gefährdung mit sich bringt. Damit ist die Zusammenfassung von Empfängern zu Kategorien von Empfängern dann ausgeschlossen, wenn der Betroffene ein Interesse daran hat, die unterschiedlichen Empfänger auseinanderhalten zu können. Dies ist immer dann der Fall, wenn besonders sensible oder nicht-standardisierte Daten weitergegeben werden oder es zulässig ist beziehungsweise damit gerechnet werden muss, dass der Empfänger die Daten selbst weitergibt. Es dürfen also nur Datensenken kategorisiert werden.

In Summa erstreckt sich dieses Privileg im Regelfall auf die listenmäßige Weitergabe personenbezogener Daten (§ 28 Abs. 3 S. 2 BDSG) und Daten, die zur Einsicht bereitgehalten werden (§ 3 Abs. 4 Nr. 3 b) Alt. 1 BDSG). Der Gesetzgeber hat sich indes anders entschieden. Die listenmäßige Übermittlung unterliegt einer Speicherpflicht des konkreten Empfängers (§ 28 Abs. 3 S. 4 i. V. m. § 34 Abs. 1a BDSG).

In der Literatur wird sogar vertreten, dass die verantwortliche Stelle gar kein Wahlrecht mehr hat, sondern immer Empfänger und ihre Kategorien mitteilen muss.[131] Ungeachtet dessen besteht bei der ex ante Angabe im Verfahrensverzeichnis immer die freie Wahl.[132]

3.7.3 Herkunft

Im Gegensatz zu den Empfängern[133] besteht keine Pflicht zur Speicherung von Herkunftsangaben.[134] Es ist nur Auskunft über die Herkunftsangaben zu leisten, die sowieso gemeinsam mit den personenbezogenen Daten gespeichert sind.

Es gibt jedoch eine Reihe von Sonderfällen, in denen die Speicherpflicht verschärft wird. Für Daten, die zum Zweck der Übermittlung gespeichert werden (insbesondere in Auskunfteien, Werbung und im Adresshandel) sind nach § 34 Abs. 1 S. 3 BDSG die verantwortlichen Stellen zur Speicherung der Herkunftsangaben verpflichtet. Werden personenbezogene Daten listenmäßig zum Zwecke der Werbung übermittelt, besteht nicht nur eine zweijährige Pflicht zur Speicherung der Herkunft (siehe Abschnitt 3.6), sondern nach § 28 Abs. 3 S. 4 BDSG müssen sogar Informationen über diejenige Stelle vorgehalten werden, die die Daten erstmalig erhoben hat.

Allerdings kann eine Herkunftsangabe gemeinsam mit dem Zeitpunkt der Übermittlung eine nähere Angabe nach § 34 Abs. 1 S. 2 BDSG sein, die den Verlangenden als Betroffenen ausweist.[135] Der verantwortlichen Stelle sei es deshalb angeraten, Herkunftsangaben über einen ähnlichen Zeitraum zu speichern wie Empfängerangaben. Nur dann kann sie gegen sie gerichteten Auskunftsansprüchen voll umfänglich gerecht werden.

[131] Meents/Hinzpeter in: Taeger/Gabel, BDSG 2013, BDSG § 34 Rn. 21.
[132] Petri in: Simitis, BDSG 2014, § 4e Rn. 10.
[133] Dix in: Simitis, BDSG 2014, § 34 Rn. 23; Gola/Schomerus, BDSG 2015, § 19 Rn. 6; siehe auch Abschnitt 3.6.
[134] Gola/Schomerus, BDSG 2015, § 34 Rn. 10 u. § 19 Rn. 5 und zustimmend Dix in: Simitis, BDSG 2014, § 34 Rn. 22.
[135] Diesen Nachweis hält das LG München II (RDV 2006, 22) für erforderlich.

In letzter Konsequenz spielt die Verpflichtung zur Speicherung der Herkunftsangaben nur zum Erhebungszeitpunkt eine eigenständige Rolle. Weitergaben zwischen Stellen innerhalb der verantwortlichen Stelle sind bereits durch die Speicherung des Empfängers auf der Herkunftsseite eindeutig identifiziert. Die Empfängerangaben der einen Seite sind die Herkunftsangaben der anderen Seite. Die verpflichtende Verfügbarkeit der Herkunftsinformationen, ausgehend vom Empfänger, ergibt sich aus der Globalität des Auskunftsanspruchs in Bezug auf das personenbezogene Datum innerhalb der verantwortlichen Stelle. Der Betroffene hat auch ein Recht darauf, verteilt abgelegte Informationszusammenhänge zu erfahren.

Wird von der verantwortlichen Stelle behauptet, dass die Herkunft personenbezogener Daten in allgemein zugänglichen Quellen liegt, trägt sie im Rechtsstreit die Darlegungslast.[136]

3.7.4 Verantwortliche Stelle

Die verantwortliche Stelle (kurz: Verantwortlicher) ist diejenige Stelle, die Erhebung, Verarbeitung und Nutzung selbst durchführt oder durch andere im Auftrag durchführen lässt (§ 3 Abs. 7 BDSG).[137]

Die Menge der verantwortlichen Stellen ist eine Teilmenge aller als Stellen bezeichneter Einheiten. Eine verantwortliche Stelle kann selbst wieder aus unterschiedlichen Funktions- und Organisationsgliedern bestehen, die für sich selbst auch Stellen, jedoch keine verantwortlichen Stellen sind. Die Ausführungen des § 2 BDSG zu verantwortlichen Stellen verwenden allein den organisatorischen Stellenbegriff. Die verantwortliche Stelle ist somit immer organisatorisch und nicht funktional zu verstehen.

Die Differenzierung in öffentliche und nicht-öffentliche Stellen, in Stellen des Bundes und der Länder ist für den auskunftsrechtlichen Stellenbegriff weitgehend unerheblich. Die Enumeration in § 2 Abs. 4 S. 1 BDSG ist nicht abschließend, sondern exemplarisch zu verstehen.

Jede natürliche Person ist verantwortliche Stelle, soweit sie für sich selbst handelt. § 1 Abs. 2 Nr. 3 BDSG nimmt den Umgang mit personenbezogenen Daten zu ausschließlich persönlichen oder familiären Zwecken aus dem Geltungsbereich des BDSG aus und legt solchen Stellen keine Pflichten auf. Dennoch können diese privilegierten natürlichen Personen Dritte im Sinne des § 3 Abs. 8 S. 2 BDSG sein.

Jede juristische Person ist Stelle. „Auf die Rechtsform der nicht-öffentlichen Stelle kommt es [..] nicht an.“[138] Jede Personengesellschaft, jeder nicht-rechtsfähige Verein ist

[136] AG Düsseldorf, Urteil vom 07.01.2009, MMR 2009, 872.

[137] Nach Art. 4 Nr. 7 DSGVO ist der Verantwortliche diejenige Stelle, die allein oder gemeinsam mit anderen über Zweck und Mittel der Erhebung, Verarbeitung und Nutzung personenbezogener Daten entscheidet.

[138] Dammann in: Simitis, BDSG 2014, § 2 Rn. 118.

Stelle.[139] Empfangende Stellen, die eigene juristische Personen sind, aber dennoch der verantwortlichen Stelle zugeordnet werden, sind die Datenverarbeiter im Auftrag.[140]

Art 26 DSGVO führt die Figur des gemeinsam für die Verarbeitung Verantwortlichen ein. Diese wurde bereits in Abschnitt 3.3 behandelt.

3.7.5 Zweck

Die Aufnahme des Zwecks in die zu beauskunftenden Daten in § 19 Abs. 1 S. 1 sowie § 34 Abs. 1 S. 1 BDSG folgt Art. 12 Lit. a erster Spiegelstrich DSRL.

Ein Zweck ist ein auf ein Ziel hin ausgerichteter Grund für die Erhebung, Verarbeitung und Nutzung personenbezogener Daten. Aus dem Zweck folgt eine zieladäquate Handlung der verantwortlichen Stelle.

Die Angabe des Zwecks ist essentiell, da anhand dessen die Rechtmäßigkeit der Verwendung personenbezogener Daten bemessen wird.[141] Nur mit Hilfe des Zwecks kann der Betroffene bestimmen, ob er weitergehende Rechte hat.

In § 28 Abs. 1 S. 2 BDSG wird der verantwortlichen Stelle die Festlegung des konkreten Zwecks der Verarbeitung und Nutzung zum Zeitpunkt der Erhebung auferlegt. Aus der automatisierten Verarbeitung personenbezogener Daten folgt gemäß § 4e S. 1 Nr. 4 BDSG sogar eine Pflicht zur noch frühzeitigeren Definition des jeweiligen Zwecks. Dieser a priori feststehende Zweck ist dem Betroffenen gegenüber zu beauskunften.

Findet eine zulässige Zweckänderung für die Übermittlung und Nutzung der personenbezogenen Daten nach § 28 Abs. 2 BDSG statt, muss auch dies im Rahmen der Auskunft sichtbar und nachvollziehbar werden.[142] Der Zweckbezug lässt sich insgesamt nur gewährleisten, wenn er auch nach einer Änderung und nach jedem Verarbeitungsschritt erkennbar bleibt.[143] Werden personenbezogene Daten zu mehreren unterschiedlichen Zwecken erhoben, verarbeitet oder genutzt, sind alle Zwecke zu beauskunften.[144]

Da für den Betroffenen nicht ersichtlich ist, ob eine Dateibezeichnung einen Bezug zum festgelegten Zweck hat oder den vollständigen Zweck umfasst, kann ein Dateiname allein den Auskunftsanspruch nicht erfüllen.[145] Die Auskunft muss zumindest einen Hinweis enthalten, dass der Zweck im Dateinamen enthalten ist. Selbst in diesem Fall kann ein Dateiname nur den Zweck einer Speicherung, nicht aber den einer anderen Verarbeitung oder Nutzung spezifizieren. Die getrennte Führung des jeweiligen Zwecks erhöht die Verständlichkeit und reduziert die Gefahr einer unbeabsichtigten Modifikation.

[139] Dammann in: Simitis, BDSG 2014, § 2 Rn. 121 f., 132 ff.
[140] Gola/Schomerus, BDSG 2015, § 34 Rn. 11.
[141] So auch ErwGr 60 DSGVO.
[142] Gola/Schomerus, BDSG 2015, § 34 Rn. 12.
[143] BVerfGE 100, 313 (360).
[144] Dix in: Simitis, BDSG 2014, § 34 Rn. 31.
[145] Für Dix in: Simitis, BDSG 2014, § 34 Rn. 31 ist die präzise, zweckspezifische Fassung des Dateinamens ausreichend.

Für die Rechtsgrundlage des Umgangs mit personenbezogenen Daten besteht keine explizite Auskunftspflicht.[146] Ist sie jedoch nicht aus dem Zweck ersichtlich, erhöht eine Nennung die Klarheit und reduziert die Wahrscheinlichkeit von Rückfragen.

Die Auskunft selbst unterliegt der strengen Zweckbindung des § 34 Abs. 5 BDSG. § 34 Abs. 5 BDSG ergänzt die besondere Zweckbindung des § 31 BDSG für die interne Datenschutzkontrolle durch eine Komponente für die externe Kontrolle und Transparenz. Die Regelung folgt damit der allgemeineren in § 6 Abs. 3 BDSG, die Daten über die Ausübung eines Betroffenenrechtes selbst der Zweckbindung unterwirft.

Die DSGVO wird die Zweckbindung etwas aufweichen. Eine Zweckänderung ist nach ErwGr 61 unter Benachrichtigung des Betroffenen möglich, soweit nach Art. 5 Abs. 1 Lit. b DSGVO der Zweck der vorgesehenen Verarbeitung mit dem Zweck der Erhebung vereinbar ist.

Auf der anderen Seite erweitert Art. 15 Abs. 1 Lit. a DSGVO die Auskunft über den Speicherzweck auf eine Auskunft über alle Verarbeitungszwecke.

3.7.6 Logischer Aufbau

Der logische Aufbau der automatisierten Verarbeitung personenbezogener Daten ist gemäß § 6a Abs. 3 BDSG bzw. Art. 12 Lit. a dritter Spiegelstrich DSRL Teil des Auskunftsanspruchs des Betroffenen. Der deutsche Gesetzgeber hat sich zu einer Minimalumsetzung der Richtlinie 95/46/EG entschieden.[147] Die in der DSRL vorgesehene Transparenz über Verarbeitungsvorgänge wird weitestmöglich beschränkt.[148] Nach dem Wortlaut der Richtlinie ist der logische Aufbau automatisierter Verarbeitung „zumindest im Fall automatisierter Entscheidungen“ zu beauskunften. Im deutschen Datenschutz ist dieses Recht im erweiterten Auskunftsanspruch des § 6a Abs. 3 BDSG verwirklicht.

Der logische Aufbau einer automatisierten Verarbeitung ist nicht selbst personenbezogenes Datum. Der diesbezügliche Auskunftsanspruch weicht insofern von den übrigen Auskunftsansprüchen aus § 19 Abs. 1 S. 1 sowie § 34 Abs. 1 S. 1 BDSG ab. Diese beziehen sich allesamt direkt auf personenbezogene Daten.

Zu beauskunften sind die tragenden Funktionsprinzipien des logischen Aufbaus, nicht technische Einzelheiten (Algorithmen).[149] Der Betroffene muss die in das Verfahren eingehenden Informationen und die Ableitung des Endergebnisses nachvollziehen können.[150]

Mit Rücksicht auf Geschäftsgeheimnisse (siehe Abschnitt 3.8) muss die Auskunft keine Angaben über die verwendete Software enthalten.[151]

[146] Dix in: Simitis, BDSG 2014, § 34 Rn. 31.
[147] Dix in: Simitis, BDSG 2014, § 34 Rn. 20.
[148] Dies wird durch Art. 15 Abs. 1 Lit. h DSGVO aufrechterhalten.
[149] Mackenthun in: Taeger/Gabel, BDSG 2013, BDSG § 6a Rn. 23.
[150] Scholz in: Simitis, BDSG 2014, § 6a Rn. 39; Roßnagel 2003, Kap. 9.2 Rn. 138.
[151] BT-Drs. 14/4329, 38.

Zu beauskunften sind die berechneten Wahrscheinlichkeitswerte beim Scoring (§ 34 Abs. 2 S. 1 Nr. 1, Abs. 4 S. 1 Nr. 1 u. 2 BDSG), ihr Zustandekommen (§ 34 Abs. 2 S. 1 Nr. 3, Abs. 4 S. 1 Nr. 4 BDSG) und gemäß § 34 Abs. 2 S. 1 Nr. 2, Abs. 4 S. 1 Nr. 3 BDSG zumindest auch die Art der für die Berechnung verwendeten Daten. Aus § 34 Abs. 2 S. 1 Nr. 3, Abs. 4 S. 1 Nr. 4 BDSG ergibt sich, dass für die konkrete Berechnung der Wahrscheinlichkeitswerte eine alleinige Angabe der Datenkategorien nicht ausreicht. Die Auskunftsverpflichtung umfasst alle eingegangenen Einzeldaten, um dem Betroffenen die Möglichkeit an die Hand zu geben, die Berechnung anhand der eingeflossenen Parameter nachzuvollziehen.[152] Auf der Output-Seite ist die Werteskala der Ergebnisse Teil des Auskunftsanspruchs.[153]

Die eigentliche Scoreformel, der Algorithmus, ist nicht zu beauskunften.[154] Ob Auskunft über die Gewichtung der Eingangsfaktoren und die Zuordnung des Betroffenen zu einer Vergleichsgruppe zu erteilen ist, ist in der Literatur umstritten.[155] Der BGH folgt jedoch der Auffassung, dass keines von beidem Teil des Auskunftsanspruchs ist.[156] Die branchenspezifische Scorecard, die einzelfallunabhängig statistische Größen, Vergleichsgruppen und Berechnungsmodalitäten festlegt, ist dem Betroffenen nicht zugänglich. Eine andere Interpretation würde das Auskunftsrecht des Betroffenen zulasten des Betriebs- und Geschäftsgeheimnisses des Scoring-Unternehmens über das vom Gesetzgeber gewünschte Maß ausdehnen.[157] Eine Auskunft über einzelfallunabhängige Elemente der Scoreberechnung würde dem Betroffenen auch keine Wahrnehmung weiterer Betroffenenrechte ermöglichen, da ihnen der Personenbezug fehlt.[158]

Der Auskunftsanspruch für das Scoring fällt damit hinter den Auskunftsanspruch über den logischen Aufbau einer Verarbeitung im Falle einer automatisierten Einzelfallentscheidung zurück. Kommen Scoring und automatisierte Entscheidung ohne weitere inhaltliche Prüfung zusammen, und kann beides als ein einheitlicher Verarbeitungsschritt verstanden werden, kommt der weitgehendere Auskunftsanspruch zum logischen Aufbau zum Zuge.[159]

Unter Berücksichtigung der abweichenden Intention der DSRL wird schon seit längerem eine Ausweitung des Auskunftsrechts zum logischen Aufbau auf jede automatisierte Datenverarbeitung, insbesondere in komplexen, arbeitsteiligen Verarbeitungsprozessen,[160]

[152] BGH, NJW 2014, 1235 (1236).
[153] Dix in: Simitis, BDSG 2014, § 34 Rn. 33.
[154] BGH, NJW 2014, 1235 (1237).
[155] Stellvertretend für die Auskunftspflicht Dix in: Simitis, BDSG 2014, § 34 Rn. 33; die Gegenauffassung wird u. a. von OLG Nürnberg, ZD 2013, 26 (27), vertreten.
[156] BGH, NJW 2014, 1235 (1237).
[157] BGH, a. a. O.
[158] BGH, a. a. O.
[159] BGH, NJW 2014, 1235 (1238).
[160] Weichert, DuD 2006, 694 (698 f.).

vorgeschlagen.[161] Die Auskunft sollte entsprechend der technischen Entwicklung nicht mehr Abbild eines statischen Zustands sein. Ein umfassender, dynamischer Begriff der Verarbeitung und daraus folgende prozessorientierte Auskunftsansprüche sind geboten.

3.7.7 Recht auf Negativauskunft

Eine Negativauskunft ist eine Bestätigung des Sachverhalts, dass durch die verantwortliche Stelle keine personenbezogenen Daten des Auskunftsersuchenden gespeichert werden oder wurden. Überwiegend wird ein Recht auf solch eine Negativauskunft zugestanden.[162] Dreh und Angelpunkt ist die Frage, ob es bei einer Negativauskunft überhaupt einen Betroffenen bzw. personenbezogene Daten gibt. Nach dem Gesetzeswortlaut wäre dann auch das Recht auf Auskunft nicht einschlägig. Weichert argumentiert jedoch zutreffend, dass die Betroffenheit im Falle der Transparenz weiter zu fassen ist.[163] Denn betroffen ist der, dessen Persönlichkeitsrechte beeinträchtigt werden. Und wie soll der einzelne sich vollkommen frei bewegen können, wenn er einer gegenwärtig nicht vorhandenen Beobachtung nicht gewahr werden kann? Auch die europarechtskonforme Auslegung lässt keinen anderen Schluss zu. Art. 12 Lit. a erster Spiegelstrich der DSRL beinhaltet einen expliziten Anspruch auf Negativauskunft.

3.8 Entgegenstehende Interessen

3.8.1 Geschäfts- und Betriebsgeheimnisse

Die Berücksichtigung von Geschäfts- und Betriebsgeheimnissen nicht-öffentlicher verantwortlicher Stellen bei der Bestimmung des Umfangs einer zu erteilenden Auskunft ist verfassungsrechtlich geboten (siehe Kapitel 2.1). Sie stützt sich europarechtlich auf Art. 13 Abs. 1 Lit. g DSRL. Danach sind Ausnahmen zum Schutz der Rechte und Freiheiten anderer Personen möglich.[164]

Der deutsche Gesetzgeber hat im Rahmen von § 34 Abs. 1 S. 4 BDSG davon Gebrauch gemacht, eine besondere Ausnahme für die Angabe von Herkunft und Empfänger einzuführen. Die Auskunft zu Herkunft und Empfänger legt Geschäftsbeziehungen und Geschäftsprozesse offen.[165] Sie „kann verweigert werden, soweit das Interesse an der Wahrung des Geschäftsgeheimnisses gegenüber dem Informationsinteresse des Betroffenen überwiegt.“ Die Regelung findet sich redundant auch in § 34 Abs. 3 S. 3 BDSG.

[161] Roßnagel/Pfitzmann/Garstka 2001, 171.

[162] Weichert, NVwZ 2007, 1004; Dix in: Simitis, BDSG 2014, § 34 Rn. 18; Mallmann in: Simitis, BDSG 2014, § 19 Rn. 23; AG Leipzig, BeckRS 2014, 14721; Gola/Schomerus, BDSG 2015, § 19 Rn. 4 – Allerdings: ablehnend in Gola/Schomerus, BDSG 2015, § 34 Rn. 5a.

[163] Weichert, NVwZ 2007, 1004 (1004).

[164] Siehe auch ErwGr 41 DSRL.

[165] BT-Drs. 4306, 51.

Die Abwägung der Interessen der verantwortlichen Stelle und des Betroffenen ist einzelfallspezifisch. Während es grundsätzlich keinen Vorrang von Grundrechten untereinander gibt, wird einfachgesetzlich das Recht auf informationelle Selbstbestimmung dem Schutz von Geschäftsgeheimnissen vorgezogen.

Die Kenntnis der Herkunft ist entscheidend, um datenschutzrechtliche Betroffenenrechte und zivilrechtliche Ansprüche (z. B. Schadenersatz) geltend machen zu können. Insbesondere wenn zum Entstehungszeitpunkt falsche Informationen oder sogar verleumdende Aussagen in die Welt gesetzt wurden, muss der Betroffene die Möglichkeit haben, gegen den ursprünglichen Verursacher vorzugehen. Entsprechende Schritte erfordern eine Kenntnis des Verantwortlichen.

Die Geheimhaltung der Geschäftsbeziehung kann nur dann verhältnismäßig sein, wenn „keine begründeten Zweifel an der Richtigkeit der Daten bestehen“[166] und in einer Einzelfallabwägung davon ausgegangen werden kann, dass das Geheimhaltungsinteresse der verantwortlichen Stelle das Auskunftsinteresse des Betroffenen aufgrund besonderer Umstände überwiegt. In Grenzfällen kann eine Begründung des Informationsinteresses vom Betroffenen eingeholt werden, um eine Letztentscheidung zu treffen.[167] Beschließt die verantwortliche Stelle im Ergebnis die Verweigerung der Auskunft über Empfänger personenbezogener Daten, muss sie diesen Beschluss in jedem Einzelfall stichhaltig begründen.[168]

Kategorien von Empfängern genießen den Schutz aus § 34 Abs. 1 S. 4 BDSG nicht.[169] Sie fassen tatsächliche Geschäftsbeziehungen so weit zusammen, dass sie als anonymisiert gelten können. Eine Ableitung der geschäftlichen Gepflogenheiten der verantwortlichen Stelle sind nur noch rudimentär möglich. Die Informationen sind insofern zumindest nicht mehr in einem Maße schützenswert, dass es gegenüber dem Interesse des Betroffenen überwiegt. Auch die geschäftlichen Interessen des Empfängers werden nicht berührt. Er ist in zu Kategorien zusammengefassten Empfängerlisten nicht mehr kenntlich.

Nicht-rechtliche Interessen von Informanten und anderen Quellen personenbezogener Daten sind nicht Teil der Interessenabwägung.[170] Wenn die Interessen eines Dritten Teil der Interessen der verantwortlichen Stelle werden, sind sie in die Interessenabwägung miteinzubeziehen.[171]

Die Erklärung von allgemeinen personenbezogenen Daten zu schützenswerten Geschäftsgeheimnissen, die in die Abwägung mit einzubeziehen sind, würde im Extremfall dazu führen, dass gar keine Auskunft möglich wäre. Gemäß § 34 Abs. 7 BDSG i. V. m. § 33 Abs. 2 Satz 1 Nr. 7 b) BDSG kann aufgrund des Geschäftsgeheimnisses eine Auskunft

[166] Dix in: Simitis, BDSG 2014, § 34 Rn. 27.
[167] Dix in: Simitis, BDSG 2014, § 34 Rn. 27; Sydow, NVwZ 2013, 467 (470).
[168] LfD Hessen, 30. Tätigkeitsbericht, LT-Drs. 15/4659, 40.
[169] Dix in: Simitis, BDSG 2014, § 34 Rn. 25.
[170] LG Ulm, Urteil vom 1.12.2004, MMR 2005, 265, 266.
[171] Dix in: Simitis, BDSG 2014, § 34 Rn. 28.

insgesamt unterbleiben, wenn die Auskunft „die Geschäftszwecke der verantwortlichen Stelle erheblich gefährden würde, es sei denn, dass das Interesse an der Benachrichtigung die Gefährdung überwiegt."

Die Fähigkeit in Kombination mit der Möglichkeit zur Profilbildung, ist der Hauptwettbewerbsvorteil[172] vieler Internetunternehmen wie Suchmaschinen oder sozialer Netzwerke. Sie war es schon immer für Anbieter von Kreditscoringverfahren. Nur wer Zugriff auf umfangreiche personenbezogene Datenmassen besitzt und gleichzeitig die notwendigen Algorithmen beherrscht, kann sie zu Profilen zusammenführen. Folglich sind neben den Algorithmen auch die personenbezogenen Rohdaten in manchen Fallkonstellationen Betriebs- und Geschäftsgeheimnisse. In der Interessenabwägung überwiegt jedoch im Allgemeinen das Auskunftsinteresse des Betroffenen.[173] Im Speziellen ist die wettbewerbliche Bedeutung der Daten miteinzubeziehen.[174] Wahrscheinlichkeitswerte im Scoring zählen, im Gegensatz zur Scoreformel selbst, nicht als Betriebs- und Geschäftsgeheimnisse.[175]

3.8.2 Persönlichkeitsrechte Dritter

Medizinische Daten Dem gesteigerten Interesse des Patienten an der Kenntnis medizinischer Daten steht das berechtigte Interessen des behandelten Arztes gegenüber, persönlich gefärbte Aufzeichnungen geheim zu halten. Auch kann die Bekanntgabe einer Diagnose oder erwarteten Patientenverhaltens Einfluss auf den Behandlungserfolg haben.[176] Deshalb schränkt die Rechtsprechung die Rechte des Betroffenen auf bewertungsneutrale, naturwissenschaftlich-objektive Aufzeichnungen ein.[177] Subjektive Wertungen des behandelnden Arztes sind ausgenommen.[178] Dies gilt jedoch nicht, sobald die Wertungen in Dateien aufgenommen werden und das datenschutzrechtliche Auskunftsrecht greift. In diesem Fall sind auch die Beurteilungen des Arztes mitzubeauskunften.[179] Möchte der behandelnde Arzt persönliche Gedanken geheim halten, hat er sie außerhalb der Patientenakte aufzubewahren und entsprechend zu kennzeichnen.[180]

[172] Alleinstellungsmerkmal eines Unternehmens im Wettbewerb; häufig auch engl. als Unique Selling Proposition bzw. Point (USP) bezeichnet.
[173] BGH, NJW 2014, 1235 (1236).
[174] Spiecker genannt Döhmann 2009, 40 f.
[175] Dix in: Simitis, BDSG 2014, § 34 Rn. 33.
[176] Eine Vortäuschung von Therapieerfolgen durch den Patienten ist dagegen kein grundrechtlich gerechtfertigter Einschränkungsgrund. – BVerfG, NJW 2006, 1116 (1121).
[177] BGH, NJW 1985, 674 (675).
[178] BGH, NJW 1983, 328 (330) bestätigt durch BVerfG, NJW 1999, 1777. Zur entstehenden Abwägungsproblematik siehe auch die Anmerkungen in Abschnitt 3.8.4.
[179] Scheiwe 1998, 324.
[180] Scheiwe 1998, 326.

Veröffentlichung personenbezogener Daten in Telemedien Werden personenbezogene Daten einer Person durch eine Organisation auf deren Webseite veröffentlicht und zum Abruf bereitgehalten, handelt es sich nach § 3 Abs. 4 Nr. 3 b) BDSG um eine Übermittlung, sobald ein Dritter auf die Daten zugreift. Dabei ist unerheblich, ob die personenbezogenen Daten aufgrund der Einwilligung des Betroffenen nach § 4a BDSG oder aufgrund einer anderen Rechtsgrundlage online gestellt werden.

Personenbezogene Daten werden von Unternehmen auf ihrem Internetauftritt präsentiert, um unter anderem Ansprechpartner zu benennen (Mitarbeiterfotos, Namen, u.ä.) oder für eine positive Außendarstellung zu sorgen (Fotos zufriedener Kunden, positive Nutzerkommentare). Staatliche Stellen haben ähnliche Motive.

Neben den veröffentlichten Daten selbst haben auch die Kommunikationsbeziehungen zwischen der verantwortlichen Stelle und den abrufenden Dritten einen Bezug zu der Person, deren Daten veröffentlicht wurden (im Folgenden: referenzierte Person). Der Umgang mit personenbezogenen Daten stellt einen faktischen Personenbezug her (siehe Abschnitt 3.7.1). Ist der Dritte eine natürliche Person, entsteht ein doppelter Personenbezug. Der Abruf der auf der Webseite veröffentlichten Daten wäre damit zunächst sowohl gegenüber dem Abrufenden, als auch gegenüber der referenzierten Person zu beauskunften.

Der Auskunftsanspruch des Abrufenden kann sich auf § 13 TMG stützen, der der referenzierten Person auf § 34 BDSG. Wäre die verantwortliche Stelle redaktionell-journalistisch tätig, würde die Auskunft den Bedingungen des § 57 Abs. 2 RStV unterliegen.[181] Der Anspruch gegenüber dem Infrastrukturanbieter über die Metadaten der elektronischen Kommunikation nach § 93 TKG soll außen vor bleiben.

Bei diesem doppelten Auskunftsanspruch entsteht eine Gemengelage unterschiedlicher Interessen. Aus Sicht des Zugreifenden besteht keine Veranlassung, seinen Abruf der Webseite zu erfassen. Eine anonyme Nutzung des Telemediendienstes gemäß § 13 Abs. 6 TMG würde seinen Interessen am ehesten gerecht. Insofern keine Erhebung personenbezogener Daten über das Faktum des Zugriffs stattfindet, wäre dementsprechend auch keine Auskunft erforderlich.

Erst durch einen möglichen Auskunftsanspruch der referenzierten Person könnte ein legitimer Zweck zur Erhebung und Verarbeitung der Metadaten des Kommunikationsvorgangs entstehen. Da der Abrufende Empfänger personenbezogener Daten des Dritten ist, ist entsprechend der Ausführungen in Abschnitt 3.7.2 ein Auskunftsanspruch zu bejahen.

Die im Recht auf informationelle Selbstbestimmung verankerten Rechte auf Anonymität und auf Kenntnisnahme zweier Grundrechtsberechtigter dürfen sich jedoch nicht gegenseitig ausschließen. Sie sind in einer Gesamtsicht in Einklang zu bringen, die die Rechte aller Beteiligter in Ausgleich bringt. Hat die referenzierte Person der Veröffentlichung ihrer personenbezogenen Daten zugestimmt oder wurde sie darüber unterrichtet, kann sie mit einer Vielzahl von Zugriffen auf ihre personenbezogenen Daten rechnen.

[181] Weichert, VuR 2009, 323 (329).

Es ist praktisch jedem Dritten möglich, von diesen personenbezogenen Daten Kenntnis zu nehmen. Ein detaillierter Auskunftsanspruch der referenzierten Person würde dieser daher ein Wissen über ihre öffentliche Sphäre zukommen lassen, das dem öffentlichen Charakter der personenbezogenen Daten nicht angemessen ist. Der Auskunftsanspruch würde zu einer mittelbaren Totalüberwachung Dritter führen. Dies widerspricht dem Geist des Rechts auf informationelle Selbstbestimmung, das immer in seinem sozialen Kontext betrachtet werden muss. Das Recht der abrufenden Person auf Anonymität wiegt hier stärker.

Den Interessen der referenzierten Person kann dadurch nachgekommen werden, dass über den Tatbestand der Veröffentlichung in allgemeiner Form informiert wird. Die Angabe von Beginn und Ende der Veröffentlichung berührt keine Interessen Dritter und stellt eine wertvolle Information dar. Soweit allgemeine Informationen über Zugreifende in statistisch-anonymisierter Form gesammelt werden, zum Beispiel gruppiert nach Ländern, sind auch diese zu beauskunften.

Wäre der Dritte jedoch keine natürliche Person, sondern eine juristische, würde das Recht auf Anonymität entfallen. Es könnte daher von der verantwortlichen Stelle gefordert werden, diejenigen Zugriffe zu speichern und zu beauskunften, die nicht durch natürliche Personen erfolgt sind. Das Ergebnis wäre jedoch, dass die verantwortliche Stelle für jede IP-Adresse überprüfen müsste, wem diese zuzuordnen ist. Im Endeffekt würde dies dazu führen, dass den zugreifenden natürlichen Personen keine anonyme oder pseudonyme Nutzung mehr möglich ist.

Anders stellt sich die Situation bei erst nach Login verfügbaren Daten dar. Ein Zugriff auf solche hat Einzelfallcharakter, ist mit einer Veröffentlichung nicht zu vergleichen und immer zu beauskunften.

Eine Protokollierung zu Datensicherungszwecken führt zu weiteren Abwägungsproblemen. In diesem Fall entsteht ein Auskunftsanspruch der zugreifenden Person. Entgegenstehende Interessen der referenzierten Person bestehen nicht. Ein Auskunftsanspruch der referenzierten Person würde jedoch auch wieder zu einer über die Datensicherung hinausgehenden Überwachungsmöglichkeit führen. Es liegt nahe, dass deshalb auch in diesem Szenario kein detaillierter Auskunftsanspruch zugebilligt wird.

3.8.3 Weitere Ausnahmen von der Auskunftspflicht

Neben den Regularien zu Geschäfts- und Betriebsgeheimnissen und den erwähnten Konflikten zwischen Persönlichkeitsrechten Dritter bestehen weitere Ausnahmen von der Auskunftspflicht nicht-öffentlicher Stellen gemäß § 34 Abs. 7 BDSG i. V. m. § 33 Abs. 2 S. 1

- Nr. 2 für die Speicherung aufgrund gesetzlicher Aufbewahrungsvorschriften bei unverhältnismäßigem Aufwand der Auskunft, sowie die Speicherung zur Datensicherung oder Datenschutzkontrolle, insbesondere Protokolldaten;

- Nr. 3 falls die Geheimhaltung wegen des überwiegenden Interesses eines Dritten erforderlich ist;[182]
- Nr. 5 in der wissenschaftlichen Forschung bei unverhältnismäßigem Aufwand der Auskunft;
- Nr. 6 falls die öffentliche Sicherheit und Ordnung oder das Wohl des Bundes beziehungsweise eines Landes gefährdet ist;
- Nr. 7 a) wenn die Speicherung für eigene Zwecke erfolgt, die Daten aus allgemein zugänglicher Quelle bezogen wurden und die Auskunft einen unverhältnismäßigem Aufwand erfordern würde.

Für öffentliche Stellen ist die Auskunft über die Übermittlung an Sicherheitsbehörden nach § 19 Abs. 3 BDSG von deren Zustimmung abhängig. Des Weiteren gibt es zum Auskunftsrecht gegenüber Sicherheitsbehörden spezialgesetzliche Regelungen.[183]

Die in § 19 Abs. 4 BDSG aufgeführten Auskunftsverbote reduzieren den Ermessensspielraum der öffentliche Stellen sogar auf Null.[184]

Analog der Regelung für nicht-öffentliche Stellen kann nach § 19 Abs. 2 BDSG das Recht auf Auskunft für personenbezogene Daten, die ausschließlich zu Zwecken der Datensicherung oder der Datenschutzkontrolle gespeichert werden, eingeschränkt werden, soweit durch die Auskunft ein unverhältnismäßiger Aufwand entstehen würde. Der Begriff der Datensicherung umfasst alle Maßnahmen der Datensicherheit nach § 9 BDSG, also auch Mißbrauchsprävention und -kontrolle.[185]

Zu einem anderen Ergebnis kommt man für die Ausnahme des § 33 Abs. 2 Satz 1 Nr. 3 und 7 a) BDSG, die ebenfalls einen unverhältnismäßig hohen Aufwand voraussetzt. Sie kann aufgrund des Einzelfallcharakters der Auskunft im Regelfall nicht geltend gemacht werden.[186] Ein erhöhter Aufwand kann im Allgemeinen nur bei der Benachrichtigung als proaktive, und hierdurch in der Frequenz häufigere, Transparenzmaßnahme auftreten. Die verantwortliche Stelle hat ihre Prozesse effizient, kostensparend und auskunftsfreundlich zu gestalten.[187] An einer solchen hypothetischen Gestaltung, nicht an den tatsächlichen Abläufen und Strukturen, ist der Aufwand zu bemessen.[188]

Die Verhältnismäßigkeitsprüfung kann bei Massenauskunftsersuchen und anderen konzentrierten Aktionen anders ausfallen, als im Regelfall.[189] Im Falle einer missbräuchlichen

[182] Bspw. im Falle der anwaltlichen Schweigepflicht – AG Köln, NJW 2015, 1701 (1701).
[183] Für die Nachrichtendienste des Bundes in § 15 BVerfSchG, auf den § 9 MADG und § 7 BNDG verweisen.
[184] Mallmann in: Simitis, BDSG 2014, § 19 Rn. 75 m. w. N. in Fn. 147.
[185] LAG Hessen, Urteil vom 29.01.2013, BeckRS 2013, 67364.
[186] BVerfGE 113, 26 (60); Dix in: Simitis, BDSG 2014, § 34 Rn. 59.
[187] Sydow, NVwZ 2013, 467 (470).
[188] BSG, NVwZ 2013, 526 (528).
[189] OVG Bremen, NJW 1987, 2393 (2396, 2398).

Verwendung des Auskunftsrechts überwiegen die Interessen der verantwortlichen Stelle an einem ungestörten Betrieb und der Verhinderung einer Ausforschung.

Die in § 33 Abs. 2 Satz 1 Nr. 2 sowie § 19 Abs. 2 BDSG mögliche Aufwandsabschätzung umfasst zusätzlich noch die in sich hohe Zahl der Operationen bei der Protokollierung. Sie wird begleitet durch die strenge Zweckbindung der Speicherung von, zur Datensicherung oder Datenschutzkontrolle erzeugten, Protokollen. Der Aufbau von Protokollen und Metadaten, insbesondere wenn sie für den Auskunftsanspruch selbst erzeugt und gespeichert werden, bietet keinen eigenen Mehrwert für den Betroffenen. Er ist implizit in der Auskunft über die Basisdaten enthalten. Die interne Struktur und Verknüpfungen zwischen Protokollen sind nicht zusätzlich zu beauskunften.[190] Gleiches gilt für den logischen Aufbau von Datenschutzauskunftssystemen.

Zusammengefasst ist festzustellen, dass die weitgehenden Ausnahmen dem eigentlichen Transparenzanspruch des Rechts auf Auskunft zuwiderlaufen. Sie verkomplizieren das Recht auf Auskunft.[191] Ansätze, die Ausnahmetatbestände stark zu reduzieren, sind bisher gescheitert.[192] Im Grundsatz ist eine restriktive Interpretation der Ausnahmen erforderlich.[193]

3.8.4 Anforderungen an den Abwägungsprozess im Einzelfall

Trifft die verantwortliche Stelle allein die Abwägung zwischen zu beauskunftenden und nicht zu beauskunftenden Informationen, entsteht ein problematisches Entscheidungsmonopol.[194] Durch die a priori vorhandene Informationsasymmetrie bleibt es für den Betroffenen unüberprüfbar, welche Informationen ihm vorenthalten wurden. Deshalb sind ablehnende Entscheidungen von der verantwortlichen Stelle immer umfangreich zu begründen.[195]

Wenn über berechtigte Interessen Dritter entschieden werden muss, sind Interessenkollisionen kaum zu vermeiden. Fürsorgepflichten gegenüber gegenwärtigen oder ehemaligen Mitarbeitern der verantwortlichen Stelle, insbesondere wenn es sich um den Schutz vor Strafverfolgung oder zivilrechtlichen Ansprüchen handelt, sind kein berechtigter Grund, um von einer Auskunft über Empfänger personenbezogener Daten abzusehen.[196]

[190] LAG Hessen, Urteil vom 29.01.2013, BeckRS 2013, 67364.
[191] Roßnagel/Pfitzmann/Garstka 2001, 172.
[192] Beispielsweise im Reformentwurf 2001, BT-Drs 14/4329, 18.
[193] Dix in: Simitis, BDSG 2014, § 34 Rn. 62.
[194] Scheiwe 1998, 316.
[195] Dix in: Simitis, BDSG 2014, § 34 Rn. 61; Gola/Schomerus, BDSG 2015, § 34 Rn. 19.
[196] LfD Baden-Württemberg, 31. Tätigkeitsbericht 2012/13, 145.

3.9 Zwischenfazit

Das Recht auf Auskunft ist eng in die Systematik des Datenschutzes, insbesondere der Transparenz- und Betroffeneninterventionsrechte, eingebunden.

Die Bestimmungen über Umfang, Reichweite und Voraussetzungen des Rechts auf Auskunft sind ausgesprochen umfangreich. Sie werden in Kapitel 4.3 wieder aufgegriffen und zusammengefasst.

Hervorzuheben ist, dass der Empfängerbegriff, bestätigt durch die DSGVO, zu einem Anspruch auf Auskunft über die vollständige, durchgängige Verarbeitungskette eines personenbezogenen Datums führt.[197] Für jede Erhebung, Verarbeitung oder Nutzung muss klar sein, woher die personenbezogenen Daten kamen und wohin sie weitergegeben wurden. In allen Schritten ist der Zweck als Prüfstein der Rechtmäßigkeit sichtbar zu machen. Der Betroffene hat das in Hypothese $\mathfrak{H}_2$ angenommenen Recht.

[197] Vgl. Abschnitt 3.7.2.

4 Datenschutzrechtliche Anforderungen an ein Datenschutzauskunftssystem

Anforderungen an ein Datenschutzauskunftssystem sind vielschichtig. Funktionalen Anforderungen, wie sie sich aus Umfang und Reichweite des Rechts auf Auskunft ergeben, stehen neben nicht-funktionale Anforderungen, hergeleitet aus den allgemeinen Prinzipien des Datenschutzrechts. Sie sind gleichzeitig zu berücksichtigen. Die strukturierte Ableitung der Anforderungen ist entscheidend, um zu einem vollständigen Anforderungsprofil zu kommen.

Dieses Kapitel stellt ein neues Modell zur Ableitung rechtlicher Anforderungen vor (Abschnitt 4.1) und präsentiert die Anwendung auf die Anforderungen an ein Datenschutzauskunftssystem. Daraus entsteht eine Aufstellung datenschutzrechtlicher Kriterien (Abschnitt 4.3) und technischer Anforderungen (Abschnitt 4.4), an der die Umsetzung des Datenschutzauskunftssystems gemessen werden kann.

4.1 Top-Down Ableitung von Anforderungen an ein Datenschutzauskunftssystem

Das Fundament der deutschen Rechtsordnung ist das Grundgesetz. Die darin kodifizierten Grundrechte bilden den Ausgangspunkt einer öffentlich-rechtlichen, in gewissem Rahmen auch einer privatrechtlichen,[1] Anforderungsanalyse.

4.1.1 Die Methode KORA

Die Methode KOnkretisierung Rechtlicher Anforderungen (KORA)[2] ist ein mögliches Werkzeug, mit dem, ausgehend von *(Grund-)Rechten*, Anforderungen stufenweise hin zu technischen Gestaltungsvorschlägen konkretisiert werden können.[3]

Nach dieser Methode ergeben sich aus den vorgeschalteten Grundrechten zunächst *rechtliche Anforderungen* (Abbildung 4.1). Solche Anforderungen sind gemäß KORA allgemeingültige Rechtsregeln, die direkt aus den Grundrechten entnommen werden kön-

[1] Siehe Abschnitt 2.1.2.
[2] Hammer/Pordesch/Roßnagel 1993, 46 ff.
[3] Schwenke 2006, 11 ff.
[4] Nach Schulz et al. 2011.

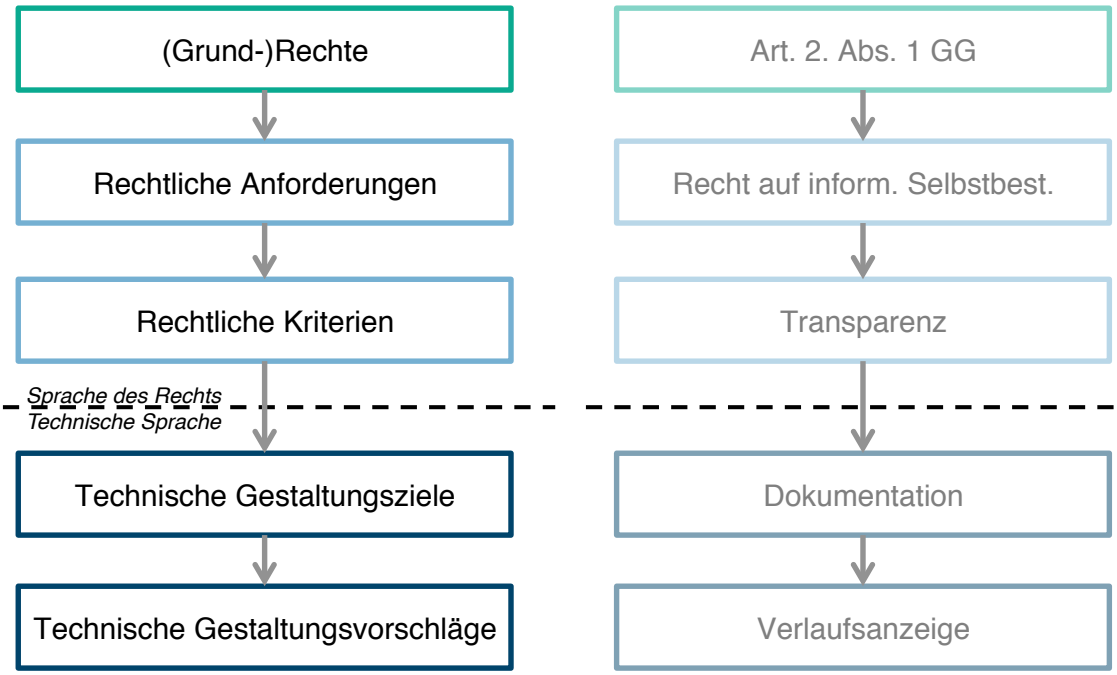


Abbildung 4.1: Die Konkretisierungsschritte der Methode KORA[4](mit Beispiel)

nen und keiner weiteren Ableitung bedürfen.[5] De facto finden sich unter der Kategorie „Grundrechte“ die jeweiligen Artikelnummern des Grundgesetzes. Unter den rechtlichen Anforderungen werden die eigentlichen, ausformulierten Grundrechte verstanden. Die auf die rechtlichen Anforderungen folgenden *rechtlichen Kriterien* sind in ihrem Abstraktionsniveau durch die Methode KORA nicht eindeutig vorgegeben. Gemeinhin werden hierunter die Datenschutz-Schutzziele (siehe Abschnitt 3.1) und rechtliche Grundgedanken (z. B. Zweckbindung) verstanden, die Lösungsvarianten offen lassen.[6]

Der Wechsel zur technischen Sprache erfolgt mit der Methode KORA bereits auf dieser sehr allgemeinen rechtlichen Ebene. Die *technischen Gestaltungsziele* der dritten Stufe beschreiben elementare Funktionen, die notwendig sind, damit ein technisches System den rechtlichen Kriterien gerecht wird.[7] Technische Gestaltungsziele werden in der Anwendung der Methode KORA sehr allgemein gehalten.[8] Ihr technischer Gehalt ist nicht immer erkennbar.

Technische Gestaltungsvorschläge als finale Stufe können nach Auffassung der Autoren als

[5]Schwenke 2006, 13.
[6]Schwenke 2006, 14.
[7]Kahlert, DuD 2014, 86 (88).
[8]Beispielsweise in Schulz et al. 2011 und Kahlert, DuD 2014, 86.

Teil eines Lastenhefts aufgefasst werden.[9] Im Gegensatz zu einem solchen werden jedoch Alternativvorschläge für mögliche, konkret umsetzbare technische Maßnahmen diskutiert, wie sie sich sonst in Pflichtenheften wiederfinden. Die Grenze zwischen Anforderung und Lösung wird an dieser Stelle durchbrochen.

Aufgrund der genannten Kritikpunkte ergibt sich der Bedarf, die Methode KORA stärker auszudetaillieren, die einzelnen Stufen klarer abzugrenzen und auf technischer Ebene funktionale und nicht-funktionale Anforderungen klar zu benennen. Deshalb wurde das erweiterte Vorgehensmodell für Anforderungen aus dem Legalbereich (EVAL) neu entwickelt. Dieses Modell wird im folgenden Abschnitt beschrieben.

4.1.2 EVAL: Erweitertes Vorgehensmodell für Anforderungen aus dem Legalbereich

Fundament des EVAL-Modells sind die *Artikel des Grundgesetzes* (Abbildung 4.2), die auch bei der Methode KORA an erster Stelle stehen. Aus ihnen ergeben sich die *Grundrechte*. Grundrechte sind enumerativ. Sie sind von Rechtsprechung und Kommentarliteratur bereits abschließend ausgearbeitet worden. Sollen datenschutzrechtliche Anforderungen erhoben werden, sind das *Recht auf informationelle Selbstbestimmung*[10] und das *Grundrecht auf Gewährleistung der Vertraulichkeit und Integrität informationstechnischer Systeme*[11] gesetzt.

Anforderungen sind nicht zwingend widerspruchsfrei. Konflikte und Konkurrenzsituationen können auf jeder Stufe auftreten. In EVAL ist es immer zulässig, Konfliktsituationen zu erkennen und auf die darunterliegende Ebene durchzureichen, soweit eine Entscheidung den Detailgrad der weiteren Stufe benötigt. Entgegenstehende Grundrechte Dritter[12] haben Einfluss auf Umfang und Reichweite des Rechts auf Auskunft und damit teilweise auch auf die technische Gestaltung. Sie sind im Abwägungsprozess mitzuberücksichtigen.

Nachfolgend werden die 4 Konkretisierungsschritte der EVAL-Anforderungsanalyse definiert, die nicht bereits enumerativ gegeben sind.

Definition 4.1. *Rechtliche Ziele sind Rahmenvorgaben, die die Grundrechte präzisieren. Rechtliche Ziele lassen sich direkt auf die Grundrechte zurückführen und liegen nicht im Ermessen des Gesetzgebers.*

Rechtliche Ziele sind durch die Rechtsprechung des Bundesverfassungsgerichts vorgegeben. Durch das Bundesverfassungsgericht vorgeschlagene Umsetzungsvarianten der rechtlichen Ziele sind nicht Teil derselben.[13] Eine einfachgesetzliche Regelung, die der

[9] Schwenke 2006, 15.
[10] Siehe Abschnitt 2.1.1.
[11] BVerfGE 120, 274 (274).
[12] Siehe Abschnitt 2.1.5.
[13] Fakultative Regelungsvorschläge finden sich in BVerfGE 65, 1 (46); BVerfGE 100, 313 (361); BVerfGE 109, 279 (364) und BVerfGE 125, 260 (325 f.).

Gesetzgeber genauso hätte nicht treffen können, kann eine rechtliche Anforderung oder ein rechtliches Kriterium sein, jedoch kein rechtliches Ziel. Für den Datenschutz sind die rechtlichen Ziele bereits abschließend durch die Literatur festgelegt. Sie finden sich in Abschnitt 4.2.

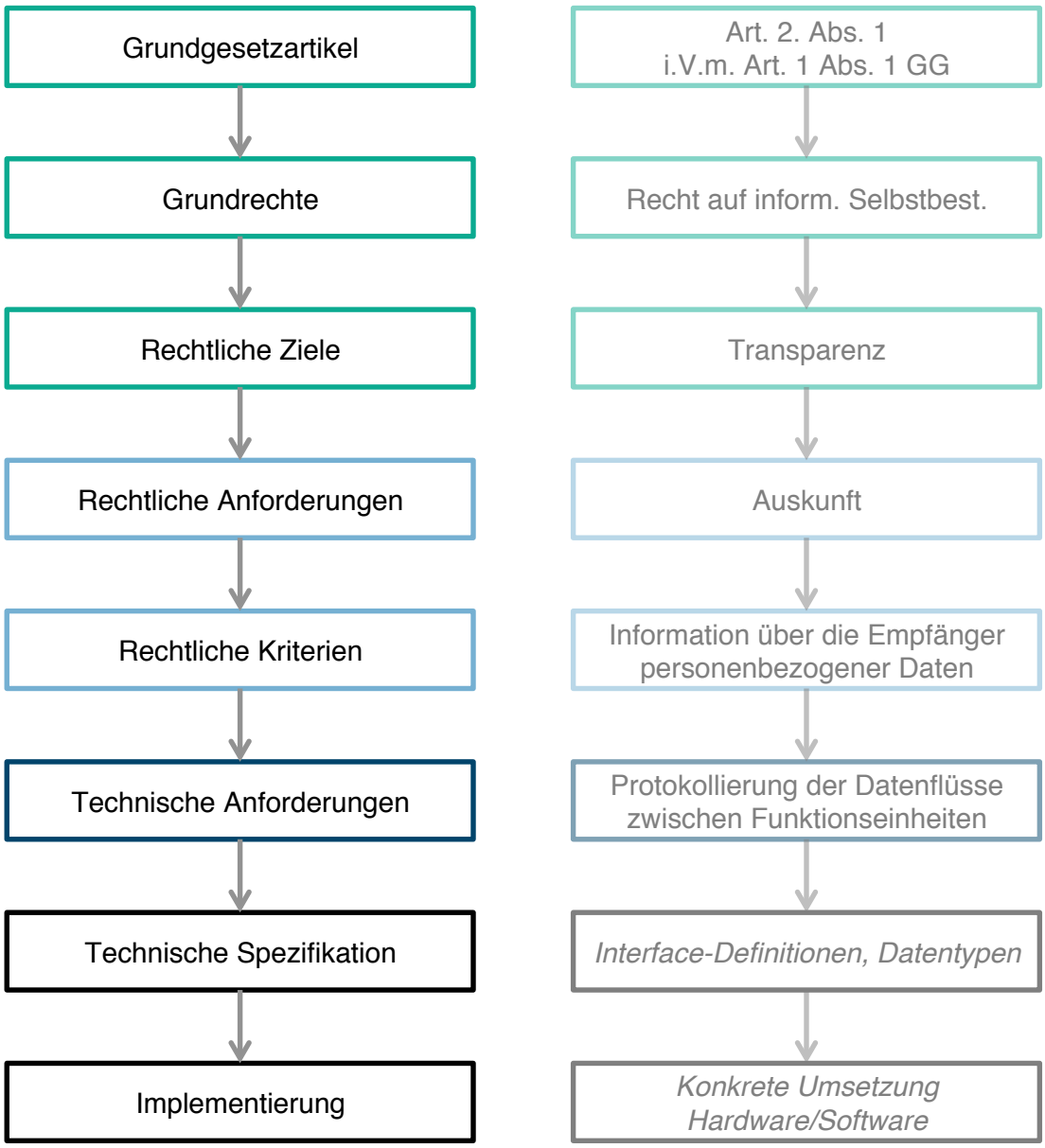


Abbildung 4.2: Die EVAL-Konkretisierungsschritte (mit Beispiel)

Die rechtspraktische Entscheidung des Gesetzgebers kommt auf der Ebene der *rechtlichen Anforderungen* zum Ausdruck und präzisiert im Rahmen der einfachen Gesetze und Verordnungen die rechtlichen Ziele.

Definition 4.2. *Rechtliche Anforderungen sind die generell-abstrakten Bezeichnungen für die Gebote und Verbote, für die sich der Gesetzgeber zur Umsetzung der rechtlichen Ziele in Rechtsnormen entschieden hat.*

Rechtliche Anforderungen leiten sich aus den rechtlichen Zielen ab. Eine rechtliche Anforderung legt dar, „was" der Normadressat zu tun oder zu unterlassen hat. Angaben über Parameter wie Häufigkeit, Zeitpunkt oder Umfang eines Tuns oder Unterlassens sind keine rechtlichen Anforderungen, sondern rechtliche Kriterien. Die in der gegenwärtigen Gesetzgebung niedergelegten datenschutzrechtlichen Anforderungen finden sich in Abschnitt 4.2. Zukünftige rechtliche Entwicklungen können diese verkürzen oder erweitern.

Definition 4.3. *Rechtliche Kriterien sind Verfahrensanweisungen zu oder Angaben über den Umfang rechtlicher Anforderungen im Einzelfall.*

Rechtliche Kriterien konkretisieren rechtliche Anforderungen. Sie beschreiben die Ausgestaltung der einzelnen Anforderungen aus rechtlicher Sicht. Während rechtliche Anforderungen die Frage nach dem „was" beantworten, geben rechtliche Kriterien vor, „wie" die Umsetzung der rechtlichen Anforderungen auszugestalten ist.

Rechtliche Kriterien befinden sich in etwa auf der Ebene von *User Requirements* oder *Soft Requirements*.[14] Sie sind aus der Perspektive des Rechts formuliert. Da eine technische oder organisatorische Umsetzung noch offen gehalten wird, sprechen rechtliche Kriterien im Regelfall nicht von konkreten Daten oder Systemen.[15] Im Gegensatz zu rechtlichen Anforderungen bedient sich die Ableitung rechtlicher Kriterien aus den Rechtsgrundlagen einer einzelfallspezifischen Exegese und Interpretation.[16]

Definition 4.4. *Technische Anforderungen sind die erwarteten technischen Eigenschaften von IT-Systemen, Kommunikationsprotokollen und Datenstrukturen.*

Organisatorische Maßnahmen und Prozesse spielen auf der Ebene technischer Anforderungen keine Rolle mehr. *Technische Anforderungen* entsprechen *System Requirements* oder *Hard Requirements*. Technische Anforderungen bündeln rechtliche Aspekte dort, wo sie aus technischer Sicht gemeinsam umgesetzt werden können. Andererseits differenzieren sich technische Anforderungen dort aus, wo ein einfacher rechtlicher Sachverhalt unterschiedliche technische Aspekte berührt. Die Bestimmung technischer Anforderungen erfolgt

[14] Maiden 2008.

[15] Anders bei Probst, DuD 2012, 439; Probst nimmt keine klare Trennung zwischen rechtlichen Anforderungen, rechtlichen Kriterien und technischen Kriterien vor.

[16] Diese wird für das Auskunftsrecht in Kapitel 3 vorgenommen.

nicht durch rechtliche Auslegung, sondern stützt sich auf die bereits in den rechtlichen Kriterien manifestierten Ergebnisse derselben.

Die technischen Anforderungen gliedern sich deshalb nicht entlang der rechtlichen Anforderungen, sondern entsprechend der üblichen Zweiteilung aus funktionalen und nicht-funktionalen Anforderungen. Die Abgrenzung ist nicht ganz trennscharf.[17] Funktionale Anforderungen sollen beschreiben, was ein Softwareprodukt tun soll. Nicht-funktionale Anforderungen sollen ergänzen, wie, also in welcher Qualität die Leistung erbracht wird.[18]

In allgemeinen Softwareprojekten gehören die aus rechtlichen Kriterien abgeleiteten technischen Anforderungen immer zu den nicht-funktionalen Anforderungen. Anders stellt sich die Situation dar, wenn die Grundfunktion des Softwareprodukts die Erfüllung einer rechtlichen Anforderung ist. Ein Datenschutzauskunftssystem hat die Grundfunktion, Datenschutzauskünfte zu erteilen. Deshalb sind die, aus der rechtlichen Anforderung „Auskunftsanspruch“ mittelbar über rechtliche Kriterien abgeleiteten, technischen Anforderungen funktionaler Art.

Um die Beziehung der Anforderungen in einem Softwareentwicklungsprozess zur Umwelt zu beschreiben, wurden in der Literatur Annahmen (Assumptions) und Fakten (Facts) eingeführt.[19] Annahmen sind für wahr gehaltene, aber unbestätigte Faktoren. Fakten sind bestätigte, objektive Rahmenbedingungen. Annahmen und Fakten sind nicht technisch beeinflussbar und können deshalb limitierend wirken. Im EVAL-Vorgehensmodell werden nicht-technische Umweltbedingungen bereits bei der Exegese der rechtlichen Kriterien berücksichtigt. Im späteren Verlauf gehen Annahmen und Fakten in die Prüfung mit ein, ob die Spezifikation die technischen Anforderungen erfüllt. Annahmen und Fakten in diesem Sinne sind von der Anforderungsanalyse unabhängig.

Bei der Formulierung von Anforderungen muss berücksichtigt werden, dass eine gute Anforderung Lösungsansätze aus dem gesamten Entwurfsraum zulässt.[20] Eine Anforderung sollte so konkret wie nötig, aber so allgemein wie möglich formuliert werden. Ob eine formelle oder informelle Beschreibung der technischen Anforderungen gewählt wird, gibt das EVAL-Modell nicht vor.[21] Dies ist unter anderem abhängig vom jeweiligen Softwareentwicklungsprozess und den organisatorischen Gegebenheiten der umsetzenden Stelle.[22]

[17] Eine Vielzahl von Definitionen im Vergleich finden sich bei Glinz 2007 und Chung/do Prado Leite 2009.
[18] Jureta/Mylopoulos/Faulkner 2008.
[19] Fabian et al. 2010.
[20] Ralph 2013.
[21] Einen Überblick über Methoden des technischen Requirements Engineering bieten Fabian et al. 2010 und Lamsweerde 2000. Eine Einordnung formaler Modellierungsansätze für rechtliche Anforderungen leistet Otto/Anton 2007. Eine Übersicht über Tools zur Formalisierung rechtlicher Anforderungen und sich ergebender Probleme liefern Kiyavitskaya/Krausová/Zannone 2008.
[22] Ob ein formaler Ableitungsprozess von Anforderungen auf Spezifikation und Implementierung praktisch möglich ist, stellen bereits Parnas/Clements 1986 in Frage, sprechen sich jedoch dafür aus, einen solchen Prozess zu simulieren.

Die technische Spezifikation und die Implementierung sind nicht mehr Teil der Anforderungsanalyse. Die technische Spezifikation ist eine formale Modellierung von Architektur, Protokollen, Datentypen und Interface-Definitionen. Die Implementierung ist die Umsetzung der technischen Spezifikation in Soft- und/oder Hardware.

Eine entlang des EVAL-Modells entwickeltes Datenschutzauskunftssystem hat den Anspruch, nicht nur rechtmäßig, sondern vorausschauend rechtsverträglich zu sein.

4.2 Datenschutz-Schutzziele und datenschutzrechtliche Anforderungen

Datenschutz-Schutzziele Für den Datenschutz erfüllen die 6 Datenschutz-Schutzziele[23] die Rolle der rechtlichen Ziele.[24] Sie konkretisieren das Recht auf informationelle Selbstbestimmung und das Grundrecht auf Gewährleistung der Vertraulichkeit und Integrität informationstechnischer Systeme ohne der Gestaltung des Datenschutzrechts im Einzelnen vorzugreifen. Sie haben den Anspruch, die Ziele der Datenschutz-Grundrechte vollständig abzubilden.[25]

Vertraulichkeit und *Integrität* sind Teil eines eigenen Grundrechts, des Grundrechts auf Gewährleistung der Vertraulichkeit und Integrität informationstechnischer Systeme. Die Forderung nach *Verfügbarkeit* ist die Ergänzung zum klassischen IT-Sicherheitsdreiklang,[26] spielt allerdings nur mittelbar eine Rolle als Verfügbarkeit transparenzsichernder Maßnahmen.[27]

Bedeutsam ist das Ziel der *Transparenz* als Recht des Einzelnen, beurteilen zu können, wer wann was über ihn weiß.[28] Die Kenntnisnahme der Maßnahmen zur Verwendung personenbezogener Daten ist grundrechtlich geschützt.[29]

In der *Intervenierbarkeit* manifestiert sich die Befugnis des Einzelnen, selbst über die Preisgabe und Verwendung seiner personenbezogenen Daten zu bestimmen.[30] Aus dem Ziel der Intervenierbarkeit ergibt sich die zwingende Beteiligung unabhängiger Datenschutzbeauftragter.[31]

[23] Rost/Pfitzmann, DuD 2009, 353.
[24] Siehe Abschnitt 3.1.
[25] Bock/Meissner, DuD 2012, 425 (426).
[26] CIA - Confidentiality, Integrity, Availability.
[27] BVerfGE 100, 313 (360); BVerfGE 109, 279 (379).
[28] BVerfGE 65, 1 (43); BVerfGE 125, 260 (334).
[29] BVerfGE 100, 313 (361); BVerfGE 109, 279 (363 f.); siehe auch Abschnitt 2.1.3.
[30] BVerfGE 65, 1 (43).
[31] BVerfGE 65, 1 (46); EuGH, NJW 2010, 1265.

Unverkettbarkeit[32] soll verhindern, dass es staatlichen Behörden und privaten Organisationen möglich ist, ein umfangreiches Persönlichkeitsprofil über jeden Einzelnen zu erstellen.[33] Bisherige Definitionsversuche wie in § 5 Abs. 1 S. 2 Nr. 5 LDSG-SH sind unscharf.[34] Aus diesem Grund wird in Kapitel 9 eine saubere Definition des Begriffs „Grad der Unverkettbarkeit“ vorgenommen.

Aus datenschutzrechtlicher Sicht sind der Unverkettbarkeit insbesondere der Zweckbindungsgrundsatz[35] sowie die informationelle Gewaltenteilung inhärent. Der Begriff der *Gewaltenteilung* ist staatsstheoretischen Ursprungs. Gewaltenteilung ist die Verteilung unterschiedlicher Machtinstrumente auf unterschiedliche Institutionen oder Organe, um Machtmissbrauch vorzubeugen. Die Institutionen oder Organe haben sich auf die ihnen zugewiesenen Aufgaben zu beschränken und sich nicht in den Aufgabenbereich anderer Stellen einzumischen.

Insbesondere Information ist Macht. *Informationelle Gewaltenteilung* ist die Aufteilung der Erhebung und Verwendung personenbezogener Daten zu unterschiedlichen Zwecken auf unterschiedliche Stellen (organisatorisch) oder IT-Systeme (technisch). Der informationellen Gewaltenteilung liegt das verwaltungsrechtliche Abschottungsprinzip zugrunde.[36] Stellen dürfen bei der Erfüllung ihrer Aufgaben nur auf die zum jeweiligen Zweck erforderlichen Informationen zugreifen.[37] Eine Stelle, die mit personenbezogenen Daten umgeht, muss räumlich, organisatorisch und personell von anderen Stellen getrennt werden.[38] Technisch verpflichtet die Gewaltenteilung dazu, dass Zugriffsrechte, Rollen sowie physische und logische Speicherorte nicht beliebig festgelegt werden. Sie sind entsprechend dem Zweck und nach dem Prinzip der Machtdistribution festzulegen. Die (Re-)Kombination personenbezogener Daten ist zu vermeiden.

Von den genannten sechs Datenschutz-Schutzzielen spielen Transparenz und Unverkettbarkeit die wichtigste Rolle für den Entwurf eines Datenschutzauskunftssystems. Die beiden Schutzziele stehen in einem Spannungsverhältnis zueinander. Transparenz erfordert eine ergänzende personenbezogene Sammlung von Protokolldaten und die Aufrechterhaltung eines Betroffenenbezuges für alle personenbezogenen Daten. Jedes Mehr an Daten

[32] Bedner/Ackermann, DuD 2010, 323 (324) ordnen die Unverkettbarkeit der Vertraulichkeit unter, definieren sie allerdings analog zum in dieser Arbeit verwendeten Begriff.

[33] BVerfGE 65, 1 (42).

[34] Unverkettbarkeit wird als ein Zustand definiert, in dem „personenbezogene Daten [..] nicht oder nur mit unverhältnismäßig hohem Aufwand für einen anderen als den ausgewiesenen Zweck erhoben, verarbeitet und genutzt werden [können].“ Dieser Definition folgt das Kompetenzzentrum für angewandte Sicherheitstechnologie (KASTEL) 2013.

[35] BVerfGE 65, 1 (43).

[36] BVerfGE 65, 1 (69); BVerfG, NJW 1988, 959 (961).

[37] Die Informationen müssen für den entsprechenden Zweck erhoben worden sein. Eine Zweckänderung ist nur eingeschränkt möglich. Insofern ist die informationelle Gewaltenteilung bereits zum Erhebungszeitpunkt zu berücksichtigen.

[38] BVerfG, NJW 1988, 959 (960).

erhöht jedoch die Gefahr der Verkettbarkeit. Unverkettbarkeit erfordert, personenbezogene Daten dauerhaft getrennt zu speichern und zu verarbeiten und möglichst keine Daten mit dem Betroffenen in Bezug zu setzen. Dann kann der Betroffene allerdings auch keine Kenntnis vom Umgang mit den Daten nehmen. Transparenz ist nicht sichergestellt.

Datenschutzrechtliche Anforderungen Die rechtlichen Anforderungen ergeben sich im BDSG in vielen Fällen direkt aus den amtlichen Überschriften der Paragraphen.[39] Manche Paragraphen fassen jedoch mehrere Anforderungen zusammen,[40] beschreiben nur die Ausgestaltung (Kriterien) einer an anderer Stelle genannten Anforderung[41] oder wenden sich gar nicht mit Anforderungen oder Kriterien an die verantwortliche Stelle.[42] Manche Anforderungen sind auch in mehreren Paragraphen verwurzelt.[43] Es gibt also keine grundsätzliche Eins-zu-eins-Beziehung zwischen Anforderung und Paragraph des maßgeblichen Gesetzes.

Eine Kurzzusammenfassung der datenschutzrechtlichen Anforderungen und der rechtlichen Ziele, die sie jeweils verfolgen, findet sich in Tabelle 4.1. Für die Synthese der einzelnen Anforderungen, ihre Korrektheit und Vollständigkeit, sei auf Anhang B.1 verwiesen.

Datenschutzrechtliche Anforderungen	Vertraulichkeit	Integrität	Verfügbarkeit	Transparenz	Intervenierbarkeit	Unverkettbarkeit
Datenvermeidung[44]	x					x
Datensparsamkeit	x					x
Anonyme und pseudonyme Dienste						x
Datenschutzfreundliche Grundeinstellungen	x			x	x	
Datengeheimnis	x					
Zutritts- und Zugangskontrolle	x					
Zugriffs- und Weitergabekontrolle	x	x				
Verfügbarkeitskontrolle			x			

[39] Beispielsweise die Betroffenenrechte der §§ 33 bis 35 BDSG.
[40] So wie § 4 und § 35 BDSG.
[41] Beispielsweise § 4e BDSG für die Meldepflicht.
[42] Beispielsweise die Straf- und Bußgeldvorschriften der §§ 43 und 44 BDSG.
[43] Häufig als Trennung zwischen öffentlichen und nicht-öffentlichen Stellen wie die §§ 19 und 34 BDSG für den Auskunftsanspruch.
[44] Einschließlich des Gebots der frühzeitigen Anonymisierung als ex post Datenvermeidung - BVerfGE 65, 1 (49).

Datenkorrektheit		x			
Revisionssichere Protokollierung (Eingabekontrolle)		x	x		
Recht auf Datenübertragbarkeit			x		
Prinzip der Verantwortlichkeit				x	
Prinzip der nicht-automatisierten Einzelentscheidung					x
Einwilligung					x
Einheitliche Ansprechpartner					x
Datenschutzkontrolle durch unabhängige BfD					x
Meldepflicht				x	x
Aufklärungspflichten				x	
Hinweis- und Kennzeichnungspflichten				x	
Unterrichtungspflichten				x	
Benachrichtigungspflichten				x	
Auskunftsanspruch				x	
Vorabkontrolle und Datenschutzfolgenabschätzung				x	x
Datenschutzaudit				x	
Informationspflichten (bei DS-Verletzungen)				x	
Führung eines Verfahrensverzeichnisses				x	
Recht auf Berichtigung					x
Recht auf Löschung	x				x
Recht auf Sperrung	x				x
Widerspruchsrecht					x
Zweckbestimmung					x
Zweckbindung und Zwecktrennung					x
Organisatorische und technische Gewaltenteilung					x
Grundsatz der Direkterhebung				x	x
Verbot mit Erlaubnisvorbehalt[45]					
Verhältnismäßigkeitsprinzip[46]					

Tabelle 4.1: Tabellarische Übersicht der datenschutzrechtlichen Anforderungen und der verfolgten Datenschutz-Schutzziele

[45] Nicht aus den Grundrechten abgeleitet, sondern rechtspolitische Entscheidung des Gesetzgebers - Rudolf in: Merten/Papier, HGR IV 2011, § 90 Rn. 28.

[46] Nicht aus den DS-Grundrechten abgeleitet, sondern aus dem Rechtsstaatsprinzip des Art. 20 Abs. 2 GG - BVerfGE 19, 342 (348 f.).

4.3 Datenschutzrechtliche Kriterien für ein Datenschutzauskunftssystem

Adäquate technische und organisatorische Vorkehrungen, um im Fall einer Betroffenenanfrage Auskunft geben zu können, sind Erfordernisse eines effektiven Grundrechtsschutzes.[47] Die DSGVO stellt in ErwGr 63 explizit den Anspruch auf, dass die verantwortliche Stelle einen elektronischen Fernzugang zu einem sicheren System bereitstellt,[48] das dem Betroffenen direkten Zugang zu seinen personenbezogenen Daten ermöglicht. Nichts anderes ist ein automatisiertes Datenschutzauskunftssystem. Denn wie in Kapitel 3.7 herausgearbeitet wurde, sind bis auf wenige Ausnahmen alle Daten, die zu beauskunften sind, personenbezogene Daten des Betroffenen. Manuelle Verfahren können die Geschwindigkeit der Informationsverarbeitung nicht bewältigen. Auskünfte, die einer Datenbank mit Stammdaten entnommen sind, geben nur einen Teilausschnitt der tatsächlichen Speicherung, Verarbeitung und Weitergabe personenbezogener Daten wieder.[49] Informationelle Zusammenhänge werden immer komplexer und bedingen eine ganzheitliche Herangehensweise. Es ist erforderlich, ein auf Datenschutz bezogenes Systemmonitoring in ein Datenschutzauskunftssystem zu integrieren, das es dem Betroffenen erlaubt, all seine Transparenz- und Interventionsrechte unmittelbar wirksam wahrzunehmen.[50]

Ausgehend von dieser Grundforderung nach einem automatisierten Datenschutzauskunftssystem entwickeln sich die datenschutzrechtlichen Kriterien für die Umsetzung eines solchen Systems entlang der datenschutzrechtlichen Anforderungen. Der Umfang des Auskunftsanspruchs und der Umfang der damit einhergehenden Protokollierung ergibt sich vorwiegend aus der einfachgesetzlichen Umsetzung des Rechts auf Auskunft. Weitere Kriterien können aus den übrigen datenschutzrechtlichen Anforderungen abgeleitet werden.

Die überwiegende Zahl der relevanten datenschutzrechtlichen Kriterien wurde bereits in Kapitel 3 hergeleitet und erörtert und wird an dieser Stelle noch einmal prägnant zusammengefasst. Die Verweise auf die jeweiligen Abschnitte im Kapitel 3 finden sich beim jeweiligen Kriterium.

Die datenschutzrechtlichen Kriterien für das Datenschutzauskunftssystem erheben keinen Anspruch auf generelle Vollständigkeit. Für die Kriterien, die sich aus den §§ 19 und 34 BDSG ergeben, stellen die Erörterungen des Kapitels 3 dennoch eine umfassende und abschließende Einschätzung des Rechts auf Auskunft dar.

Die Korrektheit der datenschutzrechtlichen Kriterien in Abschnitt 4.3.1 ergibt sich aus ihrer argumentativ schlüssigen Herleitung in Kapitel 3. Die Kriterien, die nicht dem Auskunftsanspruch entspringen, wurden in Kapitel 3 nicht alle eingehend erörtert. Einen

[47] BVerfGE 35, 79 (120); BVerfGE 56, 216 (238); siehe auch 2.1.3.
[48] Von Paal in: Paal/Pauly, DSGVO 2017, Art. 15 Rn. 14 als „Webfaces“ bezeichnet.
[49] Weichert, DuD 2006, 694 (695).
[50] Rost/Pfitzmann, DuD 2009, 353 (356 f.).

ergänzenden Einblick in ihre Herkunft gibt Anhang B.2.

Rechtliche Kriterien sind grundsätzlich als positiv-aktive Forderung formuliert. Eine Auflistung von Dingen, die nicht gefordert sind, also von den technischen Anforderungen nicht aufgegriffen werden müssen, macht keinen Sinn. Der Raum der Dinge, die ein Softwaresystem nicht können muss ist beliebig groß. Einen Sonderfall bilden bei den rechtlichen Kriterien diejenigen Kriterien, die ein vorhergehendes Kriterium einschränken. Die Kriterien 4 (Keine Auskunft bei Daten für die Datenschutzkontrolle), 16 (Kategorisierung gleichartiger Empfänger), 32 (Beschränkung der Auskunft zu Zeitangaben) und 54 (keine Auskunft bei Missbrauch) stellen Ausnahmeregelungen dar und bieten die Möglichkeit von einem vorhergehenden Kriterium abzuweichen und technische Anforderungen zu reduzieren. Ein weiterer Spezialfall ist die dem Datenschutzrecht geschuldete Prohibitivfunktionalität. Gewisse Funktionen, von denen sonst angenommen würde, dass sie aus technischer Bequemlichkeit mit umgesetzt würden, werden untersagt. Solche Kriterien sind die mit der Nummer 40 bis 44 (Datenvermeidung), 46 (Löschung von Protokolldaten nach Verarbeitung), 65 (Zweckbindung bei Datenschutzkontrolle), 72 und 73 (organisatorische und technische Gewaltenteilung).

4.3.1 Kriterien des Auskunftsanspruchs

Die einfachgesetzlichen Regelungen zum Auskunftsanspruch geben im Wesentlichen die funktionale Gestaltung und die organisatorische Einbindung eines Datenschutzauskunftssystems vor. Kern eines Datenschutzauskunftssystems ist die Fähigkeit, den Umgang mit personenbezogenen Daten zu protokollieren, die Protokolle zu verwalten und dem Betroffenen angemessen zugänglich zu machen. An diesen Aspekten richten sich auch die Kriterien des Auskunftsanspruchs aus. Die Strukturierung der Kriterien erfolgt entlang der zu protokollierenden Vorgänge und Datenarten. Wird im Folgenden von Verarbeitung gesprochen, meint dies die Verarbeitung im engeren Sinne (siehe Glossar).

Umfang der Protokollierung (Allgemein)

Die ersten sieben Kriterien geben den Rahmen vor, in dem ein Datenschutzauskunftssystem Vorgänge in informationstechnischen Systemen protokollieren muss, um Auskunftsansprüche im erforderlichen Umfang befriedigen zu können. Der Inhalt der Protokolle wird ab Kriterium 8 behandelt.

Kriterium 1. *Die Protokollierung muss sich auf alle Erhebungs-, Verarbeitungs-, Speicher-, Nutzungs- und Übermittlungsvorgänge erstrecken (Durchgängige Historie). → Abschnitt 2.1.3*

Kriterium 2. *Die personenbezogenen Daten sind so zu erheben und zu verwenden, dass eine vollständige Beauskunftung im Nachhinein möglich ist. → Abschnitt 3.4*

Kriterium 3. *Die Ablage personenbezogener Daten ist so zu gestalten, dass der Personenbezug im Zuge der Auskunft herstellbar ist, soweit er bei der Verwendung durch die verantwortliche Stelle herstellbar ist. → Abschnitt 3.4*

Kriterium 4. *Die Speicherung, interne Weitergabe und Verarbeitung von personenbezogenen Daten aus Protokollen zum Zweck der Datensicherung oder Datenschutzkontrolle müssen nicht beauskunftet werden. → Abschnitt 3.8.3*

Kriterium 5. *Eine erteilte Auskunft ist selbst zu protokollieren und zu beauskunften. → Abschnitt 3.7.1*

Kriterium 6. *Die Speicherfrist der auskunftsrelevanten Teile[51] der Protokolle muss grundsätzlich mindestens der Speicherdauer der Basisdaten entsprechen. → Abschnitt 3.6*

Kriterium 7. *Identifikatoren Auskunftsberechtigter[52] dürfen erst gelöscht werden, wenn auch alle übrigen Protokolldaten, die personenbezogene Daten des Auskunftsberechtigten betreffen, gelöscht werden. → Abschnitt 3.4*

Umfang der Protokollierung (Gespeicherte personenbezogene Daten)

Die Kriterien 8 bis 12 beschreiben, welche Informationen über gespeicherte personenbezogene Daten ein Datenschutzauskunftssystem vorhalten muss.

Kriterium 8. *Der Auskunftsanspruch umfasst die gespeicherten personenbezogenen Daten. → Abschnitt 3.7*

Kriterium 9. *Für jeden Übermittlungsvorgang ist eine Referenz auf die gespeicherten personenbezogenen Daten, die übermittelt wurden, zu protokollieren. → Abschnitt 3.7.1*

Eine eigene Speicherung der übermittelten Daten ist nicht vorgesehen. Lediglich eine Referenz auf zu anderen Zwecken gespeicherte personenbezogene Daten ist erforderlich.

Kriterium 10. *Gesperrte personenbezogene Daten sind zu beauskunften, soweit sie nicht aufgrund einer Aufbewahrungsvorschrift gespeichert werden. → Abschnitt 3.7.1*

Kriterium 11. *Der Ort der Speicherung (IT-System, Akte, Datenbank,...; besonders: Dateinamen, Struktur der Datenablage) personenbezogener Daten ist zu protokollieren und zu beauskunften. Die Protokolldaten sind so lange vorzuhalten, so lange die Speicherung anhält. → Abschnitt 3.7.1*

Kriterium 12. *Der Ort der Verarbeitung (Prozess) personenbezogener Daten ist zu protokollieren und zu beauskunften, soweit der Ort der Verarbeitung Rückschlüsse auf persönliche oder sachliche Verhältnisse des Betroffenen zulässt. Die Protokolldaten sind so lange vorzuhalten, so lange die Verarbeitung anhält. → Abschnitt 3.7.1*

[51] Welche Informationen auskunftsrelevant sind, wird in den nachfolgenden Abschnitten erläutert.
[52] Mögliche Identifikatoren werden in Kapitel 3.7.1 beschrieben.

Umfang der Protokollierung (Herkunft und Empfänger)

Die Protokollierung von Herkunft und Empfänger wird ab Kriterium 13 behandelt. Zunächst werden allgemeine Anforderungen gestellt, die unabhängig von den Kommunikationspartnern sind.

Kriterium 13. *Der Auskunftsanspruch umfasst die gespeicherten Herkunftsangaben personenbezogener Daten. → Abschnitt 3.7*

Kriterium 14. *Bei der internen Weitergabe sind Herkunftsangaben personenbezogener Daten zu protokollieren. → Abschnitt 3.7.3*

Kriterium 15. *Der Auskunftsanspruch umfasst die Empfänger personenbezogener Daten. Empfängerangaben sind zu protokollieren. → Abschnitt 3.7*

Empfänger sind alle Organisationseinheiten und alle dem Zweck nach verschiedene IT-Systeme innerhalb und außerhalb der verantwortlichen Stelle (→ Abschnitt 3.7.2). Ein IT-System ist dann einer dezidiert anderen Stelle zuzuordnen, wenn ihm eine inhärent andere Rolle und Funktionalität in der Datenverarbeitung zukommt (→ Abschnitt 3.7.2).

Kriterium 16. *Bei der Weitergabe gleichartiger Daten an gleichartige Empfänger zu gleichen Zwecken dürfen Empfänger, die Datensenken sind, im Protokoll und in der Auskunft kategorisiert werden (Betriebsgeheimnis). → Abschnitt 3.7.2*

Kriterium 17. *Die Weitergabe ist auch bei nicht-automatisierter Verarbeitung zu protokollieren. → Abschnitt 3.7.2*

Umfang der Protokollierung (Kommunikationsbeziehungen zu Dritten)

Übermittlungen weisen in ihrer rechtlichen Bewertung Besonderheiten gegenüber der internen Weitergabe auf. Diese Besonderheiten sind in den nachfolgenden Kriterien erfasst.

Kriterium 18. *Jeder Übermittlung, ausgehend von der einen verantwortlichen Stelle, muss eine protokollierte Erhebung bei einer anderen verantwortlichen Stelle gegenüberstehen. → Abschnitt 3.7*

Kriterium 19. *In der Auskunft über einen Empfänger, der Dritter ist, müssen alle Informationen, die die Übermittlung an den Empfänger referenzierbar machen (z. B. Zeitpunkt, Inhalt der Übermittlung, Gegenstelle beim Empfänger), enthalten sein. Diese Informationen sind zu protokollieren. → Abschnitt 3.7.2*

Kriterium 20. *Beginn und Ende einer Veröffentlichung sind zu protokollieren und zu beauskunften. → Abschnitt 3.8.2*

Kriterium 21. *Bei erst nach einer Authentifizierung abrufbaren personenbezogenen Daten ist jeder Abruf zu protokollieren und zu beauskunften. → Abschnitt 3.8.2*

Kriterium 22. *Protokolldaten über die Empfänger, die Dritte sind, dürfen erst dann gelöscht werden, wenn auch die Basisdaten bei Empfängern und Folgeempfängern gelöscht wurden. → Abschnitt 3.6*

Kriterium 23. *Bei der Erhebung personenbezogener Daten zum Zweck der Übermittlung sind Protokolldaten über Empfänger, die Dritte sind, noch mindestens für ein Jahr nach der letzten Übermittlung zu speichern. → Abschnitt 3.6*

Kriterium 24. *Herkunftsangaben bei der Erhebung sind für Daten, die zum Zweck der Übermittlung gespeichert werden, zu protokollieren. → Abschnitt 3.7.3*

Kriterium 25. *Bei listenmäßiger Übermittlung personenbezogener Daten zu Werbezwecken, sind Herkunftsangaben bei der Erhebung und Empfänger, die Dritte sind, noch mindestens für zwei Jahre nach der letzten Übermittlung zu speichern. → Abschnitt 3.6*

Kriterium 26. *Herkunftsangaben bei der Erhebung dürfen erst gelöscht werden, nachdem eine Löschung der Empfänger, die Dritte sind, zulässig ist, falls solche existieren. → Abschnitt 3.4*

Umfang der Protokollierung (Zweck und Zeitpunkt)

Die Zweckbindung ist eine eigenständige Anforderung des Datenschutzrechts, wirkt jedoch gemeinsam mit dem Auskunftsanspruch auch auf die Gestaltung der Protokollierung. Wie die nachfolgenden Kriterien aufschlüsseln, sind Zweckbezüge und zeitliche Zusammenhänge Teil der erforderlichen Protokollinformationen.

Kriterium 27. *Der Auskunftsanspruch umfasst den Zweck der Speicherung. Dieser ist zu protokollieren und zu beauskunften. → Abschnitt 3.7*

Kriterium 28. *Der Zweck muss nach einer Zweckänderung und in jedem Verwendungsschritt erkennbar bleiben. → Abschnitt 3.7.5*

Kriterium 29. *Werden personenbezogene Daten zu mehreren unterschiedlichen Zwecken verwendet, sind alle Zwecke zu protokollieren und zu beauskunften. → Abschnitt 3.7.5*

Kriterium 30. *Der Zweck ist getrennt und unabhängig von der Art der Speicherung der personenbezogenen Daten zu protokollieren. → Abschnitt 3.7.5*

Kriterium 31. *Der Zeitpunkt einer Erhebung oder Übermittlung personenbezogener Daten ist zu protokollieren und zu beauskunften. → Abschnitt 3.7.1*

Kriterium 32. *Der Zeitpunkt der Verarbeitung und Nutzung ist nur im Einzelfall auskunftspflichtig (Betriebsgeheimnis). → Abschnitt 3.7.1*

Umfang der Protokollierung (Funktionalität und Prozesse)

Die zwei nachfolgenden Kriterien leiten sich nicht allein aus dem Auskunftsanspruch ab, sondern beruhen auch auf dem Prinzip der nicht-automatisierten Einzelentscheidung. Dieses Prinzip erweitert den Umfang des Auskunftsanspruchs um funktionale Eigenschaften informationstechnischer Systeme.

Kriterium 33. *Die tragenden Funktionsprinzipien des logischen Aufbaus der automatisierten Verarbeitung personenbezogener Daten sind für IT-Systeme, auf die sich eine automatisierte Einzelentscheidung stützt, zu beauskunften. → Abschnitt 3.7.6*

Kriterium 34. *Die für eine automatisierte Einzelentscheidung verwendeten personenbezogenen Daten sind zu beauskunften. → Abschnitt 3.7.6*

Benutzerschnittstelle

Als Transparenzrecht ist für den Auskunftsanspruch die Interaktion des Betroffenen mit dem Datenschutzauskunftssystem von großer Bedeutung. Nur wenn der Betroffene faktisch in der Lage ist die bereitgestellten Informationen zu durchdringen und zu verstehen, wird echte Transparenz hergestellt. Entsprechend ergeben sich aus dem Auskunftsanspruch dezidierte Kriterien für die Gestaltung der Benutzerschnittstelle.

Insbesondere darf der Weg eines Auskunftsersuchens nicht auf einen einzigen Kommunikationskanal beschränkt werden. Es muss beispielsweise möglich sein, dass der betriebliche Datenschutzbeauftragte die für die Auskunft erforderlichen Informationen stellvertretend abruft und sie dem Betroffenen im Anschluss schriftlich mitteilt. Ebenso muss es möglich sein, dass ein Bevollmächtigter oder ein Erbe des Betroffenen Zugang zu den betreffenden Informationen erhält.

Kriterium 35. *Die Wahrnehmung des Auskunftsanspruchs erfordert ein aktives Handeln des Betroffenen. → Abschnitt 3.1*

Kriterium 36. *Ein stellvertretender Abruf der für die Auskunft erforderlichen Informationen muss möglich sein. → Abschnitt 3.4 und 3.3*

Kriterium 37. *Wurden keine personenbezogenen Daten erhoben oder verwendet, ist dies zu beauskunften. → Abschnitt 3.7.7*

Kriterium 38. *Die Benutzerschnittstelle muss so gestaltet sein, dass dem Betroffenen ein verständlicher – unter Verwendung klarer und einfacher Sprache – und einfacher Zugang zu den auskunftsrelevanten Informationen geboten wird. → Abschnitt 3.5*

Kriterium 39. *Die verantwortliche Stelle darf keine formalen oder technischen Bedingungen stellen, die der Betroffene zur Wahrnehmung seines Auskunftsanspruchs erfüllen muss. → Abschnitt 3.3*

4.3.2 Kriterien der übrigen Datenschutzanforderungen

Die Kriterien aus den übrigen Datenschutzanforderungen ergänzen die direkt aus dem Auskunftsanspruch abgeleiteten Kriterien.

Transparenzanforderungen, die neben dem Recht auf Auskunft stehen, beeinflussen den Entwurf eines Datenschutzauskunftssystem kaum. Sie beziehen sich auf Maßnahmen, die der Auskunft weitestgehend vorgelagert sind. Allerdings ergeben sich aus dem Auskunftsanspruch in Verbindung mit weiteren Anforderungen Kriterien für die Form der vorgelagerten Transparenzmaßnahmen. Vorgelagerte Transparenzmaßnahmen, wie beispielsweise die Benachrichtigung, sind so zu gestalten, dass sie die Auskunft im Nachhinein ermöglichen.

Organisatorische Vorgaben, wie die Etablierung eines Beauftragten für den Datenschutz (BfD), sind für ein technisches Datenschutzauskunftssystem unspezifisch und wurden deshalb in diesem Kriterienkatalog nicht erfasst.

Datenvermeidung

Kriterium 40. *Informationen über personenbezogene Daten, die für den Auskunftsanspruch nicht erforderlich sind, dürfen nicht protokolliert werden.*

Kriterium 41. *Für den Auskunftsanspruch darf keine zusätzliche allgemeine Speicherung der personenbezogenen Basisdaten vorgenommen werden.*

Kriterium 42. *Das Abrufen öffentlicher personenbezogener Daten durch einen Dritten ist nicht zu protokollieren und zu beauskunften.* → *Abschnitt 3.8.2*

Kriterium 43. *Bei der Protokollierung darf kein Mitarbeiter-/Bearbeiterbezug hergestellt werden.*

Kriterium 44. *Zeitangaben zu Verarbeitung und Speicherung dürfen nicht erhoben werden.*

Die Kriterien 43 und 44 tragen dem Mitarbeiterdatenschutz Rechnung.

Datensparsamkeit

Kriterium 45. *Zum Zwecke der Auskunft gespeicherte, nicht mehr erforderliche Informationen sind zu löschen.*

Kriterium 46. *Protokolldaten über Verarbeitungsvorgänge sind zu löschen, sobald die Verarbeitung abgeschlossen ist.*

Kriterium 47. *Protokolldaten über eine Speicherung sind zu löschen, sobald die Speicherung am entsprechenden Ort nicht mehr stattfindet.*

Anonyme und pseudonyme Dienste

Kriterium 48. *Das Auskunftsersuchen muss bei pseudonymen personenbezogenen Daten über das Pseudonym möglich sein. → Abschnitt 3.4*

Zugriffs-, Zugangs- und Weitergabekontrolle

Kriterium 49. *Die Identität des Auskunftsersuchenden ist gegen den eigenen Datenbestand zu prüfen. → Abschnitt 3.4*

Kriterium 50. *Eine der Auskunft vorausgehende Benachrichtigung soll Schlüsselkennungen (credentials) vergeben, die zur Authentifizierung bei der Auskunftserteilung verwendet werden können. → Abschnitt 3.1*

Kriterium 51. *Das Datenschutzauskunftssystem sollte eine Auskunftsanfrage auf Grundlage unverwechselbarer und exklusiver Informationen über einen Erhebungsvorgang bei einem Dritten oder dem Betroffenen bzw. einem Weitergabevorgang an Dritte oder den Betroffenen zulassen. → Abschnitt 3.4*

Kriterium 52. *Jeder Betroffene darf nur Zugriff auf Informationen zum Umgang mit personenbezogenen Daten erhalten, die sich auf ihn selbst beziehen. → Abschnitt 3.4*

Kriterium 53. *Beim Zugriff von Mitarbeitern der verantwortlichen Stelle auf zum Zwecke der Datenschutzauskunft gespeicherte personenbezogene Daten, ist das Vier-Augen-Prinzip zu wahren.*

Verfügbarkeitskontrolle

Kriterium 54. *Im Falle einer missbräuchlichen Verwendung des Auskunftsrechts überwiegen die Interessen der verantwortlichen Stelle an einem ungestörten Betrieb und der Verhinderung einer Ausforschung. Die Auskunft darf dann blockiert werden. → Abschnitt 3.8.3*

Kriterium 55. *Die Auskunft hat ohne schuldhaftes Zögern zu erfolgen.*[53] *→ Abschnitt 3.5*

Kriterium 56. *Die Auskunft hat jederzeit vollständig zu erfolgen. → Abschnitt 3.5*

Kriterium 57. *Es muss verhindert werden, dass zur Auskunft notwendige Informationen unberechtigt gelöscht oder modifiziert werden.*

[53] Gemäß Art. 12 Abs. 3 S. 1 DSGVO innerhalb eines Monats.

Revisionssichere Protokollierung

Kriterium 58. *Veränderungen an personenbezogenen Daten müssen nachvollziehbar protokolliert werden.*

Kriterium 59. *Die Integrität der Protokolldaten muss gewährleistet werden.*

Kriterium 60. *Es muss sichergestellt werden, dass die Auskunft jederzeit korrekt ist. → Abschnitt 3.5*

Recht auf Datenübertragbarkeit

Kriterium 61. *Auf der Auskunftsplattform muss eine prominent platzierte Download-Möglichkeit in einem strukturierten, gängigen und maschinenlesbaren Format zur Verfügung gestellt werden. → Abschnitt 3.5*

Einheitlicher Ansprechpartner

Kriterium 62. *Es muss eine zentrale Plattform geben, auf der das Auskunftsersuchen an die gesamte (verbundene) verantwortliche Stelle gestellt werden kann.*

Rechte auf Berichtigung, Löschung und Sperrung

Kriterium 63. *Die Auskunft muss die Wahrnehmung der Interventionsrechte Löschen, Sperren und Berichtigen ermöglichen. → Abschnitt 2.1.3*

Zweckbestimmung

Kriterium 64. *Der konkrete Zweck der Verwendung muss zum Zeitpunkt der Erhebung festgelegt werden. → Abschnitt 3.7.5*

Der Zweck einer Speicherung ergibt sich aus den nachfolgend noch zulässigen Verarbeitungs- und Nutzungszwecken.

Zweckbindung und Zwecktrennung

Kriterium 65. *Die für den Zweck der Auskunftserteilung gespeicherten Informationen dürfen nur für diesen Zweck sowie für Zwecke der Datenschutzkontrolle verwendet werden.*

Kriterium 66. *Die im Rahmen des Auskunftsersuchens durch den Betroffenen übermittelten personenbezogenen Daten dürfen nur für die Bearbeitung des Auskunftsersuchens verwendet werden und sind für alle anderen Zwecke zu sperren.*

Aus Kriterium 66 ergibt sich noch keine Löschpflicht. Die Daten können also auch in Zukunft für die Bearbeitung von Auskunftsersuchen verwendet werden. Dies gilt nicht, falls der Betroffene dagegen Einspruch erhebt oder die weitere Speicherung seinen Interessen widersprechen würde.

Kriterium 67. *Die für den Zweck der Auskunftserteilung gespeicherten Informationen sind für alle Zwecke außer der Auskunftserteilung zu sperren. → Abschnitt 3.6*

Kriterium 68. *Die Datenhaltung der Protokolldaten muss von der Datenhaltung der Basisdaten unabhängig sein.*

Kriterium 69. *Die Verarbeitung der Protokolldaten muss getrennt von der Verarbeitung der Basisdaten erfolgen.*

Organisatorische und technische Gewaltenteilung

Kriterium 70. *Zum Zweck der Auskunftserteilung erhobene Informationen dürfen zwischen Stellen nur dann ausgetauscht werden, wenn dies zur Erreichung des Auskunftszwecks notwendig ist. → Abschnitt 4.2*

Kriterium 71. *Protokolldaten dürfen nur dort gespeichert werden, wo sie anfallen (Datensubsidiarität). → Abschnitt 4.2*

Kriterium 72. *Durch das Datenschutzauskunftssystem darf keine zusätzliche Zuordnung personenbezogener Daten zueinander sowie zum Betroffenen möglich sein, soweit dies für die Auskunft nicht zwingend erforderlich ist. → Abschnitt 4.2*

Kriterium 73. *Die Verarbeitungswege und Weitergaben eines personenbezogenen Datums dürfen für eine Stelle, auch wenn sie an der Verarbeitung beteiligt ist, nicht durch das Datenschutzauskunftssystem nachvollziehbar gemacht werden. → Abschnitt 4.2*

4.4 Technische Anforderungen an ein Datenschutzauskunftssystem

Die technischen Anforderungen für ein Datenschutzauskunftssystem präzisieren die rechtlichen Kriterien auf ihre technische Umsetzung hin und beschreiben konkrete Systemeigenschaften. Daraus folgt, dass nicht jedem rechtlichen Kriterium genau eine technische Anforderung entspricht. Manche Kriterien können technisch identisch beschrieben werden, andere erfordern mehrere technische Vorgaben. Bei jeder technischen Anforderung findet sich ein Verweis auf die rechtlichen Kriterien, die ihr zugrunde liegen. Die Tabelle 4.2 gibt einen Überblick der Abbildung von rechtlichen Kriterien auf technische Anforderungen. Eine detaillierte Diskussion der Abbildung findet sich im Anhang B.3.

Der technischen Orientierung folgend, sind die technischen Anforderungen im Folgenden anders als die rechtlichen Kriterien nicht nach den datenschutzrechtlichen Anforderungen sortiert, sondern nach funktionalen und nicht-funktionalen Anforderungen. Die weitere Unterteilung der Anforderungen hat rein informativen und keinen definitorischen Charakter.

4.4.1 Funktionale Anforderungen

Die funktionalen Anforderungen geben wieder, was ein Datenschutzauskunftssystem für den Betroffenen leisten soll. Kernfunktionalität des Datenschutzauskunftssystems ist die Protokollierung des Umgangs mit personenbezogenen Daten. Deshalb machen die funktionalen Anforderungen Vorgaben an die Gestalt der Protokollinformationen, ihre Speicherung und Löschung.

Allgemeine Systemanforderungen

Anforderung 1. *Systeme, die personenbezogene Daten verarbeiten oder speichern, müssen in der Lage sein, die stattfindenden Verarbeitungs-, Speicher- und Weitergabeereignisse in all ihren Teilsystemen zu erfassen. → Kriterium 1*

Anforderung 2. *Die Erfassung der Verarbeitungs-, Speicher- und Weitergabeereignisse muss auf der niedrigsten Abstraktionsschicht stattfinden, auf der das personenbezogene Datum noch semantisch erfassbar ist.*[54] *→ Kriterium 1*

Anforderung 3. *Die Erfassung der Verarbeitungs-, Speicher- und Weitergabeereignisse muss unabhängig von der Mitwirkung des Betroffenen möglich sein. → Kriterium 39*

Anforderung 4. *Auf Grundlage der Protokollinformationen muss eine nachträgliche globale Löschung, Sperrung und Berichtigung der personenbezogenen Daten möglich sein. → Kriterium 63*

Umfang der Protokollinformationen

Anforderung 5. *Informationen über personenbezogene Daten, die in keiner Anforderung aufgeführt sind und für die es auch keine anderen zwingenden technischen Gründe gibt, dürfen nicht protokolliert werden. → Kriterien 40, 41, 43, 44*

Anforderung 6. *Jeder Umgang mit personenbezogenen Daten muss, den nachfolgend formulierten Anforderungen entsprechend, protokolliert werden. → Kriterien 1, 58*

[54] D. h., erkennbar ist, ob ein Datum ein personenbezogenes Datum ist.

Anforderung 7. *Das Protokoll zur Erhebung personenbezogener Daten muss folgendes umfassen:*

- *die Quelle der Daten als global eindeutiger Protokollbestandteil, insbesondere deren Namen*
- *den Zeitpunkt der Erhebung*
- *die Kategorie der erhobenen Daten*
- *die Organisationseinheit, in der das Datum erhoben wurde*
- *das IT-System, auf dem das Datum erhoben wurde*
- *die Zwecke der beabsichtigten Verwendung*

→ Kriterien 1, 2, 3, 13, 24, 31, 64

Anforderung 8. *Bei der Erhebung ist ein Identifikator des Betroffenen, soweit vorhanden, zu protokollieren und mit den Protokollinformationen zu verknüpfen. → Kriterium 3*

Anforderung 9. *Das Protokoll zur Übermittlung personenbezogener Daten muss folgendes umfassen:*

- *die Senke der Daten als global eindeutiger Protokollbestandteil, insbesondere deren Namen*
- *den Zeitpunkt der Übermittlung*
- *die Kategorie der übermittelten Daten*
- *die Organisationseinheit, von der das Datum übermittelt wurde*
- *das IT-System, von dem das Datum übermittelt wurde*
- *die Zwecke der Übermittlung*

→ Kriterien 1, 15, 19, 21, 28, 29, 31

Anforderung 10. *Das Protokoll zur Veröffentlichung personenbezogener Daten muss folgendes umfassen:*

- *den Beginn der Veröffentlichung*
- *das Ende der Veröffentlichung*
- *die Kategorie der veröffentlichten Daten*
- *die Organisationseinheit, von der das Datum veröffentlicht wurde*
- *das IT-System, von dem das Datum veröffentlicht wurde*

- *die Zwecke der Veröffentlichung*

→ *Kriterien 1, 20, 28, 29, 42*

Anforderung 11. *Das Protokoll zur internen Weitergabe personenbezogener Daten muss folgendes umfassen:*

- *den Zeitpunkt der Weitergabe*
- *den Sender der Daten (Organisationseinheit und IT-System)*
- *den Empfänger der Daten (Organisationseinheit und IT-System)*
- *die Zwecke der Weitergabe*

→ *Kriterien 1, 14, 17, 28, 29, 32*

Anforderung 12. *Das Protokoll zur Verarbeitung personenbezogener Daten muss folgendes umfassen:*

- *die Organisationseinheit, in der das Datum verarbeitet wurde*
- *das IT-System, auf dem das Datum verarbeitet wurde*
- *bei aussagekräftigen Spezialanwendungen: den Namen und den Typ der Anwendung*
- *die Zwecke der Verarbeitung*

→ *Kriterien 1, 12, 28, 29, 32, 44*

Das IT-System ist nur zu protokollieren, sofern es als funktional eigene Stelle betrachtet werden kann. Name und Typ der Anwendung sind nur zu protokollieren, soweit der Ort der Verarbeitung Rückschlüsse auf persönliche oder sachliche Verhältnisse des Betroffenen zulässt. Ist die funktionale Abgrenzung unterschiedlicher Stellen a-priori nicht bekannt, ist so detailliert zu protokollieren, dass die Auskunft entsprechend der funktionalen Abgrenzung a-posteriori auf jeden Fall sichergestellt werden kann.

Anforderung 13. *Im Falle der automatisierten Einzelentscheidung sind die tragenden Funktionsprinzipien der Verarbeitung zu dokumentieren und mit den Verarbeitungsprotokollen zu verknüpfen. → Kriterium 33*

Anforderung 14. *Gehen im Falle der automatisierten Einzelentscheidung mehrere personenbezogene Daten in einen Verarbeitungsvorgang mit ein, ist die gleichzeitige Verfügbarkeit dieser Daten im Prozess zu protokollieren. → Kriterium 34*

Anforderung 15. *Das Protokoll zur Speicherung personenbezogener Daten muss folgendes umfassen:*

- *die Organisationseinheit, in der das Datum gespeichert wurde*
- *das IT-System, auf dem das Datum gespeichert wurde*
- *den Dateisystempfad zu dem Datum bzw. die Metadaten (Datenbankname, Tabellenname, Attributname) der Datenbank, in der das Datum gespeichert ist*
- *die Zwecke der Speicherung*

→ Kriterien 1, 8, 11, 27, 28, 29

Anforderung 16. *Die durch den Fluss personenbezogener Daten entstehenden wechselseitigen Beziehungen von Erhebung, Verarbeitung, Speicherung, interner Weitergabe, Übermittlung und Veröffentlichung müssen über Protokollgrenzen hinweg festgehalten werden und als durchgängige Personal-Data-Provenance nachvollziehbar sein. → Kriterium 1*

Anforderung 17. *Jede protokollierte Übermittlung- und Veröffentlichung personenbezogener Daten muss, soweit verfügbar, mit einem Speicherort des Datums verknüpft sein. → Kriterium 9*

Anforderung 18. *Die Zugriffe auf die Informationen zum Umgang mit seinen personenbezogenen Daten durch den Betroffenen müssen mit Zugriffszeitpunkt und übertragenen Informationen protokolliert werden. → Kriterium 5*

Speicherung der Protokolldaten

Anforderung 19. *Die Protokolldaten sind getrennt von den personenbezogenen Daten zu speichern und zu verarbeiten. → Kriterien 30, 68, 69*

Anforderung 20. *Die Protokolldaten sind dort zu speichern, wo die protokollierten Ereignisse stattfinden. → Kriterium 71*

Anforderung 21. *Die Protokolldaten werden nur zum Zeitpunkt einer Auskunftsanfrage von den speichernden Systemen abgerufen. → Kriterium 70*

Anforderung 22. *Zum Abruf zwischengespeicherte Protokolldaten werden nach Ende der Auskunftsanfrage an zentraler Stelle wieder gelöscht. → Kriterium 45*

Speicherdauer und Löschregeln

Anforderung 23. *Das Datenschutzauskunftssystem speichert alle Protokolldaten, solange keine Löschregel zutrifft. → Kriterien 6, 57*

Anforderung 24. *Das Datenschutzauskunftssystem löscht Identifikatoren des Betroffenen, sobald diese auf keine Protokolldaten mehr verweisen. → Kriterium 7*

Anforderung 25. *Das Datenschutzauskunftssystem löscht Informationen über die Erhebung eines personenbezogenen Datums, sobald keine anderen Informationen über den Umgang mit dem personenbezogenen Datum außer über die Erhebung mehr bestehen. → Kriterien 6, 25, 26*

Anforderung 26. *Das Datenschutzauskunftssystem löscht Informationen über die Übermittlung eines personenbezogenen Datums,*

- *sobald es von allen Empfängern die Bestätigung der Löschung des personenbezogenen Datums erhalten hat,*
- *sofern die Übermittlung mindestens ein Jahr her ist,*
- *sofern die Übermittlung mindestens zwei Jahre her ist, falls die Übermittlung zum Zweck „Werbung" erfolgte und*
- *keine anderen Informationen über den Umgang mit dem personenbezogenen Datum außer über die Erhebung und Übermittlung mehr bestehen.*

→ Kriterien 6, 22, 23, 25

Anforderung 27. *Das Datenschutzauskunftssystem löscht Informationen über die interne Weitergabe eines personenbezogenen Datums, sobald keine Informationen über Verarbeitung und Speicherung mehr bestehen. → Kriterium 6*

Anforderung 28. *Das Datenschutzauskunftssystem löscht Informationen über die Verarbeitung eines personenbezogenen Datums, sobald die Verarbeitung abgeschlossen ist. → Kriterien 6, 46*

Anforderung 29. *Das Datenschutzauskunftssystem löscht Informationen über die Speicherung eines personenbezogenen Datums, sobald die personenbezogenen Daten am entsprechenden Ort nicht mehr gespeichert werden. → Kriterien 6, 47*

4.4.2 Nicht-Funktionale Anforderungen

Nicht-Funktionale Anforderungen haben bei Anwendungen, die personenbezogene Daten verarbeiten, einen besonders hervorgehobenen Stellenwert. Für ein Datenschutzauskunftssystem stellen sie insbesondere die Berücksichtigung der datenschutzrechtlichen Kriterien sicher, die sich nicht aus dem Auskunftsanspruch ergeben.

IT-Sicherheitsfeatures sind qualitative Anforderungen an ein Datenschutzauskunftssystem und stehen stellvertretend für die Breite, in der unterschiedliche Datenschutzanforderungen abgedeckt werden, die nicht zur Grundfunktionalität des Auskunftsanspruchs beitragen. Die Charakteristiken der Benutzeroberfläche sind nicht funktional, da sie nicht

vorgeben, was das Datenschutzauskunftssystem zu leisten hat, sondern wie die Ergebnisse der eigentlichen Systemlogik zu präsentieren sind. Dennoch leiten auch sie sich aus rechtlichen Kriterien ab.

In der Kombination der funktionalen mit den nicht-funktionalen Anforderungen wird der Konflikt unterschiedlicher datenschutzrechtlicher Ziele auf technischer Ebene deutlich. Die besondere Herausforderung der späteren Spezifikation und Implementierung besteht darin, sich zum Teil widersprechenden Anforderungen gerecht zu werden. Namentlich die technischen Unverkettbarkeitsanforderungen und die durch die Protokollierung zu schaffende Transparenz erfordern eine gewissenhafte Abwägung.

Authentifikation, Autorisierung und Zugriffsrechte

Anforderung 30. *Ein aktives Einloggen des Betroffenen oder seines Stellvertreters auf der Auskunftsplattform ist für den Zugriff auf die Auskunftsinformationen erforderlich. → Kriterium 35*

Anforderung 31. *Das Datenschutzauskunftssystem muss Stellvertreterzugriffsrechte ermöglichen. → Kriterium 36*

Anforderung 32. *Beim Zugriff muss eine Identifikation über Pseudonym und ein gemeinsames Geheimnis, soweit vorhanden, möglich sein. → Kriterium 48*

Anforderung 33. *Beim Zugriff muss eine Identifikation über vergebene Credentials, soweit vorhanden, möglich sein. → Kriterium 50*

Anforderung 34. *Beim Zugriff muss eine Identifikation über eine bekannte Quelle mit Informationen über Art der Daten und des Informationsflusses, Zeitpunkt und Ursprung möglich sein, soweit kein Dritter diese Informationen wissen kann. → Kriterium 51*

Anforderung 35. *Beim Zugriff soll eine Identifikation über eine bekannte Senke mit Informationen über Art der Daten und des Informationsflusses, Zeitpunkt und Adressat möglich sein, soweit kein Dritter diese Informationen wissen kann. → Kriterium 51*

Anforderung 36. *Authentifizierungsinformationen (Gemeinsame Geheimnisse, Schlüsselkennungen, elektronischer Personalausweis (nPA) oder exklusives Wissen) müssen vor dem Zugriff überprüft werden und gültig sowie korrekt sein. → Kriterium 49*

Anforderung 37. *Authentifizierungsinformationen müssen an die Autorisierung für den Datenbestand des entsprechenden Betroffenen (und keine weiteren Daten) geknüpft werden. → Kriterium 52*

Anforderung 38. *Beim Stellvertreterzugriff müssen sich zwei voneinander unabhängige Personen (z. B. Kundenbetreuer und Datenschutzbeauftragter) gemeinsam beim Datenschutzauskunftssystem authentifizieren, um für den Zugriff auf die Auskunftsinformationen autorisiert zu sein. → Kriterium 53*

Anforderung 39. *Die Protokolldaten dürfen nur vom Datenschutzauskunftssystem selbst, entsprechend den funktionalen Anforderungen, modifiziert werden. Der schreibende Zugriff muss darüber hinaus unterbunden werden. → Kriterien 57, 59*

Anforderung 40. *Auf die Protokolldaten darf nur der Betroffene (oder sein Stellvertreter) im Rahmen einer Auskunftsanfrage zugreifen. Der lesende Zugriff muss darüber hinaus unterbunden werden. → Kriterien 65, 66, 67*

Anforderung 41. *Das Datenschutzauskunftssystem darf nur im Rahmen des technisch erforderlichen und soweit möglich nur lokal lesend auf die Protokolldaten zugreifen. → Kriterium 65*

Unverkettbarkeit

Anforderung 42. *Systeme dürfen durch die bei ihnen gespeicherten Protokolldaten kein neues Wissen darüber gewinnen, auf welchen Systemen ein bestimmtes personenbezogenes Datum verarbeitet und/oder gespeichert wurde. → Kriterium 73*

Anforderung 43. *Systeme dürfen durch die bei ihnen gespeicherten Protokolldaten kein neues Wissen darüber gewinnen, welche Systeme ein bestimmtes personenbezogenes Datum an welche anderen Systeme weitergegeben haben. → Kriterium 73*

Anforderung 44. *Systeme dürfen durch die bei ihnen gespeicherten Protokolldaten kein neues Wissen darüber gewinnen, ob zwei personenbezogene Daten einen Personenbezug zum selben Betroffenen aufweisen. → Kriterium 72*

Anforderung 45. *Systeme dürfen durch die bei ihnen gespeicherten Protokolldaten kein neues Wissen darüber gewinnen, ob ein personenbezogenes Datum einen Personenbezug zu einem bestimmten Betroffenen aufweist. → Kriterium 72*

Die technischen Anforderungen an die Unverkettbarkeit werden in Kapitel 9 formalisiert und messbar gemacht.

Auskunftsplattform - Aussehen, Handhabung, Benutzbarkeit

Anforderung 46. *Der Zugriff auf die Benutzeroberfläche muss für jeden Betroffenen mit Standardsoftware (z. B. Webbrowser) möglich sein. Es darf keine Installation clientseitiger Spezialsoftware vorausgesetzt werden. → Kriterium 39*

Anforderung 47. *Über die Benutzeroberfläche muss für jeden Betroffenen ein Zugriff auf alle personenbezogenen Daten, inklusive der für die Auskunft ergänzend gespeicherten Informationen, möglich sein. → Kriterium 62*

Anforderung 48. *Die Benutzeroberfläche muss eine konsistente, durchgängige Wahrnehmung aller interaktiven Betroffenenrechte (Löschung, Sperrung, Berichtigung) ermöglichen. → Kriterium 63*

Anforderung 49. *Jedes auf der Benutzeroberfläche dargestellte Datum muss eine direkte Interaktionsmöglichkeit im Rahmen der Betroffenenrechte anbieten. → Kriterium 63*

Anforderung 50. *Die vorangegangenen erteilten Auskünfte müssen auf der Benutzeroberfläche der Auskunftsplattform sichtbar sein. → Kriterium 5*

Anforderung 51. *Auf der Benutzeroberfläche muss klar dargestellt werden, falls keine personenbezogenen Daten erhoben oder verwendet wurden. → Kriterium 37*

Anforderung 52. *Es müssen Selektionsmöglichkeiten für die Protokollinformationen nach Datenkategorie, Zweck, Herkunft und Empfänger vorhanden sein. → Kriterium 38*

Anforderung 53. *Die Protokolldaten müssen in einem strukturierten, gängigen und maschinenlesbaren Format als Ganzes herunterladbar sein. → Kriterium 61*

Zuverlässigkeit und Korrektheit

Anforderung 54. *Die Protokollierung muss ab dem Zeitpunkt der Erhebung ohne Unterbrechung erfolgen. → Kriterium 2*

Anforderung 55. *Die Auskunft muss die Protokollinformationen exakt und vollständig wiedergeben. → Kriterien 2, 56, 60*

Anforderung 56. *Die Protokolldaten müssen bei mobilen IT-Systemen[55] und IT-Systemen mit niedriger Verfügbarkeit redundant gespeichert werden. → Kriterium 56*

Anforderung 57. *Bei Angriffen auf die Auskunftsplattform oder Missbrauch derselben darf der Zugang zur Plattform vorübergehend deaktiviert werden. → Kriterium 54*

Flexibilität und Übertragbarkeit

Anforderung 58. *Die Schnittstelle des Datenschutzauskunftssystems muss dokumentiert und standardisiert sein. → Kriterien 18, 62*

Performance und Skalierbarkeit

Anforderung 59. *Eine Auskunft muss innerhalb der üblichen Ladezeiten einer Webseite mit aktiven oder multimedialen Inhalten bereitgestellt werden. → Kriterium 55*

[55] Beispielsweise Laptops oder Smartphones.

Rechtliche Kriterien	Technische Anforderungen
1	1, 2, 6, 7, 9, 10, 11, 12, 15, 16
2	7, 54, 55
3	7, 8
4	
5	18, 50
6	23, 25, 26, 27, 28, 29
7	24
8	15
9	17
10	
11	15
12	12
13	7
14	11
15	9
16	
17	11
18	58
19	9
20	10
21	9
22	26
23	26
24	7
25	25, 26
26	25
27	15
28	9, 10, 11, 12, 15
29	9, 10, 11, 12, 15
30	19
31	7, 9
32	11, 12
33	13
34	14
35	30
36	31
37	51
38	52
39	3, 46
40	5
41	5
42	10
43	5
44	5, 12
45	22
46	28
47	29
48	32
49	36
50	33
51	34, 35
52	37
53	38
54	57
55	59
56	55, 56
57	23, 39
58	6
59	39
60	55
61	53
62	47, 58
63	4, 48, 49
64	7
65	40, 41
66	40
67	40
68	19
69	19
70	21
71	20
72	44, 45
73	42, 43

Tabelle 4.2: Ableitung technischer Anforderungen aus rechtlichen Kriterien

4.5 Zwischenfazit

Dieses Kapitel hat zwei Resultate hervorgebracht: Das Modell zur Ableitung rechtlicher Anforderungen EVAL und die Anwendung dieses Modells auf das Recht auf Auskunft bzw. den Einsatz eines Datenschutzauskunftssystems. Konstruktionsziel $\mathfrak{K}1$ ist damit erreicht.

EVAL ist dem bisher bestehenden Modell KORA dahingehend überlegen, dass eine klarere Systematisierung der rechtlichen Anforderungsableitung erfolgt und die Ableitung nicht auf rechtlicher Ebene stehen bleibt, sondern der Pfad bis zur tatsächlichen Implementierung vorgezeichnet ist.

Die Aufstellung rechtlicher Kriterien für ein Datenschutzauskunftssystem baut auf den Erkenntnissen aus Kapitel 3 auf. Aus rechtlichen Kriterien ergeben sich technische Anforderungen, die auf die Implementierung des Datenschutzauskunftssystems angewandt werden können.

Teil II

Personal-Data-Provenance und Data-Usage-Control

5 Usage-Control & Provenance-Tracking: Einführung und Architektur

Aus den in den vorangegangen Kapiteln abgeleiteten Anforderungen ergibt sich, dass für ein Datenschutzauskunftssystem eine einfache Protokollierung von Ereignissen in IT-Systemen nicht ausreichend ist. Das Recht auf Auskunft fordert eine durchgängige Historie (Kriterium 1), sprich die technische Realisierung einer Personal-Data-Provenance (Anforderung 16), für den Umgang mit jedem personenbezogen Datum. Diese Datenzentriertheit, die auch die Berücksichtigung weiterer datenbezogener Anforderungen erlaubt, findet sich in bisherigen Ansätzen zur technischen Realisierung der Datenschutzauskunft nicht in ausreichendem Maße.

Das Recht auf Auskunft steht nicht allein, sondern neben weiteren Betroffenenrechten (Kriterium 4). Um die Wahrnehmung dieser weiteren Rechte zu ermöglichen, ist eine Verbindung von Personal-Data-Provenance mit Data-Usage-Control erforderlich. Usage-Control und Provenance-Tracking ergänzen sich hinsichtlich der Anforderungen des Rechts auf Auskunft.[1]

Dieses Kapitel gibt einen Überblick über die Entwicklung in den Forschungsbereichen Data-Provenance und Usage-Control sowie den noch bestehenden Defiziten im Hinblick auf eine integrierte Architektur für ein Datenschutzauskunftssystem. Anhand der Anforderungen aus Kapitel 4.4 wird eine gemeinsame Architektur für Data-Provenance und Usage-Control hergeleitet und beschrieben. Die Rollen der einzelnen Komponenten werden erläutert und ihre Schnittstellen spezifiziert.

5.1 Personal-Data-Provenance: Eine Einführung

Provenance (auch Lineage oder Pedigree) ist ein Begriff, unter dem die Sichtbarmachung der Historie eines Datums verstanden wird.[2] Data-Provenance ist die dokumentierte Ableitungshistorie eines Datums, ausgehend von seiner ursprünglichen Quelle.[3] Data-Provenance gibt Auskunft über die Herkunft von Daten sowie über die Erzeugung von Daten aus anderen Daten. Ein technisches System, das solche Fragestellungen beantworten kann, ist ein Provenance-System.

[1] Bier 2013b.
[2] Simmhan/Plale/Gannon 2005.
[3] Buneman/Khanna/Tan 2001; Miles et al. 2005; Simmhan/Plale/Gannon 2005.

Abbildung 5.1: Komponenten einer Data-Provenance-Infrastruktur

Eine Provenance-Infrastruktur muss drei wesentliche Komponenten bereitstellen (Abbildung 5.1): Einen Erhebungsmechanismus, um den Umgang mit Daten aufzuzeichnen (Provenance-Collection), die Infrastruktur zur Speicherung der Provenance-Daten (Provenance-Storage) und Funktionen, um die Daten im Nachgang adäquat nutzbar zu machen (Provenance-Dissemination).[4] In der Vergangenheit wurden Provenance-Systeme für das wissenschaftliche Rechnen,[5] Betriebssysteme,[6] Dateisysteme,[7] Geschäftsprozesse[8] und ausgewählte Anwendungen[9] entwickelt.

Anwendungsgebiete sind die Reproduzierbarkeit von wissenschaftlichen Ergebnissen, die Vertrauenswürdigkeit von Sensordaten, die Auditierung von Systemen und Prozessen, die Vorhersage zukünftiger Geschäftsprozesse, Accountability und Compliance der Datenverarbeitung in stark regulierten Branchen und die Erkennung von Beziehungen und Gemeinschaften in sozialen Netzwerken.[10] Überall dort müssen bestehende Systeme provenancefähig gemacht werden, so dass sie Provenance-Informationen publizieren können. Außerdem muss eine Speicherinfrastruktur für Provenance etabliert werden, die die Administration, die Abfrage (Querying) und das Schlussfolgern (Reasoning) von und aus Provenance-Informationen zulässt.

Data-Provenance hat drei Dimensionen, die in Provenance-Modellen abgedeckt werden müssen (Abbildung 5.2): Informationsfluss-Provenance, Funktions-Provenance[11] und Kontext-Provenance. Provenance kann sich auf einen Prozess oder auf ein Datum beziehen. Personal-Data-Provenance ist datenzentriert.

Jede Dimension von Provenance muss sich gemäß der Anforderungen aus Kapitel 4 auch in Personal-Data-Provenance wiederfinden. Gemäß der Anforderungen 7, 9, 11 und 16 muss die Informationsfluss-Provenance personenbezogener Daten, inklusive der

[4] Freire et al. 2008; Lakshmanan et al. 2011
[5] Bose/Frew 2005; Freire et al. 2008; Moreau/Groth et al. 2008.
[6] Gehani/Tariq 2012.
[7] Gehani/Lindqvist 2007; Holland et al. 2008; Sultana/Bertino 2013.
[8] Bechini et al. 2008; Curbera et al. 2008; Gadelha Jr./Mattoso 2008.
[9] Fonseca et al. 2007; Muniswamy-Reddy et al. 2009; Tariq/Ali/Gehani 2012.
[10] Moreau/Groth et al. 2008; Lakshmanan et al. 2011.
[11] Funktion im Sinne der internen Abbildungsfunktionen eines Programms, die aus einer Menge von Eingabewerten einen Ausgabewert (Resultat) bestimmt.

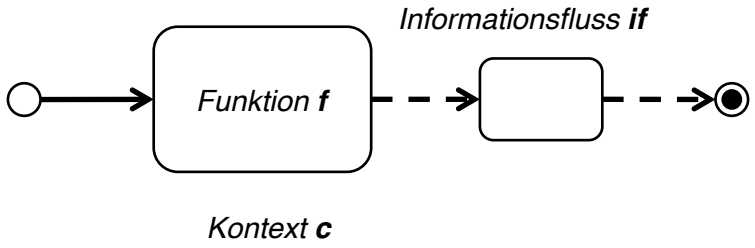


Abbildung 5.2: Data-Provenance-Dimensionen

Quellen und Senken, gesammelt werden. Funktions-Provenance ist für den Spezialfall automatisierter Einzelentscheidungen mitzuberücksichtigen (Anforderungen 13 und 14). In wenigen anderen Fällen sind die Ein- und Ausgabeparameter der Verarbeitung zu erfassen (Anforderung 12). Kontext-Provenance findet sich in Form von Zweckbindung und Zeitangaben (Anforderungen 7, 9–12, 14, 15 und 18) in der Personal-Data-Provenance.

Eine standardisiertes Daten- und Austauschformat für Provenance im wissenschaftlichen Rechnen wurde mit dem Open Provenance Model (OPM) eingeführt. Dessen Spezifikation in der Version 1.1 wurde im Jahr 2011 veröffentlicht.[12]

Personal-Data-Provenance wurde in der jüngeren Literatur eingeführt. Aldeco-Pérez und Moreau schlagen vor, Provenance für die Auditierung der Verwendung personenbezogener Daten heranzuziehen.[13] Herkenhöhner et al. stellen eine Architektur und ein Nachrichtenformat vor, um die für das Recht auf Auskunft erforderliche Provenance in organisationsübergreifenden Webservices automatisiert zu sammeln und dem Betroffenen zur Verfügung zu stellen.[14] Pulls et al. stellen ein Schema, *Insynd*, vor, um Personal-Data-Provenance zu sammeln, ohne die Verknüpfungen zwischen den einzelnen Logs zu offenbaren.[15]

Die Sicherheit von Data-Provenance im Allgemeinen wurde in der Literatur bereits intensiv diskutiert. Konzepte für die Suche in verschlüsselter Data-Provenance finden sich bei Asghar et al.[16] Davidson et al. diskutieren die Bewahrung von Anonymität in Queries auf Data-Provenance.[17] Access-Control für Provenance wurde als XACML-Erweiterung,[18] in ODRL,[19] sowie als eigenständiges Konzept[20] umgesetzt. Butin, Chicote und Le Métayer

[12] Moreau/Clifford et al. 2011.
[13] Aldeco Pérez/Moreau 2008.
[14] Herkenhöner et al. 2010.
[15] Pulls/Peeters/Wouters 2013; Pulls/Peeters 2015.
[16] Asghar et al. 2012.
[17] Davidson et al. 2011.
[18] Cadenhead et al. 2011.
[19] Grunwell/Gajanayake/Sahama 2015.
[20] Ni/Bertino/Sandhu 2009.

definieren Richtlinien für das Log-Design für Accountability-Frameworks.[21] Diese Forschungsgegenstände und Lösungsvorschläge sind orthogonal zum Thema dieser Arbeit. Ihre Anwendung auf das Datenschutzauskunftssystem kann dessen Sicherheit und damit auch den Erfüllungsgrad der nicht-funktionalen Anforderungen verbessern. Im Rahmen dieser Arbeit wird aufgrund der existierenden Forschung angenommen, dass die Personal-Data-Provenance sicher gespeichert ist und die Zugriffsrechte entsprechend des Nutzungszwecks der Provenance formuliert und durchgesetzt werden.

5.2 Usage-Control: Beschreibungssprachen und Durchsetzung

Nutzungskontrolle (engl. Usage Control (UC)) bestimmt, unter welchen Umständen Zugriff auf Daten gewährt wird und wie nachfolgend mit den Daten umgegangen werden darf. UC kann auch als eine Erweiterung von Access-Control über den Zugriffszeitpunkt hinaus betrachtet werden. Die bisherige Forschung im Themenbereich[22] hat unterschiedliche Aspekte von Usage-Control beleuchtet. Im Laufe der Zeit sind eine Vielzahl von Policy-Sprachen entstanden: $UCON$[23] und $UCON_{ABC}$,[24] sowie eine Erweiterung der EXtensible Access Control Markup Language (XACML) für $UCON$-Policies,[25] die Open Digital Rights Language (ODRL)[26] für Digital Rights Management (DRM) sowie die Obligation Specification Language (OSL),[27] die in ODRL transformierbar ist.[28] Das in dieser Arbeit vorgestellte Datenschutzauskunftssystem setzt auf der OSL auf. Grundsätzlich ist jedoch auch jede andere Beschreibungssprache mit gleicher Mächtigkeit anwendbar.

Die Durchsetzung von Policies wurde für Java,[29] die OpenNebula Cloud-Lösung,[30] Android,[31] Firefox,[32] Thunderbird,[33] Microsoft Windows,[34] OpenBSD,[35] X11[36] und die intelligente Videoüberwachung[37] sowie über Systemgrenzen hinweg[38] umgesetzt. Die

[21] Butin/Chicote/Métayer 2013.
[22] Lazouski/Martinelli/Mori 2010; Pretschner/Hilty/Basin 2006; Pretschner/Hilty/Schutz et al. 2008.
[23] Park/Sandhu 2002.
[24] Park/Sandhu 2004; X. Zhang et al. 2004.
[25] Lazouski/Martinelli/Mori 2012.
[26] Iannella 2004; https://www.w3.org/community/odrl.
[27] Hilty et al. 2007.
[28] Hilty et al. 2007.
[29] Fromm/Kelbert/Pretschner 2013.
[30] Lazouski/Mancini et al. 2014.
[31] Feth/Pretschner 2012; Rasthofer et al. 2014
[32] Kumari/Pretschner et al. 2011.
[33] Lörscher 2012.
[34] Wüchner/Pretschner 2012.
[35] Harvan/Pretschner 2009.
[36] Pretschner/Büchler et al. 2009.
[37] Birnstill/Pretschner 2013.
[38] Basin et al. 2013; Kelbert/Pretschner 2013.

Erweiterung um zustandsbasierte Policies und das dafür notwendige Informationsfluss-Tracking[39] haben das Tor zu einer Verbindung von Usage-Control und Provenance-Tracking aufgestoßen. Ein Vorteil der in diesem Kapitel vorgestellten integrierten Architektur ist die Wiederverwendbarkeit dieser bestehenden Implementierungen zur Nutzungskontrolle. Die in dieser Arbeit verwendet Beispiele greifen insbesondere auf die existierenden Usage-Control-Komponenten für Microsoft Windows zurück. Für das Shopsystem *Shopware*[40] und weitere kleine Applikationen wurden eigene Lösungen implementiert.

Parallel dazu wurden Beschreibungssprachen für den Datenschutz entwickelt, die dazu geeignet sind, die Präferenzen von Nutzern zur Datenverarbeitung in Unternehmen formalisierbar und die Bedingungen der Datenverarbeitung aushandelbar zu machen. Das Projekt Platform for Privacy Preferences (P3P),[41] das Endnutzertool *Privacy Bird*[42] und die Beschreibungssprachen Enterprise Privacy Authorization Language (EPAL),[43] S4P,[44] und A-PPL,[45] ebenfalls eine Erweiterung von XACML, sind Repräsentanten dieser Strömung. A-PPL wird im Projekt A4Cloud eingesetzt. Die aktuelle Spezifikation erlaubt die Deklaration einer Aktion `ActionLog`, um die Protokollierung des Umgangs mit personenbezogenen Daten und den Umfang des Protokolls festzuschreiben.

Diese Arbeit stellt keine neue Beschreibungssprache vor. Es werden bestehende Forschungsergebnisse aus zwei Gründen wiederverwendet: Erstens, um Provenance-Tracking über eine Tracking-Policy für ein bestimmtes Datum zu aktivieren.[46] Zweitens, um die Durchsetzung der über das Recht auf Auskunft hinausgehenden Betroffenenrechte, wie Löschen und Sperren, sicherzustellen. Der Beitrag dieser Arbeit ist die Integration von Usage-Control und Provenance-Tracking in einer gemeinsamen Architektur, um das Recht auf Auskunft sowie die übrigen Betroffenenrechte gemeinsam zu gewährleisten (Anforderung 4). Weitere Beiträge sind in den nachfolgenden Kapiteln im Themenfeld Informationsfluss-Tracking für Personal-Data-Provenance zu finden.

Im Hinblick auf OSL haben Usage-Control-Policies eine Struktur, die dem Schema „Event – Condition – Action (ECA)“ folgt.[47] Ein Event (Ereignis, siehe Kapitel 6.1) ist der Auslöser für die Anwendung der Policy. Die Condition (Bedingung) beschreibt die Voraussetzungen, die erfüllt sein müssen, damit der dritte Teil der Policy greift. Bedingungen können zeitlicher oder kardinaler Art sein oder von Umweltvariablen abhängen. Action ist die eigentliche Folge, falls die Policy greift. Vier verschiedene Mechanismen

[39] Harvan/Pretschner 2009; Pretschner/Lovat/Büchler 2011.
[40] https://de.shopware.com.
[41] https://www.w3.org/P3P.
[42] Lorrie Faith Cranor/Guduru/Arjula 2006.
[43] https://www.w3.org/Submission/2003/SUBM-EPAL-20031110.
[44] Becker/Malkis/Bussard 2010.
[45] Fernandez-Gago et al. 2015.
[46] Wird das integrierte System einzig für den Zweck der Auskunft betrieben, kann das Provenance-Tracking global für alle personenbezogenen Daten aktiviert werden. Eine Tracking-Policy ist dann nicht erforderlich.
[47] Hilty et al. 2007.

sind möglich: Eine Unterbindung des Events (Inhibit), eine zeitliche Verschiebung des Events (Delay), eine Modifikation des Events (Modify) oder die Ausführung einer Aktion zusätzlich zum Event (Execute).[48] Die Aufzeichnung der Provenance ist ein typischer Fall eines Execute-Mechanismus. Ein Sperren personenbezogener Daten lässt sich über einen Inhibit-Mechanismus realisieren.

Für die Auswertung und Durchsetzung der Policies sind die im folgenden Abschnitt vorgestellten Komponenten verantwortlich.

5.3 Integrierte generische Architektur für Usage-Control & Provenance-Tracking

Die Architekturen für Usage-Control und Provenance-Tracking sind ähnlich. Systeme für beide Konzepte benötigen eine Komponente, die Ereignisse abfängt und eine Komponente, die den Zustand der Umgebung beziehungsweise die Provenance speichert. Usage-Control und Provenance-Tracking ergänzen sich im Hinblick auf die Umsetzung und Durchsetzung der Betroffenenrechte.[49] Das Recht auf Auskunft setzt Data-Provenance voraus. Die Sicherstellung, dass die Provenance gesammelt wird, und die Durchsetzung der übrigen Betroffenenrechte erfordern Usage-Control.

Die grobe Struktur einer Provenance-Architektur wurde bereits in Abschnitt 5.1 (Abbildung 5.1) vorgestellt. Die Grundstruktur der Usage-Control-Architektur findet sich bereits in RFC 2748,[50] RFC 4261 (Common Open Policy Service (COPS) Over Transport Layer Security (TLS))[51] sowie in RFC 2904 (AAA Authorization Framework)[52] und hat sich im Rahmen von XACML etabliert.[53]

5.3.1 Architektur

Die integrierte generische Architektur umfasst alle für Usage-Control und Provenance-Tracking erforderlichen Komponenten. Diese werden nachfolgend motiviert und erläutert. Das Zusammenspiel der Komponenten wird anhand von Abbildung 5.3 deutlich gemacht.

Der Policy Enforcement Point (PEP) ist die Schnittstelle zwischen der Architektur und ihrer Umwelt. Für jede Abstraktionsschicht, die der Nutzungskontrolle unterworfen werden soll oder in der der Umgang mit personenbezogenen Daten nachvollzogen werden

[48] Pretschner/Hilty/Basin et al. 2008.
[49] Bereits vorgestellt und diskutiert in Bier 2013b.
[50] https://tools.ietf.org/html/rfc2748; PEP & PDP.
[51] https://tools.ietf.org/html/rfc4261.
[52] https://tools.ietf.org/html/rfc2904; PEP, PDP, PIP & PRP.
[53] http://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.html; PEP, PDP, PIP & PAP.

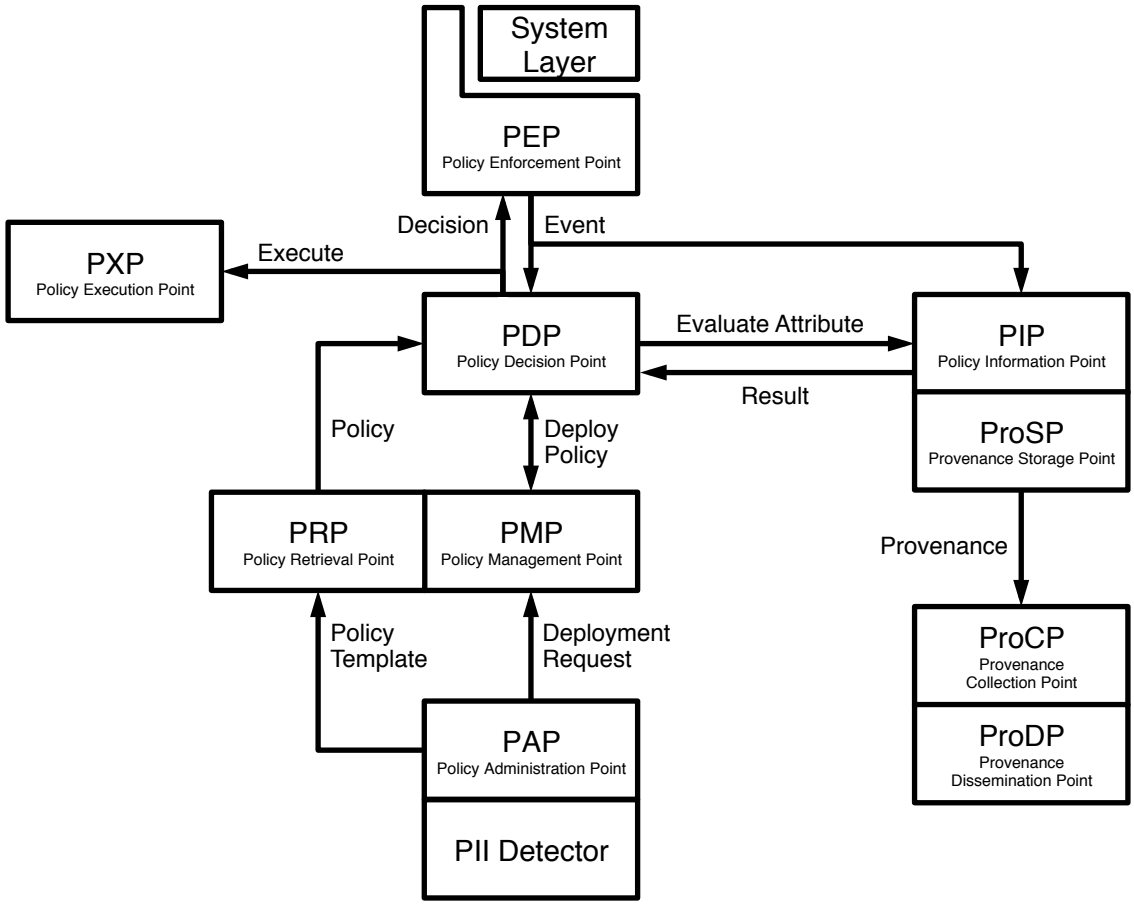


Abbildung 5.3: Logische Usage-Control- und Provenance-Architektur

soll, muss ein PEP eingebunden werden (Anforderungen 1 und 2). Abstraktionsschichten sind beispielsweise Betriebssysteme, Fenstermanager, Dateisysteme, Datenbanken, Anwendungen und Infrastrukturkomponenten wie Firewalls und Workflow-Engines. Der PEP fängt die in seiner Abstraktionsschicht stattfindenden Ereignisse ab und leitet sie an den PDP und den PIP weiter. Das Ereignis wird zunächst pausiert. Abhängig von der Entscheidung des PDP wird das Ereignis zugelassen, unterbunden oder entsprechend der Vorgaben des PDP modifiziert (siehe Abschnitt 5.2). Ein PEP, der Ereignisse nicht unterbindet oder modifiziert, wird auch als „Event-Listener“ bezeichnet.

Jeder PEP kennt die Semantik der auf seiner Abstraktionsschicht stattfindenden Ereignisse, insbesondere der daraus resultierenden Informationsflüsse (siehe Kapitel 6.2). Informationsflüsse können auch zwischen Abstraktionsschichten und IT-Systemen stattfinden. Dieser Aspekt wird in Kapitel 8.1 gesondert behandelt.

Der Policy Decision Point (PDP) hält den Zustand aller Usage-Control-Policies, insbesondere deren kardinale und zeitliche Bedingungen, vor. Wird dem PDP vom PEP ein Ereignis mitgeteilt, entscheidet er anhand des Zustands und weiterer Informationen, die der PDP vom PIP erhält, ob das Ereignis stattfinden darf, ob es modifiziert oder unterbunden werden muss oder ob das Ereignis weitere Aktionen nach sich zieht. Ein PDP ist für eine lokale Domäne, im Regelfall ein IT-System, verantwortlich.

Policy Information Point (PIP) und Provenance Storage Point (ProSP). Ein PIP hält weitere Informationen vor, die der PDP zur Entscheidungsfindung benötigt.

Ein besonderer PIP, der Informationsfluss-PIP, bildet gemeinsam mit dem ProSP den „Information-Flow-Tracking & Provenance-Storage"-Teil der Architektur.[54] Der Informationsfluss-PIP hält den Zustand des Informationsflussmodells (Kapitel 6) vor. Bei jedem Ereignis, das auf eine Policy passt, die informationsflussabhängig ist, fragt der PDP den PIP an. Der PIP berechnet den durch das Ereignis entstehenden Zustand vor und informiert den PDP auf dieser Basis. Wird ein Ereignis zugelassen, wird der PIP darüber vom PEP informiert und aktualisiert seinen internen Zustand dauerhaft. Der PIP erhält die Semantik zur Interpretation der Ereignisse vom PEP. Dazu registriert sich der PEP beim PIP, sobald er gemeinsam mit seiner Abstraktionsschicht gestartet wurde.

Der ProSP baut die Personal-Data-Provenance auf und speichert sie getrennt von den personenbezogenen Daten, für die die Provenance erzeugt wird (Anforderung 19). Personal-Data-Provenance wird also, Anforderung 20 folgend, dezentral abgelegt und für den Betroffenen zum Abruf bereitgehalten. Implementierungstechnisch können Informationsfluss-PIP und ProSP gemeinsam realisiert werden. Um beide Komponenten getrennt einsetzen zu können, beispielsweise, wenn kein Provenance-Tracking benötigt wird, werden der Informationsfluss-PIP und der ProSP nur lose gekoppelt. PIP und ProSP sind wie der PDP für eine bestimmte lokale Domäne verantwortlich.

Ein Policy Execution Point (PXP) wird vom PDP aufgerufen, sofern eine Policy gemeinsam mit einem Ereignis die Ausführung weiterer Aktionen erfordert. Ein PXP kann beispielsweise dafür verantwortlich sein, den Benutzer über die Entscheidung des PDP zu benachrichtigen.

Der Policy Management Point (PMP) ist, wie der Namen sagt, für die Verwaltung der Policies zuständig. Er nimmt neue Policy-Templates[55] entgegen, instanziiert diese und fordert den PDP dazu auf, eine Policy zu laden und anzuwenden, wenn dies von administrativer Seite gewünscht wird. Werden Policies systemübergreifend ein- und durchgesetzt,

[54] Im Rest der Arbeit wird mit dem Begriff PIP immer der Informationsfluss-PIP bezeichnet.

[55] Ein Policy-Template ist eine Policy, in der bestimmte Parameter, beispielsweise das referenzierte Datum, noch offen sind.

ist der PMP dafür zuständig, die Policy auf andere Systeme zu übertragen, bevor ein Datenfluss auf diese Systeme stattfindet. Policies und Policy-Templates werden im **Policy Retrieval Point (PRP)** gespeichert.

Ein PMP übernimmt zusätzliche Verwaltungsfunktionen. Der PMP ist für die Domänenhierarchie verantwortlich und stellt ähnlich dem Domain Name System (DNS)[56] einen Naming-Service für Domänen zur Verfügung. Details dazu finden sich in Anhang C. In jeder lokalen Domäne existiert ein PMP, der für die Verwaltung der Domäne verantwortlich ist. Ein zentraler Root-PMP steht an der Spitze der Hierarchie. Alle Komponenten registrieren sich bei ihrem lokalen PMP und können dort von anderen Komponenten angefragt werden. Ein PXP registriert außerdem die von ihm unterstützten Aktionen beim PMP, sobald er verfügbar ist. Der PDP kann dann bei Bedarf den passenden PXP beim PMP anfragen.

Mit Hilfe des Policy Administration Point (PAP) kann ein Administrator neue Policies erstellen oder dem PMP bekanntmachen. Eine neue Policy kann insbesondere auf Veranlassung einer anderen Policy erzeugt werden. In diesem Fall weist ein PXP den PAP dazu an, die entsprechende Policy zu erstellen und über den PMP in den PDP zu laden.

Der Provenance Collection Point (ProCP) aggregiert auf Anfrage die in den einzelnen ProSPs verteilt gespeicherte Provenance und stellt sie als einheitliche Datenstruktur zur Verfügung. Die Aggregation der Personal-Data-Provenance findet, in Übereinstimmung mit Anforderung 21, nur dann statt, wenn sich der Betroffene authentifiziert und eine Auskunftsanfrage gestellt hat. Der ProCP speichert selbst keine Provenance und löscht die aggregierte Provenance aus dem Zwischenspeicher, sobald sie an den Betroffenen ausgeliefert wurde (Anforderung 22). Der ProCP speichert allerdings Informationen, die zwingend zentral vorhanden sein müssen. Dazu zählen Authentifikationsmerkmale des Betroffenen, externe Stellen und das IT-System, durch das ein personenbezogenes Datum erstmalig erhoben wurde. Die Gründe werden in Kapitel 8.2 erläutert.

Da es nur einen zentralen ProCP gibt, registriert sich dieser direkt beim Root-PMP.

Der ProCP ist nicht mit dem Begriff der „Provenance-Collection“ aus Abschnitt 5.1 identisch. Die Komponente zur Erhebung der Provenance ist der PEP. Wird in den folgenden Kapiteln von „Provenance-Collection“ gesprochen, ist der ProCP gemeint.

Der Provenance Dissemination Point (ProDP) erhält die, über verteilte Systeme hinweg aggregierte, Personal-Data-Provenance vom ProCP. Er bereitet die Provenance so auf, dass sie vom Betroffenen verstanden werden kann. Er ist die Schnittstelle zwischen dem Datenschutzauskunftssystem und dem Betroffenen. Der ProDP ist die Auskunftsplattform, auch Privacy-Dashboard genannt.

[56] https://tools.ietf.org/html/rfc1034.

5.3.2 Schnittstellen und Implementierung

Bis auf zwei Ausnahmen sind alle Komponenten der Architektur in der Programmiersprache Java implementiert. Sie sind dadurch plattformunabhängig und flexibel einsetztbar. Die beiden Ausnahmen sind der PEP und der ProDP.

Die Realisierung eines PEP hängt von der Abstraktionsschicht ab, für die er eingesetzt werden soll. Soweit möglich wird ein PEP als dynamisch ladbare Erweiterung der Abstraktionsschicht implementiert. Für solche Erweiterungen sind Programmiersprache und Struktur meistens vorgegeben.

Der ProDP, die *PrivacyInsight* genannte Auskunftsplattform, ist als Webanwendung umgesetzt. *PrivacyInsight* ist in HTML 5 und JavaScript implementiert und über eine REST-Schnittstelle an den ProCP angebunden.

Die Komponenten kommunizieren über Remote Method Invocation (RMI) oder unter Austausch von JSON-Objekten über eine native TCP-Verbindung. Andere Kommunikationsprotokolle und Austauschformate sind konzeptionell nicht ausgeschlossen und durch eine modulare Implementierung explizit vorgesehen.

Die Schnittstelle des PDP ist kompakt (Listing 5.1). Die Methode notify() erlaubt dem PEP, auftretende Ereignisse an den PDP zu melden. Die drei anderen Methoden nutzt der PMP um das Laden und Zurückziehen von Policies anzusteuern.

```
Decision notify(Event event);
boolean deploy(PolicyId policyId);
boolean revokePolicy(PolicyId policyId);
ArrayList<PolicyId> listDeployedPolicies();
```

Listing 5.1: Schnittstelle des PDP

Die Schnittstelle des PIP (Listing 5.2) hat zunächst drei Methoden zur Registrierung des PEP und seiner Informationsflusssemantik (siehe Kapitel 6.2 und 8.1). Dem folgt eine Methode, um die initiale Manifestation eines Datums in einem Container festzulegen (setNewRepresentation(), siehe Kapitel 6.1) und eine Methode, die es dem PEP erlaubt, den PIP über Ereignisse zu informieren (notifyEvent()). Die übrigen Methoden erlauben unterschiedliche Zustandsabfragen durch den PDP und PMP.

```
boolean registerPEP(ComponentId pep, String ifsemantic);
boolean registerPEP(ComponentId pep, String ifsemantic, String remoteIfsemantic);
boolean unregisterPEP(ComponentId pep);
boolean setNewRepresentation(DataId data, Container container);
boolean notifyEvent(ComponentId pepID, Event newEvent);
Set<DataId> getDataOfContainer(ContainerId container);
ContainerId getContainerByDesignator(String designator);
Set<ContainerId> getContainerOfData(DataId data);
boolean representationRefinesData(ContainerId container, DataId data);
```

```
boolean representationRefinesDataAfterEvent(ContainerId container, DataId data, ComponentId
    ↪ componentID, Event event);
```

Listing 5.2: Schnittstelle des PIP für Informationsflüsse

Verbindet man den PIP mit dem ProSP sind zwei zusätzliche Methoden erforderlich, die das Provenance-Tracking aktivieren beziehungsweise deaktivieren (Listing 5.3).

```
boolean setTracking(DataId data);
boolean abandonTracking(DataId data);
```

Listing 5.3: Schnittstelle des PIP für Provenance

Intern ruft der PIP bei aktiver Provenance die Methoden createRepresentation() und terminateRepresentation() auf (siehe Kapitel 6.3). Diese Methoden werden auch gemeinsam mit der Methode addSourceMetadata() von außen zur Initialisierung der Provenance verwendet. Die letzten beiden Methoden in Listing 5.4 erlauben das Abrufen der Provenance beim ProSP.

```
ContainerId addSourceMetadata(String source, DataId data, String dataCategory, String
    ↪ dataSubject);
boolean createRepresentation(Container cont, DataId data, ContainerId source, String purpose);
boolean terminateRepresentation(ContainerId cont, DataId data);
ProvenanceGraph getProvenanceOfData(DataId data);
ProvenanceGraph getProvenanceByLocalRoot(RepresentationId rootID);
```

Listing 5.4: Schnittstelle des ProSP

Die Schnittstelle des PXP ist generisch und erlaubt unter Übergabe des Ereignisses ein Ausführen der durch den PXP unterstützten Aktionen (Listing 5.5).

```
boolean execute(ExecuteAction action, Event event);
```

Listing 5.5: Schnittstelle des PXP

Die Schnittstelle des PMP (Listing 5.6) erlaubt dem PAP die Ansteuerung des Policy-Managements, insbesondere der Instanziierung. Nicht dargestellt sind die Registrierung und das Abrufen der Komponenten sowie die Domänendienste. Der PRP erlaubt das Abspeichern (storePolicy()) und Abrufen (retrievePolicy()) von Policies.

```
PolicyId deployPolicy(URI domain, String policy);
PolicyId deployPolicyTemplate(URI domain, String policy);
boolean revokePolicy(URI domain, PolicyId policyId);
PolicyId instantiatePolicy(URI domain, PolicyId policy, DataId data);
PolicyId instantiatePolicy(URI domain, PolicyId policy, DataId data, PolicyId pId);
```

Listing 5.6: Schnittstelle des PMP

Die ersten vier Methoden des ProCP erlauben es den ProSPs, Metadaten zur Provenance zu hinterlegen, die später zum Auffinden der Provenance notwendig sind (Listing 5.7). Die Methode getProvenanceForSubject() erlaubt es dem ProDP, die vollständige Provenance für einen Betroffenen abzurufen.

```
boolean contains(DataId data);
boolean setDataCollector(DataId dataID, URI domain);
boolean addDataForSubject(DataId data, SubjectId dataSubject);
ExternalContainer getExternalContainer(String description);
ProvenanceGraph getProvenanceForSubject(SubjectId dataSubject);
```

Listing 5.7: Schnittstelle des ProCP

PAP und ProDP haben bis auf die grafische Benutzerschnittstelle keine Schnittstellen.

5.3.3 Sicherheitsannahmen

Den Überlegungen der folgenden Kapitel liegen die folgenden Annahmen zur Sicherheit des Datenschutzauskunftssystems zu Grunde:

- Das Datenschutzauskunftssystem ist korrekt implementiert und frei von Bugs.
- Das Datenschutzauskunftssystem wird vor seinem Einsatz in allen IT-Systemen korrekt aufgesetzt.
- Die verbundenen Komponenten des Datenschutzauskunftssystems laufen immer dann in einer lokalen Domäne, wenn eine Verarbeitung personenbezogener Daten stattfindet (Anforderung 54).
- Eine Modifikation des Datenschutzauskunftssystems durch die datenverarbeitenden Organisationseinheiten ist nicht möglich (Anforderung 39).
- Ein Umgehen des Provenance-Trackings ist nicht möglich oder wird organisatorisch unterbunden (Anforderung 54).
- Ein Zugriff auf die Provenance ist nur nach Authentifikation durch den Betroffenen über die implementierten Schnittstellen möglich (Anforderung 30 ff.).
- Jeder Zugriff auf die Provenance einer lokalen Domäne wird revisionssicher protokolliert. Abfragen des ProCP, die nicht auf eine Betroffenenanfrage gestützt sind, können entdeckt und sanktioniert werden (Anforderung 40).
- Die Kommunikation zwischen den Komponenten des Datenschutzauskunftssystems wird durch starke Kryptographie geschützt.

Nichtsdestotrotz ist das Datenschutzauskunftssystem grundsätzlich kein Sicherheitssystem. Der verantwortlichen Stelle muss zumindest soweit vertraut werden, dass sie keine Hintertüren in ihre eigenen IT-Systeme einbaut. Datenschutz funktioniert nur durch ein Zusammenspiel organisatorischer, rechtlicher und technischer Maßnahmen. Technische Defizite, die in der Softwareentwicklung immer vorkommen, können durch organisatorische Maßnahmen ausgeglichen werden.

5.4 Zwischenfazit

Neben einem Überblick über Provenance-Tracking und Usage-Control wurde in diesem Kapitel eine verteilte, generische Architektur für ein Datenschutzauskunftssystem vorgestellt, die beides verbindet. Provenance-Tracking und Usage-Control erlauben gemeinsam die Durchsetzung von Transparenz und Intervenierbarkeit. Die Betroffenenrechte werden ganzheitlich angegangen. Die vorgeschlagene Architektur erfüllt die in Konstruktionsziel $\mathfrak{K}2$ aufgestellten Bedingungen.

6 Ein gemeinsames Modell für Informationsfluss- & Provenance-Tracking

Um den Umgang mit personenbezogenen Daten im geforderten Umfang beauskunften zu können, muss der Fluss personenbezogener Daten vollständig nachvollzogen werden können. Kopien der erhobenen personenbezogenen Daten und von ihnen abgeleitete Daten behalten den Personenbezug und sind auskunftspflichtig.[1] Der Kern des Datenschutzauskunftssystems ist deshalb die Infrastruktur zur Erfassung des aktuellen Informationsflusszustands und zur Speicherung der Provenance. In diesem Kapitel wird zunächst das aus dem Usage-Control-Umfeld stammende Informationsflussmodel eingeführt und erweitert (Abschnitt 6.1). Gemäß dieses Modells wird der jeweils aktuelle Informationsflusszustand (Abschnitt 6.1) im PIP gespeichert.

Die Erfassung des Informationsflusszustands erfolgt anhand der durch die PEPs der einzelnen Abstraktionsschichten übermittelten Ereignisse. Die Art dieser Ereignisse unterscheidet sich je nach Abstraktionsschicht und ist dem PIP a priori unbekannt. Um unterschiedliche Systeme und Applikationen in die Datenschutzauskunft einbinden zu können, wird in Abschnitt 6.2 ein Schema zur semantischen Beschreibung von Informationsflüssen (Informationsflusssemantik) in Abhängigkeit von Ereignissen eingeführt.

Die Provenance an sich, also die datenbezogene Historie des Informationsflusses, wird getrennt vom Informationsflusszustand in einer eigenen Datenstruktur abgelegt (Abschnitt 6.3.1). Diese konzeptionelle Entscheidung hat den Vorteil, dass die Granularität der Provenance, den Anforderungen zum Umfang einer Datenschutzauskunft entsprechend (Anforderung 5 ff.), gröber als die Granularität des Informationsfluss-Trackings sein kann. Die darüber hinausgehenden Vorteile dieses Konzeptes beim Entwurf eines speicherskalierbaren Datenschutzauskunftssystems werden im Kapitel 7 besprochen.

Dennoch sind das Informationsflussmodell und das Personal-Data-Provenance-Modell eng miteinander verzahnt. Der Zustand des letzteren leitet sich aus dem des ersteren ab (Abschnitt 6.3.2).

Die wichtigsten Aspekte der Implementierung der Provenance-Datenhaltung werden am Ende des Kapitels erläutert (Abschnitt 6.4).

[1] Sofern sie nicht anonymisiert werden.

6.1 Usage-Control-Informationsflussmodell

Pretschner und Harvan haben das Informationsflussmodell für Usage-Control formal als Tupel $(\mathscr{D}, \mathscr{C}, \mathscr{F}, \Sigma, \mathscr{E}, \mathscr{T})$ eingeführt.[2]

$\mathscr{D}$ ist die Menge aller personenbezogenen Daten. $\mathscr{C}$ ist die Menge aller Container. Ein Container ist ein Ort, an dem sich ein Datum in einem IT-System manifestiert. Beispielsweise kann sich ein Profilfoto von Alice in einer Datei im Dateisystem, in einer Datenbank und gleichzeitig in einem Fenster des Fenstermanagers manifestieren. Neben Speicherorten werden je nach Abstraktionsniveau auch Prozesse, Kommunikationskanäle und ganze Stellen als Container aufgefasst. Die Unterscheidung zwischen Daten und Containern ist der zentrale Aspekt des Modells. $\mathscr{F}$ ist die Menge aller Bezeichner, die Container eindeutig identifizieren.

$\Sigma = (\mathscr{C} \to \mathcal{P}(\mathscr{D})) \times (\mathscr{C} \to \mathcal{P}(\mathscr{C})) \times (\mathscr{F} \to \mathscr{C})$ ist die Menge aller Zustände, bestehend aus den folgenden drei Abbildungen: Die *Speicherfunktion* $s : \mathscr{C} \to \mathcal{P}(\mathscr{D})$ bildet ab, welche Daten sich in welchen Containern befinden. Die *Aliasfunktion* $l : \mathscr{C} \to \mathcal{P}(\mathscr{C})$ beschreibt eine implizite Abhängigkeit des Zustands von Containern bezüglich eines Datums. Ändert sich die Speicherfunktion des Containers $c_1 \in C$, ändert sich entsprechend auch die *Speicherfunktion* eines jeden Containers $c_2 \in l(c_1)$. Die *Benennungsfunktion* $f : \mathscr{F} \to \mathscr{C}$ bildet Bezeichner auf Container ab. $\sigma_0 = (\emptyset, \emptyset, \emptyset) \in \Sigma$ ist der initiale Systemzustand.

Ereignisse (Events) $\mathscr{E}$ sind beobachtete Vorgänge, die Änderungen im Systemzustand widerspiegeln. *Ereignisse* resultieren in Änderungen der Speicher-, Alias- und Benennungsfunktion. Die Übergänge werden durch die (deterministische) Relation $\mathscr{T} \subseteq \Sigma \times \mathscr{E} \times \Sigma$ beschrieben.

Im Modell werden drei unterschiedliche Containermengen unterschieden: An erster Stelle sind reale Container $\mathscr{C}_{real}$, die Orte beschreiben, an denen sich Daten befinden können, Teil des Modells. Darüber hinaus werden abstrakte Container $\mathscr{C}_{abstr}$ modelliert. Sie fassen die unterschiedlichen realen Container in einem gemeinsamen Zustand zusammen und abstrahieren von der durch das Informationsflussmodell gegebenen Granularität auf die für Provenance erforderliche Granularität. Ein abstrakter Container vereint in sich alle Speicherfunktionen der zugrundeliegenden realen Container. Reale Container und abstrakte Container werden über eine unidirektionale Aliasrelation miteinander verbunden.[3] Als dritte Menge werden virtuelle Intermediate-Container $\mathscr{C}_\iota$ modelliert. Sie bilden eine Brücke zwischen unterschiedlichen Abstraktionsschichten innerhalb eines IT-Systems sowie zwischen IT-Systemen.[4] Ergänzend existiert der spezielle Container NIL, auf den Namen abgebildet werden, die noch keinem konkreten Container zugeordnet sind. Insgesamt ergibt sich $\mathscr{C} = \mathscr{C}_{real} \cup \mathscr{C}_{abstr} \cup \mathscr{C}_\iota \cup \{\text{NIL}\}$. $\mathscr{C}_{real}$, $\mathscr{C}_{abstr}$ und $\mathscr{C}_\iota$ sind paarweise disjunkt.

[2] Harvan/Pretschner 2009; Pretschner/Lovat/Büchler 2011.
[3] Vgl. Kapitel 7.2.
[4] Vgl. Kapitel 8.

Beispiel. *Insgesamt verarbeitet AdBokis 30 personenbezogene Daten $d \in \mathscr{D}$ ihrer beiden Kunden in 17 Datenkategorien $\theta \in \Theta$. In Tabelle 6.1 sind exemplarisch die personenbezogenen Daten von Alice Fox aufgelistet.*

d_x	θ_x	Datenkategorie	Inhalt
1	1	Vorname	Alice
2	2	Name	Fox
3	3	e-Mail	alice.fox@honigmail.de
4	4	Telefonnummer	+49 721 6091-0
5	5	Geburtsdatum	15.05.1985
6	6	Rechnungsadresse	Hansastraße 27c
7	7	Lieferadresse	Fraunhoferstr. 1
8	7	Lieferadresse	PostNummer 7654321
9	8	Wohnsitzstaat	Germany
10	9	Zustellungsart	Lieferung nach Hause
11	9	Zustellungsart	Packstation
12	10	IBAN	DE19 1234 1234 1234 1234 12
13	11	Kreditkartentyp	Master Card
14	12	Kreditkartennummer	4343 9534 4301 1007
15	13	Ablaufdatum der Kreditkarte	Dez 21
16	14	Profilbild	[nicht darstellbar]
17	15	IP-Adresse	217.146.191.19
18	15	IP-Adresse	31.130.202.80
19	16	Empfehlung	Inges Braustubenführer
20	17	Rechnung	[nicht darstellbar]

Tabelle 6.1: Personenbezogene Daten von Alice

Über das Informationsflussmodell hinaus ist es für den Anwendungsfall der Datenschutzauskunftssysteme erforderlich, ergänzend die Menge der Betroffenen $\mathscr{B}$ einzuführen. Die *Personenbezugsfunktion* $\pi : \mathscr{B} \to \mathcal{P}(\mathscr{D})$ weist jedem Betroffenen diejenigen Daten zu, die einen Personenbezug zu ihm besitzen.

Beispiel. *AdBokis hat im Minimalbeispiel zwei Kunden (Betroffene $b \in \mathscr{B}$): Alice Fox (b_1) und Peter Trollig (b_2).*

Die Architektur des Datenschutzauskunftssystems sieht vor, dass der Zustand σ im PIP vorgehalten wird. Des Weiteren sollen PEPs dynamisch zur Laufzeit hinzugefügt werden können. In der Konsequenz ist der PEP die Komponente, die den Teil der Übergangsrelation $\mathscr{T}$ spezifiziert, der die von ihm weitergereichten Ereignisse betrifft. Durch diese Trennung ist dem PIP die Übergangsrelation zunächst unbekannt.

Deshalb registriert sich ein PEP beim PIP, sobald er zu einem Datenschutzauskunftssystem hinzugefügt wird (Methode registerPEP()). Im Zuge dessen wird dem PIP vom PEP die semantische Spezifikation des Teils der Übergangsrelation bekannt gemacht, der die vom jeweiligen PEP abgefangenen und weitergereichten Ereignisse betrifft. Der nachfolgende Abschnitt erläutert das dahinterstehende Modell und die Realisierung der Spezifikation.

6.2 Semantische Beschreibung der Übergangsrelation

Der jeweilige Abschnitt der Übergangsrelation $\mathscr{T}$ ist für jeden PEP in einer Informationsflusssemantik spezifiziert. Für jedes vom PEP abgefangene Ereignis gibt die Informationsflusssemantik die Zustandsübergänge der Funktionen s, l und f an. Eine Informationsflusssemantik entspricht dem Schema in Anhang D.1. Um die semantische Spezifikation für die Praxis handhabbar zu machen, enthält sie nicht direkt die Beschreibung der Übergangsrelation, sondern benutzt generische Primitive.

6.2.1 Generische Primitive zur Beschreibung der Informationsflusssemantik

Die generischen Primitive,[5] Hilfsfunktionen zur Beschreibung von Zustandsübergängen, definieren alle Änderungen des Informationsflussmodells, die in einer Semantik spezifiziert werden können. Sie werden bei der Verarbeitung eines Ereignisses durch den PIP nacheinander gemäß der Informationsflusssemantik aufgerufen und modifizieren die Funktionen s, l und f. Sie werden im Folgenden beschrieben; ein Anwendungsbeispiel wird im Anschluss vorgestellt (Abschnitt 6.2.2).

Um die Veränderungen der Funktionen s, l und f durch Ereignisse zu beschreiben, wird auf die Notation aus den Arbeiten von Harvan und Pretschner zurückgegriffen.[6] Sei $m : A \to B$ eine beliebige Abbildung und $a \in A$ ein Element der Urbildmenge. Dann ist $m[a \leftarrow expr]_{a \in A} = m'$ mit $m' : A \to B$ definiert als

$$m'(a') = \begin{cases} expr & \text{für } a' = a \\ m(a') & sonst. \end{cases}$$

Generische Primitive zum Update der Speicherfunktion

Die Speicherfunktion hält die Abbildungen zwischen Daten und Containern vor. Sie ist die zentrale Funktion zur Modellierung von Informationsflüssen.

Das erste Primitiv für die Speicherfunktion ist das *flow*-Primitiv. Es bezeichnet den Fluss einer Menge von Daten $\{d_i\}_{1 \leq i \leq n \in \mathbb{N}}$ in einen Container c. Zusätzlich findet der Fluss in die

[5] Die Primitive wurden bereits in Birnstill/Bier et al. 2016 eingeführt.
[6] Harvan/Pretschner 2009.

referenzierten abstrakten Container statt. Das *flow*-Primitiv wird beispielsweise verwendet, um zu modellieren, dass eine neue Datei oder ein neuer Prozess erzeugt wird oder dass eine Datei kopiert wird.

Primitiv 1: $flow(\sigma, c, \{d_i\}_{1 \leq i \leq n \in \mathbb{N}})$

$(s, l, f) \leftarrow \sigma$
$s \leftarrow s[c \leftarrow s(c) \cup \{d_i\}_{1 \leq i \leq n \in \mathbb{N}}]$
forall $u \in l(c) \cap C_{abstr}$ **do**
  $s \leftarrow s[u \leftarrow s(u) \cup \{d_i\}_{1 \leq i \leq n \in \mathbb{N}}]$
end
return (s, l, f)

Findet der Fluss nicht nur in einen Container statt, sondern sind mehrere Container logisch miteinander verbunden, findet das Primitiv 2 Anwendung. Liest beispielsweise ein Prozess Daten ein und wird dieser Prozess im Fenster eines Fenstermanagers visualisiert, dann findet der Fluss nicht nur in den Speicherbereich des Prozesses, sondern auch in das Fenster statt. l^* ist die reflexiv-transitive Hülle von l. Über sie lassen sich alle Container in Ketten von Aliasrelationen adressieren.

Primitiv 2: $flow\_to\_rtc(\sigma, c, \{d_i\}_{1 \leq i \leq n \in \mathbb{N}})$

$(s, l, f) \leftarrow \sigma$
forall $u \in l^*(c)$ **do**
  $s \leftarrow s[u \leftarrow s(u) \cup \{d_i\}_{1 \leq i \leq n \in \mathbb{N}}]$
end
return (s, l, f)

Das *clear*-Primitiv 3 wird eingesetzt, wenn ein Container gelöscht oder überschrieben wird. Beispiele sind das Schließen eines Fensters oder das Löschen einer Datei.

Primitiv 3: $clear(\sigma, c)$

$(s, l, f) \leftarrow \sigma$
$s \leftarrow s[c \leftarrow \varnothing]$
forall $u \in l(c) \cap C_{abstr}$ **do**
  $s \leftarrow s[u \leftarrow \bigcup_{v \in C \mid u \in l(v)} s(v)]$
end
return (s, l, f)

Um den Löschvorgang auf die abstrakten Container zu übertragen, wird zusätzlich der Speicherzustand aller referenzierten abstrakten Container auf Grundlage der auf sie verweisenden realen Container aktualisiert.

Generische Primitive zum Update der Aliasfunktion

Die Aliasfunktion definiert die Beziehungen zwischen unterschiedlichen Containern, die zu impliziten Informationsflüssen führen können. Wann immer ein Datum in den ersten Container c_{from} fließt, wird es auch in den per Alias verbundenen Container c_{to} fließen.

Das Primitiv 4 fügt einen unidirektionalen Alias von c_{from} nach c_{to} hinzu. Beispielsweise kann so die Beziehung einer gespiegelten Datei zu einer Originaldatei, auf die nur Lesezugriff besteht, beschrieben werden.

Primitiv 4: $create\_alias(\sigma, c_{from}, c_{to})$

$(s, l, f) \leftarrow \sigma$
$l \leftarrow l[c_{from} \leftarrow l(c_{from}) \cup \{c_{to}\}]$
return (s, l, f)

Zum Aufbau bidirektionaler Aliase wird das Primitiv 5 verwendet. Es findet beispielsweise Anwendung, wenn zu einem Prozess ein Fenster erzeugt wird.

Primitiv 5: $create\_bidir\_alias(\sigma, c_{from}, c_{to})$

$(s, l, f) \leftarrow \sigma$
$l \leftarrow l[c_{from} \leftarrow l(c_{from}) \cup \{c_{to}\}]$
$l \leftarrow l[c_{to} \leftarrow l(c_{to}) \cup \{c_{from}\}]$
return (s, l, f)

Das Primitiv 6 ist das Inverse von Primitiv 4. Es entfernt genau einen Alias.

Primitiv 6: $rm\_alias\_locally(\sigma, c_{from}, c_{to})$

$(s, l, f) \leftarrow \sigma$
$l \leftarrow l[c_{from} \leftarrow l(c_{from}) \setminus \{c_{to}\}]$
return (s, l, f)

In manchen Fällen ist es notwendig, alle unidirektionalen Aliase auf einen Container zu entfernen, beispielsweise wenn dieser gelöscht wird. Dies erfolgt mit Primitiv 7.

Primitiv 7: $rm\_alias\_globally(\sigma, c_{to})$

$(s, l, f) \leftarrow \sigma$
forall $c \in C$ **do**
 | $l \leftarrow l[c \leftarrow l(c) \setminus \{c_{to}\}]$
end
return (s, l, f)

Das Primitiv 8 ist das Gegenstück zu Primitiv 7. Es entfernt alle Aliase, die von einem bestimmten Container ausgehen.

Primitiv 8: $clear\_aliases(\sigma, c)$

$(s, l, f) \leftarrow \sigma$
$l \leftarrow l[c \leftarrow \varnothing]$
return (s, l, f)

Bidirektionale Aliase werden mit Primitiv 9 entfernt. Es ist das Inverse des Primitivs 5.

Primitiv 9: $rm\_bidir\_alias\_locally(\sigma, c_{from}, c_{to})$

$(s, l, f) \leftarrow \sigma$
$l \leftarrow l[c_{from} \leftarrow l(c_{from}) \setminus \{c_{to}\}]$
$l \leftarrow l[c_{to} \leftarrow l(c_{to}) \setminus \{c_{from}\}]$
return (s, l, f)

Generische Primitive zum Update der Benennungsfunktion

Die Benennungsfunktion weist Containern Bezeichner zu. Beispielsweise werden Dateien über Dateinamen und über Datei-Handles adressiert.

Einem Container wird ein neuer Bezeichner $\phi \in \mathscr{F}$ mit Hilfe des *add_naming*-Primitivs 10 zugewiesen.

Primitiv 10: $add\_naming(\sigma, \phi, c)$

$(s, l, f) \leftarrow \sigma$
$f \leftarrow f[\phi \leftarrow c]$
return (s, l, f)

Das inverse Primitiv des *add_naming*-Primitivs ist das letzte generische Primitiv *rm_naming*.

Primitiv 11: $rm\_naming(\sigma, \phi)$
$(s, l, f) \leftarrow \sigma$
$f \leftarrow f[\phi \leftarrow \text{NIL}]$
return (s, l, f)

6.2.2 Anwendung der Informationsflusssemantik im PIP

Der PIP erzeugt die Übergangsrelation $\mathscr{T}(\sigma, e)$ anhand der Informationsflusssemantiken der bei ihm registrierten PEPs aus den generischen Primitiven.

Die Informationsflusssemantik enthält drei Arten von Informationen: (1) Welche generischen Primitive die Übergangsrelation für ein bestimmtes Ereignis beschreiben, (2) in welcher Reihenfolge sie aufzurufen sind und (3) welche Parameter eines Ereignisses auf die in der Signatur der generischen Primitive verwendeten Variablen abgebildet werden müssen.

Für jeden Parameter eines Ereignisses wird in der Informationsflusssemantik sein Typ angegeben. Mit den Parametertypen spezifiziert die Informationsflusssemantik, welcher Parameter eines Ereignisses als Datum ($d \in \mathscr{D}$), Container ($c \in \mathscr{C}$) oder Bezeichner ($\phi \in \mathscr{F}$) aufzufassen ist.[7]

Ist durch den PIP ein Ereignis auszuwerten, werden die Funktionen f und s auf jeden Parameterwert des Ereignisses so angewandt, dass der Typ der resultierenden Werte mit den Variablen in der Signatur der aufzurufenden Primitive übereinstimmt. Die generischen Primitive werden anschließend nacheinander aufgerufen.

Die konkrete Auswertung der Informationsflusssemantik sowie die einzelnen Aspekte der Ereignisverarbeitung werden nachfolgend anhand eines Beispiels erläutert.

Beispiel. *Ein Mitarbeiter der AdBokis kopiert eine Datei mit Kundendaten aus einem temporären Verzeichnis auf den Desktop.*

Nach dem Start des Betriebssystems registriert sich der Windows-PEP beim PIP und übergibt seine Informationsflusssemantik. Listing 6.1 zeigt den Ausschnitt der Informationsflusssemantik für die Kopieroperation (CopyFile). Die Semantik gibt für die beiden relevanten Parameter der Ereignisse, newFileName und oldFileName, den jeweiligen Typ an. Beide Parameter sind Bezeichner ($\mathscr{F}$). Die anschließenden *Aktionsbeschreibungen* geben für jede Ereigniskategorie die zu verwendenden Primitive und ihre Parameter an. Ein CopyFile-Ereignis erfordert die Anwendung der Primitive 2 (*flow_to_rtc*) und 3 (*clear*).

[7] Auch Datenmengen und Containermengen sind möglich. Deren Elemente werden einzeln konvertiert.

```
<?xml version="1.0" encoding="UTF-8"?>
<ifsemantic>
  <params>
    <param name="newFileName" type="DESIGNATOR" />
    <param name="oldFileName" type="DESIGNATOR" />
  </params>
  <actions>
    <action name="CopyFile">
      <operation name="SF_CLEAR">
        <left>
          <operand>newFileName</operand>
        </left>
        <right></right>
      </operation>
      <operation name="SF_FLOW_TO_RTC">
        <left>
          <operand>newFileName</operand>
        </left>
        <right>
          <operand>oldFileName</operand>
        </right>
      </operation>
    </action>
  </actions>
</ifsemantic>
```

Listing 6.1: Informationsflusssematik für das Copy-Ereignis des Windows-PEP

```
<event action="CopyFile" timestamp="2016-07-10T21:00:00">
  <parameter name="ProcessName" value="cp.exe">
  <parameter name="PID" value="2924">
  <parameter name="ProcessOwner" value="bi">
  <parameter name="TID" value="4040">
  <parameter name="oldFileName" value="C:\tmp\customers.csv">
  <parameter name="newFileName" value="C:\Users\bi\Desktop\customers.csv">
</event>
```

Listing 6.2: Copy-Ereignis des Windows-PEP

Sobald der Windows-PEP dem PIP ein CopyFile-Ereignis weiterleitet (Listing 6.2), wertet der PIP das Ereignis gemäß der gegebenen Semantik aus. Zunächst wird geprüft, ob alle in der Semantik angegebenen Parameter vorhanden sind. Ist dies der Fall, werden die Typen der Parameter mit den Signaturen der Primitive verglichen. Das Primitiv *flow_to_rtc* erwartet gemäß seiner Signatur einen Container und eine Menge von Daten. Die Para-

meter des Ereignisses sind jedoch jeweils vom Typ „Bezeichner“. Deshalb werden sie vor Ausführung der Primitive in die Zieltypen umgewandelt. Für den oldFileName sieht das wie folgt aus: $s(f(\texttt{C:\textbackslash tmp\textbackslash customers.csv})) = \{d_1, \ldots, d_n\}$. Die Ausführung der Primitive sorgt dafür, dass der Zielcontainer zunächst keine Speicherrelation mehr besitzt und ihm dann alle Daten des Quellcontainers zugewiesen werden.

Die Parameter des Ereignisses, die für die Informationsflussverarbeitung nicht relevant sind (z. B. ProcessOwner), werden ignoriert.

6.3 Provenance-Datenmodell

Das Informationsflussmodell repräsentiert den Zustand eines Systems zu einem bestimmten Zeitpunkt. Für die Datenschutzauskunft ist die durchgängige Historie der Informationsflüsse erforderlich, die Provenance der personenbezogenen Daten. Im Folgenden wird beschrieben, wie die Provenance aus dem Zustand des Informationsflussmodells und den Ereignissen abgeleitet werden kann, die Zustandsänderungen auslösen.

6.3.1 Personal-Data-Provenance-Datenmodell

Das Provenance-Datenmodell dokumentiert den Umgang mit personenbezogenen Daten über die Zeit. Es ist ein Graph, genauer ein Wald, in dem genau ein Baum (*Out-Tree*) für jedes Datum steht. Die Wurzel eines Baumes entspricht der Erhebung personenbezogener Daten bei einer Stelle außerhalb der verantwortlichen Stelle. Jeder weitere Knoten repräsentiert einen Verarbeitungsvorgang, eine Datenspeicherung, eine Datenweitergabe oder eine weitere externe Stelle. Die Kanten stehen für die Informationsflussbeziehungen.

Ein Provenance-Graph ist definiert als ein Wald $\mathscr{G} = (\mathscr{R}, \mathscr{L})$ bestehend aus den Knoten (Repräsentationen) $\mathscr{R} \subseteq \mathscr{D} \times \mathscr{C} \times \mathbb{N}$ und den gerichteten Kanten $\mathscr{L} \subseteq \mathscr{R} \times \mathscr{R}$ zwischen diesen Knoten. Ein Knoten steht für die Repräsentation eines Datums in einem Container zu einem bestimmten Zeitpunkt. Der Zeitpunkt einer Repräsentation ist ihr Entstehungszeitpunkt, also jener Moment, in dem das Datum in einen Container geflossen ist, in dem es sich zuvor nicht befunden hat. Eine Kante beschreibt den Informationsfluss eines Datums von einem Container zu einem anderen. Sie verbindet die Repräsentation, die für die Beziehung Datum-Quellcontainer zum Zeitpunkt des Informationsflusses steht, und die Repräsentation, die durch den Informationsfluss in den Zielcontainer neu entsteht.

Dass für jedes Datum genau ein Baum im Provenance-Graphen existiert, ist wie folgt begründet: Ist ein Datum bereits erhoben, ist für dessen Verwendung keine weitere Erhebung mehr erforderlich. Gibt es mehrere Erhebungen desselben Datums, kann dies nur darin begründet liegen, dass das Datum durch die erhebende Stelle nicht wiedererkannt wird. Die beiden Bäume werden dann wie die von zwei unterschiedlichen Daten behandelt. Es gibt keine geschlossenen Pfade, da ein Datum nirgendwohin fließen kann,

wo es bereits ist. Im Informationsflussmodell kann nur dann eine neue Kante zwischen Datum und Container entstehen, wenn zuvor $d \notin s(c)$ war. Somit ist die Baumstruktur der Provenance sichergestellt.

Auf einer Repräsentation ist ergänzend die Funktion $term : \mathscr{R} \rightarrow \mathbb{N} \cup \text{NIL}$ definiert. Sie weist jeder Repräsentation einen Terminierungszeitpunkt zu. Der Terminierungszeitpunkt ist der Zeitpunkt, zu dem sich das repräsentierte Datum seit dem Entstehungszeitpunkt erstmalig nicht mehr im referenzierten Container befindet. Solange die Repräsentation besteht, verweist *term* auf NIL.

6.3.2 Verbindung von Informationsfluss- und Provenance-Datenmodell

Die Aktualisierung des Provenance-Graphen zur Laufzeit hängt direkt mit den Veränderungen des Systemzustands Σ im Informationsflussmodell zusammen. Änderungen der Speicherfunktion, sprich Informationsflüsse und Informationslöschungen, resultieren in Änderungen des Provenance-Graphen. Im Folgenden werden die beiden möglichen Modifikationen der Speicherfunktion und ihre Auswirkungen auf den Provenance-Graphen formal dargestellt.

Die erste mögliche Änderung der Speicherfunktion wird im Informationsflussmodell durch das Primitiv *flow* ($s[c \leftarrow s(c) \cup d]$), dem Fluss eines (neuen) Datums in einen Container, ausgelöst.[8] Sei $\mathscr{G}_t = (\mathscr{R}_t, \mathscr{L}_t)$ der Provenance-Graph zum Zeitpunkt $t \in \mathbb{N}$, dann entstehen neue Repräsentationen im Graphen $\mathscr{G}_{t+1}$ in Abhängigkeit von den Zuständen $\sigma_t = (s_t, l_t, f_t)$ und $\sigma_{t+1} = (s_{t+1}, l_{t+1}, f_{t+1})$. Für jedes Datum $d \in \mathscr{D}$, das einem Container $c \in \mathscr{C}$ neu zufließt, wird im Provenance-Graphen eine neue Repräsentation erzeugt:[9]

$$\forall d \in \mathscr{D}\ \forall c \in \mathscr{C} : d \notin s_t(c) \wedge d \in s_{t+1}(c) \wedge l_t(c) \cap \mathscr{C}_{abstr} = \emptyset \longrightarrow (d, c, t+1) \in \mathscr{R}_{t+1}$$

Da immer nur dann eine neue Repräsentation für eine Datum-Container-Kombination in $t+1$ zum Provenance-Graph hinzugefügt wird, wenn im zugrundeliegenden Informationsflussmodell zum vorhergehenden Zeitpunkt t noch keine Relation zwischen diesen beiden Elementen bestand ($d \notin s_t(c)$), kann immer nur maximal eine Repräsentation zu einer Datum-Container-Kombination existieren, die noch keinen Terminierungszeitpunkt besitzt.

Ebenso wie das Erzeugen der Repräsentationen beruht das Hinzufügen dazwischenliegender Kanten auf den Änderungen des Informationsflussmodells durch die generischen Primitive. Eine Kante geht von einer Repräsentation aus, die noch keinen Terminierungszeitpunkt aufweist, hin zur neu entstandenen Repräsentation. Eine Kante beschreibt den Informationsfluss des Datums d von Container c_1 zum Container c_2 zum Zeitpunkt $t+1$.

[8] Die Primitive *flow* und *flow_to_rtc* sind in diesem Grundverhalten äquivalent.

[9] Flüsse in Container, die auf abstrakte Container ($\mathscr{C}_{abstr}$) verweisen, werden nicht in die Provenance mit aufgenommen; siehe dazu Kapitel 7.2.

Nur wenn ein Informationsflussereignis in seinen Parametern die Quelle des Informationsflusses, einen Container, mitangibt ($s[c_1 \leftarrow s(c_2) \cup s(c_1)]$)[10] bleibt der Provenance-Baum für jedes Datum verbunden und es werden Kanten zu $\mathscr{L}$ hinzugefügt. Diese Kanten verbinden im aktualisierten Provenance-Graphen $\mathscr{G}_{t+1}$ für alle $d \in s_t(c_1)$ die Repräsentationen (d, c_1, z) mit den neuen Repräsentationen $(d, c_2, t+1)$.

Beispiel. *Ein Fluss personenbezogener Daten wirkt sich wie folgt aus: Für eine Datei mit Kundendaten $d_1, \ldots, d_n$ im temporären Verzeichnis c_1 existieren vor Beginn der Kopieroperation die Repräsentationen $(d_1, c_1, t_1), \ldots, (d_n, c_1, t_1)$. t_1 ist der Zeitpunkt, zu dem die Kundendaten im temporären Verzeichnis abgelegt wurden. Durch die Kopieroperation zum Zeitpunkt t_2 entstehen, nach Interpretation des Primitivs flow_to_rtc, die neuen Repräsentationen $(d_1, c_2, t_2), \ldots, (d_n, c_2, t_2)$ in der Datei auf dem Desktop c_2. Neue und alte Repräsentationen sind durch die Kanten $((d_1, c_1, t_1), (d_1, c_2, t_2)), \ldots, ((d_n, c_1, t_1), (d_n, c_2, t_2))$ miteinander verbunden. Dadurch wird in der Provenance ersichtlich, dass die Daten in t_2 vom temporären Verzeichnis auf den Desktop gelangt sind. Im Informationsflussmodell ist diese Information nicht vorhanden.*

Die zweite mögliche Modifikation der Speicherfunktion wird im Informationsflussmodell durch das Primitiv *clear*, der Entfernung aller Daten aus einem Container, ausgelöst. Eine Repräsentation erhält einen Terminierungszeitpunkt, wenn die Relation zwischen ihrem Datum und ihrem Container aus dem Zustand des Informationsflussmodells entfernt wird ($s[c \leftarrow s(c) \backslash \{d\}]$):

$$\forall d \in \mathscr{D}\ \forall c \in \mathscr{C}\ \exists! z \in \mathbb{N} : d \in s_t(c) \wedge d \notin s_{t+1}(c) \Rightarrow term((d, c, z)) = t + 1$$

Als Nebenbedingung gilt, dass eine Repräsentation ihren Terminierungszeitpunkt nicht vor ihrem Entstehungszeitpunkt haben kann.

6.4 Implementierung der Provenance-Datenhaltung

Die Personal-Data-Provenance wird im ProSP gespeichert und gemäß dem Schema aus dem vorherigen Abschnitt aus dem Informationsflussmodell des PIP generiert. Die Datenstruktur ist in Abbildung 6.1 als UML-Klassendiagramm dargestellt. Die Kanten zwischen den Repräsentationen werden als Vorgängerattribut predecessor und Menge von Nachfolgern successor abgespeichert.

Für jede Repräsentation wird festgelegt, welche Rolle sie im datenschutzrechtlichen Sinne inne hat. Sie wird dadurch so klassifiziert, dass die datenschutzrechtlichen Anforderungen aus Kapitel 4.4 auf sie abgebildet werden können. Die Rolle einer Repräsentation wird in ihrem Repräsentationstyp festgelegt. Mögliche Repräsentationstypen sind Herkunft (SourceRepresentation), Empfänger (SinkRepresentation), Erhebung (CollectionRepresentation), Übermittlung (TransferRepresentation), interne Weitergabe

[10] Die Anwendung von s wird durch die Informationsflusssemantik spezifiziert (vgl. Abschnitt 6.2.2).

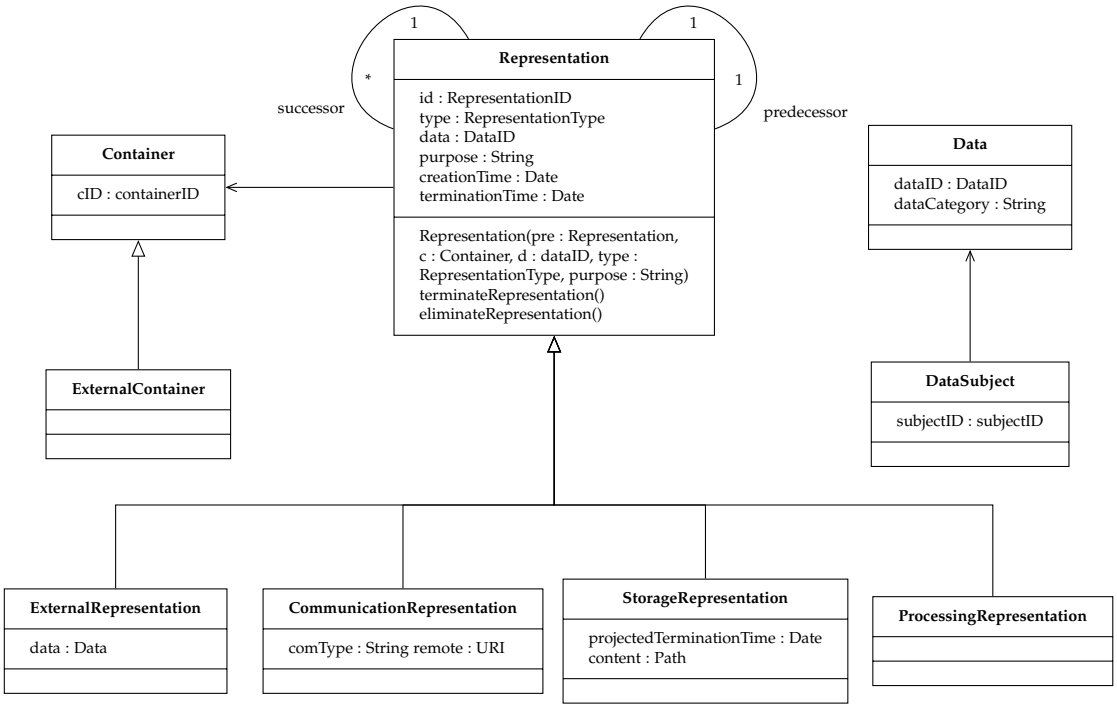


Abbildung 6.1: Datenstruktur der Personal-Data-Provenance

(InternalCommunicationRepresentation), Speicherung (StorageRepresentation) oder Verarbeitung (Process−Representation). Repräsentationstypen sind containerabhängig. Deshalb haben zwei Repräsentationen in ein und demselben Container denselben Repräsentationstyp. Die Repräsentationstypen sind, wo nötig, durch Vererbung und zusätzlich durch ein eigenes Attribut realisiert.

Zusätzlich zu den zeitlichen Angaben einer Repräsentation sowie ihrem Typ werden noch weitere Metadaten gesammelt. Der Bedarf für jedes Attribut der Repräsentation leitet sich aus den technischen Anforderungen an den Umfang der Protokollierung in einem Datenschutzauskunftssystem ab.

Die Anforderungen 7, 9, 10, 11, 12 und 15 geben den Zweck jeder Verwendung personenbezogener Daten sowie den Zeitpunkt von Erhebung, Weitergabe und Übermittlung als Teil der Provenance vor. Beides ist deshalb in der Datenstruktur als Attribut einer Repräsentation vorhanden.

Um die gespeicherten personenbezogenen Daten selbst und die Metadaten der Speicherung beauskunften zu können, ist in Anforderung 15 die Protokollierung des Speicherortes

festgelegt. Dieser findet sich als weiteres Attribut in der Speicherrepräsentation.

Die Kategorisierung der Daten ist gemäß der Anforderungen 7, 9 und 10 für die Außenbeziehungen der verantwortlichen Stelle vorgesehen. Deshalb findet sich die Datenkategorie (Bsp.: Anschriften) in der Klasse „Data“ wieder, auf die in externen Repräsentationen, den Stellvertretern für Herkunft und Empfänger, verwiesen wird. Zum Datum selbst und zum Betroffenen wird ansonsten nur eine ID als Pseudonym gespeichert.

Die Provenence wird im ProSP im Arbeitsspeicher sowie wahlweise in der graphenbasierten Datenbank Neo4J gespeichert.[11] Neo4J erlaubt hochperformante Abfragen auf Graphen, wie dem Personal-Data-Provenance-Datenmodell.[12]

6.5 Zwischenfazit

Das vorangegangene Kapitel hat gezeigt, wie sich die für die Datenschutzauskunft erforderliche Provenance aus dem UC-Informationsflussmodell ableiten lässt und wie alle für die Auskunft relevanten Informationen in einer Provenance-Datenstruktur abgelegt werden können. Die ergänzend eingeführte semantische Beschreibungssprache für Ereignisse (Informationsflusssemantik) erlaubt das Hinzufügen von PEPs zum Datenschutzauskunftssystem zur Laufzeit. Dadurch können Geschäftsprozesse zur Verarbeitung personenbezogener Daten jederzeit flexibel umgestaltet werden.

Die wissenschaftlichen Beiträge dieses Kapitels lassen sich wie folgt zusammenfassen: (1) Eine Erweiterung des Informationsflussmodells um den Betroffenen und unterschiedliche Containertypen sowie die Ergänzung der Informationsflusssemantik, (2) ein Datenmodell für Personal-Data-Provenance und (3) die Verbindung zwischen Informationsflussmodell und Personal-Data-Provenance-Modell.

Auf Grundlage dieser Ergebnisse sind, wie in Konstruktionsziel $\mathfrak{K}3$ postuliert, eine generische, datenzentrierte und semantisch konfigurierbare Erfassung von Ereignissen und eine Speicherung von Personal-Data-Provenance in einem Datenmodell gemäß der datenschutzrechtlichen Anforderungen möglich.

Die Trennung zwischen Informationsflussmodell und Personal-Data-Provenance erlaubt, die Granularität des Informationsflusstrackings und der Provenance unterschiedlich zu gestalten. Dieses Potential wird im folgenden Kapitel dazu genutzt, eine speichereffiziente und datenminimale Provenance-Datenhaltung umzusetzen.

[11] https://neo4j.com.
[12] Vicknair et al. 2010.

7 Datenschutzgerechtes und skalierbares Provenance-Tracking

Eine Datenschutzauskunft muss vollständig sein, also insbesondere alle Weitergaben personenbezogener Daten umfassen. Um dies zu erreichen, ist ein detailliertes Informationsfluss-Tracking auf Betriebssystems- und Anwendungsebene notwendig. Nur so lässt sich eine unvollständige oder fehlerhafte Annahme über den Fluss personenbezogener Daten weitestgehend vermeiden.

Auf der anderen Seite würden extrem detaillierte Protokolle jeden Umgangs mit personenbezogenen Daten den Grundsatz der Datenminimierung verletzen (Kriterium 40) und die mit der Datenverarbeitung betrauten Mitarbeiter einem besonderen Überwachungsrisiko aussetzen. Nahezu jeder Mausklick würde festgehalten.

Darüber hinaus führt eine vollständige Protokollierung aller Ereignisse zu einem exponentiellen Wachstum des Speicherbedarfs im Worst-Case, sprich $\varphi \in \mathcal{O}(|\mathscr{D}|^{|\mathscr{E}|})$.[1] Selbst wenn man von atomaren Ereignissen ausgeht, von jedem Ereignis also nur ein Quellcontainer betroffen ist, liegt der Speicherbedarf in der Größenordnung von $\varphi \in \mathcal{O}(|\mathscr{E}| \cdot |\mathscr{D}|)$, wächst also linear.

Aus diesem Grund wurde im letzten Kapitel die konzeptionelle Entscheidung getroffen, dass das Informationsfluss- und das Personal-Data-Provenance-Modell getrennt werden. Der Umfang der Protokollierung kann so gröber als der Detailgrad des Informationsfluss-Trackings festgelegt werden. Im Informationsflussmodell wird ausschließlich der aktuelle Zustand eines Systems und keine Historie abgebildet. Sein Speicherverbrauch kann deshalb nicht über die Menge der Zustände des laufenden Systems hinausgehen. Die Historie, die einzig in der Provenance-Datenstruktur vorliegt, kann allerdings mit jedem weiteren Ereignis wachsen. Optimierungsversuche können sich deshalb auf die Größe der Provenance-Datenstruktur beschränken.

Um sowohl der Datenminimierungsproblematik als auch der Speicherproblematik in der Provenance-Datenstruktur Herr zu werden, werden in diesem Kapitel zwei aufeinander aufbauende Ansätze vorgestellt. Abschnitt 7.1 stellt Löschregeln für eine zielgerichtete und datenminimale Personal-Data-Provenance vor. Im Abschnitt 7.2 wird das Konzept abstrakter Container als Teil des Informationsflussmodells erläutert. Abstraktionsregeln erlauben, die Provenance bereits zum Entstehungszeitpunkt noch kompakter zu gestalten. Beide Ansätze werden in Abschnitt 7.3 evaluiert.

[1] Beispiel: Fork aller datenverarbeitender Prozesse bei jedem Ereignis.

7.1 Zielgerichtete und datenminimale Personal-Data-Provenance

Von den in den technischen Anforderungen festgelegten Löschregeln (Anforderungen 24 ff.), sind nur die Regeln für Speicher- und Verarbeitungsrepräsentationen (Anforderungen 28 und 29) ausschließlich von den Speicher- und Verarbeitungsvorgängen selbst abhängig. Eine Verarbeitungsrepräsentation muss aus der Provenance gelöscht werden, sobald die Verarbeitung beendet ist. Gleiches gilt für eine Speicherrepräsentation und die Speicherung. Im gemäß dieser Anforderungen konzipierten Datenschutzauskunftssystem läuft dementsprechend beim Aufruf der Methode terminateRepresentation() einer Repräsentation[2] in der Provenance-Datenstruktur des ProSP ein Löschalgorithmus ab.

Der Algorithmus operiert auf zwei Attributen der Repräsentation, dem predecessor und dem successorSet. Der predecessor ist der Vorgänger einer Repräsentation in der Baumstruktur des Provenance-Graphen. Das successorSet ist die Menge der Nachfolger.[3] Für alle anderen außer den oben erwähnten Repräsentationen wird durch den Algorithmus nur der Terminierungszeitpunkt (terminationTime) gesetzt. Eine Speicher- oder Verarbeitungsrepräsentation wird dagegen aus dem Provenance-Graph entfernt.

Dazu werden zunächst neue Kanten unter Umgehung der zu entfernenden Repräsentation gebildet. Die Nachfolger der Repräsentation werden zu den Nachfolgern ihres Vorgängers. Gleichzeitig wird ihr Vorgänger zum Vorgänger ihrer Nachfolger. Indem abschließend alle Kanten von und zu der Repräsentation gelöscht werden, ist die Repräsentation vollständig aus dem Provenance-Graphen ausgelöst. Der Speicherplatz wird freigegeben.

7.2 Abstrakte Container und Abstraktionsregeln

Das Recht auf Auskunft verpflichtet die verantwortliche Stelle nicht dazu, die Funktionalität ihrer IT-Systeme im Detail wiederzugeben. Datenflüsse in technisch bedingten Systemprozessen sind gänzlich uninteressant.[4] Gleiches kann für Speicherorte gelten. Beispielsweise ist die Speicherung personenbezogener Daten in einer Kundendatenbank zu beauskunften, aber nicht, in welcher Zelle einer bestimmten Tabelle die Daten abgelegt sind. Auf der anderen Seite ist ein feingranulares Tracking personenbezogener Daten erforderlich, um nicht die Spur eines Datums zu verlieren oder dem Problem der Überapproximation zu erliegen.[5] Würde eine Datenbank insgesamt als Container aufgefasst und nur der Fluss von einer Datenbank in eine andere getrackt, müsste bei jedem dieser Flüsse angenommen werden, dass alle personenbezogenen Daten in der Datenbank geflossen

[2] Vgl. Diagramm 6.1.
[3] Vgl. auch Kapitel 6.4.
[4] Beispielsweise von einer Windows-Explorer-Instanz in eine andere.
[5] Lovat/Oudinet/Pretschner 2014.

sind. Datenschutzauskunftssysteme müssen so aufgebaut werden, dass sie mit diesen gegensätzlichen Anforderungen umgehen können. Ein Ansatz sind abstrakte Container.

Abstrakte Container reduzieren den Detailgrad, in dem Provenance-Daten gesammelt werden. Sie fassen reale Container zusammen. Ein abstrakter Container vereint in sich alle Speicherfunktionen der zugrundeliegenden realen Container. Reale Container und abstrakte Container werden über eine unidirektionale Aliasrelation miteinander verbunden.

Abstrakte Container optimieren die Sammlung von Provenance-Daten und weisen eine Reihe von Vorteilen gegenüber einer Selektion und Gruppierung realer Container zum Abfragezeitpunkt auf:

- Reduzierung des Speicherbedarfs für Provenance-Daten: Für zugrundeliegende reale Container werden keine eigenen Repräsentationen gespeichert.
- Verbesserung der Laufzeit des Provenance-Systems: Die Anzahl der Methodenaufrufe auf dem ProSP ist deutlich niedriger. Objekterzeugung und Datenbankaufrufe fallen weg.
- Datenvermeidung entsprechend der datenschutzrechtlichen Anforderungen wird umgesetzt:
  - Es werden nur die Datum-Realcontainer-Beziehungen des gegenwärtigen Systemzustandes gespeichert.
  - Es werden keine ergänzenden Parameter (z. B. Zeit) gespeichert.

Als Beispiel für die Wirkweise abstrakter Container sollen die parallelisierten Prozesse in Multitasking-Betriebssystemen dienen. Die einzelnen Container der Teilprozesse[6] sind für die Datenschutzauskunft nicht von Interesse und werden auf einen abstrakten Container *Prozesse* abgebildet. Der abstrakte Container enthält alle Daten der Container der Teilprozesse. Jedes Mal, wenn die Speicherfunktion eines realen Containers modifiziert wird, wird auch die Speicherfunktion des abstrakten Containers aktualisiert. Bei einem Datenaustausch zwischen Teilprozessen würde sich jedoch die Gesamtsicht auf die Prozesse nicht verändern. Es fließt kein neues Datum in den abstrakten Container *Prozesse*. Seine Speicherfunktion ändert sich nicht. Ein neuer Provenance-Datensatz wird nicht erzeugt.

Aus modelltheoretischer Sicht kann mit einem abstrakten Container genauso umgegangen werden, wie mit jedem anderen Container. Faktisch gibt es keinen direkten Bezug von Ereignissen auf abstrakte Container. Die Speicherfunktion abstrakter Container wird nur implizit über eine Aliasfunktion verändert.

[6] Der Container eines Prozesses bezeichnet den Speicherbereich im Hauptspeicher und im Hauptprozessor, den ein Prozess belegt.

7.2.1 Abstraktionsregeln

Welche realen Container durch welche abstrakten Container zusammengefasst werden, wird durch Abstraktionsregeln ausgedrückt. Abstraktionsregeln bilden anhand der Benennungsfunktion Äquivalenzklassen von realen Containern.

Es ist notwendig, die Benennung von Containern hierarchisch auszudifferenzieren, um Abstraktionsregeln zu formulieren. Nur so kann ausgedrückt werden, auf welcher Ebene abstrahiert wird.

Der ausdifferenzierte Bezeichner lautet $\mathscr{F} \subseteq Dom \times LoA \times Type \times Hndl$. Auf der obersten Ebene steht die lokale Domäne (*Dom*), in der sich ein Container befindet. Sie beschreibt auf der niedrigsten Hierarchieebene in der Regel ein konkretes IT-System. In einer Organisation bestimmt sie den lokalen Zuständigkeitsbereich der meisten Usage-Control- und Provenance-Komponenten.[7] Unterhalb der Domäne steht die Abstraktionsschicht (Layer of Abstraction, *LoA*). Sie beschreibt die Quelle der Ereignisse und damit den Zuständigkeitsbereich eines PEP. Häufig entspricht eine Abstraktionsschicht einer Applikation oder einem Element der Datenhaltung. Dadurch wird sie als Abgrenzungsmerkmal für Abstraktionsregeln interessant. Des Weiteren hat jeder Container einen Typ (*Type*). Der Typ entspricht einer Container-Klasse, wie sie im Usage-Control-PSM-Modell definiert ist.[8] Ein Beispiel für einen Containertyp wäre *File* auf der Abstraktionsschicht eines Dateisystems. Letzter Bestandteil des Bezeichners ist der konkrete Identifikator (*Hndl*), mit dem ein einzelner Container in einer Abstraktionsschicht referenziert wird. Beispiele dafür sind der Pfadname in einem Dateisystem oder der *Window Handle* auf Betriebssystemebene.

Dom, *LoA* und *Type* sind für jeden Container eindeutig. Sprechen zwei unterschiedliche IT-Systeme oder Abstraktionsschichten vom gleichen Container, wird dies durch eine bidirektionale Aliasbeziehung zwischen zwei im Modell getrennten Containern deutlich gemacht. Deshalb kann vereinfachend für die Benennungsfunktion eine komponentenweise Umkehrfunktion $f_i = pr_i \circ f^{-1}$ für $i \in \{Dom, LoA, Type\}$ mit der Projektion $pr_i(a_1, \ldots, a_n) = a_i$ angegeben werden.

Die erweiterte Benennungsfunktion erfüllt gleichzeitig die durch die Anforderungen 7, 9, 10, 11, 12 und 15 verlangte Protokollierung von Organisationseinheit und IT-System sowie die von Anforderung 12 in Einzelfällen geforderte Angabe des Typs einer Anwendung.

Aus der erweiterten Benennungsfunktion ergeben sich drei verschiedene Äquivalenzrelationen R auf der Menge der realen Container:

- Die Domänenäquivalenz mit $c_1 \sim_{Dom} c_2 \Leftrightarrow f_{Dom}(c_1) = f_{Dom}(c_2)$
- Die Abstraktionsschichtäquivalenz mit

[7] Insbesondere des PDP, des PIP, des lokalen PMPs und des ProSP. Die Bezeichner der Usage-Control-Domänen werden in Anhang C beschrieben.

[8] Siehe Kumari/Pretschner 2013.

$c_1 \sim_{LoA} c_2 \Leftrightarrow f_{Dom}(c_1) = f_{Dom}(c_2) \wedge f_{LoA}(c_1) = f_{LoA}(c_2)$

- die Typäquivalenz mit
  $c_1 \sim_{Type} c_2 \Leftrightarrow f_{Dom}(c_1) = f_{Dom}(c_2) \wedge f_{LoA}(c_1) = f_{LoA}(c_2) \wedge f_{Type}(c_1) = f_{Type}(c_2)$

Eine Abstraktionsregel beschreibt, welche Äquivalenzklasse $[c]_R \in \mathcal{P}(\mathscr{C}_{real})$ auf welchen abstrakten Container $c_{abstr} \in \mathscr{C}_{abstr}$ abgebildet wird. Eine Abstraktionsregel ist durch einen Vertreter der Äquivalenzklasse, die Art der Äquivalenzrelation (*Dom*, *LoA* oder *Type*) und den referenzierten abstrakten Container vollständig charakterisiert.

Abstraktionsregeln sind containerzentriert und nicht datenzentriert. Insofern ist es im Rahmen des Modells nicht möglich, für unterschiedlich sensible Daten den Detailgrad des Trackings zu variieren.

Algorithmus 12 zeigt, wie das Informationsflussmodell modifiziert wird, wenn eine Abstraktionsregel hinzugefügt wird.

Algorithmus 12: $addAbstraction(\sigma, [c]_R, c_{abstr})$

```
1 (s, l, f) ← σ
2 forall u ∈ [c]_R do
3 |   l ← l[u ← l(u) ∪ {c_abstr}]
4 end
5 (s, l, f) ← update_abstr_container((s, l, f), c_abstr)
6 return (s, l, f)
```

Entsprechend stellt Algorithmus 13 die Veränderungen dar, die vorgenommen werden, wenn eine Abstraktionsregel wieder entfernt wird. Die Hilfsfunktion *update_abstr_container* wird in Algorithmus 14 beschrieben. Sie aktualisiert die Speicherfunktion für einen abstrakten Container anhand der auf diesen verweisenden Aliasbeziehungen.

Algorithmus 13: $rmAbstraction(\sigma, [c]_R, c_{abstr})$

```
1 (s, l, f) ← σ
2 forall u ∈ [c]_R do
3 |   l ← l[u ← l(u) \ {c_abstr}]
4 end
5 (s, l, f) ← update_abstr_container((s, l, f), c_abstr)
6 return (s, l, f)
```

Abstraktionsregeln, die auf ein Informationsflussmodell angewendet werden, dürfen je abstraktem Container nur disjunkte Äquivalenzklassen enthalten. Würden Überschneidungen auftreten, könnten Abstraktionsregeln nicht mehr unabhängig von anderen Regeln entfernt werden (vgl. Zeile 3 in Algorithmus 13). Nur wenn jeder reale Container

ausschließlich in einer der Äquivalenzklassen auftaucht, bleiben die Vorgänger- und Nachfolgerbeziehungen bei Informationsflüssen erhalten.

Abstraktionsregeln führen zu einer Ergänzung des Informationsflussmodells um die Speicher- und Aliasrelationen abstrakter Container. Es gehen keine Informationen verloren, die bereits ohne Abstraktionsregeln Teil des Informationsflussmodells sind. Abstraktionsregeln reduzieren ausschließlich die Anzahl der im Provenance-Modell zu speichernden Repräsentationen.

Algorithmus 14: $update\_abstr\_container(\sigma, c_{abstr})$

1 $(s, l, f) \leftarrow \sigma$
2 $s \leftarrow s[c_{abstr} \leftarrow \bigcup_{c \in C \mid c_{abstr} \in l(c)} s(c)]$
3 **return** (s, l, f)

7.2.2 Umsetzung der Abstraktionsregeln in der Implementierung

Die Vorgabe, welche Abstraktionsregeln greifen, kann entweder global festgelegt oder in UC-Policies definiert sein. Da die Relevanz eines Containertyps, einer Abstraktionsschicht oder einer ganzen Domäne für den Auskunftsanspruch nicht vom einzelnen Datum, sondern von generellen Überlegungen abhängig ist, sind Abstraktionsregeln nicht datenspezifisch.

Abstraktionsregeln werden im PIP konfiguriert. Deshalb wird die Schnittstelle des PIP um die drei in Listing 7.1 gezeigten Methoden erweitert. Mit der ersten Methode lässt sich ein abstrakter Container, mit der zweiten eine Abstraktionsregel hinzufügen. Die dritte Methode entfernt Abstraktionsregeln.

```
boolean addAbstractContainer(AbstractContainer abstractContainer);
RuleId addAbstractionRule(String dom, String loa, String type, ContainerId abstractContainer);
boolean removeAbstractionRule(RuleId rule);
```

Listing 7.1: Schnittstelle des PIP für Abstraktionsregeln

Beispiel. *Eine Abstraktionsregel, die alle Prozesse auf einem Arbeitsplatzrechner zusammenfasst, wird durch die Domäne* `urn:ucn:adbokis:sales:workspace23`*, die Abstraktionsschicht* `os:process` *und den Typ* `process` *beschrieben.*

Die Primitive aus Kapitel 6.2 erfüllen bereits alle Voraussetzungen zur Verarbeitung abstrakter Container. Auch in der Verbindung von Informationsfluss- und Provenance-Modell in Kapitel 6.3.2 wurden abstrakte Container bereits berücksichtigt. Vom ProSP wird ein abstrakter Container wie jeder andere Container behandelt.

7.3 Evaluation der Skalierbarkeit

Zu Beginn dieses Kapitels wurde die Frage aufgeworfen, ob und wie ein skalierbares, insbesondere speicherskalierbares, Datenschutzauskunftssystem konstruiert werden kann. Angenommen wurde ein exponentielles Wachstum des Speicherbedarfs im Worst-Case.

Die vorhergehenden Abschnitte haben zwei theoretische Konzepte zur Adressierung der Problemstellung eingeführt: Löschregeln und Abstraktionsregeln. Ohne eine Evaluation der Konzepte am realen System kann jedoch keine Aussage darüber getroffen werden, ob die Konzepte praktisch wirksam sind.

In diesem Abschnitt werden der Aufbau und die Ergebnisse eines Lasttestes zur Evaluation der Laufzeit und des Speicherverbrauchs des Datenschutzauskunftssystems vorgestellt. Sowohl Löschregeln als auch Abstraktionsregeln sollten sich aus konzeptioneller Sicht positiv auf den Speicherverbrauch des ProSP auswirken. Bei den Abstraktionsregeln sind potentiell negative Auswirkungen auf den Speicherverbrauch des PIP möglich.

Die Zielsetzung dieses Abschnitts ist, zu zeigen, dass (1) die Komponenten des Datenschutzauskunftssystems die Laufzeit der Verarbeitung personenbezogener kaum beeinflussen (konstanter Faktor) und (2) Löschregeln und Abstraktionsregeln den Speicherbedarf des Datenschutzauskunftssystems im Verhältnis zur Menge der in jedem Schritt verarbeiteten personenbezogenen Daten maximal linear und im Verhältnis zur Menge der Operationen sublinear steigen lassen.

7.3.1 Testaufbau und Konfiguration

Im Lasttest zu obigen Konzepten werden sequentiell Dateien kopiert. Die Kopierereignisse auf Betriebssystemebene werden durch das Datenschutzauskunftssystem abgefangen. Sie werden durch den PIP und den ProSP verarbeitet und gespeichert. Der gemessene Speicherverbrauch der Komponenten lässt Rückschlüsse auf die Wirksamkeit der Löschregeln und Abstraktionsregeln zu. Eine Kopieroperation ist die einfachste Operation die alle drei Primitive der Speicherfunktion abdeckt. Welche Ereignisse dafür verantwortlich sind, wird weiter unten ausgeführt.

Nach einer Erläuterung der verwendeten Hard- und Software werden im Folgenden der genaue Testablauf und die verwendeten Parameter und Modi dargestellt.

Hardware und Betriebssystem Die Laufzeit des Datenschutzauskunftssystems und die Auswirkungen von Löschregeln und Abstraktionsregeln auf den Speicherverbrauch wurden auf einem System unter Microsoft Windows 7 x64 mit einem Intel Core i7-4790 Prozessor @ 3,6 GHz und 8 GB Arbeitsspeicher getestet. Alle Hintergrundprozesse, die sich störend auf die Messungen auswirken könnten, wie beispielsweise der Update-Mechanismus des Betriebssystems, wurden deaktiviert.

Für die Messungen wurde auf Ereignisse des Betriebssystems zurückgegriffen. Der Vorteil gegenüber Ereignissen aus Applikationen ist, dass damit das gesamte Prozessgeschehen abgedeckt wird und die Messungen repräsentativ für eine Vielzahl von realen Einsatzszenarien eines Datenschutzauskunftssystems sind. Der wesentliche Nachteil ist, dass das Abfangen von Ereignissen auf Betriebssystemebene „teuer" ist, also viel Rechenzeit kostet. Dies wirkt sich negativ auf die Gesamtlaufzeit der Tests aus.

Verwendete Komponenten des Datenschutzauskunftssystems Grundlage des für Windows 7 verwendeten PEPs ist das Hooking-Framework von Wüchner.[9] Es setzt auf Deviare auf,[10] um sich in die Windows API einzuhängen.[11] Die Einbindung des Hooking-Frameworks in das Datenschutzauskunftssystem übernimmt ein ergänzend entwickelter Adapter.

In den Tests standen die Komponenten PIP und ProSP im Mittelpunkt. Informationsflüsse durch die folgenden Ereignisse sind Teil der Informationsflusssemantik des Windows-PEP und werden überwacht: CreateProcess, KillProcess, CreateFile, CreateEmptyFile,[12] TruncateFile,[13] DeleteFile, MoveFile, CopyFile, ReadFile, WriteFile, CreateWindow, SetClipboardData, GetClipboardData, CreateDC,[14] Send und Recv.[15] Alle anderen abgefangenen Ereignisse werden dem PIP zwar vom PEP mitgeteilt, sie werden jedoch nicht ausgewertet. Alle Komponenten des Datenschutzauskunftssystems stehen auf einer Whitelist. Die von ihnen verursachten Ereignisse werden vom PEP nicht weitergeleitet.

Als weitere Komponenten kamen ein ProCP, ein PMP und ein PAP zum Einsatz. Ein PDP wurde nicht verwendet, da keine UC-Policies eingesetzt wurden. Das Provenance-Tracking wurde für jedes Datum direkt vom PAP aktiviert. Der PAP gab außerdem dem PIP die Abstraktionsregeln vor.

Der ProCP lief auf demselben System wie die übrigen Komponenten (lokale Konfiguration). Der ProCP wird nur beim Bekanntmachen neuer personenbezogener Daten aufgerufen und spielt im weiteren Verlauf der Erfassung von Informationsflüssen keine Rolle. Sein Einfluss auf die Laufzeit des Datenschutzauskunftssystems ist somit minimal. Der Speicherbedarf des ProCP wächst im ungünstigsten Fall linear mit der Anzahl der berücksichtigten personenbezogenen Daten. Er wurde deshalb im Experiment nicht gesondert betrachtet.

Die Komponenten sind alle in Java implementiert und laufen dementsprechend in einer Java Virtual Machine (JVM). Der Garbage Collector der JVM führt zu instabilen

[9] Wüchner/Pretschner 2012.
[10] http://www.nektra.com/products/deviare-api-hook-windows.
[11] kernel32.dll, user32.dll, Gdi32.dll und Ws2_32.dll.
[12] CreateFile mit dem Parameter dwCreationDisposition=CREATE_ALWAYS.
[13] CreateFile mit dem Parameter dwCreationDisposition=TRUNCATE_EXISTING.
[14] Ausgabe auf ein Gerät, z. B. einen Drucker.
[15] Send und Recv bezeichnen Zugriffe auf den Netzwerkstack von Windows.

Messungen des Speicherverbrauchs der Komponenten. Um diesen Effekt zu reduzieren, wurde der inkrementelle Garbage Collector (Parameter −Xincgc) aktiviert. Der inkrementelle Garbage Collector läuft dauerhaft parallel zur Programmausführung, während der reguläre Garbage Collector nur dann Speicher freigibt, wenn ein bestimmter Schwellwert des Speicherverbrauchs überschritten wurde. Da der Garbage Collector in einem eigenen Prozess läuft, sind die Auswirkungen des inkrementellen Garbage Collectors auf die gemessene Laufzeit des Datenschutzauskunftssystems im verwendeten Testsystem mit Mehrkernprozessor vernachlässigbar.

Traces und Aktivitäten Von der entwickelten Testumgebung werden sogenannte *Traces* ausgeführt. Ein *Trace* ist eine sequentielle Abfolge atomarer Aktivitäten, die ein Benutzer ausführen kann, wie beispielsweise der Aufruf eines Kommandozeilentools zum Kopieren von Dateien (Kopieroperation). Die Länge eines Traces wird durch die Testparameter bestimmt. Für die Lasttests wurden zwei Aktivitäten gewählt: CallPAP und CopyFile. Die Aktivität CopyFile ist nicht identisch mit dem Ereignis CopyFile der Windows-API, sondern ist in der Implementierung der Testumgebung ein *Bash*-Skript, das das Kommandozeilentool cp aufruft.[16]

Im einfachsten, für die Lasttests gewählten Fall wird CallPAP nur zu Beginn eines Traces ausgeführt. Das Testverzeichnis enthält zu Beginn nur eine Datei zufälligen Inhalts, deren Größe durch die Testparameter bestimmt wird. Die Aktivität CallPAP besteht aus genau einem Aufruf des PAP. Der PAP weist der initialen Datei (Container) eine durch die Testparameter vorgegebene Anzahl personenbezogener Daten zu.[17] Für jedes Datum wird das Provenance-Tracking aktiviert.

Alle anderen Aktivitäten des Traces sind CopyFile-Aktivitäten. Die Parameter einer CopyFile-Aktivität sind eine zufällige Datei im Testverzeichnis (zu Beginn die initiale Datei) als Quelle und eine neue Datei, in die die Quelle kopiert werden soll. Dateien werden somit niemals gelöscht oder überschrieben.[18]

Ein Trace der Länge 5 würde aus den Aufrufen [CallPAP, CopyFile, CopyFile, CopyFile, CopyFile] bestehen.[19] Jeder Aufruf von CopyFile führt zu einer Reihe nicht in Gänze deterministischer Ereignisse auf der Windows-API. Diese Ereignisse werden nachfolgend erläutert.

[16] Verwendet werden die *Bash* und GNU cp als Teil von *MSYS* bzw. *MinGW*, http://mingw-w64.org.

[17] Ein personenbezogenes Datum wird unabhängig vom Inhalt der Datei als ID ausgedrückt (siehe Anhang C). Die Daten sind alle dem selben Betroffenen zugeordnet. Eine Zuweisung zu unterschiedlichen Betroffenen führt zu einmaligem konstantem Overhead je Betroffenem und spielt daher für die Aussage zur Last- und Speicherskalierbarkeit keine Rolle.

[18] Ein Überschreiben oder Löschen würde den Speicherbedarf von PIP und ProSP verringern. Dies ist im Lasttest unerwünscht.

[19] Bei den in der Auswertung genannten Längenangaben wird CallPAP ignoriert. Die Angabe „Tracelänge 800“ bedeutet also tatsächlich das 800-fache Ausführen von CopyFile und eine Länge des Traces von 801 incl. CallPAP.

Ereignisse von *cp* Bei jedem Aufruf von cp wird zuerst der cp-Prozess erzeugt. Dann wird aus einer zufälligen Datei im Testverzeichnis gelesen. Anschließend wird der Inhalt der Datei in eine neue Datei geschrieben. Zuletzt wird der cp-Prozess wieder beendet. Daraus ergeben sich für cp unter anderem die Ereignisse CreateProcess, KillProcess, CreateFile, ReadFile und WriteFile. Die Ereignisse decken die drei Primitive der Speicherfunktion ab,[20] die zu einer Modifikation der Provenance führen.[21]

Der sequentielle Wechsel aus Erzeugung und Beendigung von cp-Prozessen sowie der Fluss von Daten aus Dateien in Prozesse und umgekehrt führen zu einer hohen Last für den PIP und den ProSP. Andere Arten von Ereignissen würden entweder gar keine Änderung des Informationsflussmodells verursachen oder ausschließlich den PIP belasten. Insofern ist der gewählte Lasttest zwar nicht repräsentativ für alle Vorgänge in datenverarbeitenden Systemen. Er stellt aber ein Hochlastszenario dar, das geeignet ist, die Skalierbarkeit des Datenschutzauskunftssystems zu überprüfen.

Parameter und Modi Insgesamt sind bei den Lasttests drei Parameter zu berücksichtigen: (1) die Anzahl der personenbezogenen Daten in der initialen Datei, (2) die Tracelänge, also die Anzahl der Kopiervorgänge, und (3) die Größe der Datei.

Jede betrachtete Konfiguration der Parameter wurde mindestens 30 Mal getestet. Die Boxplots in den Abbildungen dieses Abschnitts zeigen die beobachtete Schwankungsbreite der Ergebnisse.

Die Testumgebung kennt 7 Modi:

[0] Durchlauf des Traces ohne aktives Datenschutzauskunftssystem

[1] das Hooking-Framework ist aktiv

[2] der PEP mit integriertem Hooking-Framework ist aktiv

[3] PEP, PMP und PIP sind aktiv

[4] PEP, PMP, PIP, ProSP & ProCP sind aktiv – ProSP ohne Löschregeln und Abstraktionsregeln

[5] PEP, PMP, PIP, ProSP & ProCP sind aktiv – ProSP mit Löschregeln und ohne Abstraktionsregeln

[6] PEP, PMP, PIP, ProSP & ProCP sind aktiv – ProSP mit Löschregeln und Abstraktionsregeln für Prozesse (os:process/process) und Dateien (os:filesystem/file)

[20] *flow*, *flow_to_rtc* und *clear*; vgl. die Informationsflusssemantik im Anhang D.4.
[21] Vgl. Kapitel 6.3.2.

Modus 0 gibt die Vergleichswerte vor. Die Modi 1 und 2 sind in nahezu allen Aspekten äquivalent und zeigen die Laufzeitbeeinträchtigungen durch das Hooking-Framework. Modus 3 ist der Vergleichswert für eine reine UC-Infrastruktur. Die Modi 4 bis 6 liefern Vergleichsdaten für das komplette Datenschutzauskunftssystem, konfiguriert mit den verschiedenen, in diesem Kapitel vorgestellten, Möglichkeiten.

7.3.2 Ergebnisse der Laufzeitmessungen

Die Bewertung des Datenschutzauskunftssystems hängt von der verursachten Prozessorlast, überprüfbar anhand der Laufzeit eines Traces in den verschiedenen Modi, und vom Speicherbedarf des PIP und des ProSP ab. Fraglich ist, welche Faktoren Laufzeit und Speicherverbrauch bestimmen und ob die Komponenten des Datenschutzauskunftssystems in Dateigröße, Tracelänge und Anzahl der personenbezogenen Daten skalieren. In diesem Unterabschnitt wird zunächst die Laufzeit diskutiert.

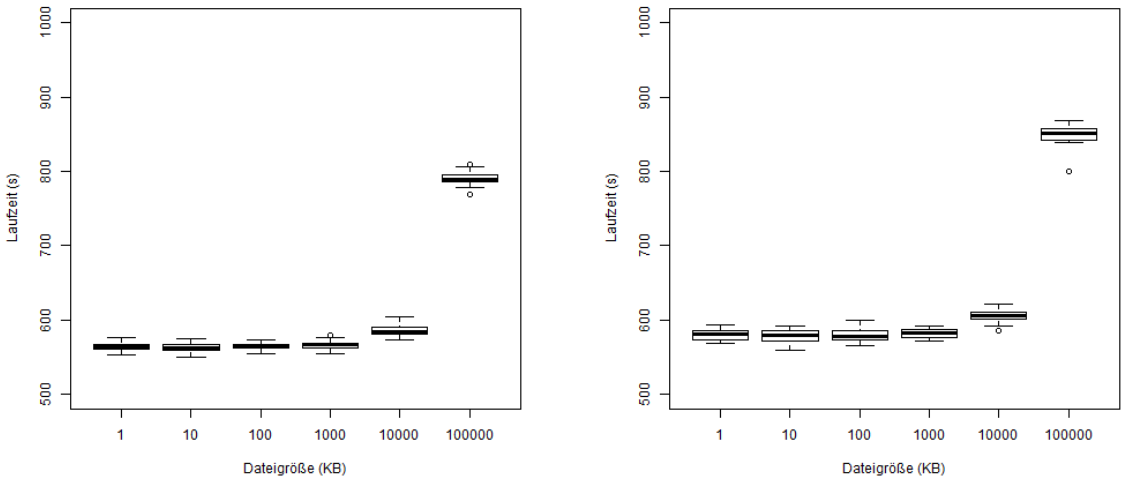
Dateigröße Zunächst ist festzustellen, dass bei großen Dateien der Aufwand für die eigentlichen Kopieroperation steigt. Dateigrößen von 1 KB bis 10 000 KB führen im Modus 0 (nicht abgebildet) nur zu Laufzeiten zwischen 1 s und 5 s. Bei einer Größe der initialen Datei von 100 000 KB steigt die Laufzeit dagegen auf durchschnittlich 57 s an. Bei einer Tracelänge von 100 werden insgesamt 10 GB kopiert.

Die Größe der initialen Datei hat unter anderem deshalb bis 1 000 KB kaum Einfluss auf die Laufzeit eines Traces. Sehr große Dateien verlängern hingegen die Laufzeit (vgl. Abbildung 7.1).

Die Schwelle zum Laufzeiteinfluss bei 1 000 KB ist anhand einer Auswertung der vom Hooking-Framework abgefangenen Ereignisse erklärbar. Beim Sprung von 100 KB auf 1 000 KB verändert sich die Anzahl der beobachteten Ereignisse. Bei 1 KB, 10 KB und 100 KB und einer festen Tracelänge von 100 werden jeweils im Durchschnitt 1 685 Ereignisse beobachtet. Bei 1 000 KB steigt die Anzahl der Ereignisse auf 2 271. Bei 10 000 KB und 100 000 KB werden bereits 9 367 beziehungsweise 78 319 Ereignisse erreicht. Zum Kopieren einer Datei mit einer Größe von 1 000 KB werden also erstmals mehr Aufrufe der Windows-API erforderlich als für Operationen auf kleineren Dateien. Dies erklärt den in Modus 2 beobachteten deutlichen Anstieg der Laufzeit (vgl. Abbildung 7.1a). Das Hooking-Framework muss jedes einzelne Ereignis abfangen und an den PIP weiterleiten. Das Abfangen eines Ereignisses kostet Rechenzeit.

Dagegen skalieren der PIP und der ProSP gut mit der Anzahl der zu verarbeitenden Ereignisse. Selbst bei einer initialen Dateigröße von 100 000 KB steigt die Laufzeit im Modus 4 gegenüber dem Modus 2 nur um 8 %.

Sowohl im PIP als auch im ProSP haben die zusätzlichen Ereignisse keinen Einfluss auf die Datenstrukturen. Die stattfindenden Informationsflüsse werden mehrfach redundant



(a) Modus 2

(b) Modus 4

Abbildung 7.1: Laufzeit abhängig von der Größe der Datei (10 Daten, Tracelänge 100)

detektiert. Bereits die ersten ReadFile- und WriteFile-Ereignisse beschreiben den Zustandsübergang der Speicherfunktion vollständig. Aus diesem Grund – und zugunsten einer effizienten Testdurchführung – wurde für die übrigen Tests eine Dateigröße von 100 KB gewählt.

Tracelänge Die Laufzeit eines Traces bewegt sich im Modus 0, abhängig von der Länge des Traces, zwischen 0 und 34 Sekunden. Die Schwankungsbreite ist allerdings im Vergleich zur Gesamtdauer groß. Die Laufzeit eines Traces der Länge 1 000 schwankt selbst schon zwischen 15 und 34 Sekunden.

Wird das Hooking-Framework zugeschaltet, steigt die Laufzeit der Traces um Größenordnungen (Abbildung 7.2b). Zusätzlich zur Ereignisverarbeitung bremst das Hooking eines jeden neuen Prozesses das Betriebssystem spürbar aus. Ein Trace mit etwa 78 300 Ereignissen und 100 neuen Prozessen (Dateigröße von 1 000 KB in Abbildung 7.1a) hat eine durchschnittliche Laufzeit von 790 s. Ein Trace mit etwa 16 800 Ereignissen und 1 000 neuen Prozessen (Tracelänge 1 000 in Abbildung 7.2b) liegt bereits bei einer durchschnittlichen Laufzeit von 5 751 s. Dies ist mehr als das Siebenfache.

Mit Aktivierung des PEP, des PMP, des PIP, des ProSP und des ProCP steigt die Laufzeit in Relation zur Gesamtlaufzeit nur marginal (Abbildung 7.2c). Der Median der Laufzeit erhöht sich je nach Tracelänge um 1,5 bis 2 Prozent. Löschregeln und Abstraktionsregeln

haben keinen nennenswerten Einfluss auf die Laufzeit eines Traces (Abbildung 7.2d).

Über alle Modi hinweg verhält sich die Laufzeit der Traces linear proportional zur Länge des Traces (Abbildung 7.2b). Die verschiedenen Komponenten des Datenschutzauskunftssystems verlängern die Laufzeit nur um einen konstanten Faktor. Wie am Hooking-Framework für Windows 7 ersichtlich wird, hängt der Laufzeitoverhead des Datenschutzauskunftssystems im Wesentlichen von der Implementierung des Hooking-Teils des PEP ab.

Anzahl personenbezogener Daten Wird statt der Tracelänge die Anzahl der personenbezogenen Daten in der initialen Datei variiert, stellt sich heraus, dass die Laufzeit eines Traces vollkommen unabhängig von diesem Parameter ist (Abbildungen 7.3a bis 7.3d). Die Anzahl neuer Prozesse und die Anzahl der zu verarbeitenden Ereignisse sind unabhängig von der Anzahl der personenbezogenen Daten. Der Aufwand zur Modifikation der Datenstrukturen fällt nicht ins Gewicht.

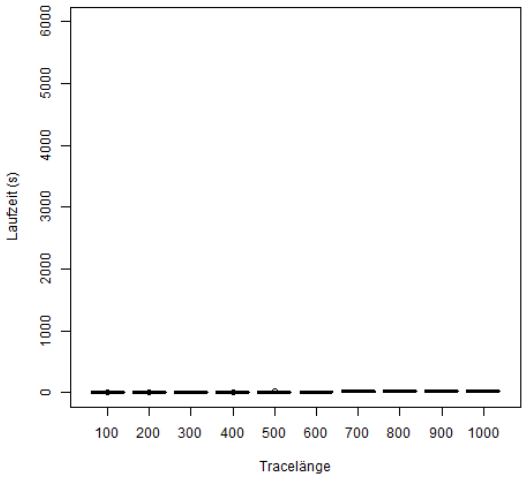
Daraus folgt, dass das Datenschutzauskunftssystem ausreichend robust für den Einsatz in Datenverarbeitungsszenarien mit intensiver Nutzung personenbezogener Daten ist. Die Anwendung der Lösch- und Abstraktionsregeln ist auch bei einer Vielzahl personenbezogener Daten effizient durchführbar.

7.3.3 Ergebnisse zur Speicherskalierbarkeit

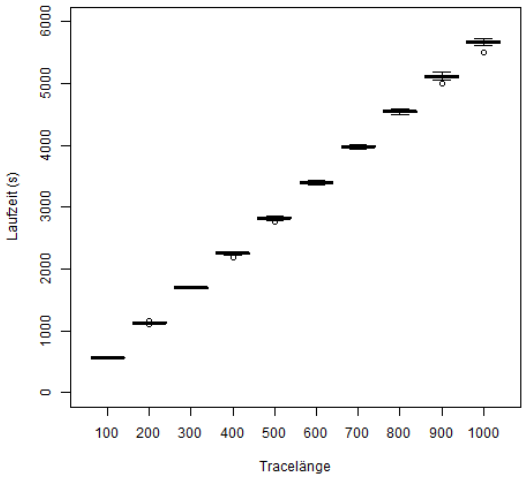
Der Speicherbedarf des Datenschutzauskunftssystems setzt sich aus dem Speicherbedarf des PIP für das Informationsflussmodell und dem Speicherbedarf des ProSP für die Provenance zusammen. Der Speicherbedarf des ProCP ist für jedes personenbezogene Datum konstant und damit unabhängig von den eigentlichen Datenverarbeitungsvorgängen. In den Lasttests ist der Speicherbedarf des ProCP deshalb vernachlässigbar.

Das Informationsflussmodell des PIP wächst mit jedem Ereignis, das die Speicher-, Alias- oder Benennungsfunktion um neue Relationen erweitert. Im gleichen Maße sinkt jedoch der Speicherbedarf des Informationsflussmodells, wenn Ereignisse Relationen aus der Speicher-, Alias- oder Benennungsfunktion entfernen (Primitive *clear*, *rm_alias_locally*, *rm_alias_globally*, *clear_aliases*, *rm_bidir_alias_locally* und *rm_naming*). Letzteres geschieht beispielsweise, wenn Prozesse beendet oder Dateien gelöscht werden.

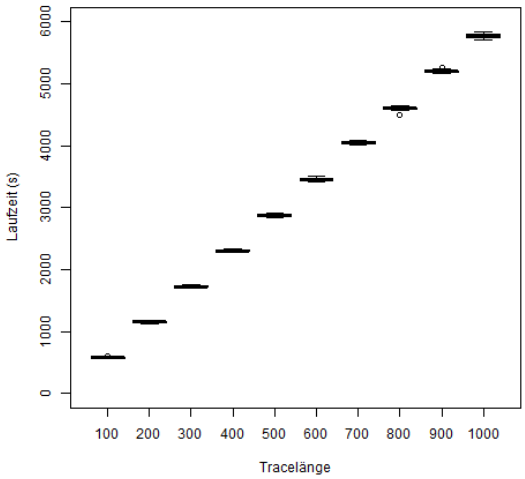
Für den Speicherbedarf des ProSP sieht die Situation anders aus. Die Provenance wächst mit jedem Informationsflussereignis. Die Historie vergisst in der Grundkonfiguration nichts. Der Umgang mit personenbezogenen Daten muss durchgängig nachvollziehbar sein (Anforderung 16). Löschregeln reduzieren potentiell den Speicherbedarf, indem Verarbeitungsschritte aus der Provenance entfernt werden, die abgeschlossen und für die Datenschutzauskunft nicht erforderlich sind. Abstraktionsregeln verhindern von Anfang an, dass das Informationsflussmodell in vollem Detailgrad in die Provenance übertragen wird. Abstraktionsregeln werden im Informationsflussmodell über Aliasrelationen



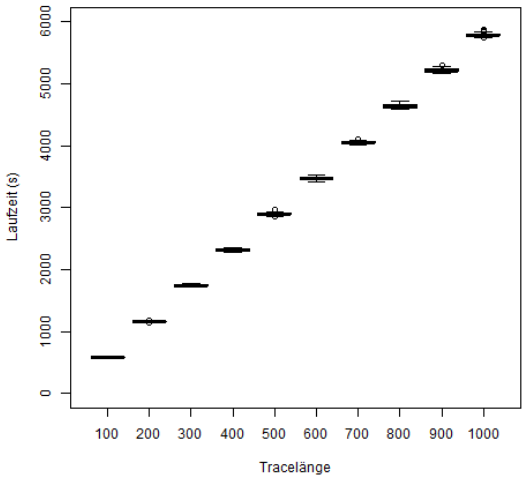
(a) Modus 0



(b) Modus 2

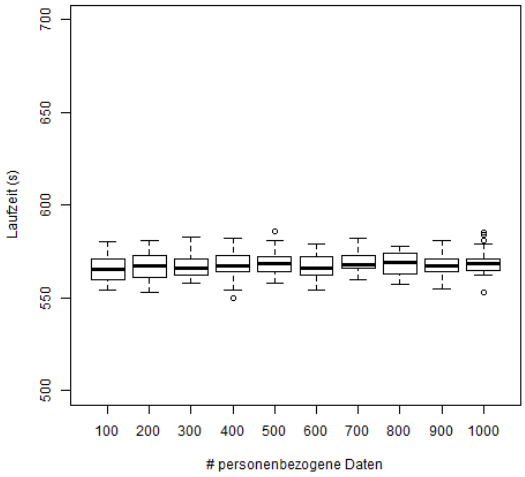


(c) Modus 4

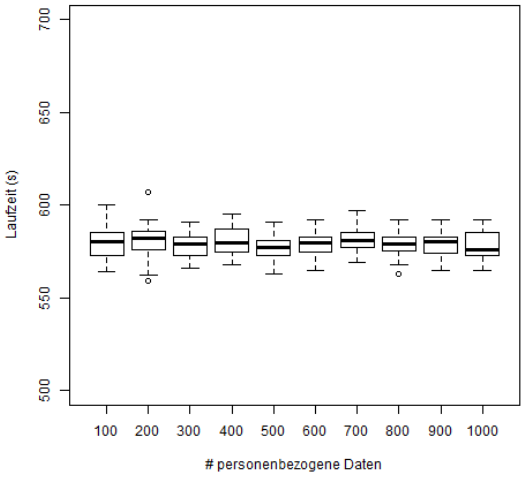


(d) Modus 6

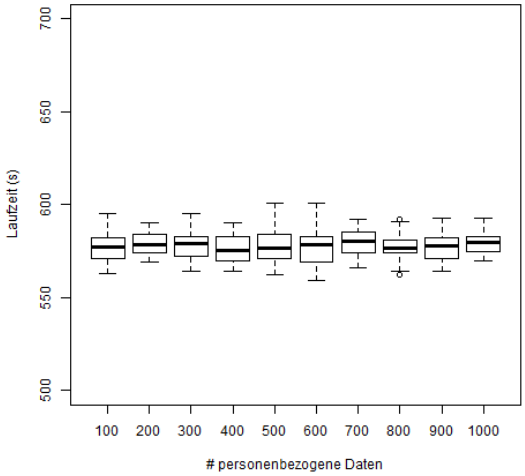
Abbildung 7.2: Laufzeit in Abhängigkeit von der Tracelänge (10 personenbezogene Daten, 100 KB Dateien)



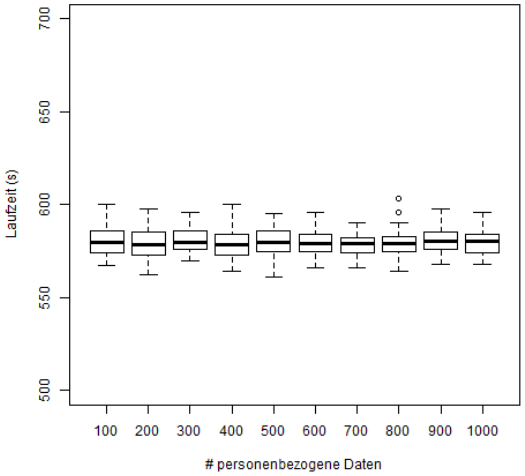
(a) Modus 3



(b) Modus 4



(c) Modus 5



(d) Modus 6

Abbildung 7.3: Laufzeit in Abhängigkeit von der Anzahl der personenbezogenen Daten (Tracelänge 100, 100 KB Dateien)

hinterlegt. Aus diesem Grunde erhöhen sie potentiell den Speicherbedarf des PIP.

In den Abbildungen 7.4 bis 7.7 ist der zusätzliche Speicherbedarf der Komponenten, der über den initialen Speicherverbrauch durch die Anwendungslogik der Komponenten hinaus auftritt, abhängig von den gewählten Parametern und Modi dargestellt.

PIP Betrachtet man den tatsächlichen Speicherbedarf des PIP, wird ersichtlich, dass Abstraktionsregeln bei kurzen Tracelängen im Gesamtspeicherverbrauch des PIP untergehen. Unabhängig von der Anzahl der personenbezogenen Daten ist der Speicherverbrauch nach einem Trace der Länge 100 mit (Modus 6) und ohne (Modus 4) Abstraktionsregeln fast identisch (Abbildung 7.5). Die für einen abstrakten Container in der Datenstruktur je Datum zusätzlich erforderliche Speicherrelation scheint nicht ins Gewicht zu fallen. Die notwendigen Aliasrelationen sind von der Anzahl der durch die Kopieroperationen berührten Container, und damit von der Tracelänge, abhängig.

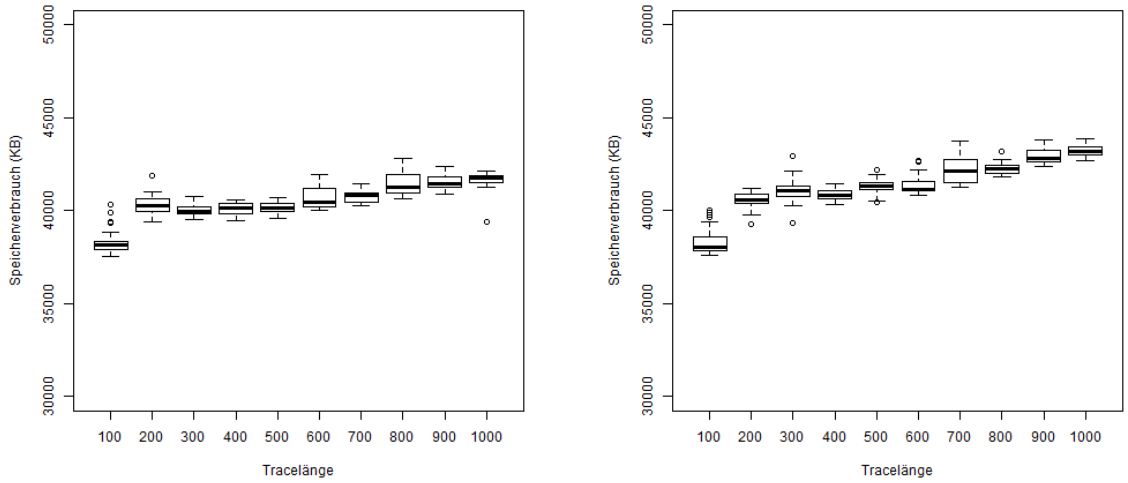
Bei längeren Traces steigt allerdings der Speicherbedarf des PIP (Abbildung 7.4). Jede Kopieroperation in eine neue Datei erfordert zwei neue Aliasrelationen, eine zum Container des Kopierprozesses und eine zum Dateicontainer. Der resultierende Speicherbedarf des Modus 6 bewegt sich jedoch in vergleichbaren Dimensionen wie der des Modus 4. Einflüsse des Garbage Collectors können nicht ausgeschlossen werden.

ProSP Der Speicherbedarf des ProSP steigt ohne Lösch- und Abstraktionsregeln (Modus 4) sowohl linear mit der Tracelänge (Abbildung 7.6a) als auch linear mit der Anzahl der personenbezogenen Daten (Abbildung 7.7a).

Mit Löschregeln (Modus 5) bleibt der Speicherverbrauch unverändert (Abbildungen 7.6b und 7.7b). Die Löschregeln sind wirkungslos. Einmal für die Provenance vergebener Speicher wird nicht wieder freigegeben, obwohl die Repräsentationen der personenbezogenen Daten in den Kopierprozessen im Laufe eines Traces wieder aus der Datenstruktur entfernt werden. Der inkrementelle Garbage Collector ist entweder für diesen Anwendungsfall nicht leistungsfähig genug oder die Speicherfreigabe geschieht nur innerhalb des Prozesses der JVM und ist von Außen nicht sichtbar.

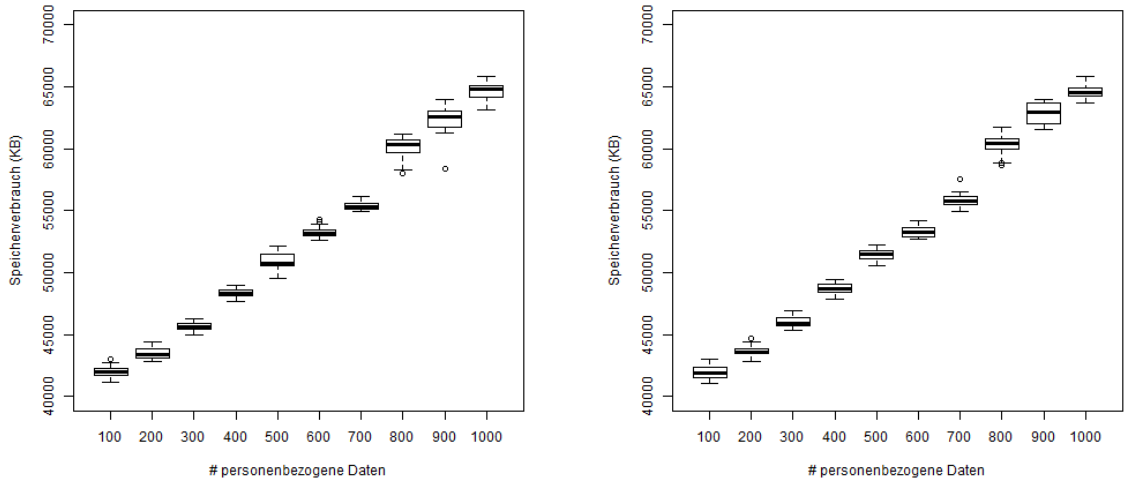
Ganz anders stellt sich die Situation mit aktivierten Abstraktionsregeln dar (Modus 6). Eine Steigerung der Tracelänge führt zu keiner Veränderung des Speicherbedarfs der Provenance (Abbildung 7.6c). Der Speicherbedarf bleibt konstant, weil die Provenance nur bezüglich der beiden abstrakten Container, je einer für die Prozesse und einer für das Dateisystem, gespeichert wird. Die Kopieroperationen führen zu keinen neuen Einträgen in der Provenance.

Wird die Anzahl der personenbezogenen Daten erhöht, steigt der Speicherbedarf auch weiterhin, jedoch um Größenordnungen weniger (Abbildungen 7.7c und 7.7d). Zusätzliche personenbezogene Daten erfordern zusätzliche Repräsentationen in der Provenance. In Abbildung 7.7d wird erkennbar, dass der Speicherbedarf sublinear steigt.



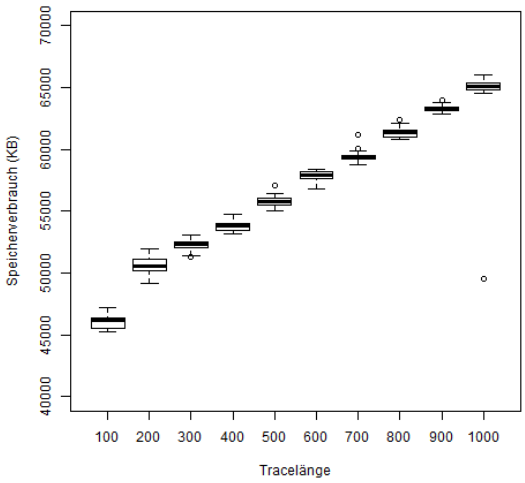
(a) Modus 4 (b) Modus 6

Abbildung 7.4: Speicherbedarf des PIP in Abhängigkeit von der Tracelänge (10 personenbezogene Daten, 100 KB Dateien)

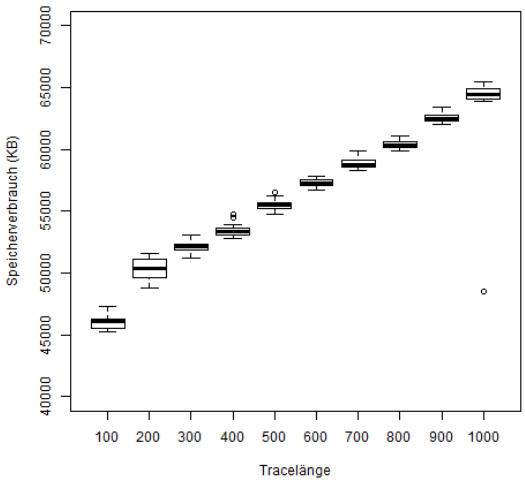


(a) Modus 4 (b) Modus 6

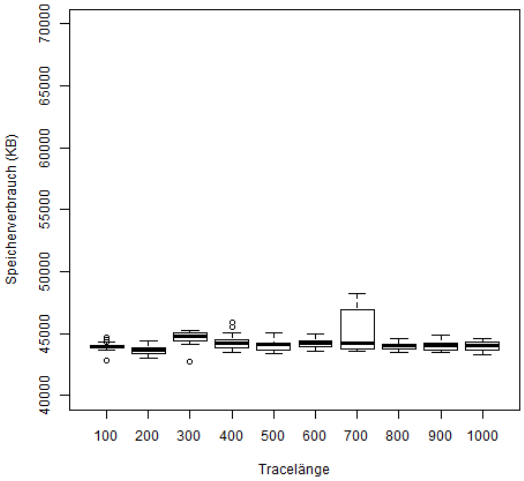
Abbildung 7.5: Speicherbedarf des PIP in Abhängigkeit von der Anzahl der personenbezogenen Daten (Tracelänge 100, 100 KB Dateien)



(a) Modus 4

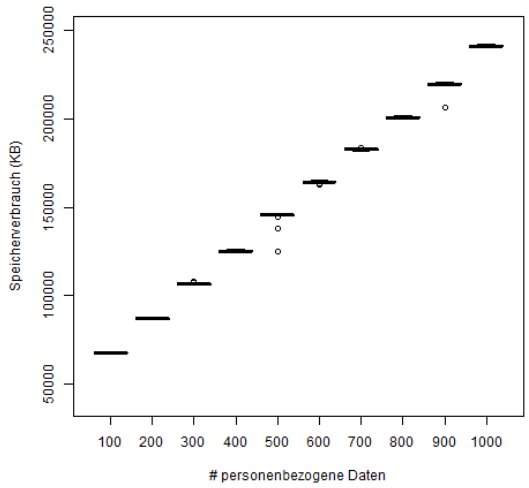


(b) Modus 5

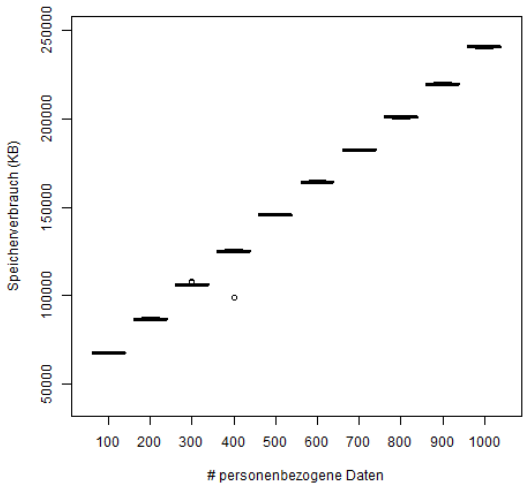


(c) Modus 6

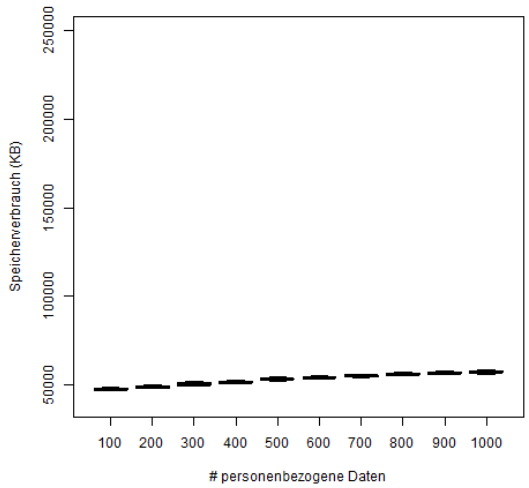
Abbildung 7.6: Speicherbedarf des ProSP in Abhängigkeit von der Tracelänge (10 personenbezogene Daten, 100 KB Dateien)



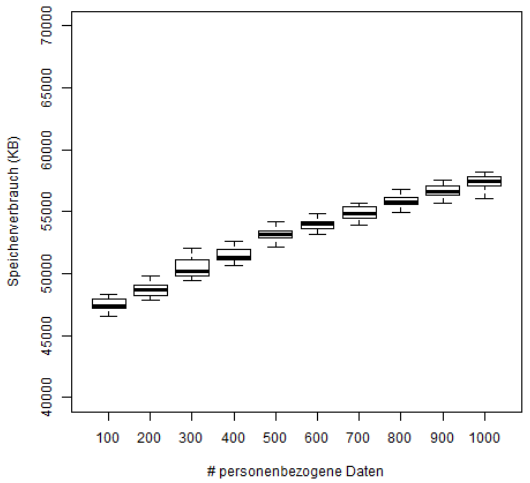
(a) Modus 4



(b) Modus 5



(c) Modus 6



(d) Modus 6 (Wertebereich bis 70 000 KB)

Abbildung 7.7: Speicherbedarf des ProSP in Abhängigkeit von der Anzahl der personenbezogenen Daten (Tracelänge 100, 100 KB Dateien)

Die Speicherskalierbarkeit des Datenschutzauskunftssystems wird durch die Abstraktionregeln signifikant verbessert. Der leichte Speichermehrbedarf im PIP wird durch die Speichereinsparungen im ProSP mehr als aufgewogen.

7.4 Zwischenfazit

Zwei Maßnahmen sollen zum datenminimalen und skalierbaren Einsatz des Datenschutzauskunftssystems beitragen: Löschregeln und Abstraktionsregeln. Konzeptionell führen diese Maßnahmen zu einer passgenauen Datensammlung und einer Provenance, die exakt dem durch den Auskunftsanspruch geforderten Umfang entspricht.

Die vorliegende Evaluation wurde als Lasttest für Kopieroperationen durchgeführt. Kopieroperationen sind ein wesentlicher Faktor für den Austausch personenbezogener Daten. Sie decken die Primitive der Speicherfunktion im Informationsflussmodell vollständig ab. Kopieroperationen stellen ein Hochlastszenario für das Datenschutzauskunftssystem dar und testen das Datenschutzauskunftssystem in einer Extremsituation. Dennoch sind die Aussagen des Lasttests nicht generell auf beliebige andere Szenarien übertragbar. Beispielsweise könnten Ereignisse mit komplexeren Informationsflusssemantiken den Einfluss des PIP auf die Laufzeit des Systems erhöhen.

Die Evaluation der Laufzeit des Datenschutzauskunftssystems zeigt, dass der Flaschenhals des Datenschutzauskunftssystems die PEPs sind. Sie haben den bedeutendsten Einfluss auf die Laufzeit des Systems. Mit dem Einsatz effizienterer Hooking-Mechanismen würde die Gesamtlaufzeit des Datenschutzauskunftssystems merklich sinken. Der nahezu konstante Overhead der übrigen Komponenten des Datenschutzauskunftssystems lässt den Schluss zu, dass das Datenschutzauskunftssystem an sich einem produktiven Einsatz gewachsen ist.

Die Analyse des Speicherbedarfs des ProSP ergibt, dass Löschregeln bei der in Java vorgenommenen Implementierung keine Vorteile im Speicherverbrauch ergeben. Löschregeln haben allerdings weiterhin dort ihre Berechtigung, wo Speicherorte in der Provenance persistiert werden müssen. Sie dienen dann dem Ziel der nachträglichen Datenminimierung.

Abstraktionsregeln haben in dem gegebenen Szenario stark positive Effekte. Die Größe der Provenance steigt im Verhältnis zur Menge der in jedem Schritt verarbeiteten personenbezogenen Daten sublinear und im Verhältnis zur Menge der Operationen gar nicht mehr. Es kann demnach bestätigt werden, dass eine datenminimale und skalierbare Implementierung von Provenance-Tracking möglich ist (ℌ3).

Eine Einschränkung ergibt sich daraus, dass Abstraktionsregeln nur wirksam sein können, wenn ein hoher Abstraktionsgrad der Provenance akzeptabel ist. Werden einzelne Verarbeitungsschritte zu vollständig unterschiedlichen Zwecken durchgeführt, müssen diese Schritte in der Provenance erkennbar sein. Ein höherer Speicherbedarf ist dann die Folge.

8 Verteiltes Provenance-Tracking

Provenance-Tracking für ein Datenschutzauskunftssystem erfolgt in zweierlei Hinsicht verteilt: (1) Beim Austausch personenbezogener Daten zwischen unterschiedlichen Domänen und (2) bei der Speicherung und Aggregation von Personal-Data-Provenance.

Der Austausch personenbezogener Daten zwischen unterschiedlichen Domänen ist als systemübergreifendes Informationsfluss- und Provenance-Tracking gestaltet. Nach einer Einführung in die Thematik werden im Abschnitt 8.1.1 die notwendigen Erweiterungen der Informationsflusssemantik aus Kapitel 6.2 behandelt.[1] Im anschließenden Abschnitt 8.1.2 wird der Ablauf des systemübergreifenden Provenance-Trackings dargestellt und an einem Beispiel erläutert.

Die Konzepte zur Aggregation von Personal-Data-Provenance werden im Abschnitt 8.2 umrissen. Insbesondere wird die Rolle des ProCP und der Aufbau einer vollständigen Personal-Data-Provenance beleuchtet.

8.1 Schichten- und systemübergreifendes Informationsfluss- und Provenance-Tracking

Das im Kapitel 6.1 vorgestellte Informationsflussmodell unterstützt Informationsflüsse innerhalb einer Abstraktionsschicht. Die PEPs der einzelnen Abstraktionsschichten können auf Grundlage dieses Modells die Semantik ihrer Ereignisse dynamisch beim zuständigen PIP hinterlegen. Das Modell ist allerdings noch nicht in der Lage, schichten- und systemübergreifende Informationsflüsse abzubilden. Schichtenübergreifende Informationsflüsse meint Informationsflüsse zwischen unterschiedlichen Abstraktionsschichten, z. B. dem Betriebssystem und einer Anwendung. Mit systemübergreifenden Informationsflüssen werden Informationsflüsse zwischen IT-Systemen, die jeweils einen eigenen PDP, PIP und ProSP haben, bezeichnet.

Bereits existierende Ansätze Ein Modell für schichtenübergreifende Informationsflüsse wurde bereits von Lovat entworfen.[2] Dieses im Abschnitt 8.1.1 beschriebene Modell wird in diesem Kapitel auf systemübergreifende Informationsflüsse verallgemeinert und einer Beschreibung in der Informationsflusssemantik (siehe Kapitel 6.2) zugänglich gemacht.

[1] Die Ideen der Abschnitte wurden bereits in Birnstill/Bier et al. 2016 veröffentlicht.
[2] Lovat 2015.

Ein Modell für Nutzungskontrolle in verteilten Systemen unter Berücksichtigung von Informationsflüssen wurde von Kelbert entwickelt.[3] Kelbert legt einen starken Fokus auf die verteilte Auswertung und Durchsetzung von UC-Polices und auf die effiziente Kommunikation zwischen PDPs und PIPs. Das nachfolgend vorgestellte Modell für systemübergreifende Informationsflüsse übernimmt viele Ideen aus diesem Modell, unterscheidet sich allerdings in einigen wesentlichen Punkten. Kelberts Modell setzt einen Adressraum voraus, der eine Funktion von Adressen auf IT-Systeme bereitstellt. In der von Kelbert vorgestellten Instanziierung ist dies TCP/IP. Dieser Adressraum ist fest in die Implementierung integriert. Systemübergreifende Informationsflüsse werden, darauf aufbauend, anhand der TCP/IP-Verbindung identifiziert.

Während alle gängigen Protokolle der Anwendungsschicht auf TCP/IP aufsetzen, ist es in der Praxis häufig nicht möglich, das Informationsfluss-Tracking direkt durch ein Monitoring des Netzwerk-Stacks durchzuführen. Eine Ergänzung von Betriebssystemen um Komponenten eines automatisierten Datenschutzes wird zwar bereits seit Jahren gefordert,[4] ist jedoch gegenwärtig noch nicht Realität. Deshalb ist es für Geschäftsanwendungen erstrebenswert, PEPs als Plug-Ins in Applikationen zu integrieren, um nicht ins Betriebssystem eingreifen zu müssen.

Das im Folgenden vorgestellte Modell erlaubt systemübergreifendes Informationsflusstracking auf Anwendungsebene. Die Verknüpfung der Informationsflüsse in den Anwendungen erfolgt dynamisch anhand der Informationsflusssemantik, die auf eine beliebige Kennung (Scope) verweist. Eine feste Implementierung im PIP ist nicht notwendig.[5]

Die Architektur von Kelbert erlaubt die Synchronisierung der Zustände in den Policy-Mechanismen (PDP) und des Zustands des Informationsflussmodels (PIP). Implementierungsseitig wird dafür eine verteiltes Apache-Cassandra-Datenbanksystem[6] verwendet. Da verteilte Nutzungskontrollentscheidungen für Data-Provenance und die Betroffenenrechte nicht erforderlich sind, werden sie von der hier vorgestellten Architektur nicht unterstützt. Ein IT-System wirkt für den PIP eines anderes IT-Systems wie ein PEP. Ein globaler Zustand über alle Datennutzungen würde dem Prinzip der informationellen Gewaltenteilung zuwiderlaufen.[7]

Neben dem Ansatz von Kelbert existieren noch eine Reihe von Konzepten zur Propagierung von Taintmarken über Systemgrenzen hinweg. SeeC speichert je Prozess eines IT-Systems Taintmarken in einem Schattenspeicher und erzeugt für jeden Socket eine *write queue* für Taintmarken zu Datenflüssen über TCP/IP.[8] NEON ist ein Monitor für

[3] Kelbert/Pretschner 2013; Kelbert/Pretschner 2014; Kelbert/Pretschner 2015.

[4] Wächter, DuD 1996, 272 (272).

[5] In der Implementierung von Kelbert ist generell ein bidirektionaler Alias zwischen lokalem Socket und Remote-Socket vorgesehen.

[6] http://cassandra.apache.org.

[7] In Cassandra haben dagegen alle Knoten dieselbe Rolle und damit gleichberechtigte Zugriffsrechte auf die gesamte verteilte Datenbank.

[8] Kim et al. 2009.

virtuelle Maschinen und unterstützt Taintmarken zum Tracken von Informationsflüssen auf Byteebene.[9] Muniswamy-Reddy et al. integrieren in ihrem Framework Provenance-Collection und -Storage über Systemschichten hinweg.[10] Gleiches erlaubt GARM mit Hilfe von Application-Rewriting.[11]

Motivation einer Erweiterung des Informationsflussmodells Sowohl bei schichten- als auch bei systemübergreifenden Informationsflüssen gibt es für jedes Ereignis, das einen ausgehenden Informationsfluss anzeigt (*ausgehendes Ereignis*), ein korrespondierendes Ereignis, das einen eingehenden Informationsfluss signalisiert (*eingehendes Ereignis*). Diese beiden Ereignisse müssen durch die Semantik in Deckung gebracht werden, um eine durchgängige Provenance zu gewährleisten.

Beispiel. *Im Online-Shopsystem von Adbokis sind die Kundenstammdaten sowie die Daten zu allen vorangegangen Bestellungen hinterlegt. Das Backend des Shopsystems ist intern per Webbrowser erreichbar. Es lässt einen Export und Download der Kundenstammdaten durch Mitarbeiter von Adbokis zu. Werden Kundenstammdaten exportiert, muss der Zweckbezug und die Herkunft der Daten erkennbar bleiben.*

Sowohl auf Senderseite als auch auf Empfängerseite müssen in den Anwendungen, dem Shopsystem bzw. dem Browser, PEPs als Plug-Ins integriert sein, die die eingehenden beziehungsweise ausgehenden Ereignisse abfangen können. Prototypisch wurden Plug-Ins für das Online-Shopsystem Shopware[12] und den Browser Chrome[13] in PHP bzw. JavaScript implementiert. Darüber hinaus muss auf beiden Systemen eine vollständige lokale Infrastruktur des Datenschutzauskunftssystems bereitstehen.

Der PEP des Shopsystems beobachtet nun zunächst ein Ereignis, das den Start eines Downloads signalisiert. Ein Download ergibt einen Informationsfluss von einem lokalen Datenobjekt hin zu einem Container, der die Downloadverbindung verkörpert. Das Startereignis wird dem PIP mitgeteilt. Ohne eine Erweiterung des Modells würde der PIP das Ereignis nicht besonders interpretieren und die Downloadverbindung wie jeden anderen Container behandeln. Das empfangende System würde nicht informiert, die Provenance an dieser Stelle unterbrochen. Das auf Clientseite beobachtete Downloadereignis hätte folglich keine Datenflüsse zur Folge. Zur Erkennung und Verarbeitung systemübergreifender Informationsflüsse ist es deshalb erforderlich, dass die Ereignisse auf beiden Seiten durch die PIPs gemäß einer (Remote-)Informationsflusssemantik interpretiert werden.

[9] Q. Zhang et al. 2010.
[10] Muniswamy-Reddy et al. 2009.
[11] Demsky 2009; Demsky 2011.
[12] https://de.shopware.com.
[13] https://www.google.com/chrome.

8.1.1 Scope-Spezifikation und Scope-Verarbeitung

Der Scope ist eine Kennzeichnung, die auf allen am Informationsfluss beteiligten Schichten oder Systemen bekannt ist. Eine Scope-Spezifikation kennzeichnet innerhalb einer Informationsflusssemantik, ob und in welcher Weise ein Ereignis zu einem Informationsfluss zwischen Abstraktionsschichten oder IT-Systemen gehört. Die Ereignisse, die zum selben Informationsfluss gehören, werden anhand des Scopes, gegeben in einem Parameter der Ereignisse, ein und demselben Informationsfluss zugeordnet. Der Parameter des Ereignisses kann auf unterschiedlichen Schichten allerdings durchaus unterschiedlich heißen. Die Zuordnung wird durch den Scopenamen in der Semantik sichergestellt.

Das Informationsflussmodell wird demgemäß um eine Menge von Scopes SCOPE und Intermediate-Containern $\mathscr{C}_\iota$ erweitert. Ein Intermediate-Container ($c_\iota \in \mathscr{C}_\iota$) steht für eine bestimmte Verbindung zwischen zwei Schichten oder IT-Systemen. Intermediate-Container unterschiedlicher IT-Systeme sind nicht identisch, selbst wenn sie sich auf denselben Scope ($\varsigma \in$ SCOPE) beziehen. Jedes Ereignis gehört zu maximal einem schichtenübergreifenden oder systemübergreifenden Scope. Der Zustand wird außerdem um die folgenden beiden Abbildungen ergänzt:[14] Die *Intermediate-Container-Funktion* ι : SCOPE $\to \mathscr{C}_\iota$ bildet jeden Scope auf einen Intermediate-Container ab. Die *Scope-Zustandsfunktion* ζ : SCOPE $\to \{\textsc{activated}, \textsc{deactivated}\}$ bezeichnet die gegenwärtig aktiven, d. h. geöffneten, Scopes.

Insgesamt ergibt sich ein Zustand $\Sigma = (\mathscr{C} \to \mathcal{P}(\mathscr{D})) \times (\mathscr{C} \to \mathcal{P}(\mathscr{C})) \times (\mathscr{F} \to \mathscr{C}) \times (\text{SCOPE} \to \mathscr{C}_\iota) \times (\text{SCOPE} \to \{\textsc{activated}, \textsc{deactivated}\})$. Im initialen Systemzustand σ_0 gibt es einen Intermediate-Container c_ι für jeden Scope ς und $\zeta(\varsigma)$ ist DEACTIVATED für alle $\varsigma \in$ SCOPE.

Drei Attribute einer Scope-Attributspezifikation in der Informationsflusssemantik definieren, wie der der Systemzustand von einem Ereignis modifiziert wird:

$$\chi : \Sigma \times E \to \text{SCOPE} \times \text{DELIMITER} \times \text{BEHAVIOR} \times \text{INTER}$$
$$\text{DELIMITER} = \{\textsc{open}, \textsc{close}, \textsc{none}\}$$
$$\text{BEHAVIOR} = \{\textsc{in}, \textsc{out}, \textsc{intra}\}$$
$$\text{INTER} = \{\textsc{xlayer}, \textsc{xsystem}\}$$

Der *delim* $\in$ DELIMITER gibt an, ob ein Ereignis einen neuen Informationfluss einleitet. Ist der Delimiter OPEN, dann ändert sich der Zustand des Scopes auf ACTIVATED. Ist der Delimiter CLOSE, dann ändert sich der Zustand des Scopes auf DEACTIVATED. Bei NONE ändert sich nichts. Das Verhalten *behav* $\in$ BEHAVIOR beschreibt, ob ein Ereignis einen ausgehenden Fluss (OUT) oder einen eingehenden Fluss aus einer anderen Abstraktionsschicht oder einem anderen System (IN) signalisiert. Das Verhalten eines Ereignisses beeinflusst die Verarbeitung der generischen Primitive in der Übergangsrelation. INTRA

[14] Lovat 2015.

bedeutet, dass das Ereignis zu keinem übergreifenden Informationsfluss gehört. In diesem Fall hat das Ereignis keinen Scope[15] oder Delimiter. Die Primitive werden nicht modifiziert. $inter \in$ INTER unterscheidet zwischen schichtenübergreifenden (XLAYER) und systemübergreifenden (XSYSTEM) Informationsflüssen. Das XML-Schema in Anhang D.1 berücksichtigt bereits all diese Aspekte.

Wie in obiger Gleichung sichtbar, ist die Funktion χ nicht nur vom Ereignis, sondern auch vom aktuellen Zustand des Informationsflussmodells abhängig. Deshalb müssen die Scope-Attribute zur Laufzeit anhand der Informationsflusssemantik identifiziert werden. Die Aktionsbeschreibung eines Ereignisses in der Semantik beinhaltet die möglichen Scope-Attributspezifikationen für dieses Ereignis in einer geordneten Liste. Scope-Attributspezifikationen beinhalten, außer für behavior="INTRA", einen Scopenamen, der auf den Ereignisparameter verweist, der den Scope enthält.

Der PIP prüft bei der Verarbeitung eines Ereignisses in der durch die Aktionsbeschreibung gegebenen Reihenfolge die folgenden drei Bedingungen:

1. Falls vorhanden, stimmt der Scopename mit dem Namen eines Parameters des Ereignisses überein?
2. Falls die Scope-Attributspezifikation delimiter="OPEN" vorsieht, ist der in diesem Parameter referenzierte Scope DEACTIVATED?
3. Falls die Scope-Attributspezifikation delimiter="NONE" oder delimiter="CLOSE" vorsieht, ist der in diesem Parameter referenzierte Scope ACTIVATED?

Treffen alle Bedingungen zu, wird die entsprechende Scope-Attributspezifikation ausgewählt, die Attribute werden ausgelesen und die Übergangsrelation wird auf deren Grundlage modifiziert.

Algorithmus 15 beschreibt, wie die Übergangsrelation $\mathscr{T}$ modifiziert wird, um $\mathscr{T}_{mod}$, die Übergangsrelation für XLAYER- und XSYSTEM-Informationsflüsse, zu erzeugen. $\mathscr{T}[links \xrightarrow{\text{subst.}} rechts]$ bedeutet, dass der Term von $\mathscr{T}$ auf der linken Seite durch den Term auf der rechten Seite ersetzt wird.

Zunächst wird geprüft, ob überhaupt ein Scope vorliegt (Zeile 3). Ohne einen Scope wird die Übergangsrelation nicht modifiziert. Falls der Delimiter OPEN ist, wird der Scope aktiviert (Zeile 6). Ist der Delimitier CLOSE wird der Scope nach Abarbeitung des Ereignisses geschlossen (Zeile 23). Dazwischen wird abhängig vom Verhalten entweder das linke (Quelle, Zeile 15 ff.) oder das rechte (Ziel, Zeile 8 ff.) Argument der Primitve der Speicherfunktion durch den Intermediate-Container des Scopes ersetzt. $\mathscr{T}_{mod}$ wird zum Abschluss auf den Zustand σ angewandt (Zeile 22). Der Zustandsübergang ist vollzogen.

Ein $inter =$ XSYSTEM führt zu keiner Modifikation der Übergangsrelation. Stattdessen löst es aus, dass der PIP dem PIP des anderen Systems eine spezielle Semantik, die Remote-

[15] Abbildung auf NIL.

Semantik weiterleitet. Diese Semantik lässt den Informationsfluss auf Empfängerseite wie einen schichtübergreifenden Informationsfluss behandeln. Entsprechende Modifikationen werden vorgenommen. Die Details werden im nachfolgenden Beispiel behandelt.

Algorithmus 15: $\mathscr{T}_{inter}(\mathscr{T}, \sigma, e)$

$(ş, delim, behav, inter) \leftarrow \chi(\sigma, e)$
$(s, l, f, \iota, \zeta) \leftarrow \sigma$
if $ş \neq$ NIL **then**
  $c_\iota \leftarrow \iota(ş)$
  $\mathscr{T}_{mod} \leftarrow \mathscr{T}$
  if $delim =$ OPEN **then**
    $\sigma \leftarrow (s, l, f, \iota, \zeta[ş \leftarrow \text{ACTIVATED}])$
  if $behav =$ OUT $\wedge$ $inter =$ XLAYER **then**
    $\mathscr{T}_{mod} \leftarrow \mathscr{T}_{mod}\big[s[c \leftarrow s(c) \cup \{d_i\}_{1\leq i\leq n\in\mathbb{N}}]$
      $\xRightarrow{\text{subst.}} s[c_\iota \leftarrow s(c_\iota) \cup \{d_i\}_{1\leq i\leq n\in\mathbb{N}}]\big]$
    $\mathscr{T}_{mod} \leftarrow \mathscr{T}_{mod}\big[\forall u \in l(c) : s[u \leftarrow s(u) \cup \{d_i\}_{1\leq i\leq n\in\mathbb{N}}]$
      $\xRightarrow{\text{subst.}} \forall u \in l(c_\iota) : s[u \leftarrow s(u) \cup \{d_i\}_{1\leq i\leq n\in\mathbb{N}}]\big]$
    $\mathscr{T}_{mod} \leftarrow \mathscr{T}_{mod}\big[\forall u \in l^*(c) : s[u \leftarrow\ s(u) \cup \{d_i\}_{1\leq i\leq n\in\mathbb{N}}]$
      $\xRightarrow{\text{subst.}} \forall u \in l^*(c_\iota) : s[u \leftarrow\ s(u) \cup \{d_i\}_{1\leq i\leq n\in\mathbb{N}}]\big]$
  if $behav =$ IN $\wedge$ $inter =$ XLAYER **then**
    $\mathscr{T}_{mod} \leftarrow \mathscr{T}_{mod}\big[s[c \leftarrow s(c) \cup \{d_i\}_{1\leq i\leq n\in\mathbb{N}}]$
      $\xRightarrow{\text{subst.}} s[c \leftarrow s(c) \cup s(c_\iota)]\big]$
    $\mathscr{T}_{mod} \leftarrow \mathscr{T}_{mod}\big[\forall u \in l(c) : s[u \leftarrow s(u) \cup \{d_i\}_{1\leq i\leq n\in\mathbb{N}}]$
      $\xRightarrow{\text{subst.}} \forall u \in l(c) : s[u \leftarrow s(u) \cup s(c_\iota)]\big]$
    $\mathscr{T}_{mod} \leftarrow \mathscr{T}_{mod}\big[\forall u \in l^*(c) : s[u \leftarrow\ s(u) \cup \{d_i\}_{1\leq i\leq n\in\mathbb{N}}]$
      $\xRightarrow{\text{subst.}} \forall u \in l^*(c) : s[u \leftarrow\ s(u) \cup s(c_\iota)]\big]$
  $\sigma \leftarrow \mathscr{T}_{mod}(\sigma, e)$
  if $delim =$ CLOSE **then**
    $(s, l, f, \iota, \zeta) \leftarrow \sigma$
    $\sigma \leftarrow (s[c_\iota \leftarrow \varnothing], l, f, \iota, \zeta[ş \leftarrow \text{DEACTIVATED}])$
else
  $\sigma \leftarrow \mathscr{T}(\sigma, e)$
return σ

8.1.2 Systemübergreifende Informationsflüsse

Zur Verarbeitung eines systemübergreifenden Informationsflusses gehören vier Schritte: (1) Benachrichtigung der empfangenden Seite über die Behandlung der fließenden Daten, (2) Verarbeitung des ausgehenden Informationsflusses auf Sender und Empfängerseite, (3) Verarbeitung des eingehenden Informationsflusses auf Empfängerseite und (4) Abschluss der Informationsflussbeziehung.

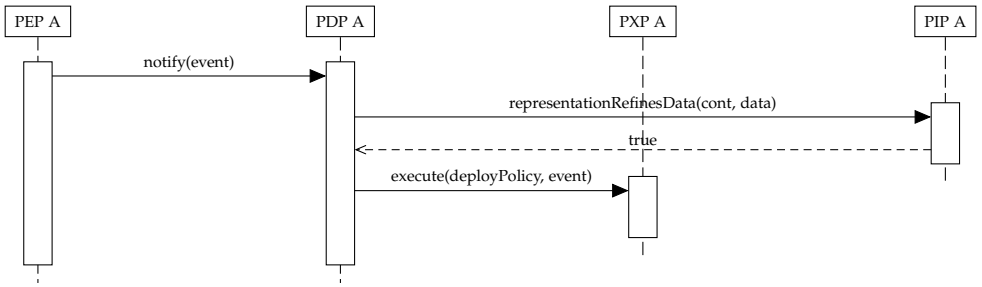


Abbildung 8.1: Systemübergreifendes Usage-Control (1)

Der erste Schritt beginnt damit, dass der PEP des sendenden IT-Systems (PEP A in Abbildung 8.1; Bsp.: Shopware-PEP) dem PDP seines Systems (PDP A) das Ereignis des beginnenden ausgehenden Informationsflusses mitteilt (downloadFile_start) und das Stattfinden des Ereignisses zunächst blockiert. Für jede Policy im PDP, die dieses Ereignis als Auslöser enthält, fragt der PDP beim PIP über die Methode representationRefinesData() anhand des Ausgangscontainers ab, ob das in der Policy referenzierte Datum von diesem Ereignis betroffen ist. Jede Tracking-Policy, für die das der Fall ist (Bsp.: Alle Policies für die Daten aus dem Kundendatensatz), löst die Execute-Action deployPolicy aus.

Im Folgenden wird o. B. d. A. nur von einem Datum ausgegangen. Der PXP[16] bezieht das in der Policy referenzierte Policy-Template für dieses Datum vom PMP/PRP (Abbildung 8.2). Anschließend wird das Policy-Template auf dem PMP des Empfängersystems (PMP B) abgelegt. Dazu wird der lokale PMP (PMP A) über die Methode deployPolicyTemplate() instruiert. Anhand der übergebenen Domäne (dom) wird beim Root-PMP[17] der empfangende PMP abgefragt (lookupPMP()) und dort dieselbe Deployment-Methode aufgerufen.

Ist das Template auf dem Empfängersystem hinterlegt, wird es mit der ID des fließenden Datums instanziiert (Abbildung 8.3).Neben der ID des Datums wird die ID des Templates zweifach übergeben. Einmal, um auf das Template zu verweisen, das instanziiert werden

[16]Im Allgemeinen implementiert der PMP die entsprechende ExecuteAction.
[17]Vgl. Kapitel 5.3.1

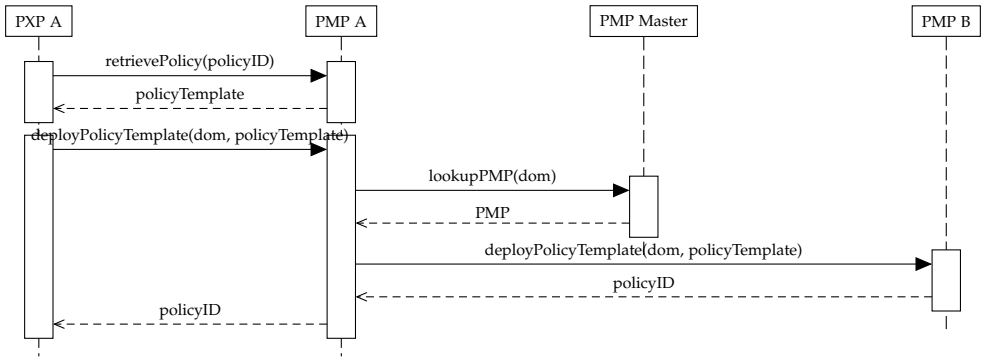


Abbildung 8.2: Systemübergreifendes Usage-Control (2)

soll (templateID) und einmal als eine ID, die zur Instanziierung in das Template geschrieben werden soll (policyID). Dadurch kann das Template auch bei Auswertung der neu entstandenen Policy auf Empfängerseite für einen weiteren Informationsfluss wiedergefunden werden.

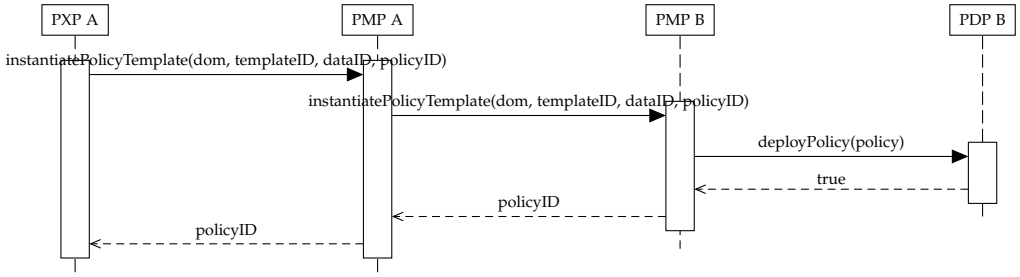


Abbildung 8.3: Systemübergreifendes Usage-Control (3)

Damit ist der erste Schritt abgeschlossen. Der PDP gibt die Entscheidung zurück, dass der Informationsfluss stattfinden darf.

Wird Provenance-Tracking nicht auf Basis von Tracking-Policies betrieben, sondern dauerhaft in allen PIPs aktiviert, entfällt der erste Schritt vollständig.

Im zweiten Schritt wird zunächst der PIP der Senderseite (A) vom PEP über den nun zugelassenen Informationsfluss informiert (Abbildung 8.4, lookup-Aufrufe beim PMP sind zur Vereinfachung nicht dargestellt). Das Ereignis (Listing 8.1) wird auf Grundlage der bekannten Informationsflusssemantik des PEP (Listing D.1) durch den PIP interpretiert.

Die Semantik des Ereignisses gibt einen systemübergreifenden Informationsfluss vor (behavior="OUT" delimiter="OPEN" intersystem="TRUE"). Dies bedeutet, dass das Ereignis auf

Senderseite konventionell, das heißt ohne die Anwendung der Modifikationen in Algorithmus 15, verarbeitet wird. Der Scope wird allerdings aktiviert.

```
<event action="downloadFile_start" timestamp="2016−07−10T21:45:00">
  <complexParameter name="network">
    <parameter name="domain" value="urn:ucn:adbokis:sales:online">
    <parameter name="loa" value="app:shopware">
    <parameter name="type" value="net">
    <parameter name="name" value="192.168.10.7:80;192.168.10.9:6927">
  </complexParameter>
  <parameter name="network_name" value="192.168.10.7:80;192.168.10.9:6927">
  <parameter name="applicationObject_name" value="cdb_connect_475908">
  <parameter name="URL" value="http://192.168.10.7:80/shopware/download/export_12345.csv">
</event>
```

Listing 8.1: Ereignis zum Start des Downloads beim Shopware-PEP

Im Informationsflussmodell wird der Verbindung ein Bezeichner zur späteren Referenzierung zugewiesen (NF_ADD_NAMING). Anschließend werden der Speicherfunktion des Netzwerk-Containers alle Daten zugewiesen, die sich zuvor im für den Download zuständigen Verarbeitungsobjekt von Shopware befunden haben (SF_FLOW).

Des Weiteren wird der ProSP der Senderseite über den Informationsfluss informiert. Für jedes Datum wird die Provenance aktualisiert. Eine Repräsentation der Netzwerkverbindung wird mit der Repräsentation im Verarbeitungsobjekt als Vorgänger erzeugt. Da es sich um einen systemübergreifenden Informationsfluss handelt, wird außerdem der ProSP der Gegenstelle (ProSP B) über die neue Repräsentation in Kenntnis gesetzt. Die Repräsentation wird ohne gesetztes Vorgängerattribut übertragen, so dass auf Empfängerseite keine direkte Verbindung zur Provenance auf Senderseite besteht. Um die gesamte Provenance im Nachhinein herstellen zu können, wird in der CommunicationRepresentation auf beiden Seiten das Attribut remote mit der Domäne der jeweiligen Gegenstelle gesetzt (vgl. Abbildung 6.1).

Anschließend macht der PIP der Senderseite dem PIP der Empfängerseite die Speicherfunktion des Netzwerkcontainers bekannt. Dies geschieht über Aufrufe der Funktion setNewRespresentation(). Über diesen Vorgang wird auch der ProSP B vom PIP B informiert. Da die jeweiligen Repräsentationen bereits bekannt sind, findet keine Veränderung der Provenance auf Empfängerseite statt.

In der Folge leitet PIP A das unveränderte Ereignis (downloadFile_start). an den PIP B weiter. Ist das weiterzuleitende Ereignis das erste Ereignis dieser Art, registriert der PIP A stellvertretend für den PEP A die Remote-Informationsflusssemantik des Ereignisses beim PIP B. Vom PIP B wird das Ereignis nun anhand der Remote-Informationsflusssemantik (Listing D.2) interpretiert. Für die Empfängerseite ist ein systemübergreifender Informationsfluss äquivalent mit einem von der Transportschicht *eingehenden* schichtenüber-

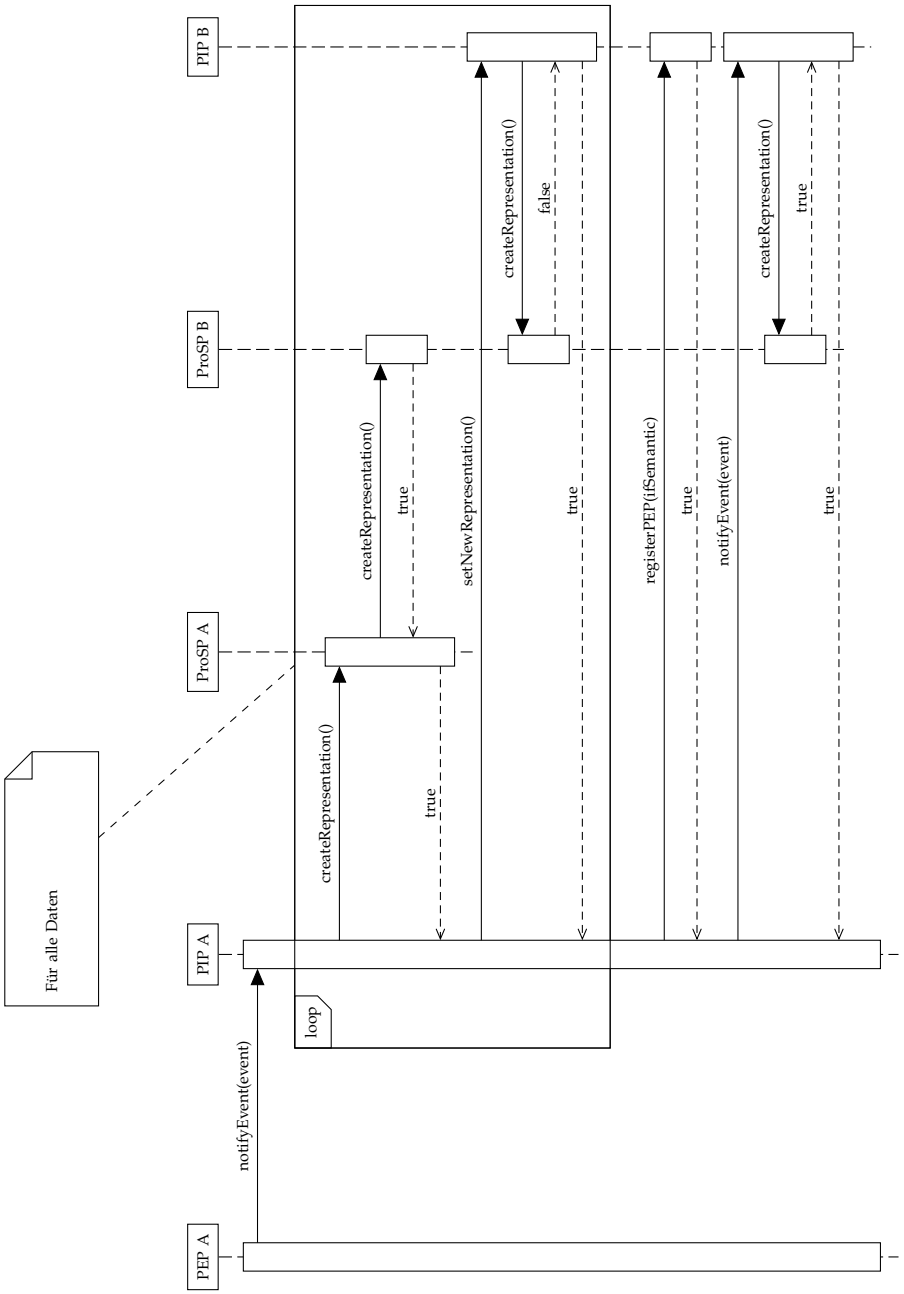


Abbildung 8.4: Systemübergreifende Informationsflüsse (1)

greifenden Informationsfluss. Deshalb sind die Operanden des Primitivs SF_FLOW in der Remote-Informationsflusssemantik getauscht. Nach Anwendung der Scope-Semantik (behavior="OUT" delimiter="OPEN" intersystem="FALSE") findet der Fluss aus dem Netzwerk-Container in den Intermediate-Container des Scopes statt (Zeile 15 in Algorithmus 15). Dieser Informationsfluss wird auch dem ProSP mitgeteilt und führt zu einer entsprechenden Aktualisierung der Provenance.

Der Scope ist nun auf Sender- und auf Empfängerseite geöffnet. Ein erster Informationsfluss hat stattgefunden. Weitere könnten im Rahmen des selben Scopes folgen. Allerdings ist dies für den einzelnen Download nicht notwendig.

Am Ende des zweiten Schrittes hebt der senderseitige PEP die Blockade auf dem Ereignis in der Abstraktionsschicht auf und der Informationsfluss findet tatsächlich statt.

Im dritten Schritt findet das duale Ereignis zum ausgehenden Informationsfluss statt: Der PEP der Empfängerseite (PEP B) fängt den eingehenden Informationsfluss ab (Abbildung 8.5). Das Ereignis wird zunächst an den PDP weitergegeben und dort ohne weitere Konsequenzen ausgewertet.

Im Anschluss erhält der PIP B die Mitteilung über das Ereignis (notifyEvent()) und interpretiert das Ereignis (onCreated) anhand der Informationsflusssemantik des Browsers (im Beispiel die des Chrome-Download-Managers, Listing D.3). Da der Scope bereits von der Senderseite geöffnet wurde (behavior="IN" delimiter="NONE"), findet der Informationsfluss aus dem Intermediate-Container des Scopes in den downloadItem-Container des Browsers statt (SF_FLOW). Entsprechend wird die Provenance im ProSP aktualisiert.

Im vierten und letzten Schritt erhält der PIP der Senderseite vom PEP die Information (Abbildung 8.6), dass der Download beendet wurde (downloadFile_end-Ereignis). Das Ereignis wird auch an den PIP B weitergegeben. Auf beiden Seiten wird gemäß der (Remote-)Informationsflusssemantik die Speicherfunktion des Netzwerk-Containers im Informationsflussmodell geleert (SF_CLEAR) und der zum Container gehörende Bezeichner entfernt (NF_RM_NAMING). Entsprechend wird die Repräsentation im Provenance-Modell auch auf Sender- und Empfängerseite als terminiert gekennzeichnet. Auf der Empfängerseite wird außerdem der Scope geschlossen (behavior="OUT" delimiter="CLOSE" ↪ intersystem="FALSE") . Daraus folgt eine Bereinigung des Intermediate-Containers im Informationsflussmodell. Die Repräsentationen des Intermediate-Containers werden nicht weiter benötigt und aus der Provenance entfernt.

Anstelle einer Schließung des Scopes von einem Kommunikationspartner aus, kann, sofern die Verbindung unkontrolliert abgebaut wurde, der Scope auch von jeder Seite einzeln geschlossen werden. In der Semantik müsste dafür jeweils ein gleichlautender „Intra-Close“ hinterlegt werden.

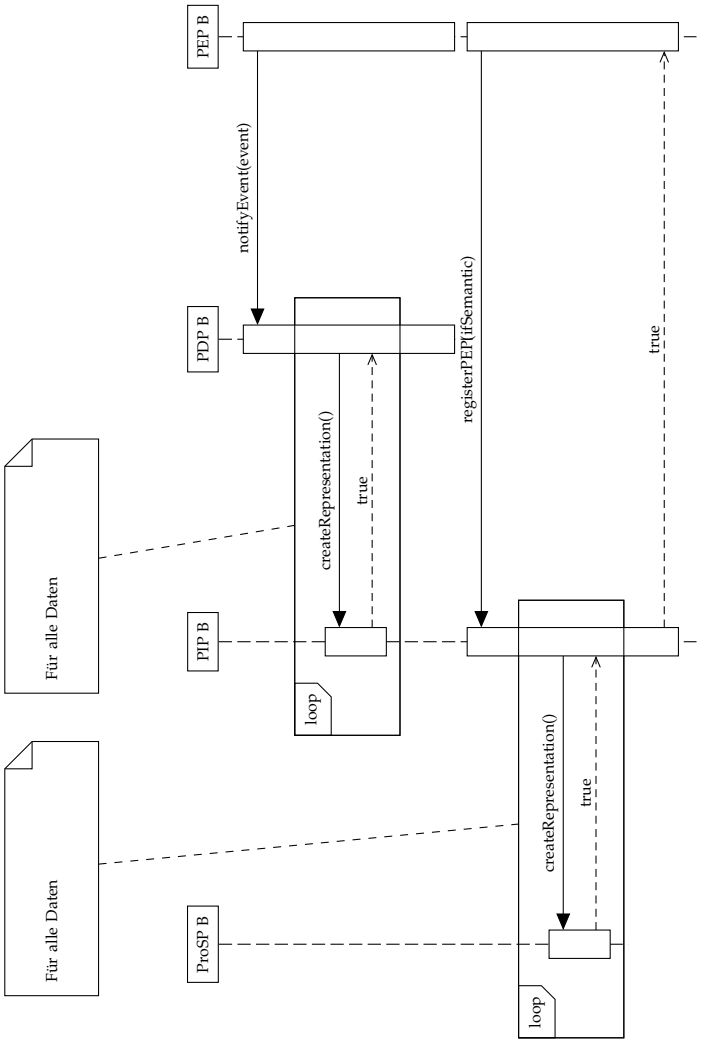


Abbildung 8.5: Systemübergreifende Informationsflüsse (2)

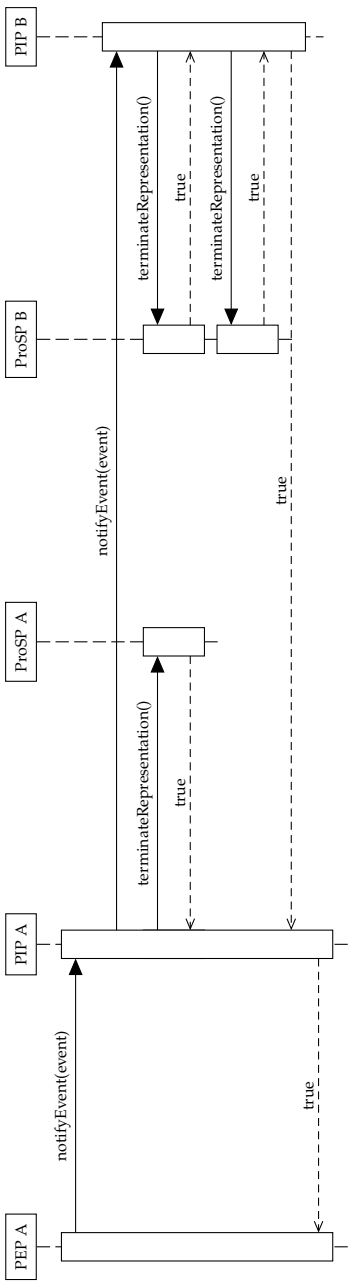


Abbildung 8.6: Systemübergreifende Informationsflüsse (3)

8.2 Aggregation verteilter Personal-Data-Provenance

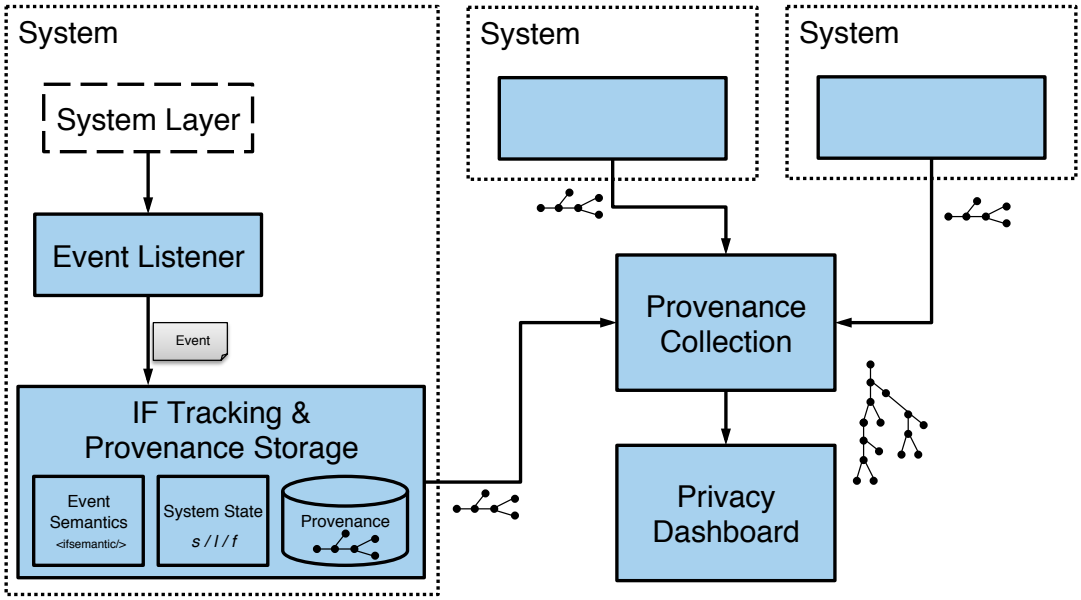


Abbildung 8.7: Aggregation verteilter Personal-Data-Provenance

Die Personal-Data-Provenance wird über alle Systeme verteilt gespeichert. Jeder ProSP hält in Übereinstimmung mit Anforderung 20 nur genau die Teile der Provenance vor, die sich auf Ereignisse der Domäne, die er verantwortet, beziehen. Ferner wird aus dem gleichen Grund bei systemübergreifenden Informationsflüssen, mit Ausnahme einer Repräsentation, keine Provenance ausgetauscht (Schritt 2 in Abschnitt 8.1.2).

Der ProCP, als für die Aggregation verantwortliche Komponente der Architektur, wird nur zum Erhebungszeitpunkt involviert (Anforderung 21). Er erhält Informationen zum Betroffenen und der Domäne der erhebenden Stelle (Kapitel 5.3.2). Dies ist auch der Einstiegspunkt der Aggregation: Über die Methode getProvenanceOfData() ruft der ProCP beim ProSP des erhebenden Systems den ersten Teil der Provenance, die Wurzel des Provenance-Baums, ab. Diesen Provenance-Abschnitt durchsucht der ProCP nach Communication-Repräsentationen. Anhand der in diesen Repräsentationen verzeichneten Remote-Domänen kann der ProCP die nächsten Abschnitte der Provenance über die Methode getProvenanceByLocalRoot() abrufen. So traversiert der ProCP den gesamten Provenance-Baum und rekonstruiert die vollständige Provenance eines Datums (Abbildung 8.7).

Die Provenance aller personenbezogenen Daten eines Betroffenen ergibt gemeinsam die Personal-Data-Provenance des Betroffenen. Diese wird zur grafischen Aufbereitung

gesammelt an den ProDP *PrivacyInsight* gegeben. Die konkreten personenbezogenen Daten, auf die sich die Provenance bezieht, sind noch nicht Teil der Provenance. Diese Daten werden von *PrivacyInsight* bei Bedarf einzeln von den speichernden Systemen abgerufen. Hat der Betroffene kein Interesse an den konkreten Daten, erhält auch der ProCP/ProDP keine Kenntnis von ihnen (Anforderung 21).

Die Personal-Data-Provenance wird nur so lange im ProCP/ProDP gespeichert, solange der Betroffene Zugriff auf die Auskunft nimmt (Anforderung 22). *PrivacyInsight* sieht dafür eine Logout-Möglichkeit und einen Timeout vor.

Die verteilte Speicherung der Personal-Data-Provenance in den einzelnen ProSPs führt zu einer reduzierten Verfügbarkeit der Provenance. Insbesondere mobile Systeme und Arbeitsplatzsysteme sind nicht zwingend zum Abfragezeitpunkt erreichbar. Dies widerspricht dem Erfordernis einer jederzeit vollständigen Datenschutzauskunft (Anforderung 55). Personal-Data-Provenance sollte zu diesem Zweck immer ausreichend redundant gespeichert werden (Anforderung 56). Ein erster Lösungsansatz, angelehnt an verteilte Hashtabellen, findet sich bei Stritzke.[18]

8.3 Zwischenfazit

Personal-Data-Provenance kann mit Hilfe der in Abschnitt 8.1 beschriebenen Informationsflusssemantik und Architektur systemübergreifend erhoben werden. Die Provenance wird in den ProSPs der einzelnen Domänen gespeichert und, entsprechend Konstruktionsziel $\mathfrak{K}4$, erst zum Abfragezeitpunkt entlang der Baumstruktur der Provenance eines personenbezogenen Datums aggregiert. Die vollständige Provenance wird dem Betroffenen auf der Auskunftsplattform *PrivacyInsight* zur Verfügung gestellt.

Das hier vorgestellte Konzept hat im Vergleich zum Ansatz von Kelbert[19] den Vorteil, dass systemübergreifende Informationsflüsse von beliebigen Anwendungsschichten aus gestaltet werden können. Außerdem dient das Konzept dem Erhalt der informationellen Gewaltenteilung. Allerdings führt die höhere Komplexität des Informationsaustauschs bei der Übertragung von personenbezogenen Daten potentiell zu Performanceeinbußen. Insbesondere beim ersten Kontakt zweier Systeme kann der Austausch der Informationsflusssemantik zu Verzögerungen führen.

[18] Stritzke 2014.
[19] Kelbert/Pretschner 2015.

Teil III

Abwägung zwischen Transparenz und Unverkettbarkeit

9 Eine Metrik für Unverkettbarkeit

Das Datenschutz-Schutzziel der Unverkettbarkeit soll die Bildung von Persönlichkeitsprofilen verhindern. Unverkettbarkeit steht damit im Konflikt zur Transparenz.[1] Dieses Kapitel führt eine Metrik ein, die die Auswirkungen eines Datenschutzauskunftssystems auf die Unverkettbarkeit sichtbar macht.[2]

Unverkettbarkeit wird sowohl in der juristischen als auch in der technischen Fachliteratur auf unterschiedlichste Art definiert, zum Teil nur grob umrissen. In Abschnitt 9.1.1 werden deshalb existierende Begriffsbestimmungen vorgestellt. Formale Modellierungsansätze werden in Abschnitt 9.1.2 diskutiert. Welche Aspekte für eine Unverkettbarkeitsdefiniton grundsätzlich erforderlich sind, wird in Abschnitt 9.2 ausgeführt, um in den Abschnitten 9.3 bis 9.7 zu einer allgemeinen Metrik für Unverkettbarkeit und ihrer Instanziierung in vier Varianten zu kommen. Im anschließenden Abschnitt 9.8 wird die für die Metrik erforderliche Modellierung des Angreiferwissens erläutert. Dabei sind insbesondere die Annahmen zum Hintergrundwissen des Angreifers von entscheidender Bedeutung. Eine Heuristik zur effizienten Berechnung der Metrik wird in Abschnitt 9.9 vorgestellt. Ihre praktische Anwendbarkeit wird demonstriert und in Abschnitt 9.10 diskutiert.

9.1 Existierende Begriffsbestimmungen und Modelle für Unverkettbarkeit

Der Begriff der Unverkettbarkeit wurde in der Literatur bereits vielfach aufgegriffen. Im Folgenden werden gängige Begriffsbestimmungen und Modelle vorgestellt. Ein besonderer Schwerpunkt wird auf die Bestimmung von Unverkettbarkeit über eine Metrik gelegt.

9.1.1 Existierende Begriffsbestimmungen für Unverkettbarkeit

Unverkettbarkeit als Zielvorstellung des Datenschutzes wird indirekt anhand der möglichen Methoden zur Zielerreichung, Zweckbindung, Zwecktrennung sowie der informationellen und organisatorischen Gewaltenteilung umrissen.

[1] Vgl. Kapitel 4.1.2

[2] Wesentliche Teile dieses Kapitels wurden bereits in Bier 2016 veröffentlicht.

Bis dato wurde eine präzisere Definition in unterschiedlichster Art und Weise versucht. Als Systemeigenschaft formuliert, erfüllt ein System dann Unverkettbarkeit, wenn personenbezogene Daten nicht über Domänengrenzen hinweg, die durch Zweck und Kontext vorgegeben sind, „zusammengeführt" werden können.[3] In der Definition nach ISO/IEC 15408-2:2008 ist Unverkettbarkeit die Eigenschaft, dass andere nicht in der Lange sind, festzustellen, ob zwei Operationen in einem System vom selben Benutzer verursacht wurden. Als konkreter, gewünschter Endzustand wird Unverkettbarkeit als die Unmöglichkeit des In-Bezug-Setzens personenbezogener Daten definiert.[4] Verkettung ist, darauf bezogen, das „Zusammenführen" personenbezogener Daten mittels identifizierender Merkmale des Betroffenen.[5]

Die genannten Definitionen sind zum einen einschränkend, weil sie beispielsweise die Methode des Zusammenführens oder die relevanten Entitäten vorgeben, zum anderen unklar in dem Sinne, dass beispielsweise nicht gesagt wird, wer der Verketter ist und was „Zusammenführen" eigentlich bedeutet. Festzustellen ist, dass es nicht eine Unverkettbarkeit, sondern eine Menge von Unverkettbarkeiten, abhängig von der betrachteten Verkettung, gibt.[6]

9.1.2 Bereits existierende absolute und relative Unverkettbarkeitsmodelle

Unverkettbarkeit kann absolut oder relativ zu einem vorhergehenden Zustand definiert werden. Absolut gesehen sind zwei oder mehrere Entitäten[7] aus Sicht eines Angreifers dann unverkettbar, wenn der Angreifer nicht feststellen kann, ob die Entitäten innerhalb des definierten Modells in einem bestimmten Verhältnis zueinander stehen oder nicht.[8] Daraus ergeben sich Modelle, die die Unverkettbarkeit von Entitäten danach bestimmen, ob ein Angreifer grundsätzlich die Möglichkeit hat, etwas über eine vorher festgelegte Verkettungsrelation zwischen diesen Entitäten zu erfahren („all-or-nothing"-Prinzip). Die Modelle werden unter dem Begriff der Relationenunterscheidbarkeit zusammengefasst.

Relative Unverkettbarkeit vergleicht die Unsicherheit eines Angreifers, welche Verkettungsrelation tatsächlich vorliegt, nach Interaktion mit einem System, mit der Unsicherheit, die bereits vor der Interaktion mit dem modellierten System bestand. Die darauf basierenden Metriken führen zu einem Wertekontinuum für den Grad der Unverkettbarkeit.

[3] Hansen/Jensen/Rost 2015.
[4] ULD/TU Dresden 2007, 19 f.
[5] ULD/TU Dresden 2007, 49.
[6] So in der Tendenz bereits Pfitzmann/Hansen 2010, 12 sowie Bedner/Ackermann, DuD 2010, 323 (324).
[7] Welche Entitäten betrachtet werden, variiert von Definition zu Definition.
[8] Pfitzmann/Hansen 2010, 12.

Absolute Aussagen zur Unverkettbarkeit: Relationenunterscheidbarkeit

Die Relationenunterscheidbarkeit ist ein formales Angreifermodell, mit Hilfe dessen evaluiert werden kann, ob ein Verfahren oder Protokoll die Unverkettbarkeit von Entitäten vollständig aufrechterhält oder nicht.

Das Modell von Bohli und Pashalidis definiert, inspiriert von einem Standardexperiment für die Definition sicherer kryptographischer Schemata, der IND-CPA (Ciphertext indistinguishability – Chosen-plaintext Attack), einen Chosen-Relation-Angriff (CRA).[9] Hat ein Angreifer keinen, nicht vernachlässigbaren, Vorteil (Unterscheidungsvorteil), hält das untersuchte Verfahren oder Protokoll die Unverkettbarkeit der Entitäten aufrecht. Die betrachteten Entitätsmengen sind Identifikatoren und Benutzer.

Ein ähnliches Modell, in dem der Angreifer versucht zwischen zwei Kommunikationsmatrizen zu unterscheiden, findet sich bei Hevia und Micciancio.[10] Es ist spezifisch für Kommunikationssysteme mit der Relation „Sender-Nachricht-Empfänger".

Die Ideen von Hughes und Shmatikov orientieren sich wie die von Hevia und Bohli am „all-or-nothing"-Ansatz.[11] Abweichend werden nicht nur zweistellige, sondern auch dreistellige Relationen zwischen sogenannten Agenten und Markern einbezogen.

In einer Situation, in der ein gewisser Wissenszuwachs unabdingbar ist und in Kauf genommen wird, wie bei Auskunftssystemen, ist ein „all-or-nothing"-Ansatz nicht hilfreich. Die Metrik würde jederzeit trivial messen, dass die Unverkettbarkeit nicht gewahrt bleibt. Deshalb untersucht diese Arbeit relative Ansätze im Sinne entropiebasierter Unverkettbarkeitsmetriken.

Relative Aussagen zur Unverkettbarkeit: Entropiebasierte Unverkettbarkeitsmetriken

In der Literatur hat sich die informationstheoretische Bestimmung relativer Unverkettbarkeit etabliert. Die gängigsten Unverkettbarkeitsmetriken, die einen kontinuierlichen Werteraum besitzen, basieren auf dem Verhältnis zwischen der Entropie des A-posteriori-Wissens des Angreifers und der maximalen Entropie. Die A-priori-Situation wird also als eine Situation ohne jedes Wissen festgelegt.

Die Idee, Anonymität informationstheoretisch zu beschreiben, wurde bereits von Serjantov und Danezis ins Spiel gebracht.[12] Sie überführen das klassische „Anonymity Set" auf ein nach den Wahrscheinlichkeiten der einzelnen Elemente der Menge gewichtetes Maß. Diaz et al. ergänzen die Normierung des Anonymitätsmaßes.[13] Die Arbeiten von Steinbrecher und Köpsell übertragen den informationstheoretischen Ansatz auf Unver-

[9] Bohli/Pashalidis 2011.
[10] Hevia/Micciancio 2008.
[11] Hughes/Shmatikov 2004.
[12] Serjantov/Danezis 2003.
[13] Diaz et al. 2003.

kettbarkeit.[14] Der Ansatz wird von Franz et al.[15] aufgegriffen und von Pashalidis[16] von Äquivalenzrelationen auf homogene Relationen verallgemeinert. Franz et al. sowie Pashalidis lassen bewusst offen, welche Entitäten gemeint sind, während zuvor noch explizit von Benutzern, Nachrichten oder Aktivitäten gesprochen wird.

Die Beschränkung auf Äquivalenzrelationen bzw. homogene Relationen und auf bestimmte Entitätsmengen sowie die Nichtberücksichtigung des tatsächlichen A-priori-Wissens des Angreifers sind die wesentlichen Nachteile der existierenden Ansätze.[17]

In den folgenden Abschnitten wird eine Unverkettbarkeitsmetrik vorgestellt, die die genannten Beschränkungen überwindet und dennoch zugleich konkreter ist, als die rein begrifflichen Definitionen. Sie berücksichtigt A-priori-Wissen und ist offen für beliebige, auch mehrstellige, Relationen auf beliebigen Entitätsmengen. Als geeignete Metrik für ein Datenschutzauskunftssystem wird sie mit vier Relationen auf drei Entitätsmengen instanziiert und operationalisiert.

9.2 Aspekte einer Unverkettbarkeitsdefinition

Aus einer Analyse der bereits existierenden Unverkettbarkeitsvorstellungen ergibt sich, dass die Festlegung der folgenden vier Aspekte erforderlich ist, um zu einer Unverkettbarkeitsdefinition zu kommen. Sie legen den Ausschnitt der Realität fest, in dem Unverkettbarkeit gemessen werden soll. Die vier Aspekte sind im Einzelnen:

- die betrachteten Entitäten ($\mathcal{E}$)
- ihre Beziehungen zueinander (Verkettungsrelationen R)
- der Verketter, in Analogie zur IT-Sicherheit auch als Angreifer $\mathscr{A}$ bezeichnet, aus dessen Perspektive die Unverkettbarkeit bestimmt wird
- die Annahmen zum System $\Sigma^{\mathscr{A}}$,[18] in dem die Verkettung stattfinden kann

Die vier Aspekte werden im Laufe dieses Kapitels mit Bezug auf ein Datenschutzauskunftssystem mit Inhalt gefüllt. Sie wurden bereits von Pfitzmann und Hansen in ihrer

[14] Steinbrecher/Köpsell 2003.
[15] Franz/Meyer/Pashalidis 2007.
[16] Pashalidis 2008.
[17] Die Privatsphäre wird nicht nur durch die Clusterung einzelner Merkmale gefährdet, sondern durch das Relationenprodukt unterschiedlicher Relationen auf unterschiedlichen Entitätsmengen. Eine Formalisierung des Relationenproduktes findet sich in Anhang E.1.1.
[18] Σ steht in diesem Kapitel nicht für einen Systemzustand, sondern für das System und sein Verhalten insgesamt.

Definition verwendet,[19] jedoch nicht formal beschrieben. Diese Arbeit erweitert insbesondere die Möglichkeiten zur Formulierung von Verkettungsrelationen und die Berücksichtigung von Hintergrundwissen im Angreifermodell.

Eine Verkettungsrelation R ist eine Teilmenge des kartesischen Produkts von $n \geq 2$ Teilmengen $E_1, \ldots, E_n \subseteq \mathcal{E}$ ($R \subseteq E_1 \times \cdots \times E_n$). Um zu einer aussagekräftigen Definition zu kommen, werden sowohl in der Literatur als auch in dieser Arbeit Teilmengen $E_1, \ldots, E_n$ gleichartiger Entitäten (Entitätsmengen der Entitätsklassen) gewählt. Entitäten $\varepsilon \in \mathcal{E}$ sind gleichartig, wenn sie bestimmte Eigenschaften gemeinsam haben.

Beispiel. *Zwei Entitätsklassen wurden in dieser Arbeit bereits vorgestellt: Die Betroffenen und ihre personenbezogenen Daten mit den dazugehörigen Entitätsmengen $\mathscr{B}$ und $\mathscr{D}$.*[20] *Der Personenbezug lässt sich durch die Verkettungsrelation $R^{<} \subseteq \mathscr{D} \times \mathscr{B}$ beschreiben.*[21] *$(d, b) \in R^{<}$ gilt, falls das Datum $d \in \mathscr{D}$ einen Personenbezug auf den Betroffenen $b \in \mathscr{B}$ besitzt.*

Einstellige Relationen (Prädikate) führen zu keinem Verkettungsausdruck und können deshalb nicht Teil einer Unverkettbarkeitsdefinition sein.

9.3 Eine allgemeine Metrik für Unverkettbarkeit

Eine allgemeine Metrik für Unverkettbarkeit lässt sich bereits definieren, ohne sich auf die Entitäten, die Verkettungsrelation, den Angreifer und die Annahmen zum System festzulegen. Diese Festlegung wird erst für die Instanziierung der Metrik in den nachfolgenden Abschnitten vorgenommen.

In den folgenden Abschnitten dieses Kapitels werden Wahrscheinlichkeiten gemäß des Bayesschen Wahrscheinlichkeitsbegriffes als „Grad (vernünftiger) Glaubwürdigkeit/persönlicher Überzeugung“ (*degree of belief*) verwendet.

Wie bereits angedeutet, ist *relative Unverkettbarkeit* der Vergleich der Unsicherheit des Angreifers $\mathscr{A}$ bezüglich der wahren Verkettungsrelation R_{T} nach Interaktion mit dem Gesamtsystem $\Sigma^{\mathscr{A}}$ mit der Unsicherheit, die bereits vor der Interaktion mit dem modellierten System bestand.[22] Die Unsicherheit vor Interaktion ist vom Hintergrundwissen (A-priori-Wissen) des Angreifers abhängig. Die Interaktion mit $\Sigma^{\mathscr{A}}$ lässt den Angreifer Beobachtungen machen (Beobachtungsereignis I). Das kombinierte Wissen des Angreifers

[19] Deren Definition lautet im Original „Unlinkability of two or more items of interest (IOIs, e.g., subjects, messages, actions, ...) from an attacker's perspective means that within the system (comprising these and possibly other items), the attacker cannot sufficiently distinguish whether these IOIs are related or not.“; Pfitzmann/Hansen 2010, 12, aktualisierte Fortschreibung von Pfitzmann/Köhntopp 2001.

[20] Vgl. Kapitel 6.1.

[21] Im Folgenden wird $R^{<}$ als die *Identifikationsrelation* bezeichnet.

[22] Die Verkettungsrelation R_{T} muss nirgendwo gespeichert sein. Sie ist gleichbedeutend mit den realen Beziehungen der Entitäten gemäß der Semantik der Verkettungsrelation.

aus Hintergrundwissen und Beobachtungen wird auch als A-posteriori-Wissen bezeichnet.

Sei X eine Zufallsvariable über der endlichen Menge der Kandidatenrelationen $\mathcal{R}$. Kandidatenrelationen sind alle Relationen, die zur Beschreibung der Beziehungen zwischen den betrachteten Entitäten potentiell in Frage kommen. Sowohl vor als auch nach Interaktion mit dem Gesamtsystem $\Sigma^{\mathscr{A}}$ weist der Angreifer $\mathscr{A}$ allen Kandidatenrelationen $R \in \mathcal{R}$ einen Wahrscheinlichkeitswert $\mathbb{P}(X = R)$ zu.[23] $\mathbb{P}(X = R)$ ist die angenommene Wahrscheinlichkeit, dass R die tatsächliche Relation $R_\top$ zwischen den Entitäten aus $E_1, \ldots, E_n$ ist.

Beispiel. *Angenommen es gibt nur die Betroffenen $b_1, b_2 \in \mathscr{B}$ (Alice und Peter) sowie das Datum $d_1 \in \mathscr{D}$ (Vorname), das zu einem der beiden Betroffenen gehört. Dann ist die Menge der Kandidatenrelationen $\mathcal{R} = \{\{(b_1, d_1)\}, \{(b_2, d_1)\}\}$ und die tatsächliche Relation lautet $R_\top = \{(b_1, d_1)\}$. Ein Angreifer ohne weiteres Hintergrundwissen würde den beiden Kandidatenrelationen a priori die Wahrscheinlichkeitswerte $\mathbb{P}(X = \{(b_1, d_1)\}) = \mathbb{P}(X = \{(b_2, d_1)\}) = 0{,}5$ zuweisen.*

Dann ergibt sich die Entropie des A-priori- bzw. A-posteriori-Wissens des Angreifers als:

$$H(X) = -\sum_{R \in \mathcal{R}} \mathbb{P}(X = R) \log_2 \mathbb{P}(X = R) \qquad [H] = bit$$

Wobei $\mathbb{P}(X = R) \log_2 \mathbb{P}(X = R) = 0$ für $\mathbb{P}(X = R) = 0$ angenommen wird. Die Entropie misst die Informationsmenge, die $\mathscr{A}$ noch braucht, um $R_\top$ vollständig zu identifizieren.

Der *Grad der Unverkettbarkeit* ($\Delta(X,I)$) ergibt sich als Verhältnis zwischen A-priori- und A-posteriori-Entropie (mit dem Beobachtungsereignis I):

$$\Delta(X,I) = \frac{H(X \mid I)}{H(X)}$$

Der Grad der Unverkettbarkeit beschreibt das Verhältnis zwischen der Situation nach und der Situation vor der Interaktion des Angreifers $\mathscr{A}$ mit dem Gesamtsystem $\Sigma^{\mathscr{A}}$ bezüglich des noch benötigten Wissens zur vollständigen Aufdeckung der Relation.

In bisherigen Arbeiten werden meist die A-priori-Situation und die maximale Unverkettbarkeit entsprechend der Maximum-Entropie-Methode gleichgesetzt (Maximum-Entropie-Prior, $H(X) = H_{\max}(X) = \log_2(|\mathcal{R}|)$). Bei allgemeinen Relationen bestimmt sich die Mächtigkeit der Menge der Kandidatenrelationen aus der Größe der Potenzmenge des kartesischen Produkts: $|\mathcal{R}| = |\mathcal{P}(E_1 \times \cdots \times E_n)|$.

Ein Maximum-Entropie-Prior macht im diskutierten Szenario indes keinen Sinn. Eine Welt, in der es ein Datenschutzauskunftssystem gibt, kann nicht hinter das zurücktreten, was der zu beschreibende Angreifer schon im Voraus weiß. Es ist einzig und

[23] Ähnlich bereits bei Pashalidis 2008.

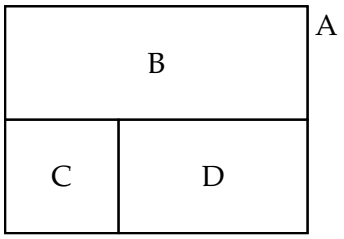


Abbildung 9.1: Verhältnis der unterschiedlichen Wissensstände

allein von Interesse, welchen negativen Einfluss der Einsatz des Datenschutzauskunftssystems auf die Unverkettbarkeit hat. Würde das Hintergrundwissen in der Berechnung der A-priori-Entropie nicht mitberücksichtigt, sondern ausschließlich in die A-posteriori-Entropie miteinbezogen, würde die entstehende, initial maximale, Entropie die tatsächliche Wirkung der Einführung eines Datenschutzauskunftssystems verzerren. Die hinzukommende Verkettbarkeit würde systematisch überschätzt. Als Konsequenz werden a priori die Beobachtungen aus Datenverarbeitungsvorgängen ohne den Einsatz eines Datenschutzauskunftssystems vorausgesetzt. A posteriori werden die Beobachtungen aus denselben Datenverarbeitungsvorgängen unter Berücksichtigung des Einsatzes eines Datenschutzauskunftssystems ins Angreiferwissen mit aufgenommen. Es handelt sich also um ein Gedankenexperiment und keine tatsächliche Vorher-Nachher-Betrachtung. Statt eines Maximum-Entropie-Priors ist der Vergleichszustand (subjektiver Prior) schon ein Zustand mit partiellem Wissen.

Abbildung 9.1 stellt die Situation anschaulich dar. Die Fläche A repräsentiert das gesamte Wissen, das zur vollständigen Aufdeckung der Unverkettbarkeitsrelation erforderlich ist. Die Fläche B ist das A-priori-Wissen des Angreifers. Die Fläche C ist das durch die Interaktion mit $\Sigma^{\not A}$ gelernte Wissen. Die Flächen B und C bilden gemeinsam das A-posteriori-Wissen. Die Fläche D ist das auch a posteriori unbekannte Wissen. Der Grad der Unverkettbarkeit ergibt sich aus dem Verhältnis zwischen D und D+C.

9.4 Anforderungen an Unverkettbarkeitsmetriken

Da Anonymität als Spezialfall der Unverkettbarkeit aufgefasst wird,[24] wurden die Anforderungen an gute Anonymisierungsmetriken von Andersson und Lundin[25] wie folgt für Unverkettbarkeitsmetriken verallgemeinert:

1. Eine Unverkettbarkeitsmetrik sollte ihre Analyse auf Wahrscheinlichkeiten stützen.

[24] Ausführungen dazu in Anhang E.1.2.
[25] C. Andersson/Lundin 2008.

2. Eine Unverkettbarkeitsmetrik muss wohldefinierte und intuitive Endpunkte für ihre Skala besitzen.[26]

3. Eine Unverkettbarkeitsmetrik muss so konstruiert sein, dass der Grad der Unverkettbarkeit um so höher ist, je näher die Verteilung der Kandidatenrelationen an der Gleichverteilung ist.

4. Die Elemente im Wertebereich der Metrik sollten wohldefiniert sein.

5. Der Wertebereich der Metrik sollte geordnet und nicht zu grob sein.

Die oben definierte Metrik erfüllt all diese Anforderungen mit Ausnahme der zweiten: Die Metrik hat zwar einen wohldefinierten unteren Endpunkt (0), kann aber nach oben beliebig groß werden. Denn sei $H(X) \neq H_{\max}(X)$ (kein Maximum-Entropie-Prior), dann ist bei Beobachtungen, die der A-priori-Annahme entgegengesetzt sind, hypothetisch ein $\Delta(X,I) > 1$ möglich. Korrigierte Fehleinschätzungen können für den Angreifer zu einer subjektiv höheren Entropie führen. Die Wahl eines realistischeren A-priori-Bezugspunktes wird also dadurch erkauft, dass die Skala von $\Delta(X,I)$ keinen wohldefinierten oberen Endpunkt besitzt. Allerdings ist als Maß der Unverkettbarkeit nicht der Grad der Unverkettbarkeit bezüglich eines bestimmten Angreifers von Interesse, sondern der niedrigste Grad der Unverkettbarkeit über alle Angreifer. Trivial lässt sich immer ein Angreifer konstruieren, der kein Hintergrundwissen hat ($H(X) = H_{\max}(X)$) und durch seine Beobachtungen nichts dazulernen kann ($H(X \mid I) = H(X)$). Dessen Grad der Unverkettbarkeit ist immer $\Delta(X,I) = 1$. Somit ist der *normierte globale Grad der Unverkettbarkeit* $\|\Delta\| = \min_{\mathscr{A}}(\{\Delta(X, I_{\mathscr{A}})\}) \in [0;1]$.[27]

Zuletzt zwei wichtige Hinweise zum Verständnis der Metrik: Ein Grad der Unverkettbarkeit von 1 bedeutet nicht, dass Unverkettbarkeit bewahrt wird, sondern einzig und allein, dass sich am Status quo nichts ändert. Waren dem Angreifer schon zuvor fast alle Relationen bekannt, sind sie es ihm auch weiterhin. Umgekehrt spielt der Nenner (Prior) in der Formel zum Grad der Unverkettbarkeit keine Rolle, wenn der Zähler (Posterior) 0 ist. Ein Grad der Unverkettbarkeit von 0 bedeutet vollständiges Wissen des Angreifers ohne etwas über den Wissenszuwachs auszusagen.

9.5 Provenance-Systemmodell

Zwei mögliche Entitätsmengen für die Instanziierung der Metrik im Kontext eines Datenschutzauskunftssystems wurden bereits in den Beispielen eingeführt: Die Betroffenen $\mathscr{B}$ und die personenbezogenen Daten $\mathscr{D}$. Die dritte Entitätsmenge sind die *Systeme* $\mathscr{S}$. Ein

[26] Beispielsweise 1 und n oder 0 und 1.

[27] Das Optimierungsproblem ist unter der Nebenbedingung zu lösen, dass $\mathscr{A}$ ein Angreifer aus der Menge aller Angreifer ist.

System $\mathscr{S}$ hat nichts mit dem angegriffenen Gesamtsystem $\Sigma^{\mathscr{A}}$ zu tun. $\mathscr{S}$ ist eine Entität in den noch zu bestimmenden Verkettungsrelationen. $\Sigma^{\mathscr{A}}$ bezeichnet in der Instanziierung alle datenverarbeitenden Verfahren sowie das Datenschutzauskunftssystem.

Konkret ist ein System eine Ansammlung technischer (z. B. ein Cluster) oder organisatorischer (z. B. eine Abteilung) Entitäten, denen ein gemeinsames Wissen unterstellt wird. Interne Datenflüsse und verarbeitete Daten sind allen beteiligten Entitäten bekannt. Sofern eine Ansammlung organisatorischer Entitäten gemeint ist, korrespondiert der Systembegriff mit dem organisatorischen Stellenbegriff. Entspricht ein System einem konkreten IT-System, so hat dies technische, aber keine konzeptionellen Gründe.

Im Szenario eines Datenschutzauskunftssystems versucht ein Angreifer unter anderem Informationen über die Beziehungen zwischen Systemen und personenbezogenen Daten zu bekommen. Diese Informationen sind Teil der Provenance. Die Differenzierung zwischen einzelnen Repräsentationen in einem System ist für die Formulierung der Verkettungsrelationen mit Bezug auf die in Abschnitt 9.6 berücksichtigten Angreifer nicht erforderlich. Aus diesem Grund wird das Provenance-Datenmodell aus Kapitel 6.3 auf ein Provenance-Systemmodell reduziert.

Ein Provenance-System-Graph $\mathscr{G}^{\Sigma} = (\mathscr{S}, \mathscr{L}^{\Sigma}, col)$ ist ein kantengefärbter, gerichteter Graph mit den Knoten (Systemen) $\mathscr{S}$, den Kanten $\mathscr{L}^{\Sigma} \subseteq \mathscr{S} \times \mathscr{S}$ und der Färbung $col : \mathscr{L}^{\Sigma} \to \mathcal{P}(\mathscr{D})$. Die Kanten verbinden Systeme abhängig von den Kanten zwischen den Repräsentationen im Provenance-Datenmodell. Die Kantenfärbung enthält die Information, welche Daten von einem System in ein anderes geflossen sind. Die genaue Ableitung des Provenance-System-Graphen aus dem Provenance-Graphen findet sich im Anhang E.3.

Beispiel. *Eine Betroffene wie Alice hat nach § 34 BDSG Anspruch auf Auskunft über die zu ihrer Person gespeicherten personenbezogenen Daten, deren Herkunft, Empfänger und den Zweck der Speicherung. Empfänger können der Betroffene, Dritte, Auftragsdatenverarbeiter und Stellen innerhalb der verantwortlichen Stelle sein. Der Stellenbegriff ist funktional und organisatorisch definiert (siehe Kapitel 3.7.2).*

Die Kunden Alice Fox (b_1) und Peter Trollig (b_2) sind gleichzeitig auch Systeme nach obiger Definition. Die CloudyCloud GmbH ist als Auftragsdatenverarbeiter für AdBokis tätig. Außerdem übermittelt AdBokis im Rahmen ihrer Geschäftsprozesse personenbezogene Daten an die PayPortal Inc. und die Bonus Card GmbH. Intern spielen bei der Datenverarbeitung die Abteilungen Kundenbetreuung, Vertrieb, IT und Infrastruktur und Recht eine Rolle. In der Abteilung Vertrieb wird neben dem System für den Onlineverkauf auch ein Archivserver betrieben. Zudem gibt es Arbeitsplatzsysteme, die im Vertrieb normalerweise nicht für die Verarbeitung personenbezogener Daten vorgesehen sind. Exemplarisch ist deshalb im Minimalbeispiel der Workspace23 enthalten. Alle diese Entitäten werden als Systeme $\varsigma \in \mathscr{S}$ bezeichnet. Tabelle 9.1 gibt einen vollständigen Überblick und weist der Übersicht halber jedem System eine Nummer zu.

Der vollständige Provenance-System-Graph für die Betroffene Alice ist in Abbildung 9.2 zu

ς_x	Bezeichnung
1	Alice Fox
2	Peter Trollig
3	CloudyCloud GmbH
4	PayPortal Inc.
5	Bonus Card GmbH
6	Kundenbetreuung
7	Vertrieb – Onlineverkauf
8	Vertrieb – Archiv
9	Vertrieb – Workspace23
10	IT und Infrastruktur
11	Rechtsabteilung

Tabelle 9.1: Systeme

sehen. Die Knoten sind die oben erwähnten Systeme. Die Kanten zeigen die Datenflüsse aller Daten aus Tabelle 6.1. Die jeweiligen IDs der Daten sind als Färbung auf den Kanten aufgetragen.

9.6 Der Angreifer 𝒜

Das Angreifermodell ist eine der Grundvoraussetzungen für die Bestimmung von Unverkettbarkeit. Es gibt die Leitlinien vor, an denen sich die Analyse der A-priori- und der A-posteriori-Situationen orientieren kann. Das Angreifermodell ist entsprechend der datenschutzrechtlichen Rahmenbedingungen auszugestalten.

Der Angreifer kann Teil der datenverarbeitenden Organisation (verantwortliche Stelle) sein oder außerhalb der Organisation zu finden sein. Mögliche externe Angreifer sind:

- unberechtigt Auskunftsersuchende
- Cyberkriminelle
- staatliche Stellen

Ein *unberechtigt Auskunftsersuchender* stellt eine Auskunftsanfrage ohne nach den datenschutzrechtlichen Vorschriften auskunftsberechtigt zu sein. Der unberechtigt Auskunftsersuchende gibt sich als ein anderer Betroffener aus oder meint fälschlicherweise selbst Betroffener zu sein. Sein Angriff besteht in der Verschleierung der eigenen Identität und Berechtigungen bei der Auskunftsanfrage.

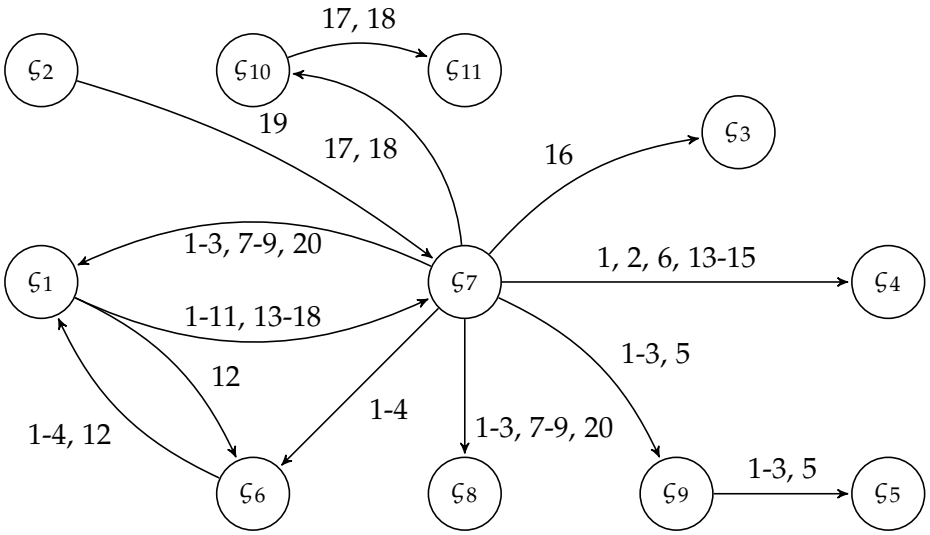


Abbildung 9.2: Tatsächliche Datenflüsse für Alice Fox (Daten als Kantenbeschriftung)

Annahme 1. *Bei jedem Auskunftsersuchen wird die Authentizität des Ersuchenden und dessen Autorisierung sicher überprüft. Auskünfte werden nur an authentische Betroffene und ausschließlich bezüglich deren personenbezogener Daten erteilt.*

Die Überprüfung der Authentizität eines Auskunftsersuchenden kann beispielsweise anhand eines gemeinsamen Geheimnisses erfolgen. Der unberechtigt Auskunftsersuchende kann unter obiger Annahme keine Beobachtungen machen.

Ein *Cyberkrimineller* versucht illegitim an gespeicherte personenbezogene Daten zu gelangen, um diese für Erpressung, Betrug oder beliebige andere Aktivitäten zu verwenden. Der Cyberkriminelle versucht aktiv, unter Ausnutzung der Schwachstellen von IT-Systemen, an die Daten zu gelangen.

Annahme 2. *Nur autorisierte Systeme haben im Rahmen legitimer Prozesse Zugriff auf die Provenance. Sie wird über Kommunikationskanäle weitergegeben, deren Sicherheit äquivalent zur Sicherheit der zugrundeliegenden personenbezogenen Daten ist.*

Aufgrund dieser Annahme kann ein Cyberkrimineller aus dem Datenschutzauskunftssystem nicht mehr Informationen ziehen, als wenn er die Organisation ohne existierendes Datenschutzauskunftssystem angreifen würde.

Eine Sonderrolle spielen *staatliche Stellen*. Sofern ihr Zugriffswunsch auf personenbezogene Daten legitim und verhältnismäßig ist, kann sich eine Organisation dem Abgriff von Informationen nicht entgegenstellen. Ist der Zugriffswunsch illegitim, können sie analog

zu Cyberkriminellen behandelt werden. Staatliche Stellen werden deshalb im Folgenden nicht gesondert betrachtet.

Neben den externen Angreifern gibt es auch eine Reihe möglicher interner Angreifer:

- Vorgesetzte, die die Produktivität ihrer Mitarbeiter überwachen wollen
- Organisationseinheiten (Systeme $\varsigma \in \mathscr{S}$) mit (wirtschaftlichem) Interesse an den Daten der Betroffenen (Kunden) – genannt *Systemangreifer* ($\mathscr{A}^{\varsigma}$)
- Betreiber der zentralen Infrastruktur für die Datenschutzauskunft ($\mathscr{A}^{c}$)

Der *Vorgesetzte* will das Datenschutzauskunftssystem einer Zweitverwendung zuführen, um seine eigenen Mitarbeiter zu kontrollieren. Da die Personal-Data-Provenance teilweise auch Zeitinformationen enthält, könnte sie hypothetisch zur Überwachung der Produktivität der Mitarbeiter verwendet werden. Der Mitarbeiterdatenschutz ist jedoch nicht im Fokus dieser Ausarbeitung und wird nachfolgend nicht berücksichtigt.

Der *Systemangreifer* ($\mathscr{A}^{\varsigma}$) verarbeitet möglicherweise selbst personenbezogene Daten. Er möchte aber Wissen über weitere Verarbeitungsvorgänge gewinnen. Vorstellbar ist beispielsweise eine Marketingabteilung, die wissen möchte, in welchem Maße und unter Preisgabe welcher Informationen ein Kunde bisher den Kundenservice angefragt hat.

Der *zentrale Angreifer* ($\mathscr{A}^{c}$) entsteht in seiner Sonderstellung erst durch das Datenschutzauskunftssystem. Er entspricht im Datenschutzauskunftssystem dem ProCP. Als Einstiegspunkt für den Abruf der gesamten Provenance-Kette erhält er die in Kapitel 8.2 beschriebenen Metadaten als zusätzliches Wissen aus dem Betrieb des Datenschutzauskunftssystems.

Die zentrale IT und die verantwortliche Stelle an sich sind als Angreifer auszunehmen und als modelltechnisch vertrauenswürdig anzunehmen. Eine Vertrauenswürdigkeit dieser mächtigen Instanzen muss durch organisatorische Maßnahmen sichergestellt werden. Wären sie nicht vertrauenswürdig, könnten sie im Zweifelsfall auf allen IT-Systemen eine Protokollinfrastruktur installieren, die dieselben Daten sammelt wie das Datenschutzauskunftssystem, nur ohne datenschutzrechtliche Vorgaben zu berücksichtigen. Kann der Organisation nicht vertraut werden, ergibt auch ein Datenschutzauskunftssystem wenig Sinn, da dann der Auskunft an sich nicht vertraut werden kann. Im weiteren Verlauf werden deshalb zwei Angreifertypen betrachtet: Der *Systemangreifer* $\mathscr{A}^{\varsigma}$ und der *zentrale Angreifer* $\mathscr{A}^{c}$.

$\mathscr{A}^{\varsigma}$ und $\mathscr{A}^{c}$ werden als passive Angreifer angenommen. Sie halten die festgelegten Kommunikationsprotokolle des Datenschutzauskunftssystems vollständig ein. Eine Missachtung der Kommunikationsprotokolle kann von den Kommunikationspartnern festgestellt und organisatorisch verfolgt werden. Beide Angreifer haben nur auf die bei ihnen gespeicherten Daten Zugriff. Der Systemangreifer hat Zugriff auf die von ihm erhobene Provenance, der zentrale Angreifer hat Zugriff auf die unabhängig von Auskunftsanfragen zentral gespeicherten Daten. Arbeiten mehrere Systeme $\varsigma \in \mathscr{S}$ zusammen, werden

sie gemeinsam als ein Systemangreifer $\mathscr{A}^{\cup\varsigma}$ mit gemeinsamem Hintergrundwissen und gemeinsamen Beobachtungen aufgefasst. Sie werden analog zu einem einzeln agierenden Systemangreifer behandelt.

Das Wissen der beiden betrachteten Angreifer und weitere Modellannahmen werden im nächsten Abschnitt beschrieben.

Annahmen zum Hintergrundwissen der Angreifer $\mathscr{A}^{\varsigma}$ und $\mathscr{A}^{c}$ sowie zu den Eigenschaften des modellierten Gesamtsystems $\Sigma^{\mathscr{A}}$ Die Verarbeitung personenbezogener Daten findet entlang etablierter Verarbeitungsprozesse statt. Diese sind von der verantwortlichen Stelle gemäß § 4g Abs. 2 S. 1 BDSG i.V.m § 4e Satz 1 BDSG in einem internen Verfahrensverzeichnis zu dokumentieren. Teil dieser Dokumentation sind die verarbeiteten Datenkategorien, eine Beschreibung der Verarbeitungsprozesse sowie die möglichen Empfänger der Daten. Aus letzterer Information ergibt sich die Möglichkeit der Verkettung einzelner Verarbeitungsprozesse.

Es ist jedoch möglich, dass Daten auch außerhalb der vorgesehenen Prozesse verwendet werden. Dies ist entweder der Fall, wenn es sich um einen Einzelfall handelt, oder wenn ein Verstoß gegen das datenschutzrechtliche Zweckbindungsgebot vorliegt.

Wichtigster Baustein des A-priori-Hintergrundwissens der Angreifer ist das interne Verfahrensverzeichnis. Neben dem internen Verfahrensverzeichnis sind dem Angreifer auch allgemeine Unternehmensstatistiken bekannt. Solche Statistiken enthalten Informationen zur Anzahl der Kunden und zur Menge der verarbeiteten Daten. Daraus ergeben sich die folgenden drei Annahmen:

Annahme 3. *Dem Angreifer ist die Art und die Anzahl aller Systeme $\varsigma \in \mathscr{S}$ a priori bekannt.*

Annahme 4. *Dem Angreifer ist die Anzahl der von der Datenverarbeitung betroffenen Kunden $|\mathscr{B}|$ a priori bekannt.*

Annahme 5. *Dem Angreifer ist die Anzahl der verarbeiteten personenbezogenen Daten $|\mathscr{D}|$ a priori bekannt.*

Über das Verfahrensverzeichnis hinaus ist für die Angreifer $\mathscr{A}^{\varsigma}$ und $\mathscr{A}^{c}$ die Anzahl der Systeme auch aus der Datenverarbeitung an sich zu schließen. Voraussetzung für die Datenverarbeitung der Angreifer $\mathscr{A}^{\varsigma}$ ist, mit allen datenverarbeitenden Systemen innerhalb der verantwortlichen Stelle kommunizieren zu können. Die zentrale Infrastruktur für die Datenschutzauskunft ($\mathscr{A}^{c}$) muss für ein Datenschutzauskunftssystem alle anderen Systeme erreichen können (Provenance-Collection) und muss von allen Systemen erreichbar sein. Die Anzahl der Systeme ließe sich also für beide Angreifertypen über eine Art „Netzwerkscan“ erschließen.

Die Angaben zur Anzahl der verarbeiteten personenbezogenen Daten ist im Allgemeinen nur als grobe Schätzung verfügbar. Sind die Zahlen groß genug, hat der genaue Wert jedoch keinen merklichen Einfluss auf die Bestimmung der Unverkettbarkeitsmetriken.

Ein Datum kann potentiell personenbezogenes Datum mehrerer Betroffener sein. Dies ist in der Provenance abbildbar und von den im nachfolgenden Abschnitt definierten Verkettungsrelationen erfassbar. Um die Erläuterungen in Abschnitt 9.8 zu vereinfachen, wird dennoch o. B. d. A. angenommen, dass ein Datum nur einen Personenbezug zu einem Betroffenen haben kann.

Annahme 6. *Das Verhältnis von personenbezogenen Daten und Betroffenen ist eine n:1-Beziehung.*

Die beiden letzten Annahmen 7 und 8, sind wichtige Annahmen zur Unabhängigkeit von Datenflüssen. Sie sind eine entscheidende Voraussetzung für die Berechenbarkeit der Unverkettbarkeitsmetriken. Beide Annahmen werden in der Realität nicht in jedem Fall vollständig eingehalten. Die durch sie induzierte Ungenauigkeit kann jedoch nur zu einem Unterschätzen des A-priori-Wissens des Angreifers führen. Das Delta zur A-posteriori-Situation wird dann größer. Die Gefährdung für den Datenschutz wird überschätzt. Deshalb sind die Annahmen vom Ergebnis her gedacht sinnvoller, als unbelegte Annahmen über die Abhängigkeit von Datenflüssen zu treffen, welche zu einem Unterschätzen des Datenschutzrisikos führen könnten.

Annahme 7. *Das A-priori-Wissen zu Datenflüssen (Verarbeitungsprozessen) ist nur von der Kategorie der Daten, nicht von den Daten selbst abhängig.*

Es ist für den Datenfluss beispielsweise egal aus welchen Ziffern die IBAN des Betroffenen besteht oder welchen Vornamen er hat. Eine Abhängigkeit kann in der Realität (s.o.) immer dann vorkommen, wenn ein Prozess mehrere Verarbeitungsalternativen vorgibt, deren Wahl abhängig vom Inhalt der Daten (z. B. Bonität, Alter, Lieferart) ist. Auf Datenkategorieebene kann man dies nur über die allgemeine Wahrscheinlichkeit der Alternativen fassen. Dies kann problematisch sein, falls angenommen wird, dass der Angreifer den Inhalt der Daten kennenlernen kann.

Annahme 8. *Die Flüsse zweier Daten sind stochastisch unabhängig voneinander.*

Die Wahrscheinlichkeit für ein Datum d für einen bestimmten Fluss ist gleich hoch, unabhängig davon, ob ein Datum d' entsprechend der Verfahren oder abweichend geflossen ist.

Was jedoch nicht abgedeckt wird, ist die Aneinanderreihung von Verfahren und die Berücksichtigung von Verarbeitungsalternativen. In diesen beiden Fällen ist es durchaus so, dass Daten einen gemeinsamen Weg nehmen. Fließen Daten immer gemeinsam, sollten die Daten verbunden als ein Datum behandelt werden. Solch ein zusammengefasstes Datum ist dann wieder unabhängig von den Datenflüssen anderer Daten.

Darüber hinaus sind auch Situationen denkbar, in denen Daten zwar nicht immer gemeinsam verarbeitet werden, jedoch von der Verarbeitung des einen Datums auf eine

höhere Wahrscheinlichkeit der Verarbeitung eines anderen Datums geschlossen werden kann. Deshalb bleibt die Unabhängigkeitsannahme letztendlich eine Vereinfachung, die dem Vorgehen vorzuziehen ist, mit Abhängigkeiten zu arbeiten, die nicht bekannt sind.

9.7 Instanziierung der Unverkettbarkeit als Gegenspielerin der Transparenz

Die relevanten Entitäten ergeben sich aus den Teilinformationen der Datenschutzauskunft. Es sind die Systeme $\varsigma \in \mathscr{S}$, die personenbezogenen Daten $d \in \mathscr{D}$ und die Betroffenen $b \in \mathscr{B}$. Gleiches gilt für die Verkettungsrelationen, die über diesen Entitäten definiert sind. Sie ergeben sich außerdem aus den technischen Anforderungen in Kapitel 4.4. Sie bilden das Interesse des Angreifers an den zu einem Betroffenen gespeicherten personenbezogenen Daten ($R^{\lessdot}$, Anforderung 45), an der Herkunft und den Empfängern personenbezogener Daten ($R^{\rhd}$, Anforderung 43) und am zweckbestimmten Verarbeitungsort personenbezogener Daten ($R^{\triangledown}$, Anforderung 42) ab. $R^{\equiv}$ ergibt sich aus dem Gebot der Zwecktrennung (Anforderung 44). Wird die Zwecktrennung überwunden, kann ein Persönlichkeitsprofil des Betroffenen hergestellt werden, unabhängig davon, ob schon klar ist, wer er ist. Alle Relationen stellen immer nur die Situation zu einem bestimmten Zeitpunkt dar. Werden Daten neu erhoben oder weitergegeben, ändern sich die Relationen. Die vier genannten Relationen sind wie folgt definiert:

- Die *Identifikationsrelation* $R^{\lessdot} \subseteq \mathscr{D} \times \mathscr{B}$ gibt an, ob das Datum $d \in \mathscr{D}$ einen Personenbezug auf den Betroffenen $b \in \mathscr{B}$ besitzt.

- Die *Verknüpfungsrelation* $R^{\equiv} \subseteq \mathscr{D} \times \mathscr{D}$ gibt an, ob zwei Daten d, $d' \in \mathscr{D}$ einen Personenbezug auf denselben (aber unbekannten) Betroffenen besitzen.

- Die *Speicher- und Verarbeitungsrelation* $R^{\triangledown} \subseteq \mathscr{S} \times \mathscr{D}$ gibt für alle Systeme $\varsigma \in \mathscr{S}$ an, ob sie das Datum $d \in D$ verarbeitet und/oder gespeichert haben.

- Die *Datenflussrelation* $R^{\rhd} \subseteq \mathscr{S} \times \mathscr{S} \times \mathscr{D}$ gibt für zwei Systeme ς, $\varsigma' \in \mathscr{S}$ an, ob sie für ein bestimmtes personenbezogenes Datum $d \in \mathscr{D}$ in einer direkten Vorgänger-Nachfolger-Beziehung stehen.

Damit ist die Definition der Unverkettbarkeit für ein Datenschutzauskunftssystem vollständig.

Definition 9.1. *Der Grad der Unverkettbarkeit ist die Unsicherheit eines Angreifers $\mathscr{A}^{\varsigma}$ oder $\mathscr{A}^{c}$ über die wahre Verkettungsrelation in $\mathcal{R}^{\lessdot}$, $\mathcal{R}^{\equiv}$, $\mathcal{R}^{\triangledown}$ oder $\mathcal{R}^{\triangleright}$ im Gesamtsystem $\Sigma^{\mathscr{A}}$, nachdem das Beobachtungsereignis I eingetreten ist, im Vergleich zur Unsicherheit mit seinem vorherigen Wissensstand.*

Bei manchen Unverkettbarkeitsmetriken ist es möglich, die Verkettungsrelation auf unterschiedlichen Entitätsteilmengen zu bestimmen. $\Delta(X^{\lessdot}, I)$ und $\Delta(X^{\equiv}, I)$ sind globale Metriken. Bei der Bestimmung des Grads der Unverkettbarkeit sind alle Betroffenen $\mathscr{B}$ und alle personenbezogenen Daten $\mathscr{D}$ miteinzubeziehen.

Anders stellt sich die Situation bei $\mathcal{R}^{\triangleright}$ und $\mathcal{R}^{\triangledown}$ dar. Der Grad der Unverkettbarkeit bezüglich dieser Mengen ist global und lokal bestimmbar. Lokal meint die Fokussierung auf bestimmte Systeme $\varsigma \in \mathscr{S}$ oder Betroffene $b \in \mathscr{B}$. Im Kontext der Datenschutzauskunft ist für einen Betroffenen nur relevant, wie sich die Unverkettbarkeit der Flüsse seiner personenbezogenen Daten entwickelt, um zu entscheiden, ob er für oder gegen technische Transparenzmaßnahmen plädiert. Deshalb werden im Abschnitt 9.8.2 nur die Daten in der Teilmenge $\mathscr{D}_b \subseteq \mathscr{D}$, im Beispiel die personenbezogenen Daten von Alice $\mathscr{D}_{Alice} \subseteq \mathscr{D}$, betrachtet. Im Text wird dennoch im Sinne einer allgemeingültigen Darstellung von $\mathscr{D}$ gesprochen. Gleichzeitig wird im Abschnitt 9.8.2 angenommen, dass dem Angreifer a priori bekannt ist, welche Identifikatoren von personenbezogenen Daten (aber nicht welcher Kategorie) zu welchem Betroffenen gehören. Die Unsicherheit über dieses Faktum wird bereits durch den Grad der Unverkettbarkeit von $\mathcal{R}^{\lessdot}$ gemessen.

Beispiel. *Für die Bestimmung des Grads der Unverkettbarkeit wird das A-priori-Wissen der Angreifer in Bezug zum sich aus den tatsächlichen Datenflüssen ergebenden A-posteriori-Wissen gesetzt. In den Beispielen des Folgeabschnitts werden die tatsächlichen Datenflüsse von Alice verwendet. Alice werden in der Auskunft die in Abbildung 9.2 bereits dargestellten tatsächlichen Datenflüsse mitgeteilt.*

9.8 Bestimmung des Grads der Unverkettbarkeit unter Berücksichtigung des A-priori- und A-posteriori-Wissens der Angreifer

Um den Grad der Unverkettbarkeit bezüglich der vier genannten Relationen zu bestimmen, ist es erforderlich, das Hintergrundwissen der Angreifer und den Wissenszuwachs durch die Einführung der Datenschutzauskunft messbar zu machen. Das Hintergrundwissen der Angreifer geht in die A-priori-Wahrscheinlichkeiten $\mathbb{P}(X = R)$ mit ein. Der Wissenszuwachs der Angreifer wird durch das Beobachtungsereignis I und die daraus resultierenden A-posteriori-Wahrscheinlichkeiten $\mathbb{P}(X = R \mid I)$ erklärt. A-priori- und A-posteriori-Wahrscheinlichkeitsverteilungen aller vier Relationen werden in diesem Abschnitt erläutert und anhand des Minimalbeispiels aus Kapitel 1.6 bestimmt.

9.8.1 Bestimmung der Wahrscheinlichkeitsverteilungen von $X^{\leqslant}$ und $X^{\equiv}$

Für die beiden Relationen $R^{\leqslant}$ und $R^{\equiv}$ gilt, dass es, mit Ausnahme der Mächtigkeit der Mengen $\mathscr{B}$ und $\mathscr{D}$, kein globales Hintergrundwissen gibt.

Ein Systemangreifer $\mathscr{A}^{\varsigma}$ kann möglicherweise über den Inhalt der personenbezogenen Daten, die er selbst verarbeitet, auf deren Personenbezug schließen (Einfluss auf die Wahrscheinlichkeitsverteilung von $X^{\leqslant}$ und $X^{\equiv}$). Zudem kann die gemeinsame Verarbeitung von Daten unterschiedlicher Datenkategorien für solch einen Angreifer ein Indiz sein, dass zwei Daten zum gleichen Betroffenen gehören (Einfluss auf die Wahrscheinlichkeitsverteilung von $X^{\equiv}$). Beides ist sehr spezifisch für den jeweiligen Angreifer. Letztendlich ist die Bestimmbarkeit des Unverkettbarkeitspriors jedoch nur dann relevant, wenn er sich potentiell vom Posterior unterscheiden kann. Dies ist für Systemangreifer nicht der Fall. Durch das vorgestellte Datenschutzauskunftssystem werden auf den Systemen nur solche Teile der Provenance vorgehalten, die auf Ereignisse im jeweilige System zurückzuführen sind. Daraus folgt für solche Angreifer $\Delta(X^{\leqslant}, I) = \Delta(X^{\equiv}, I) = 1$.

Der zentrale Angreifer $\mathscr{A}^{c}$ verarbeitet selbst keine personenbezogenen Daten. Für ihn sind $H(X^{\leqslant}) = H_{\max}(X^{\leqslant}) = \log_2 |\mathcal{R}^{\leqslant}|$ und $H(X^{\equiv}) = H_{\max}(X^{\equiv}) = \log_2 |\mathcal{R}^{\equiv}|$. Die A-Priori-Entropie des Angreifers hängt damit einzig von der Mächtigkeit der Menge der beiden Kandidatenrelationen ab.

Die Mächtigkeit der Menge der Kandidatenrelationen ist unter Berücksichtigung von Annahme 6 für die Identifikationsrelation durch $|\mathcal{R}^{\leqslant}| = |\mathscr{B}|^{|\mathscr{D}|}$ gegeben.

Beispiel. *Die Anzahl der Kandidatenrelationen für* $|\mathscr{D}| = 30$ *und* $|\mathscr{B}| = 2$ *ist* $2^{30} = 1\,073\,741\,824$.

$R^{\equiv}$ ist die über den Mittler[28] $\mathscr{B}$ aus $R^{\leqslant}$ abgeleitete Äquivalenzrelation.[29] Bei Äquivalenzrelation ist die Anzahl möglicher Relationen durch die Bellsche Zahl $\mathcal{B}_{|\mathscr{D}|}$,[30] die Anzahl der Partitionen (Zerlegungen in disjunkte nichtleere Teilmengen) der zugrundeliegenden Menge $\mathscr{D}$, gegeben. Die Bellsche Zahl lässt sich über die Stirling-Zahl zweiter Art bestimmen:

$$\mathcal{B}_n = \sum_{k=0}^{n} \mathcal{S}_{n,k}$$
$$\text{mit } \mathcal{S}_{n,k} = \mathcal{S}_{n-1,k-1} + k \cdot \mathcal{S}_{n-1,k}$$
$$\text{sowie } \mathcal{S}_{n,n} = 1 \text{ und } \mathcal{S}_{n,k} = 0 \text{ für } k = 0 < n \vee n < k.$$

[28] Auf die Elemente eines Mittlers werden in der ursprünglichen Relation die Äquivalenzklassen der zweiten Entitätsmenge abgebildet. Genaueres findet sich in Anhang E.1.1.

[29] Herleitung und Beweis finden sich in Anhang E.1.3. Aufgrund von Annahme 6 ist $R^{\leqslant}$ rechtseindeutig. Da Betroffene, für die keine personenbezogene Daten verarbeitet werden, auf NIL abgebildet werden können, ist $R^{\leqslant}$ auch linkstotal.

[30] Bell 1934.

Da $\mathcal{R}^{\equiv}$ auf $\mathcal{R}^{\leqslant}$ zurückzuführen ist, sind die tatsächlichen Kandidatenrelationen durch die Anzahl der Betroffenen $|\mathscr{B}|$ beschränkt. Nur k-Partitionen mit $k \leq |\mathscr{B}|$ sind möglich. Obige Formel ist deshalb in korrigierter Form anzuwenden:

$$|\mathcal{R}^{\equiv}| = \sum_{k=0}^{\min(|\mathscr{B}|,|\mathscr{D}|)} \mathcal{S}_{|\mathscr{D}|,k}$$

Beispiel. *Die Anzahl der Kandidatenrelationen für* $|\mathscr{D}| = 30$ *und* $|\mathscr{B}| = 2$ *ist* $|\mathcal{R}^{\equiv}| = \sum_{k=0}^{2} \mathcal{S}_{30,k} = 536\,870\,912$. *In diesem Spezialfall mit zwei Betroffenen ist* $|\mathcal{R}^{\equiv}| = \frac{1}{2} \cdot |\mathcal{R}^{\leqslant}|$, *da* $\mathcal{S}_{n,2} = 2^{n-1} - 1$.

A posteriori, also unter Einsatz des Datenschutzauskunftssystems, erhält der zentrale Angreifer weitere Informationen I. Bei jeder Erhebung eines personenbezogenen Datums wird ihm ein pseudonymer Identifikator für das erhobene Datum zusammen mit Informationen zum Betroffenen übermittelt. Daraus kann der Angreifer nicht schließen, welches personenbezogene Datum oder welche Kategorie personenbezogener Daten erhoben wurde. Allerdings kann er bestimmen, wie viele personenbezogenen Daten für jeden einzelnen Betroffenen erhoben wurden. Kandidatenrelationen, die keine entsprechende Struktur aufweisen, kann er ausschließen.

Seien die dem Angreifer bekannt gewordenen k-Partitionen für die Menge der Daten $\mathscr{D}$ von der Größe $j_1, j_2, \ldots, j_k$. Dann sind

$$|(R^{\equiv} \in \mathcal{R}^{\equiv} \mid \mathbb{P}(X^{\equiv} = R^{\equiv} \mid I) \neq 0)| = \binom{|\mathscr{D}|}{j_1}\binom{|\mathscr{D}| - j_1}{j_2} \cdots \binom{|\mathscr{D}| - (j_1 + j_2 + \cdots + j_{k-1})}{j_k}$$

und

$$|(R^{\leqslant} \in \mathcal{R}^{\leqslant} \mid \mathbb{P}(X^{\leqslant} = R^{\leqslant} \mid I) \neq 0)| = k! \cdot |(R^{\equiv} \in \mathcal{R}^{\equiv} \mid \mathbb{P}(X^{\equiv} = R^{\equiv} \mid I) \neq 0)|.$$

Unter Weiterbestehen der Gleichverteilungsannahme ergibt sich die A-posteriori-Entropie direkt aus der Mächtigkeit der obigen beiden Mengen.

Beispiel. *Für Alice wurden 20 personenbezogene Daten erhoben, für Peter 10. Damit sinkt die Anzahl der möglichen Relationen* $\mathcal{R}^{\equiv}$ *auf* $|(R^{\equiv} \in \mathcal{R}^{\equiv} \mid \mathbb{P}(X^{\equiv} = R^{\equiv} \mid I) \neq 0)| = \binom{30}{20} = \binom{30}{10} = 30\,045\,015$. *Folglich ist der resultierende Grad der Unverkettbarkeit* $\Delta(X^{\equiv}, I) = \frac{\log_2 30\,045\,015}{\log_2 2^{29}} \approx \frac{24{,}8406}{29} \approx 0{,}8566$.
Die Anzahl der verbleibenden Identifikationsrelation ist $|(R^{\leqslant} \in \mathcal{R}^{\leqslant} \mid \mathbb{P}(X^{\leqslant} = R^{\leqslant} \mid I) \neq 0)| = 2! \cdot 30\,045\,015 = 60\,090\,030$. *Entsprechend ist der resultierende Grad der Unverkettbarkeit* $\Delta(X^{\leqslant}, I) = \frac{\log_2 60\,090\,030}{\log_2 2^{30}} \approx \frac{25{,}8406}{30} \approx 0{,}8614$.

9.8.2 Bestimmung der Wahrscheinlichkeitsverteilungen von $X^{\triangleright}$ und $X^{\triangledown}$

Unter der Annahme, dass die Weitergaben unterschiedlicher personenbezogener Daten voneinander unabhängig sind (Annahme 8), kann die Wahrscheinlichkeit $\mathbb{P}(X^{\triangleright} = R^{\triangleright})$ für eine Kandidatenrelation $R^{\triangleright} \in \mathcal{R}^{\triangleright}$ aus den Wahrscheinlichkeiten für die Teilrelationen je Datum $\mathbb{P}(X_d^{\triangleright} = R_d^{\triangleright})$ mit $R_d^{\triangleright} \subseteq \mathscr{S} \times \mathscr{S}$ berechnet werden. Es gilt $\mathbb{P}(X^{\triangleright} = R^{\triangleright}) = \prod_{d \in \mathscr{D}} \mathbb{P}(X_d^{\triangleright} = R_d^{\triangleright})$.

Das Wissen des Angreifers setzt sich aus drei Teilen zusammen: Dem Wissen über die Kategorie eines personenbezogenen Datums, dem Wissen über die Herkunft eines personenbezogenen Datums und dem Wissen über die Verarbeitungsverfahren.

Datenkategorie Das Wissen eines Angreifers wird zunächst dadurch charakterisiert, inwiefern ihm die Kategorie des personenbezogenen Datums bekannt ist. Die Kategorie des Datums bestimmt dessen Herkunft und Verarbeitung gemäß Verfahrensverzeichnis. Jedem Datum ist seine Datenkategorie über die Funktion $\vartheta : \mathscr{D} \to \Theta$ zugewiesen. Die Wahrscheinlichkeitsverteilung $\mathbb{P}(X_\theta = \theta)$ gibt die Wahrscheinlichkeit der Datenkategorien $\theta \in \Theta$ für ein gegebenes Datum an. Ist dem Angreifer das Datum inhaltlich bekannt, ist ihm zwangsläufig auch dessen Kategorie bekannt.

$$\mathbb{P}(X_d^{\triangleright} = R_d^{\triangleright}) = \sum_{\theta \in \Theta} \mathbb{P}(X_d^{\triangleright} = R_d^{\triangleright} \mid X_\theta = \theta) \mathbb{P}(X_\theta = \theta)$$

Beispiel. *Die Zuordnung zwischen Daten und Datenkategorien ist einem Systemangreifer für diejenigen Daten bekannt, die er selbst verarbeitet. So ist $\mathscr{A}_{\varsigma_3}$ bekannt, dass $\vartheta(d_{16}) = \theta_{14}$ (Profilbild) ist. Für alle anderen Daten sowie grundsätzlich für den zentralen Angreifer muss entsprechend der Maximum-Entropie-Methode die Gleichverteilung $\mathbb{P}(X_\theta = \theta_i) = \frac{1}{|\Theta|}$ für $i \in \{1, \ldots, |\Theta|\}$ angenommen werden.*

Herkunft Das Wissen eines Angreifers wird außerdem dadurch charakterisiert, inwiefern ihm die Herkunft der personenbezogenen Daten bekannt ist. Die Herkunft der personenbezogenen Daten ist von der Kategorie der Daten abhängig. Die Wahrscheinlichkeiten der Zufallsvariable X_ς für bestimmte Startsysteme $\varsigma \in \mathscr{S}$, abhängig von der Datenkategorie, sind $\mathbb{P}(X_\varsigma = \varsigma \mid X_\theta = \theta)$. Häufig ist die Herkunft von personenbezogenen Daten einer Kategorie eindeutig, in den meisten Fällen ist es der Betroffene selbst.[31] Ist das Herkunftssystem $\varsigma_1 \in \mathscr{S}$, dann ist die Wahrscheinlichkeit für die Herkunft des Datums aus dem jeweiligen System $\mathbb{P}(X_\varsigma = \varsigma_1 \mid X_\theta = \theta) = 1$ sowie $\mathbb{P}(X_\varsigma = \varsigma_i \mid X_\theta = \theta) = 0$ für $i \in \{2, \ldots, |\mathscr{S}|\}$. Kommen mehrere Systeme in Frage, dann ist die Gesamtwahrscheinlichkeit

$$\mathbb{P}(X_d^{\triangleright} = R_d^{\triangleright} \mid X_\theta = \theta) = \sum_{\varsigma \in \mathscr{S}} \mathbb{P}(X_d^{\triangleright} = R_d^{\triangleright} \mid X_\varsigma = \varsigma, X_\theta = \theta) \mathbb{P}(X_\varsigma = \varsigma \mid X_\theta = \theta).$$

[31] O. B. d. A. ist das System, das den Betroffenen repräsentiert, das System $\varsigma_1 \in \mathscr{S}$.

Beispiel. *Alle personenbezogenen Daten, bis auf jene der Kategorien Rechnung und Empfehlung, werden ausschließlich beim Betroffenen selbst erhoben. Für diese ist das Herkunftssystem ς_1 bekannt.*
Eine Rechnung wird immer durch den Rechnungsgenerator im Onlineshopsystem des Vertriebs erzeugt. Dies ist den Angreifern ebenfalls aus dem Verfahrensverzeichnis bekannt. Das Herkunftssystem für die Rechnung ist damit aus Sicht der Angreifer eindeutig ς_7.
Über die Quelle einer Empfehlung (θ_{16}) ist hingegen nur bekannt, dass sie von einem Kunden kommen muss. Da über das Verhältnis der Kunden untereinander nichts im Verfahrensverzeichnis enthalten ist, ist die A-priori-Wahrscheinlichkeit der Angreifer für das Herkunftssystem über alle vorhanden Kunden gleichverteilt. Im Beispiel wird der Einfachheit halber nur von zwei Kunden ausgegangen. Dann ergeben sich die Wahrscheinlichkeiten

$$\mathbb{P}(X_\varsigma = \varsigma_i \mid X_\theta = \theta_{16}) = \begin{cases} 0{,}5 & \textit{für } i = 1 \vee i = 2 \\ 0 & \textit{sonst.} \end{cases}$$

Einschub Jede zweistellige[32] Relation R über endlichen Mengen kann als binäre bzw. boolsche Matrix $R^{\boxdot} = (r_{ij})$, $r_{ij} \in \{0,1\}$ dargestellt werden. Die Einträge der Matrix r_{ij} stehen für die Realisationen der wie folgt definierten Zufallsvariablen X_{ij}:

$$X_{ij} : \mathcal{R} \to \{0,1\}$$
$$R \mapsto \begin{cases} 1 & \textit{für } (\varepsilon_i, \varepsilon_j) \in R \\ 0 & \textit{sonst.} \end{cases}$$

Der Eintrag bzw. das Ereignis 1 bedeutet, dass die durch den Index gegebenen Elemente in Relation zueinander stehen, der Eintrag 0, dass keine Beziehung vorliegt. Im Folgenden wird deshalb zur Vereinfachung nicht zwischen der Relation R und ihrer Matrixdarstellung $R^{\boxdot}$ differenziert.

$$r_{ij} = 1 \Leftrightarrow X_{ij} = 1 \Leftrightarrow (\varepsilon_i, \varepsilon_j) \in R$$

Daraus abgeleitet wird folgende Kurzschreibweise verwendet:

$$R_{r_{i_1,j_1}, r_{i_2,j_2}, \ldots, r_{i_k,j_k}} :\Leftrightarrow R = \{(\varepsilon_{i_1}, \varepsilon_{j_1}), (\varepsilon_{i_2}, \varepsilon_{j_2}), \ldots, (\varepsilon_{i_k}, \varepsilon_{j_k})\}$$

Beispiel. *Ein Datum $d \in \mathscr{D}$ wird nur durch den Betroffenen selbst gespeichert.*

$$R^{\triangleright}_{d, r^{\triangleright}_{d11}} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix}$$

[32] Gilt grundsätzlich auch für mehrstellige Relationen.

Verarbeitungsverfahren Bezüglich der konkreten Datenflüsse stützen sich die Angreifer auf die Angaben des Verfahrensverzeichnisses. Dieses hinterlegt für alle Daten die vorgesehenen Verarbeitungsprozesse. Das Wissen der Angreifer wird als Matrix der bedingten Flusswahrscheinlichkeiten W modelliert. Der Eintrag w_{ij} gibt die Wahrscheinlichkeit an, mit der ein Fluss von ς_i nach ς_j, angenommen wird, unter der Bedingung, dass das Datum bereits in ς_i verarbeitet wurde:

$$W_{\theta,\varsigma} : \{1,\ldots,m\} \times \{1,\ldots,m\} \to [0,1] \\ (i,j) \mapsto w_{ij} = \mathbb{P}(X^{\triangleright}_{dij} = 1 \mid X^{\triangleright}_{dii} = 1, X_\varsigma = \varsigma, X_\theta = \theta)$$

mit $m = |\mathscr{S}|$. Implizit ist $w_{ii} = 1$. Der reflexive Fluss, gleichbedeutend mit der Speicherung und Verarbeitung im System ($\forall d,i : \mathbb{P}(r^{\triangleright}_{dii}) = \mathbb{P}(r^{\triangledown}_{di})$), ist vollständig durch die eingehenden Flüsse erklärt:

$$\mathbb{P}(X^{\triangleright}_{dii} = 1 \mid \exists j \in \{1,\cdots,i-1,i+1,\cdots,m\} : X^{\triangleright}_{dji} = 1, X_\varsigma = \varsigma, X_\theta = \theta) = 1$$
$$\mathbb{P}(X^{\triangleright}_{dii} = 1 \mid \forall j \in \{1,\cdots,i-1,i+1,\cdots,m\} : X^{\triangleright}_{dji} = 0, X_\varsigma = \varsigma, X_\theta = \theta) = 0$$

Die inverse Matrix $\overline{W}_{\theta,\varsigma} = \mathbf{1} - W_{\theta,\varsigma}$ mit $\mathbf{1} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$ enthält die jeweiligen Gegenwahrscheinlichkeiten $\overline{w}_{ij} = \mathbb{P}(X^{\triangleright}_{dij} = 0 \mid X^{\triangleright}_{dii} = 1, X_\varsigma = \varsigma, X_\theta = \theta)$. Außerdem gilt für alle Datenflussrelationen $\mathbb{P}(X^{\triangleright}_{dij} = 0 \mid X^{\triangleright}_{dii} = 0, X_\varsigma = \varsigma, X_\theta = \theta) = 1$ und $\mathbb{P}(X^{\triangleright}_{dij} = 1 \mid X^{\triangleright}_{dii} = 0, X_\varsigma = \varsigma, X_\theta = \theta) = 0$. Es kann keine ausgehenden Flüsse geben, falls es keinen eingehenden Fluss gibt. Damit ist der Wahrscheinlichkeitsbaum für die Datenflussrelation vollständig erklärt.

Die in den bedingten Flusswahrscheinlichkeiten zum Ausdruck kommenden Pfade ergeben sich aus dem Verfahrensverzeichnis. Bezüglich der Abläufe der Verfahren wird davon ausgegangen, dass ein Angreifer auf grobes Erfahrungswissen zurückgreifen kann, das sich in zwei zentralen Parametern ausdrücken lässt.

Zunächst hängt die Wahrscheinlichkeit von Flüssen in linearen Verfahren maßgeblich von der *Fortschrittsquote* $\varpi \in (0{,}5;1]$ eines Prozesses ab (Wie viele Daten erreichen anteilig den nächsten Prozessschritt?). Dieser Parameter wird überall dort in der Flussmatrix eingesetzt, wo ein Fluss einem Prozessschritt entspricht. Die Fortschrittsquote sollte immer größer als 0,5 sein, da ein Prozessschritt, der nicht im Regelfall stattfindet, kein etabliertes Verfahren sein kann. Verzweigungen in einem Verfahren sind als Sonderfall davon ausgenommen. In sich verzweigenden Prozessen ist punktuell der Anteil der Daten zu berücksichtigen, die auf die eine oder die andere Weise weiterverarbeitet werden.

Unvorhergesehene Abweichungen vom Verfahren werden durch eine *Fehlerwahrscheinlichkeit* $\varrho \in [0;0{,}5)$ beschrieben. Mit dieser Fehlerwahrscheinlichkeit finden Flüsse zu und zwischen Systemen außerhalb des vorgesehenen Prozessablaufs statt. Je nach Detailgrad

der Modellierung des Hintergrundwissens können Fortschrittsquote und Fehlerwahrscheinlichkeit pro Verfahren, pro Datentyp oder global angegeben werden. Bei detaillierteren plausiblen Annahmen über das Hintergrundwissen eines Angreifers können die Wahrscheinlichkeiten für bestimmte Flüsse individuell festgelegt werden.

Beispiel. *Die AdBokis Buchclub GmbH hat folgende Verfahren etabliert: Registrierung, Bestellung, Zahlungsabwicklung, Kundendatenarchivierung, Missbrauchsbekämpfung und Kundenservice (Abbildung 9.3).*

Die Fortschrittsquote wird mit 90 % ($\varpi = 0{,}9$) und die Fehlerwahrscheinlichkeit mit $\varrho = 0{,}02$ angenommen. Betrachten wir exemplarisch das Profilfoto θ_{14}. Es ist Teil von zwei Verarbeitungsprozessen, der Registrierung und der Kundendatenarchivierung. Im Regelfall werden alle Profilfotos, die bei der Registrierung (ς_7) übermittelt werden, auch in der Cloud (ς_3) gespeichert. Herkunftssystem ist immer der Kunde (Betroffene) ς_1. Dann ergibt sich folgende Flussmatrix:

$$W_{\theta_{14},\varsigma_1} = \begin{bmatrix} 1 & 0{,}02 & 0{,}02 & 0{,}02 & 0{,}02 & 0{,}02 & 0{,}9 & 0{,}02 & 0{,}02 & 0{,}02 & 0{,}02 \\ 0{,}02 & 1 & 0{,}02 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0{,}02 \\ 0{,}02 & 0{,}02 & 1 & \ddots & & & & & & & \vdots \\ 0{,}02 & \vdots & \ddots & \ddots & \ddots & & & & & & \\ 0{,}02 & \vdots & & \ddots & \ddots & \ddots & & & & & \\ 0{,}02 & \vdots & 0{,}9 & & \ddots & \ddots & \ddots & & & & \\ 0{,}02 & \vdots & & & & \ddots & \ddots & \ddots & & & \\ 0{,}02 & \vdots & & & & & \ddots & \ddots & \ddots & & \\ 0{,}02 & \vdots & & & & & & \ddots & 1 & 0{,}02 & \\ 0{,}02 & 0{,}02 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0{,}02 & 1 & \end{bmatrix}$$

Anmerkung Aus der Definition bedingter Wahrscheinlichkeiten erhält man durch vollständige Induktion die Kettenregel[33]

$$\mathbb{P}(X_d^{\triangleright} = R_d^{\triangleright} \mid X_\varsigma = \varsigma, X_\theta = \theta) =$$

$$\prod_{i \in \mathscr{S}} \prod_{j \in \mathscr{S}} \begin{cases} \mathbb{P}\left(X_{dij}^{\triangleright} = r_{dij}^{\triangleright} \mid X_{d1i}^{\triangleright} = r_{d1i}^{\triangleright}, \ldots, X_{d,i-1,i}^{\triangleright} = r_{d,i-1,i}^{\triangleright}, \right. \\ \qquad \left. X_{d,i+1,i}^{\triangleright} = r_{d,i+1,i}^{\triangleright}, \ldots, X_{dmi}^{\triangleright} = r_{dmi}^{\triangleright}, X_\varsigma = \varsigma, X_\theta = \theta\right), & i = j \\ \mathbb{P}\left(X_{dij}^{\triangleright} = r_{dij}^{\triangleright} \mid X_{dii}^{\triangleright} = r_{dii}^{\triangleright}, X_\varsigma = \varsigma, X_\theta = \theta\right), & i \neq j \end{cases}$$

[33] Russell/Norvig 2010, 514.

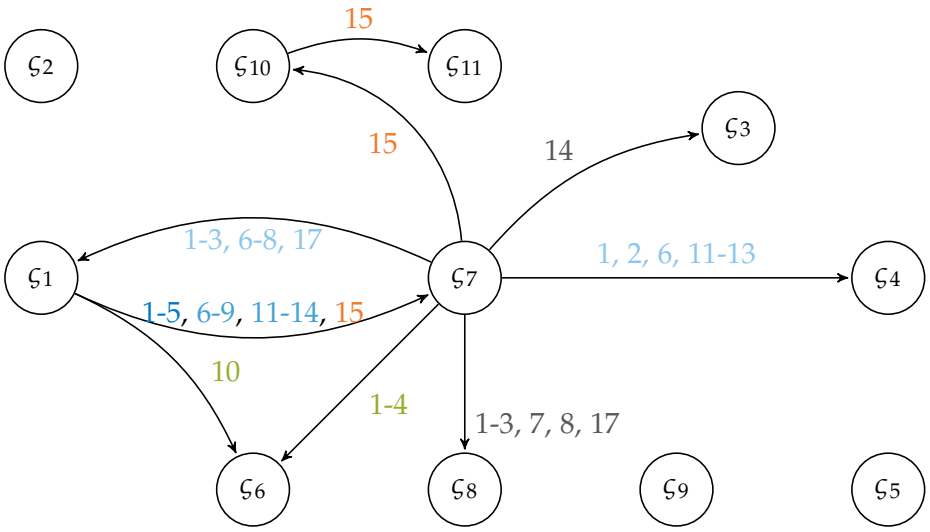


Abbildung 9.3: Verfahren (Datenkategorien als Kantenbeschriftung)

Auf dieser Grundlage kann die A-priori-Wahrscheinlichkeit für die einzelnen Kandidatenrelationen iterativ berechnet werden (mit W als der zur Relation gehörigen Flussmatrix):

$$\mathbb{P}(X_d^{\triangleright} = R^{\triangleright}_{d,r_{11}=1} \mid X_\varsigma = \varsigma_1, X_\theta = \theta) = w_{11}\overline{w}_{12}\cdots\overline{w}_{mm} = w_{11}\overline{w}_{12}\cdots\overline{w}_{1m} = p$$

$$\mathbb{P}(X_d^{\triangleright} = R^{\triangleright}_{d,r^{\triangleright}_{d11}=1,r^{\triangleright}_{d12}=1,r^{\triangleright}_{d22}=1} \mid X_\varsigma = \varsigma_1, X_\theta = \theta) = \frac{p}{\overline{w}_{12}} w_{12}w_{21}w_{22}\overline{w}_{23}\cdots\overline{w}_{2m}$$

$$\ldots$$

Berechenbarkeit Die Komplexität der vollständigen Berechnung der Wahrscheinlichkeiten aller möglichen Kandidatenrelationen ist allerdings in $\mathcal{O}(2^{|\mathscr{S}|\cdot|\mathscr{S}|})$. Selbst bei wenigen Systemen ist somit die Berechenbarkeit der Wahrscheinlichkeitsverteilung nicht mehr gegeben. Deshalb ist nur eine heuristische Lösung entsprechend dem in Abschnitt 9.9 beschriebenen Verfahren möglich.

Beispiel. *Für obige Flussmatrix ergibt sich die Wahrscheinlichkeit, dass das Datum nur im Herkunftssystem ς_1 verarbeitet wird, sofern es von der Datenkategorie θ_{14} ist, als*

$$\mathbb{P}(X_d^{\triangleright} = R^{\triangleright}_{d,r^{\triangleright}_{d11}=1} \mid X_\varsigma = \varsigma_1, X_\theta = \theta_{14}) = 1 \cdot 0{,}1 \cdot 0{,}98^9 \approx 0{,}0834$$

Würde man kombinatorisch aus den Wahrscheinlichkeiten $\mathbb{P}(X_d^{\triangleright} = R_d^{\triangleright})$ die Gesamtwahrscheinlichkeit $\mathbb{P}(X^{\triangleright} = R^{\triangleright}) = \prod_{d\in\mathscr{D}} \mathbb{P}(X_d^{\triangleright} = R_d^{\triangleright})$ berechnen, hätte dies eine Komplexität in $\mathcal{O}(|\mathscr{D}|^{|\mathscr{S}|\cdot|\mathscr{S}|})$. Erfreulicherweise gilt für unabhängige Teilsysteme (Teilrelationen),

dass die Entropie eine additive Größe ist ($H(X^{\triangleright}) = H(X^{\triangleright}_{d_1}) + \cdots + H(X^{\triangleright}_{d_n})$ mit $n = |\mathscr{D}|$).[34]

Somit lässt sich die Gesamtwahrscheinlichkeit aus den approximierten Teilwahrscheinlichkeiten bestimmen.

Die Wahrscheinlichkeitsverteilung für $X^{\triangledown}$ lässt sich auf Grundlage der Wahrscheinlichkeitsverteilung von $X^{\triangleright}$ ermitteln:

$$\mathbb{P}(X^{\triangledown} = R^{\triangledown}) = \sum_{R^{\triangleright} \in \mathcal{R}^{\triangleright} | R^{\triangleright} \equiv_{\triangledown} R^{\triangledown}} \mathbb{P}(X^{\triangleright} = R^{\triangleright})$$

$$\text{mit} \quad R^{\triangleright} \equiv_{\triangledown} R^{\triangledown} \Leftrightarrow$$

$$\forall d \in \mathscr{D}, \varsigma \in \mathscr{S} : ((d,\varsigma,\varsigma) \in R^{\triangleright} \wedge (d,\varsigma) \in R^{\triangledown}) \vee ((d,\varsigma,\varsigma) \notin R^{\triangleright} \wedge (d,\varsigma) \notin R^{\triangledown})$$

Zu diesem allgemeinen Hintergrundwissen kommen noch die jeweiligen Beobachtungen der Angreifer hinzu. Ein Systemangreifer $\mathscr{A}^{\varsigma}$ kann die Datenflüsse durch sein System überwachen. Die Likelihood $\mathbb{P}(I \mid R^{\triangleright})$ ist für solche Beobachtungen sicher 1 oder 0. Die A-posteriori-Wahrscheinlichkeit beträgt

$$\mathbb{P}(R^{\triangleright} \mid I) = \frac{\mathbb{P}(I \mid R^{\triangleright})\mathbb{P}(R^{\triangleright})}{\mathbb{P}(I)} = \frac{\mathbb{P}(I \mid R^{\triangleright})\mathbb{P}(R^{\triangleright})}{\sum_{R^{\triangleright\prime} \in \mathcal{R}^{\triangleright}} \mathbb{P}(I \mid R^{\triangleright\prime})\mathbb{P}(R^{\triangleright\prime})}$$

und damit entweder 0 oder $\frac{\mathbb{P}(R^{\triangleright})}{\sum_{R^{\triangleright\prime} \in \mathcal{R}^{\triangleright}} \mathbb{P}(I,R^{\triangleright\prime})}$. Es findet also eine Normierung auf die Summe der Wahrscheinlichkeiten der Relationen, die die Beobachtung des Angreifers zulassen, statt.

Beobachtungen durch das Datenschutzauskunftssystem Der zentrale Angreifer $\mathscr{A}^{c}$ kann ohne das Datenschutzauskunftssystem keine Beobachtungen machen, sondern muss sich vollständig auf das Hintergrundwissen auf Grundlage des Verfahrensverzeichnisses verlassen. Er ist jedoch der einzige Angreifer, der mit Hilfe des Datenschutzauskunftssystems weitere Beobachtungen machen kann (Beobachtungsereignis I'). Der Systemangreifer $\mathscr{A}^{\varsigma}$ hat nur Einsicht in die Provenance der sowieso bei ihm stattfindenden Verarbeitungsprozesse. Für ihn ist $\Delta(X^{\triangledown}, I') = \Delta(X^{\triangleright}, I') = 1$.

Bei der Registrierung neu erhobener personenbezogener Daten im zentralen Verzeichnis lernt der zentrale Angreifer die Quelle der personenbezogenen Daten und den Ort der ersten Verarbeitung im Unternehmen kennen. Sei ς_0 die Quelle des Datums und ς_κ das erhebende System. Analog zu I ergibt sich eine Likelihood von

$$\mathbb{P}(I' \mid R^{\triangleright}_d) = \begin{cases} 1, & \text{für } (\varsigma_0, \varsigma_\kappa) \in R^{\triangleright}_d \\ 0, & \text{für } (\varsigma_0, \varsigma_\kappa) \notin R^{\triangleright}_d \end{cases}$$

[34] Der Beweis findet sich in Anhang E.2.

Beispiel. *Die meisten Daten (d_1 bis d_{11} und d_{13} bis d_{18}) wurden durch den Onlineverkauf ς_7 beim Betroffenen ς_1 erhoben. Die IBAN wurde dagegen vom Kundenservice ς_6 beim Betroffenen erhoben. Nur die Empfehlung d_{19} kam aus einer anderen externen Quelle ς_2 zum Onlineverkauf. Die Rechnung d_{20} wurde wie im Verfahrensverzeichnis vorgesehen im Onlineverkauf erzeugt. Die Beobachtung bestätigt nur bereits vorhandenes Wissen.*

Als Ergebnis lassen sich jeweils die A-posteriori-Wahrscheinlichkeiten und der abgeleitete Grad der Unverkettbarkeit für $X^{\triangleright}$ und $X^{\triangledown}$ nach dem im Abschnitt 9.9 beschriebenen Verfahren bestimmen:

$$\mathbb{P}(R^{\triangleright} \mid I') = \frac{\mathbb{P}(I' \mid R^{\triangleright})\mathbb{P}(R^{\triangleright})}{\mathbb{P}(I')} = \frac{\mathbb{P}(I' \mid R^{\triangleright})\mathbb{P}(R^{\triangleright})}{\sum_{R^{\triangleright\prime} \in \mathcal{R}^{\triangleright}} \mathbb{P}(I' \mid R^{\triangleright\prime})}$$

$$\mathbb{P}(R^{\triangledown} \mid I') = \sum_{R^{\triangleright} \in \mathcal{R}^{\triangleright} \mid R^{\triangleright} \equiv_{\triangledown} R^{\triangledown}} \mathbb{P}(R^{\triangleright} \mid I')$$

$$\Delta(X^{\triangleright}, I') = \frac{\sum_{R^{\triangleright} \in \mathcal{R}^{\triangleright}} \mathbb{P}(R^{\triangleright} \mid I') \log_2 \mathbb{P}(R^{\triangleright} \mid I')}{\sum_{R^{\triangleright} \in \mathcal{R}^{\triangleright}} \mathbb{P}(R^{\triangleright}) \log_2 \mathbb{P}(R^{\triangleright})}$$

$$\Delta(X^{\triangledown}, I') = \frac{\sum_{R^{\triangledown} \in \mathcal{R}^{\triangledown}} \mathbb{P}(R^{\triangledown} \mid I') \log_2 \mathbb{P}(R^{\triangledown} \mid I')}{\sum_{R^{\triangledown} \in \mathcal{R}^{\triangledown}} \mathbb{P}(R^{\triangledown}) \log_2 \mathbb{P}(R^{\triangledown})}$$

9.9 Implementierung

Wie bereits im vorherigen Abschnitt erwähnt, ist die Wahrscheinlichkeitsverteilung für die Relation $R_d^{\triangleright}$ auch schon bei wenigen Systemen, Daten und Datentypen nicht mit akzeptablem Aufwand an Zeit und Speicher vollständig berechenbar. Allerdings ist eine approximative Lösung möglich. Sind die Wahrscheinlichkeitsmatrizen nur spärlich mit Fortschrittswahrscheinlichkeiten belegt, ballt sich die Wahrscheinlichkeitsmasse bei denjenigen Kandidatenrelationen, die einen Fluss entlang des Verarbeitungsprozesses vorsehen. Kandidatenrelationen, die kaum Flüsse im Verarbeitungsprozess vorsehen, bilden den „Long Tail“ der Verteilung. Ihr Gewicht bei der Berechnung der Entropie ist gering. Würde man die Entropie anhand einer Stichprobe numerisch abschätzen, würden die Elemente des „Long Tail“ nur sehr selten auftreten. Ein naiver Schätzer $\hat{H}(X) = -\sum_i \hat{p}_i \log_2 \hat{p}_i$, mit $\hat{p}_i = \frac{n_i}{N}$ und n_i als der Häufigkeit von $R_i^{\triangleright}$ in der Stichprobe und $N = \sum_i n_i$ als deren Größe, würde die Entropie systematisch unterschätzen und sich der tatsächlichen Entropie von unten annähern.[35]

Diesen Umstand kann man sich bei der Berechnung der Entropie aus den Wahrscheinlichkeitsmatrizen zunutze machen, indem man systematisch zuerst die wahrscheinlicheren

[35] Bonachela/Hinrichsen/Mũoz 2008.

Kandidaten in die Berechnung aufnimmt und den „Long Tail“ nur bis zu einem gegebenen *Schwellwert* τ erschließt. Die Wahrscheinlichkeiten ergeben sich aus dem Wahrscheinlichkeitsbaum (Abbildung 9.4). Durch Tiefensuche in diesem Baum kann die Entropie, ausgehend vom Startsystem, approximativ erschlossen werden.

Algorithmus 16: computeℙ$(R_d^{\triangleright} \mid \varsigma, \theta)$

```
1  node ← DecisionTree(ς, θ).getRootNode
2  while node.hasNext do
3  |  while node.hasChild ∧ node.getProb > τ do
4  |  |  node ← node.pollNextNode
5  |  end
6  |  node.touchChildNodes
7  |  result ← result ∪ {getMostProbableR (node.getState), node.getProb}
8  |  node ← node.pollNextNode // poll parent node
9  end
10 return result
```

Algorithmus 16 zeigt das grundsätzliche Vorgehen. Der Baum wird solange durchlaufen, bis der gegebenen Schwellwert unterschritten wird (Zeile 3). Ist noch kein Blatt erreicht, wird der übrige Teilbaum nicht weiter durchsucht (Zeile 6). Die gesuchte Wahrscheinlichkeit ergibt sich direkt aus dem Baum. Bei Blättern gilt das gleiche für die Relation. Wurde die Schleife aufgrund des Schwellwerts abgebrochen, muss noch die Relation mit der höchsten Wahrscheinlichkeit als Repräsentant des ganzen Teilbaums ausgewählt werden (Zeile 7). Der übrige Teilbaum wird nicht weiter berücksichtigt (Verzweigungen unterhalb der orangenen Knoten in Abbildung 9.4).

Beispiel. *Tabelle 9.2 enthält die berechneten Ergebnisse für unterschiedliche Schwellwerte. Die Annäherung von unten ist deutlich sichtbar. Dem Betroffenen könnte ein Mindestgrad an Unverkettbarkeit von* 0,94 *bzw.* 0,90 *garantiert werden.*

9.10 Interpretation und Anwendung der Metrik

Die folgenden Abschnitte verdeutlichen (1) die Bedeutung des Hintergrundwissens für die Metrik, (2) die Relevanz der Metrik für das Datenschutzauskunftssystem und (3) die Anwendung der Metrik zum Systemvergleich.

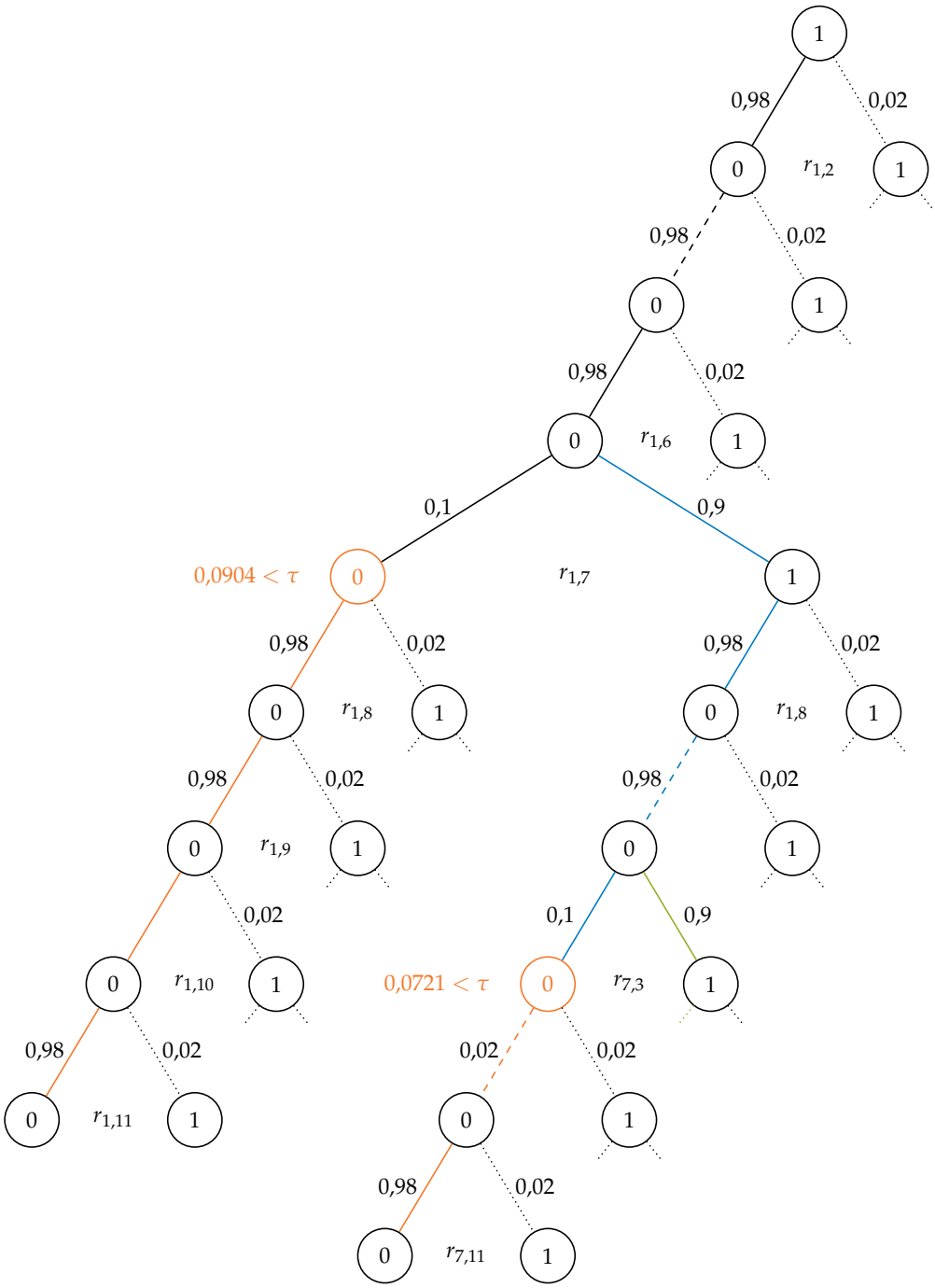


Abbildung 9.4: Wahrscheinlichkeitsbaum für θ_{14}, ausgehend von ς_1, mit angedeuteter Tiefensuche (**erste**, zweite und dritte Iteration sowie die Suche nach dem jeweiligen Repräsentanten) für $\tau = 0{,}1$

τ	$H(X^{\triangleright})$	$H(X^{\triangleright}, I')$	$\Delta(X^{\triangleright}, I')$	$H(X^{\triangledown})$	$H(X^{\triangledown}, I')$	$\Delta(X^{\triangledown}, I')$
10^{-1}	125,8201	114,9513	0,8927	96,2104	85,8917	0,9136
10^{-2}	139,5569	128,8865	0,9235	103,6033	92,4655	0,8925
10^{-3}	149,4219	139,3821	0,9328	107,1627	96,1635	0,8974
10^{-4}	156,1207	147,2097	0,9429	109,1815	98,5819	0,9029
10^{-5}	157,3259	148,4618	0,9436	109,4951	98,8992	0,9032
10^{-6}	159,4152	150,9697	0,9470	109,9215	99,4427	0,9046
10^{-7}	159,6693	151,2301	0,9471	109,9721	99,4967	0,9047

Tabelle 9.2: Approximierte Werte für die Entropie und den Grad der Unverkettbarkeit der Relationen $R^{\triangleright}$ und $R^{\triangledown}$ für unterschiedliche Schwellwerte τ

9.10.1 Berücksichtigung des Hintergrundwissens

Im Abschnitt 9.3 wurde bereits ausgeführt, dass der negative Einfluss des Datenschutzauskunftssystems auf die Unverkettbarkeit bei Nichtberücksichtigung des Hintergrundwissens als Teil des A-Priori-Wissens systematisch überschätzt würde. Alle Beobachtungen und das mögliche Angreiferwissen würden auf das A-posteriori-Wissen verlagert. Der Grad der Unverkettbarkeit würde viel zu niedrig angegeben werden.

Die Verlagerung des Hintergrundwissens in die Modellierung des A-posteriori-Wissens wäre außerdem Grundsätzlich fragwürdig. Hintergrundwissen fällt nicht bei Einsatz eines Datenschutzauskunftssystems plötzlich vom Himmel.

Wird das Hintergrundwissen des Angreifers weder a priori noch a posteriori berücksichtigt, gibt es sogar Verarbeitungssituationen, in denen die Unverkettbarkeitsmetriken zu einer grundsätzlich falschen Interpretation verleiten würden. Aus Gründen der Vereinfachung wird zur Erläuterung folgendes Minimalbeispiel verwendet:

Beispiel. *Gegeben sei eine Situation mit einem personenbezogenen Datum $d = d_5$ der einzigen Kategorie $\theta = \theta_5$ (Geburtsdatum) und den drei Systemen $\varsigma_a = \varsigma_1$ (Alice Fox), $\varsigma_b = \varsigma_7$ (Vertrieb – Onlineverkauf) und $\varsigma_c = \varsigma_6$ (Kundenbetreuung).*

Der zentrale Angreifer $\mathscr{A}^c$ lernt demnach durch das Datenschutzauskunftssystem, ob das Geburtsdatum vom Vertrieb oder vom Kundenservice bei Alice Fox erhoben wurde. Dies ist gleichbedeutend damit, ob ein Fluss von ς_a nach ς_b oder von ς_a nach ς_c stattgefunden hat ($r_{12} = 1$ bzw. $r_{13} = 1$).

Ohne jedes Hintergrundwissen ergibt sich,[36] egal welchen der beiden Flüsse der An-

[36] Würde man das Hintergrundwissen als A-posteriori-Wissen miteinbeziehen, würde eine Differenzierung zwischen den beiden beobachteten Flüssen zustande kommen. Resultat wäre ein systematisch unterschätzter Grad der Unverkettbarkeit von $\Delta(X^{\triangleright}, I') \approx 0{,}0484$ beziehungsweise $\Delta(X^{\triangleright}, I'') \approx 0{,}1119$.

greifer lernt, ein Grad der Unverkettbarkeit für die Datenflussrelation von

$$\Delta(X^{\triangleright},I) = \frac{H(X^{\triangleright} \mid I)}{H_{\max}(X^{\triangleright})} = \frac{\log_2(2^{|\mathscr{S}|^2-1})}{\log_2(2^{|\mathscr{S}|^2})} = \frac{3^2-1}{3^2} = \frac{8}{9} \approx 0{,}8889.$$

Ganz anders fällt die Bewertung aus, wenn Hintergrundwissen gemäß den Erläuterungen der vorangegangenen Abschnitte miteinbezogen wird.

Beispiel. *ς_a ist das Herkunftssystem von θ. Das Verfahrensverzeichnis sieht einen Datenfluss von ς_a zu ς_b vor. Fortschrittsquote und Fehlerwahrscheinlichkeit werden, wie gehabt, mit $\varpi = 0{,}9$ und $\varrho = 0{,}02$ angenommen. Die Flussmatrix sieht dann wie folgt aus:*

$$W_{\theta,\varsigma_a} = \begin{bmatrix} 1 & 0{,}9 & 0{,}02 \\ 0{,}02 & 1 & 0{,}02 \\ 1 & 0{,}02 & 1 \end{bmatrix}$$

Daraus ergibt sich eine A-priori-Entropie für den Angreifer von $H(X^{\triangleright}) \approx 0{,}8757$.

Die A-posteriori-Entropie hängt davon ab, welche Beobachtung gemacht wurde. Wird ein Fluss von von ς_a nach ς_b festgestellt (I'), ergibt sich $H(X^{\triangleright}, I') \approx 0{,}4355$ und resultierend $\Delta(X^{\triangleright},I') \approx 0{,}4974$.

Wird dagegen ein Fluss von ς_a nach ς_c erfasst (I''), steigt die A-posteriori-Entropie sogar. Die Beobachtung ist der A-priori-Annahme entgegengesetzt. Aus einem Wert von $H(X^{\triangleright}, I'') \approx 1{,}007$ ergibt sich ein Grad der Unverkettbarkeit von $\Delta(X^{\triangleright},I'') \approx 1{,}15$. Nach der Normierung gemäß Abschnitt 9.4 bleibt ein normierter globaler Grad der Unverkettbarkeit von $\|\Delta\| = 1$.

Folglich unterscheidet sich der resultierende Grad der Unverkettbarkeit für den Betroffenen, je nach tatsächlichem Datenfluss, deutlich. In einem größeren Szenario wäre der Unterschied weitaus geringer, da eine einzelne Bobachtung nicht so sehr ins Gewicht fällt. Nichtsdestoweniger gilt jedoch, dass das A-priori-Hintergrundwissen eine Differenzierung erlaubt, die sonst nicht möglich wäre. Der Grad der Unverkettbarkeit ist ohne Hintergrundwissen, unabhängig von der tatsächlichen Qualität der Beobachtung des Angreifers, immer gleich.

Zusammenfassend lässt sich feststellen, dass die modellierte Einbeziehung des Hintergrundwissens nicht nur einen quantitativen Unterschied macht, sondern auch einen qualitativen.

9.10.2 Einbindung der Metrik in das Datenschutzauskunftssystem

Der Grad der Unverkettbarkeit gemäß der vier Metriken kann dem Betroffenen als Anhaltspunkt dafür dienen, welchen Einfluss ein Datenschutzauskunftssystem auf die Profilbildungsmöglichkeiten innerhalb des datenverarbeitenden Unternehmens hat. Die präzise

Angabe des Grads der Unverkettbarkeit bietet dem Betroffenen die Möglichkeit, selbst darüber zu entscheiden ob er der Transparenz (dem Auskunftssystem) oder der Unverkettbarkeit seiner Daten einen höheren Stellenwert einräumt. Zu diesem Zweck werden die Werte der Metriken auf der Auskunftsplattform *PrivacyInsight* visualisiert. Kapitel 10 beschreibt, wie die Metriken eingebunden werden und welche Optionen sich dem Betroffenen bieten. In der Nutzerstudie in Kapitel 12 wird geschildert, wie sich Probanden bei einer Konfrontation mit der Metrik verhalten.

Neben der implementierten Einbindung auf der Auskunftsplattform ist es auch denkbar, die Metrik bereits bei der Erhebung personenbezogener Daten einzubinden. Eine clientseitige Vorausberechnung könnte in einer Webanwendung realisiert werden. Die Metrik würde dann den Wissenszuwachs der am Datenschutzauskunftssystem beteiligten Stellen durch die Datenerhebung repräsentieren. Der Betroffene könnte a priori entscheiden, ob er zunächst das Datenschutzauskunftssystem für sich deaktivieren will oder nicht.

Die Metriken stellen jedoch nur den Wissenszuwachs durch die Provenance, nicht den Wissenszuwachs durch die erhobenen personenbezogenen Daten selbst dar. Bei Berücksichtigung des Inhalts müsste der Einfluss desselben auf die Verkettungsmöglichkeiten bestimmt werden. Eine solche Inhaltsanalyse ist nur schwer in allgemeiner Form umzusetzen und wird durch die vorgestellten Verfahren nicht abgedeckt.

9.10.3 Systemvergleich mit Hilfe der Metrik

Die Metrik ist auch nur eingeschränkt für den Vergleich von Architekturen vor der Realisierung eines Datenschutzauskunftssystems geeignet. Da immer ein konkreter Prior und Posterior herangezogen wird, ist das Ergebnis der Metrik vom konkreten Szenario und Betrachtungszeitpunkt abhängig. Für in ihrer Struktur deutlich unterschiedliche Architekturen kann die Metrik dennoch eine Abschätzung bieten.

Beispielhaft könnte eine Alternativarchitektur zur, in dieser Arbeit vorgestellten,[37] Architektur vorsehen, die gesamte Provenance auf einem zentralen Server zu speichern (Alternative 1). Da dieses Serversystem Einblick in alle Datenflüsse hätte, würde der Grad der Unverkettbarkeit, die Unsicherheit des Angreifers, für die Verarbeitungsrelation und die Datenflussrelation bei 0 liegen. Solange weiterhin nicht die konkreten personenbezogenen Daten, sondern nur Pseudonyme zentral gespeichert würden, würde der Grad der Unverkettbarkeit für die Identifikationsrelation und die Verknüpfungsrelation zwar niedriger als in der tatsächlich realisierten Architektur liegen,[38] jedoch zumindest nicht bei 0.

Die gleiche Situation ergibt sich bei der redundanten Speicherung der Provenance gemäß der beiden in Kapitel 8 erwähnten Konzepte Cassandra-Cluster und verteilte Hashtabelle (Alternative 2). Bei einem Cassandra-Cluster kann jedes System auf die gesamte

[37] Vgl. Kapitel 8.2.

[38] Anhand der Datenflüsse kann auf die Kategorie der personenbezogenen Daten geschlossen werden.

Provenance zugreifen. Beim Konzept der verteilten Hashtabelle kann der zentrale Server, der die Schlüssel für die gesamte Provenance kennt, auf die gesamte Provenance, ohne Einwirkung des Systems, aus dem die Provenance ursprünglich stammt, zugreifen. Der zentrale Server hat somit effektiv den gleichen Wissensstand wie bei einer zentralen Speicherung.

Die interessanteste Vergleichsarchitektur für die, in dieser Arbeit vorgestellte, Architektur ist die von *Insynd* (Alternative 3).[39] Aufgrund der Verschlüsselung der Provenance auf den Logservern und der Schlüsselgewalt durch den Betroffenen, würde der Grad der Unverkettbarkeit für die Verarbeitungsrelation und die Datenflussrelation bei 100 % liegen. Da alle Systeme den öffentlichen Schlüssel des Betroffenen kennen müssen, ist die Situation für die Identifikationsrelation und die Verknüpfungsrelation differenzierter. Diese Metriken können nur für ein konkretes Szenario zu einem festgelegten Zeitpunkt bestimmt werden. Zum Vergleich wird wie gehabt das durchgängige Minimalbeispiel gewählt.

Angreifer ist bei der *Insynd*-Architektur nur $\mathscr{A}^{\varsigma}$. Einen zentralen Angreifer gibt es nicht. Auch wenn es einen zentralen Logserver gibt, liegt die Provenence dort nur verschlüsselt vor. Der Systemangreifer, in dessen System die meisten personenbezogenen Daten verarbeitet werden, ist der Angreifer der durch die Schlüsselzuordnung im ungünstigsten Fall am meisten lernen kann.

Im besten Fall[40] ist durch den Inhalt der personenbezogenen Daten in einem datenverarbeitenden System direkt klar, zu welchem Betroffenen die Daten gehören. Die Schlüsselverwaltung spielt in diesem Fall keine Rolle. Der Grad der Unverkettbarkeit liegt für die Identifikationsrelation und die Verknüpfungsrelation bei jeweils 100 %.

Im ungünstigsten Fall ist durch den Inhalt und die Struktur der Datenspeicherung keine Identifikation des Betroffenen möglich. Ebenso ist es für den Angreifer ausgeschlossen, herauszufinden, welche personenbezogenen Daten zum selben Betroffenen gehören. In diesem Fall wird der Angreifer beim Einsatz eines Datenschutzauskunftssystems mit *Insynd* neu lernen, welche personenbezogenen Daten zum selben Betroffenen gehören.[41] Begründet liegt dies in dem Erfordernis, alle personenbezogenen Daten eines Betroffenen mit dessen öffentlichem Schlüssel zu verschlüsseln.

Beispiel. *Das System, das bei AdBokis die meisten personenbezogenen Daten verarbeitet, ist der Onlineverkauf im Vertrieb (ς_7). Dort werden 18 personenbezogene Daten von Alice (d_1 bis d_{19} mit Ausnahme von d_{12}) verarbeitet.*[42] *Bei der Berechnung der Anzahl der A-posteriori-Kandidatenrelationen für die Verknüpfungsrelation steht die Zusammengehörigkeit der 18 Daten*

[39] Pulls/Peeters 2015.

[40] Im Hinblick auf die Auswirkungen eines Datenschutzauskunftssystems mit *Insynd*, nicht für den Betroffenen.

[41] Sind die Schlüssel nicht pseudonymisiert, sondern mit einem Identifikator verknüpft, ist sogar klar, zu welchem Betroffenen die Daten gehören.

[42] Peter bleibt in diesem Beispiel zur Vereinfachung außen vor.

bereits im Vorhinein fest. Sie können wie ein Datum behandelt werden. Damit reduziert sich die Anzahl der Kandidatenrelationen auf $|\mathcal{R}^{\equiv}| = \sum_{k=0}^{2} \mathcal{S}_{13,k} = 2^{12} = 4096$. *Folglich ist der resultierende Grad der Unverkettbarkeit* $\Delta(X^{\equiv}, I) = \frac{\log_2 2^{12}}{\log_2 2^{29}} = \frac{12}{29} \approx 0{,}4138$.[43]

Das Vorgehen für die Identifikationsrelation ist äquivalent. Die zusammengehörigen Daten können nur gemeinsam zu einem Betroffenen gehören. Die reduzierte Anzahl der Kandidatenrelationen liegt bei $|\mathcal{R}^{\leq}| = 2^{13} = 8192$ *und der resultierende Grad der Unverkettbarkeit ist* $\Delta(X^{\leq}, I) = \frac{\log_2 2^{13}}{\log_2 2^{30}} = \frac{13}{30} \approx 0{,}4333$.

Damit sind beide Werte im ungünstigsten Fall deutlich niedriger als bei dem in dieser Arbeit verwendeten System. Tatsächlich würden die Werte jedoch irgendwo zwischen 100 % und den oben berechneten Werten liegen.

Zusammenfassend kann festgehalten werden, dass man sich für *Insynd* entscheiden würde, falls Prozesswissen der entscheidende Faktor ist und man den Betroffenen in den Schlüsselerzeugungsprozess einbinden kann. Zieht man die Verknüpfung unterschiedlicher personenbezogener Daten eines Betroffenen als Hauptfaktor heran, hängt die Bewertung davon ab, welche Schlüsse sich bereits ohne das Datenschutzauskunftssystem in den erhebenden Systemen ziehen lassen. Im Beispiel fließen fast alle Daten des Betroffenen gemeinsam nach ς_7. Allein aufgrund dieser Strukturinformation ist davon auszugehen dass der Systemangreifer bereits a priori mit hoher Sicherheit sagen kann, welche personenbezogenen Daten zusammengehören. Es kann indes nicht abschließend festgestellt werden, ob *Insynd* oder die in dieser Arbeit verwendete Architektur die bessere ist.

9.11 Zwischenfazit

Die wissenschaftlichen Beiträge dieses Kapitels lassen sich wie folgt zusammenfassen: (1) Die Formulierung einer Unverkettbarkeitsmetrik unter Berücksichtigung beliebiger Relationen, (2) die Berücksichtigung und Bestimmung des Hintergrundwissens und der Beobachtungen des Angreifers („Verketters“), (3) eine Instantiierung der Metrik für das Szenario eines Datenschutzauskunftssystems und (4) eine Heuristik zur Berechnung der Metrik in der Praxis.

Die in diesem Kapitel vorgestellte Unverkettbarkeitsmetrik berücksichtigt die vier Faktoren Entitäten, Verkettungsrelation, Verketter und betrachtetes System. Sie beruht auf informationstheoretischen Überlegungen und erfüllte alle in Abschnitt 9.4 aufgeführten formalen Anforderungen an Metriken für Unverkettbarkeit.

Im Kontext eines Datenschutzauskunftssystems wurde die Metrik in vier Varianten instanziiert, die in Übereinstimmung mit Konstruktionsziel $\mathfrak{K}5$ alle Aspekte der Verarbeitung der Personal-Data-Provenance abdecken.

[43] Vgl. Abschnitt 9.8.1.

Es ist nicht realistisch, nur Angreifer ohne Hintergrundwissen anzunehmen. Deshalb wurden für die Überlegungen in diesem Kapitel Angreifer mit Hintergrundwissen eingeführt. Das Hintergrundwissen stützt sich vor allem auf die Annahme, dass unternehmensinternen Angreifern das Verfahrensverzeichnis bekannt ist. Das Hintergrundwissen wurde formalisiert und die darauf basierende Bestimmung der A-priori- und A-posteriori-Entropie erläutert.

Die resultierende Metrik ermöglicht eine Differenzierung abhängig von den Beobachtungen des Angreifers. Berücksichtigt wurden ein Systemangreifer und ein zentraler Angreifer. Weitere Angreifer wurden nicht betrachtet. Im Allgemeinen sind solche Beobachtungen relevant, die Rückschlüsse auf die betrachteten Relationen zulassen. Im Szenario des Datenschutzauskunftssystems sind dies Informationsflüsse und Metainformationen über den Personenbezug.

Die Annahme unabhängiger Datenflüsse ist wesentliche Voraussetzung für die Berechenbarkeit der Metrik. Für die Bestimmung der Entropie je Datum wurde eine Heuristik entwickelt, die sich das asymptotische Verhalten von Entropieschätzern zu Nutze macht. Dadurch kann die Metrik in konkreten Praxisszenarien bestimmt werden.

Eingeschränkt kann die Unverkettbarkeitsmetrik für den Vergleich von Systemarchitekturen verwendet werden. Wichtiger ist jedoch die Information des Betroffenen. Die Unverkettbarkeitsmetrik macht die Auswirkungen des Datenschutzauskunftssystems für den Betroffenen erkennbar. Er kann sich selbst entscheiden, ob er für das Mehr an Transparenz bereit ist, den zusätzlichen Verlust an Unverkettbarkeit zu akzeptieren. Das nachfolgende Kapitel wird diesen Gedanken weiter ausführen.

10 Betroffenenautonomie zwischen Transparenz und Unverkettbarkeit

Ziel eines integrierten Datenschutzauskunftssystems ist es, den Betroffenen mit seiner vollen Entscheidungsautonomie in den Mittelpunkt zu stellen. Zur Entscheidungsfreiheit des Betroffenen gehört auch, aus Bedenken bezüglich der Verkettung von Daten und Verarbeitungsbeziehungen, auf das Provenance-Tracking und daraus resultierend auf die elektronische Auskunftserteilung zu verzichten. In Abschnitt 10.1 wird erörtert, ob eine solche freie Entscheidung zwischen Transparenz und Unverkettbarkeit im Rahmen des Datenschutzrechts möglich ist.

Ein Datenschutzauskunftssystem könnte beispielsweise eine Funktion vorsehen, mit der der Betroffene das Provenance-Tracking jederzeit abschalten kann. Dadurch reduziert sich für ihn die Transparenz, aber gleichzeitig erhöht sich für ihn die Unverkettbarkeit zukünftiger Datenverarbeitungsvorgänge. Wie die Auskunftsplattform *PrivacyInsight* aufgebaut ist und an welcher Stelle eine solche Option integriert ist, stellt Abschnitt 10.3 vor.

Es ist außerdem Teil der Betroffenenautonomie, selbst zu bestimmen, wer die eigenen personenbezogenen Daten zu einem bestimmten Zweck verarbeitet. Die DSGVO hat aus diesem Grund ein Recht auf Datenübertragbarkeit neu eingeführt. Der Zusammenhang zwischen Datenübertragbarkeit und Auskunft wird in Abschnitt 10.2 erläutert.

10.1 Freie Entscheidung zwischen Transparenz und Unverkettbarkeit

Ein Datenschutzauskunftssystem ist ein entscheidendes Werkzeug, um die Transparenz des Umgangs mit personenbezogenen Daten zu verbessern. Auf der anderen Seite führt ein Datenschutzauskunftssystem zu einer zweckgebundenen, aber doch zusätzlichen Verwendung personenbezogener Daten. Für die Auskunft muss die Personal-Data-Provenance so verarbeitet und gespeichert werden, dass sie zum Zeitpunkt der Auskunftserteilung zu einem Gesamtbild des Betroffenen zusammengefügt werden kann.[1] Aus diesem Grund kann es durchaus Betroffene geben, die lieber auf die Möglichkeit einer elektronischen Auskunftserteilung verzichten würden. Abhängig von der Bewertung der

[1] Leucker, PinG 2015, 195 (198) sieht bereits Risiken durch die Strukturierung für das Recht auf Datenübertragbarkeit.

Ergebnisse der Unverkettbarkeitsmetriken aus Kapitel 9 könnten sie zum Schluss kommen, dass eine informative Auskunft die Profilbildungsmöglichkeiten durch das Provenance-Tracking nicht aufwiegt.

Allerdings scheint solch ein Verzicht auf das Recht auf Auskunft im Datenschutzrecht nicht vorgesehen. § 6 Abs. 1 BDSG kodifiziert die Unabdingbarkeit des Rechts auf Auskunft.

Das Recht des Betroffenen auf Auskunft kann nicht durch Rechtsgeschäft ausgeschlossen oder beschränkt werden. Die Dispositionsbefugnis des Betroffenen wird im Interesse des Betroffenen eingeschränkt.[2] Es soll verhindert werden, dass der Betroffene seine Rechte für das sprichwörtliche Linsengericht aufgibt.[3]

Die Unabdingbarkeit des Rechts auf Auskunft gilt unabhängig von der Form des Rechtsgeschäfts.[4] „Ein Rechtsgeschäft ist auf die Herbeiführung eines rechtlichen Erfolges gerichtet, der nach der Rechtsordnung eintritt, weil er gewollt ist.“[5] Ein Rechtsgeschäft besteht aus einer oder mehreren Willenserklärungen, die, gegebenenfalls in Verbindung mit weiteren Tatbestandsmerkmalen, die Rechtsfolge herbeiführen.[6] Darunter fällt auch die einseitige Willenserklärung des Betroffenen zum freiwilligen Verzicht auf die Nutzung des Datenschutzauskunftssystems.

Eine Beschränkung des Rechts auf Auskunft ist immer eine Änderung zum Nachteil des Betroffenen.[7] Einer Erweiterung oder Verstärkung des Auskunftsanspruchs zu Gunsten des Betroffenen steht die Vorschrift nicht entgegen.[8] Die Wahl zwischen Auskunftsplattform und Unverkettbarkeit ist keine solche Erweiterung, da sie gerade der Abwägung unterschiedlicher Nachteile durch den Betroffenen bedarf. Der Verzicht auf die elektronische Form der Auskunft ist zwar kein Verzicht auf die Auskunft an sich. In komplexen Informationssystemen ist jedoch ohne ein, wie in den Kapiteln 5 bis 8 dargestelltes, Informationsflusstracking keine vollständige Auskunft möglich. Insofern ist ein Rückfall auf die manuelle Teilauskunft eine Beschränkung des Rechts auf Auskunft.

Die Einschränkung des Dispositionsrechts geht davon aus, dass dem Betroffenen sein Recht abgekauft wird oder er, aufgrund anderer datenschutzferner Erwägungen, auf seine Rechte verzichtet. Die Regelung erwartet nicht, dass dem Betroffenen Nachteile in der Wahrnehmung anderer Datenschutzrechte entstehen. Insofern ist die Abwägung zwischen Transparenz und Unverkettbarkeit regelungsuntypisch.

Auch im Datenschutzrecht gilt der Grundsatz der Privatautonomie. Die Privatautonomie als Teil der allgemeinen Handlungsfreiheit aus Art. 2 Abs. 1 GG schützt die Selbst-

[2] BT-Drs. 11/4306, 41.
[3] Tangens 2008, 140 nach 1. Mose 25, 29-34.
[4] Dix in: Simitis, BDSG 2014, § 6 Rn. 15.
[5] Völzmann-Stickelbrock/Ahrens in: Prütting/Wegen/Weinreich, BGB 2015, Vor § 116 ff Rn. 1.
[6] Völzmann-Stickelbrock/Ahrens in: Prütting/Wegen/Weinreich, BGB 2015, Vor § 116 ff Rn. 3.
[7] Dix in: Simitis, BDSG 2014, § 6 Rn. 21.
[8] Dix in: Simitis, BDSG 2014, § 6 Rn. 22; Gola/Schomerus, BDSG 2015, § 6 Rn. 4.

bestimmung privater Akteure im Zivilrechtsverkehr.[9] Eingriffe in den Schutzbereich der Privatautonomie sind zulässig, um strukturellen Macht- oder Informationsasymmetrien zu begegnen, die dazu führen, dass eine Partei den Vertragsinhalt einseitig bestimmt.[10] Ziel dieser Eingriffe muss der im Sozialstaatsprinzip (Art. 20 Abs. 1, Art. 28 Abs. 1 GG) wurzelnde Schutz vor Fremdbestimmung sein.[11]

Die Regelungen des nicht-öffentlichen Datenschutzes gehen von solch einer Machtasymmetrie aus.[12] Ist der Betroffene jedoch, analog zur Einwilligung, fundiert *informiert*, und ist seine Entscheidung *frei* und *unbeeinflusst*, dann ist ein Eingriff in die Privatautonomie nicht mehr gerechtfertigt. Dem Betroffenen ist ein Dispositionsrecht zuzugestehen.[13]

Ein Opt-out aus der elektronischen Auskunftserteilung durch ein Datenschutzauskunftssystem muss sich an diesen Kriterien messen lassen.

Die Entscheidung über das Opt-out ist analog zu § 4a Abs. 1 S. 2 BDSG und Art. 4 Nr. 11 i. V. m. ErwGr 32 DSGVO nur möglich, wenn sie *informiert* erfolgt. Die Entscheidung erfolgt informiert, wenn der Betroffene rechtzeitig und umfassend über die Konsequenzen der Handlungsalternativen unterrichtet wurde.[14] Der Anlass und die Folgen der Entscheidung müssen ersichtlich werden.[15] Das Datenschutzauskunftssystem informiert den Betroffenen durch den in Abbildung F.10 dargestellten Text über die Konsequenzen der Auskunft und die Bedeutung der, in Kapitel 9 vorgestellten, Unverkettbarkeitsmetriken. Die Unverkettbarkeitsmetriken sind ein objektiver Maßstab für das Profilbildungsrisiko durch das Datenschutzauskunftssystem. Den entstandenen Transparenzgewinn kann der Betroffene durch die Nutzung des Datenschutzauskunftssystems evaluieren.

Der *freien* Entscheidung liegt die Vorstellung vom Menschen als selbstbestimmtem, geistig-sittlichem Wesen zu Grunde.[16] Der Betroffene entscheidet frei,[17] wenn er im Vollbesitz seiner geistigen Kräfte[18] und ohne Zwang[19] aus mindestens zwei Handlungsalternativen[20] auswählt. Beides ist beim vorgesehenen Opt-out grundsätzlich gegeben.

Zur freien Entscheidung gehört auch, dass der Betroffene jederzeit berechtigt ist, eine einmal gewählte Option nachträglich, mit Wirkung für die Zukunft, wieder zu ändern.[21] Allerdings scheidet nach Beginn einer Verarbeitung eine Änderung dann aus, wenn es

[9] BVerfGE 89, 214 (231); Dreier in: Dreier, GG 2013, Art. 1 I Rn. 62.
[10] BVerfGE 81, 242 (255); BVerfGE 89, 214 (232); BVerfGE 103, 89 (100 f.).
[11] Dreier in: Dreier, GG 2013, Art. 1 I Rn. 63.
[12] Vgl. Kapitel 2.1.2 zur Drittwirkung der Grundrechte.
[13] In der DSGVO findet die grundsätzliche Unabdingbarkeit der Auskunft gar keine Erwähnung mehr.
[14] Simitis in: Simitis, BDSG 2014, § 4a Rn. 70.
[15] Simitis, a. a. O.
[16] BVerfGE 45, 187 (227).
[17] Art. 4 Nr. 11 DSGVO.
[18] Analog zu den Anforderungen an eine Willenserklärung in § 105 Abs. 2 BGB.
[19] Art. 2 Lit. h DSRL.
[20] Die „echte Wahl“ des ErwGr 42 DSGVO.
[21] Simitis in: Simitis, BDSG 2014, § 4a Rn. 94.

der verantwortlichen Stelle objektiv nicht mehr möglich oder nicht zuzumuten ist, die Verarbeitung der Daten zukünftig gemäß der neuen Handlungsvorgabe durchzuführen. Dies kann beispielsweise bei anonymisierten oder pseudonymisierten Daten der Fall sein.[22]

Dem Opt-out aus der elektronischen Auskunftserteilung ist somit ein nachträgliches Opt-In beizugesellen. Ein solches Opt-In führt dazu, dass zukünftige Verarbeitungsvorgänge personenbezogener Daten wieder elektronisch beauskunftet werden können. Da während der Opt-Out-Phase Speicherorte entstehen können, die sich einem zukünftigen Tracking entziehen, ist es der verantwortlichen Stelle objektiv nicht mehr möglich, die Vollständigkeit der Auskunft bezüglich bereits erhobener personenbezogener Daten sicherzustellen. Für zukünftig erhobene Daten kann allerdings eine Vollständigkeitsgarantie gegeben werden.

Eine Entscheidung ist *unbeeinflusst*, wenn sie frei von Einwirkungen getroffen wird, die nichts mit dem datenschutzrechtlichen Gehalt der Entscheidung zu tun haben. Jenseits von Zwangsmaßnahmen können sowohl die verantwortliche Stelle als auch Dritte durch flankierende Maßnahmen Einfluss auf die Entscheidung des Betroffenen nehmen. Eine Variante ist die Kopplung der Entscheidung an einen Vertragsschluss oder an monetäre Anreize. Eine solche Kopplung ist grundsätzlich unzulässig.[23] Eine ebenfalls unzulässige Maßnahme ist die tendenziöse Darstellung der möglichen Handlungsalternativen.[24] Alle Handlungsalternativen müssen neutral und mit ihren jeweiligen Folgen dargestellt werden.[25] Ergänzend führt ein Abhängigkeitsverhältnis zwischen Betroffenem und verantwortlicher Stelle zu einer mittelbaren Beeinflussung des Entscheidungsprozesses.[26] Ein Opt-Out in der Arbeitnehmerdatenverarbeitung kann deshalb nicht erwogen werden.

In der Gesamtschau ist ein informiertes Opt-out aus der elektronischen Auskunftserteilung durch ein Datenschutzauskunftssystem zulässig. Notwendige Voraussetzungen sind die objektive und vollständige Unterrichtung über die Konsequenzen der Alternativen sowie die Ermöglichung eines nachträglichen Opt-In.

Entscheidend ist außerdem, dass die Konsequenz eines Opt-out kein vollständiger Verzicht auf jede Transparenz ist. Unterrichtungen und Benachrichtigungen können und müssen weiterhin erfolgen. Des Weiteren kann der Betroffene eine manuelle Teilauskunft verlangen, wenn er nach § 34 Abs. 1 S. 2 BDSG die Art und den Speicherort der personenbezogenen Daten, zu denen er Auskunft verlangt, näher bezeichnet. Eine vollständige Auskunft ist nach dem Opt-out jedoch ausgeschlossen, da die dafür erforderliche Verknüpfung fehlt.

[22] Simitis in: Simitis, BDSG 2014, § 4a Rn. 98 und explizit Art. 7 Abs. 3 DSGVO.
[23] Simitis in: Simitis, BDSG 2014, § 4a Rn. 63 und explizit Art. 7 Abs. 4 DSGVO.
[24] Beliebt im Zuge des Soft-Paternalismus.
[25] Wie in Kapitel 12.4 dargestellt, kann eine neutrale, stark formalisierte Metrik nicht von jedem Betroffenen einfach verstanden werden. Der Aspekt der Informiertheit leidet.
[26] Simitis in: Simitis, BDSG 2014, § 4a Rn. 62.

10.2 Das Recht auf Datenübertragbarkeit

Das Recht auf Datenübertragbarkeit ist ein vollständige Neuschöpfung der DSGVO.[27] Da mit diesem Recht neue Vorstellungen einhergehen, wird, abweichend vom sonstigen Vorgehen, exponiert die künftige Rechtslage erörtert.

Das Recht auf Datenübertragbarkeit ist in Art. 20 DSGVO geregelt und hängt eng mit dem Recht auf Auskunft zusammen, ist mit ihm jedoch nicht identisch.[28] Gemäß Art. 15 Abs. 3 DSGVO steht dem Betroffenen eine Kopie seiner personenbezogenen Daten zu. Im Auskunftsanspruch ist allerdings nicht festgelegt, in welcher Form diese Kopie auszuhändigen ist. Auf der anderen Seite umfasst das Recht auf Auskunft die in Kapitel 3.7 geschilderten Informationen. Das Recht auf Datenübertragbarkeit betrifft nur die gespeicherten personenbezogenen Daten, die der verantwortlichen Stelle vom Betroffenen bereitgestellt wurden. Ein personenbezogenes Datum wurde dann vom Betroffenen bereitgestellt, wenn die verantwortliche Stelle das Datum bei ihm erhoben hat.[29]

Art. 20 Abs. 1 DSGVO fordert, dass der Betroffene die Kopie der personenbezogen Daten in einem strukturierten, gängigen und maschinenlesbaren Format erhält.[30] Ein Format ist strukturiert, wenn es eine festgelegte, für einen Computer oder einen Menschen verständliche Syntax besitzt. Ein Format ist gängig, wenn es standardisiert wurde und im Markt gebräuchlich ist.[31] Ein Format ist maschinenlesbar, wenn es automatisiert durch einen Computer interpretiert werden kann.[32] Kandidaten, die diese Voraussetzungen erfüllen, sind die Extensible Markup Language (XML)[33] und die JavaScript Object Notation (JSON).[34]

Das Recht auf Datenübertragbarkeit gilt gemäß Art. 20 Abs. 1 Lit. a DSGVO nur dann, wenn die Rechtsgrundlage der Datenverarbeitung Einwilligung oder Vertrag sind. Ist die

[27] Laue/Nink/Kremer 2016, § 4 Rn. 59; Kamlah in: Plath, BDSG/DSGVO 2016, Art. 20 Rn. 1; Paal in: Paal/-Pauly, DSGVO 2017, Art. 20 Rn. 28.

[28] Kamlah in: Plath, BDSG/DSGVO 2016, Art. 20 Rn. 2 sieht Art. 20 DSGVO als einen Spezialfall des Art. 15 DSGVO.

[29] Kamlah in: Plath, BDSG/DSGVO 2016, Art. 20 Rn. 6 fasst darunter nur die Stammdaten und in Rn. 7 weitergehend keine Daten, die der Betroffene zur Nutzung von technischen Features selbst eingegeben hat. Dieser engen Sichtweise kann nicht gefolgt werden. Sie widerspricht der wettbewerbs- und datenschutzrechtlichen Zielsetzung der Norm.

[30] Der Gesetzgeber lässt die Definitionen offen und auch die Kommentarliteratur äußert sich noch nicht – Laue/Nink/Kremer 2016, § 4 Rn. 66; Kamlah in: Plath, BDSG/DSGVO 2016, Art. 20 Rn. 8; Paal in: Paal/Pauly, DSGVO 2017, Art. 20 Rn. 19. Insofern stellen die hier eingeführten Definitionsversuche ein Novum dar.

[31] Nach ErwGr 55 DSGVO folgt daraus keine Pflicht zu technisch kompatiblen Datenverarbeitungssystemen.

[32] Entspricht dem Begriff „elektronisch“ in Art. 15 Abs. 3 S. 3 DSGVO, so Schätzle, PinG 2016, 71 (74). Darüber hinaus kann vertreten werden, dass aus dem Begriff folgt, dass die Semantik des Datenformats definiert sein muss.

[33] https://www.w3.org/XML/Core.

[34] https://tools.ietf.org/html/rfc7159 und http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf.

verantwortliche Stelle durch Gesetz zur Datenverarbeitung verpflichtet, gilt das Recht auf Datenübertragbarkeit nicht.

Anstelle dessen, dass der Betroffene eine Kopie seiner personenbezogenen Daten erhält, sollte es ihm nach Art. 20 Abs. 2 DSGVO ergänzend möglich sein, die personenbezogenen Daten direkt von einer verantwortlichen Stelle zu einer anderen übermitteln zu lassen. Dieses weitere Recht steht unter dem Vorbehalt der technischen Machbarkeit. Diese ist objektiv anhand des Stands der Technik zu bestimmen.[35]

Die Direktübertragung und nahtlose Weiternutzung bei einer anderen verantwortlichen Stelle ist nur dann möglich, wenn die verwendeten Formate der verantwortlichen Stellen kompatibel sind. Deshalb fordert ErwGr 68 DSGVO die Verantwortlichen dazu auf, gemeinsame interoperable Formate zu entwickeln. Inkonsequenterweise hält ErwGr 68 gleichzeitig fest, dass sich daraus keine Pflicht begründen lässt „technisch kompatible Datenverarbeitungssysteme zu übernehmen oder beizubehalten." Dadurch sind Vermeidungsstrategien unwilliger verantwortlicher Stellen Tür und Tor geöffnet.

Für das Datenschutzauskunftssystem bedeutet das Recht auf Datenübertragbarkeit zwei Dinge: Zum einen muss auf der Auskunftsplattform eine Downloadmöglichkeit der personenbezogenen Daten in einem Format, das die obigen Bedingungen erfüllt, angeboten werden. Zum anderen muss die Datenstruktur so aufgebaut werden, dass sie sich einfach an mögliche zukünftige Standards anpassen lässt.

10.3 Möglichkeiten des Betroffenen auf der Auskunftsplattform *PrivacyInsight*

Es ist nicht ausreichend, dem Betroffenen immer mehr Informationen und Wahlmöglichkeiten anzubieten. Ausufernde Konvolute textueller Informationen können und wollen die meisten Betroffenen nicht lesen. Nur durch eine knappe und übersichtliche Darstellung der Informationen in einem abgestuften Modell wird das Verständnis des Betroffenen gefördert.[36] Die im Rahmen dieser Arbeit entwickelte[37] Auskunftsplattform *PrivacyInsight*[38] bietet zum Einstieg eine knappe Übersicht und erlaubt bei Interesse ein stufenweises Eintauchen in weitere Informationen. Die Kernfunktion von *PrivacyInsight* ist die Visualisierung der *Informationsflüsse*, der Herkunft-Empfänger-Ketten.

Der Betroffene soll seine Rechte auf Auskunft, Berichtigung, Löschung, Sperrung und Datenübertragung sowie die genannte Wahlmöglichkeit zwischen Transparenz und Unverkettbarkeit an einer Stelle gesammelt wahrnehmen können (Anforderung 47). Eine solche integrierte Sicht ist ein weiterer Aspekt der *PrivacyInsight* auszeichnet.

[35] Laue/Nink/Kremer 2016, § 4 Rn. 65.
[36] Robrecht 2015, 70.
[37] Ein frühes Konzept findet sich bereits in Sommerfeld 2013.
[38] Im Kontext von Kühne 2016 entstanden.

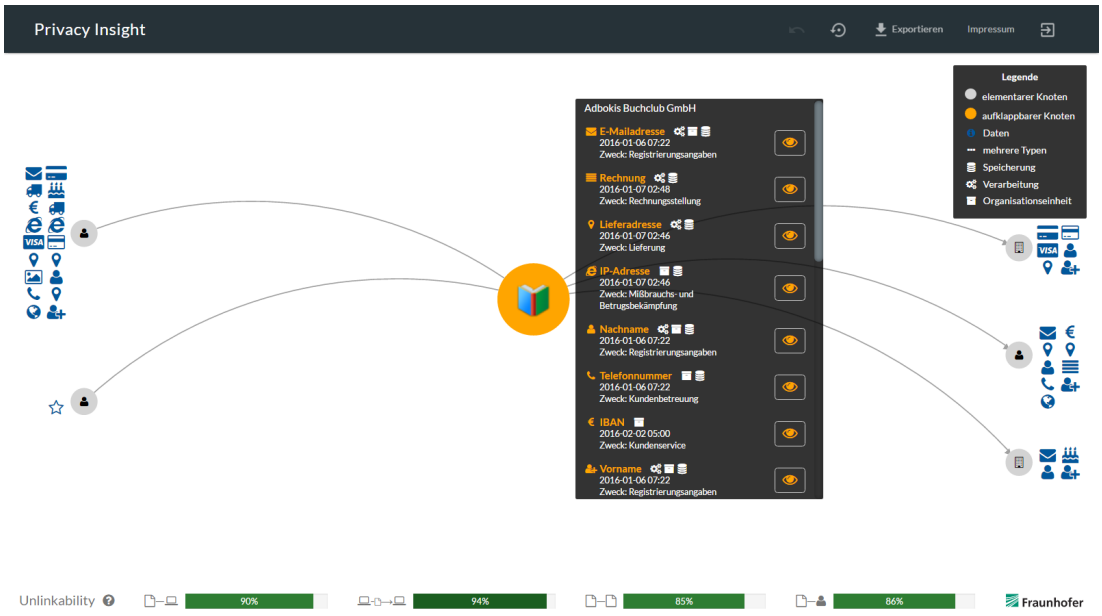


Abbildung 10.1: Die Auskunftsplattform *PrivacyInsight*

PrivacyInsight ist eine Webanwendung.[39] Sie besteht von oben nach unten aus den drei Teilen Navigationsleiste, Visualisierung des Informationsflusses und Statusleiste (Abbildung 10.1). Nach Anmeldung über die vorgelagerte Anmeldemaske wird der Informationsfluss zunächst mit minimalen Details angezeigt. Am linken Fensterrand sind alle Datenquellen, am rechten Rand alle Senken dargestellt. Die verantwortliche Stelle wird als ein Knoten in der Fenstermitte repräsentiert.

Links neben den Quellen und rechts neben den Senken befinden sich jeweils Icons für alle erhobenen beziehungsweise übermittelten personenbezogenen Daten. Die Icons sind abhängig von der Datenkategorie gestaltet. Dies soll das Wiederfinden bestimmter personenbezogener Daten erleichtern. Bei der Auswahl eines Datums wird der Provenance-Baum für dieses Datum hervorgehoben (Anforderung 52). Wird die Detailsicht für ein Datum geöffnet, ergibt sich dort die Möglichkeit das Datum einzusehen, zu exportieren, zu berichtigen, zu sperren oder zu löschen (Anforderungen 47, 48 und 49). Das eigentliche personenbezogene Datum wird erst beim Öffnen dieser Detailansicht aus den speichernden Systemen geladen. Die letzten beiden Optionen werden durch UC-Policies umgesetzt.

Wie oben erwähnt sind in der Informationsflussvisualisierung zu Beginn nur die Erhebung und Übermittlung personenbezogener Daten dargestellt. Verarbeitung, Speicherung und interne Weitergabe sind hinter dem Knoten der verantwortlichen Stelle verborgen. Durch Anwählen (Anklicken oder Antippen) des Knotens öffnet sich die nächste Ebene

[39] Bereits veröffentlicht in Bier/Kühne/Beyerer 2016.

in der Domänenhierarchie,[40] beispielsweise eine Sicht auf die einzelnen Abteilungen und die Informationsflüsse zwischen ihnen. Abhängig von den Vorgaben der verantwortlichen Stelle, insbesondere unter Berücksichtigung ihrer Betriebsgeheimnisse, kann die vollständige Domänenhierarchie bis auf IT-Systemebene aufgerufen oder sogar in Anwendungen hineingeschaut werden. Von besonderem Interesse für den Betroffenen sind Informationsflüsse zu Auftragsdatenverarbeitern. Diese werden im Provenance-Graphen gesondert hervorgehoben.

Für jeden Knoten ist es möglich, ein Kontextmenü aufzurufen, um die in der entsprechenden Domäne, dem System oder der Abstraktionsschicht verarbeiteten oder gespeicherten Daten, gemeinsam mit dem jeweiligen Zweck der Verwendung, einzusehen. Kleine Icons machen deutlich, was mit den Daten im jeweiligen Knoten geschieht. In diesem Kontextmenü kann, wie über die Datenicons am Fensterrand, der Provenance-Baum eines einzelnen Datums ausgewählt werden. Auch die Detailansicht für ein Datum, mit den weiteren Betroffenenrechten, kann dort direkt aufgerufen werden.

Die *Navigationsleiste* bietet Interaktionsmöglichkeiten, die sich nicht direkt auf einzelne Elemente des Informationsflusses beziehen. Sie erlaubt den letzten Schritt beim Eintauchen in den Graphen rückgängig zu machen oder den Graphen auf den Ausgangszustand zurückzusetzen. Außerdem sind in der Navigationsleiste Informationen zur verantwortlichen Stelle hinterlegt. Die Möglichkeit, im Rahmen des Rechts auf Datenübertragbarkeit, die gesamte Auskunft in einem maschinenlesbaren Format (JSON) zu exportieren, findet sich ebenfalls dort (Anforderung 53). Exportmöglichkeiten für die konkreten Daten finden sich bei den jeweiligen Datenkategorien im Hauptfenster.

Die *Statusleiste* visualisiert die vier Unverkettbarkeitsmetriken aus Kapitel 9. Die Unverkettbarkeit wird als prozentualer Statusbalken, farblich codiert von grün für 100 % bis tiefrot für 0 %, dargestellt (Abbildung 10.2). Über einen Informationsbutton sind Erläuterungen zur Metrik zugänglich. Am Ende der Erläuterungen findet sich auch der Opt-Out-Button zur Deaktivierung des automatisierten Provenance-Trackings für alle zukünftigen Erhebungen und Verwendungen personenbezogener Daten durch AdBokis. So ist sichergestellt, dass der Betroffene beim Opt-Out eine informierte Entscheidung trifft.

Beispiel. *Alice wird bei der Abfrage ihrer Auskunft über PrivacyInsight mit den Werten für den Grad der Unverkettbarkeit gemäß der vier Metriken (Identifikationsrelation 85 %; Verknüpfungsrelation 86 %; Speicher- und Verarbeitungsrelation 94 %; Datenflussrelation 90 %) in Form einer Balkengrafik konfrontiert. Mit Unterstützung der hinterlegten Erläuterungen kann sie daraus beispielsweise ableiten, dass es eine Stelle im Unternehmen gibt, bei der sich allein durch die Existenz des Datenschutzauskunftssystems die Unsicherheit, welche personenbezogenen Daten zu ihr gehören, um 15 % reduziert hat. AdBokis bietet ihr über das Frontend die Möglichkeit, in Zukunft auf das Datenschutzauskunftssystem zu verzichten und dafür die Gefahr der Profilbildung zu verringern. Aufgrund der exzellenten und umfangreichen Information in PrivacyInsight ver-*

[40] Siehe Anhang C.

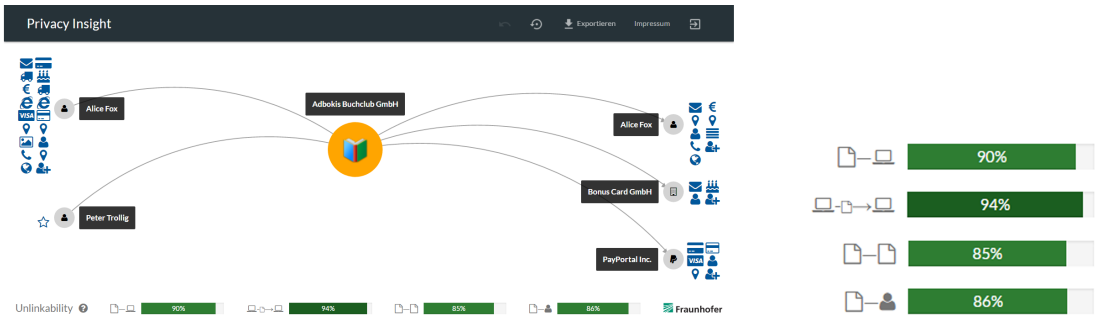


Abbildung 10.2: Die Auskunftsplattform *Privacy Insight* (links) mit integrierter Metrik für Unverkettbarkeit (rechts hervorgehoben)

traut Alice AdBokis und nimmt die zusätzliche Verkettung im Austausch für eine transparente Datenverarbeitung auch in Zukunft in Kauf.

10.4 Zwischenfazit

Im Rahmen der Betroffenenautonomie ist es zulässig, dem Betroffenen eine Opt-Out-Möglichkeit für die elektronische Auskunftserteilung anzubieten (Bestätigung von Hypothese $\mathfrak{H}4$). Dies gilt unter der Voraussetzung, dass der Betroffene anhand der Metriken für Unverkettbarkeit eine freie, informierte und von Erwägungen außerhalb der Datenschutzziele unabhängige Entscheidung treffen kann.

Verzichtet der Betroffene auf die elektronische Auskunftserteilung, ist davon das Recht auf Datenübertragbarkeit mitbetroffen. Die beiden Funktionen sind technisch voneinander abhängig.

Die Auskunftsplattform *PrivacyInsight* ermöglicht dem Betroffenen die Wahrnehmung der Auskunft in einem übersichtlichen, abgestuften Modell. So ist die Information vollständig, ohne die Aufnahmefähigkeit des Betroffenen zu überfordern.

Auf der Auskunftsplattform sind die Metriken für Unverkettbarkeit, das informierte Opt-Out, der Datenexport und die Wahrnehmung der Betroffenenrechte auf Berichtigung, Löschung und Sperrung an einer Stelle integriert. Wie gut Betroffene mit diesen vielfältigen Möglichkeiten arbeiten können, wird in Kapitel 12 überprüft.

Teil IV

Bewertung und Ausblick

11 Rechtliche Bewertung des Datenschutzauskunftssystems

In den Kapiteln 2 bis 4 wurden Kriterien für ein datenschutzgerechtes Auskunftssystem entwickelt. Auf Grundlage der daraus abgeleiteten technischen Anforderungen wurde in den Kapiteln 5 bis 8 ein Datenschutzauskunftssystem konzipiert.

Dieses Kapitel diskutiert, inwiefern die entstandene Implementierung den entwickelten Kriterien des Auskunftsanspruchs gerecht wird (Abschnitt 11.1). Zusätzlich wird dargestellt, welche Konsequenzen die Nichteinhaltung der datenschutzrechtlichen Vorgaben zum Auskunftsanspruch hat (Abschnitt 11.2).

11.1 Erfüllung der datenschutzrechtlichen Kriterien des Auskunftsanspruchs

Die in Kapitel 4.3 aufgestellten datenschutzrechtlichen Kriterien wurden als Leitlinie für die Konzeption des in dieser Arbeit vorgestellten Datenschutzauskunftssystems verwendet. Nach erfolgter technischer Umsetzung stellt sich die Frage, inwiefern das Ergebnis den Kriterien entspricht.

Nachfolgend werden die Kriterien des Auskunftsanspruchs aus Kapitel 4.3.1 diskutiert.

Die ersten Kriterien des Auskunftsanspruchs umreißen, auf welche Art und in welchem Umfang die für die Erfüllung der Auskunft erforderlichen Informationen zu sammeln sind. In der Architekturbeschreibung des Datenschutzauskunftssystems ist hinterlegt, dass PEPs in alle Applikationen integriert werden sollen, die personenbezogene Daten verarbeiten. Wird dies umgesetzt, werden die Ereignisse aller Erhebungs-, Verarbeitungs-, Speicher-, Nutzungs- und Übermittlungsvorgänge erfasst (Kriterium 1). Das Datenschutzauskunftssystem sieht Techniken vor, um personenbezogene Daten so zu erfassen, dass eine vollständige Beauskunftung möglich ist (Kriterium 2). Im Rahmen von Vorarbeiten wurden Verfahren entwickelt, um E-Mails mit unstrukturierten personenbezogenen Daten zu identifizieren[1] sowie strukturierte personenbezogene Daten in einem Webshopsystem zu erfassen.[2] Dadurch ist der Personenbezug ab dem Erhebungszeitpunkt für die

[1] Bier/Prior 2014.
[2] Helwig 2016.

Kriterium	1	2	3	4	5	6	7	8	9	10	11	12	13
Entwurf	X	X	X	X	X	X	X	X	X	X	X	X	X
Implementierung	Y	Y	Y	X		X	X	Y	Y	X	X	X	X

Kriterium	14	15	16	17	18	19	20	21	22	23	24	25	26
Entwurf	X	X	X			X	X	X	X	X	X	X	X
Implementierung	X	X	X				X			X	Y	X	X

Kriterium	27	28	29	30	31	32	33	34	35	36	37	38	39
Entwurf	X	X	X	X	X	X			X	X	X	X	X
Implementierung	X	X	X	X	X	X			X	X	X	X	X

Tabelle 11.1: Erfüllung (X) und teilweise Erfüllung (Y) der datenschutzrechtlichen Kriterien des Auskunftsanspruchs in Entwurf und Implementierung

Auskunft verfügbar (Kriterium 3).[3]

Der Umfang der Provenance ist für die einzelnen Arten des Umgangs mit personenbezogenen Daten im Detail in den Kriterien festgelegt. Der Umfang der Provenance ergibt sich aus dem Umfang des Auskunftsanspruchs.

Erster Aspekt des Auskunftsanspruchs sind die gespeicherten personenbezogenen Daten (Kriterium 8). Sie sind nicht Teil der Provenance, sondern werden in dieser referenziert. Dies ermöglicht, dass die konkreten personenbezogenen Daten erst bei Erteilung einer Auskunft auf Wunsch des Betroffenen aggregiert werden. Die Auskunftsplattform ist so gestaltet, dass der Betroffene individuell entscheiden kann, welche personenbezogenen Daten er einsehen möchte. Dadurch, dass die Provenance die komplette Verarbeitungskette erfasst, kann von jedem Informationspunkt in der Auskunftsplattform, inklusive der Übermittlungen, auf die gespeicherten personenbezogenen Daten zugegriffen werden (Kriterium 9). Um einen Schnellzugriff zu ermöglichen, werden die personenbezogenen Daten in der Auskunftsplattform als Icons hervorgehoben. Die Informationen zu Speicher- und Verarbeitungsvorgängen werden genau so lange gespeichert, wie die Speicherung oder Verarbeitung andauert (Kriterium 6). Diese Logik ist direkt im Datenmodell der Provenance hinterlegt. Jede Repräsentation im Provenance-Modell enthält in ihren Attributen Angaben zur Domäne, also dem Ort der Speicherung oder Verarbeitung, und zu Applikationen und Speicherpfaden, die für die Auskunft von Relevanz sind (Kriterium 11 und 12). Welche Applikationen und Speicherpfade Teil der Provenance sind, wird über Abstraktionsregeln gesteuert. Eine Sperrung personenbezogener Daten wirkt sich, gesteu-

[3] Mit Ausnahme der in Kriterium 4 vorgesehenen Fälle.

ert durch entsprechende UC-Policies, nur auf die Verwendung durch die verantwortliche Stelle, nicht auf die Auskunft aus (Kriterium 10).

Datenverarbeitungsvorgänge und die in sie eingehenden Daten werden ebenso erfasst wie die Speicherung personenbezogener Daten (Kriterium 34). Der logische Aufbau einer Datenverarbeitung, im Falle der automatisierten Einzelentscheidung, wird in der Provenance allerdings nicht erfasst (Kriterium 33).

Für die Herkunft und externe Empfänger personenbezogener Daten sind spezielle Repräsentationen vorgesehen (Kriterium 13, 15, 19 und 24). In ihnen werden die Außenbeziehungen der verantwortlichen Stelle gespeichert, so dass sie dem Betroffenen als Teil der Auskunft mitgeteilt werden können. Auf der Auskunftsplattform sind Dritte sowie der Betroffene besonders hervorgehoben. Die internen Datenflüsse sind in den Kanten des Provenance-Graphen und in der Auskunft enthalten (Kriterium 14 und 15). Die Kanten des Provenance-Graphen werden auf der Auskunftsplattform dargestellt. Eine Kategorisierung personenbezogener Daten findet nicht statt (Kriterium 16).

Das Datenschutzauskunftssystem kann nur die automatisierte Datenverarbeitung erfassen. Insofern müssen nicht-automatisierte Weitergaben nach Kriterium 17 händisch zur Provenance hinzugefügt werden. Die Information, wann in der Vergangenheit eine Auskunft erteilt wurde, muss nicht durch das Datenschutzauskunftssystem erfasst werden (Kriterium 5). Es ist ausreichend, wenn in die Auskunftsplattform ein Zugriffsprotokoll integriert wird, auf das der Betroffene nach seinem Log-In zugreifen kann. Das Datenschutzauskunftssystem nutzt für den Log-In keine auskunftsspezifischen Identifikatoren, zeichnet aber die bei der Erhebung mitgeteilten auf (Kriterium 7).

Jede verantwortliche Stelle ist nur für ihre eigene Datenverarbeitung verantwortlich. Insofern kann Kriterium 18, die beidseitige Protokollierung einer Übermittlung, nur für die jeweils eigene Rolle im Übermittlungsvorgang eingefordert werden. Darüber hinausgehend können sich verantwortliche Stellen nach Art. 26 DSGVO zu einer gemeinsamen verantwortlichen Stelle zusammenschließen. Indem die Domänenhierarchie des Datenschutzauskunftssystems verknüpft wird, können dann auch Übermittlungsvorgänge vollständig nachvollzogen werden. Solange kein gemeinsames Datenschutzauskunftssystem etabliert ist, ist unbekannt, wann Dritte personenbezogene Daten löschen. Deshalb werden Herkunfts- und Empfängerangaben im in dieser Arbeit präsentierten Datenschutzauskunftssystem gar nicht gelöscht (Kriterium 22 und 26).[4]

Bei einer Veröffentlichung personenbezogener Daten sind Informationen über Beginn und Ende der Veröffentlichung (Kriterium 20) durch die Repräsentation der Daten auf der Veröffentlichungsplattform gegeben. Der Abruf personenbezogener Daten über eine zugangsgeschützte Plattform kann wie jede Übermittlung vom Datenschutzauskunftssystem erfasst werden (Kriterium 21).

Für jeden Umgang mit personenbezogen Daten sind die jeweiligen Zwecke in die Aus-

[4]Die speziellen Zeitangaben der Kriterien 23 und 25 sind damit obsolet.

kunft mit aufzunehmen (Kriterium 27 und 29). Die Datenstruktur der Provenenance lässt die Aufnahme der Zwecke zu. Eine Zweckänderung ist in jedem Datenverwendungsschritt möglich (Kriterium 28). Vorhergehende Zwecke sind anhand des Provenance-Graphen weiterhin rückverfolgbar. Wird kein neuer Zweck definiert, erbt ein Vorgang seinen Zweck vom vorherigen Vorgang. Da der Zweck Teil der Provenance ist, ist er unabhängig von der Speicherung der personenbezogenen Daten (Kriterium 30).

Der Zeitpunkt jeder Erhebung und jedes Datenverwendungsvorgangs ist Teil der Provenance (Kriterium 31). Durch Abstraktionsregeln kann verhindert werden, dass Zeitpunkte für zu detaillierte Datenverarbeitungsvorgänge aufgezeichnet werden (Kriterium 32).

Da eine Auskunft die Transparenz der Datenverarbeitung für den Betroffenen stärken soll, ist eine angemessene und verständliche Darstellung von großer Bedeutung. Die letzten Kriterien des Auskunftsanspruchs machen deshalb Vorgaben für die Benutzerschnittstelle einer Auskunftsplattform. *PrivacyInsight* erfüllt alle diese Kriterien.

PrivacyInsight ist eine Webapplikation die von jedem aktuellen Endgerät aus aufgerufen werden kann (Kriterium 39). Der Log-In ist nicht streng an den Betroffenen gebunden. Benutzername und Passwort sind nur eine Variante der Authentifikation und Autorisierung. Über weitere Zugriffskanäle ist deshalb auch ein Stellvertreterzugriff möglich (Kriterium 36). *PrivacyInsight* stellt den Betroffenen und seine Wünsche in den Mittelpunkt. Selbst nach dem Log-In geschieht jede Bereitstellung personenbezogener Daten nur auf Veranlassung des Betroffenen (Kriterium 35). Alle Interaktions- und Wahlmöglichkeiten sind in einer Oberfläche integriert. Für die Benutzerfreundlichkeit von *PrivacyInsight* wird auf Kapitel 12 verwiesen (Kriterium 38). Steht für den Betroffenen keine Auskunft zur Verfügung, da keine personenbezogenen Daten erhoben wurden, erhält er eine entsprechende Meldung (Kriterium 37).

Das Datenschutzauskunftssystem erfüllt die rechtlichen Kriterien des Auskunftsanspruchs bis auf wenige Ausnahmen. Auskünfte über nicht-automatisierte Verarbeitungsvorgänge gemäß Kriterium 17 müssen manuell und einzelfallbezogen behandelt werden. Auskünfte über vorangegangene Auskünfte (Kriterium 5) sind im Prototypen von *PrivacyInsight* noch nicht integriert, aber technisch trivial. Informationen zum logischen Aufbau der Verarbeitung bei automatisierten Einzelentscheidungen nach Kriterium 33 werden vom, in dieser Arbeit vorgestellten, Datenschutzauskunftssystem nicht unterstützt. Geeignete Konzepte aus dem wissenschaftlichen Rechnen sind in der Literatur vorhanden.[5] Die Rechtsprechung legt den dahingehenden Auskunftsanspruch allerdings so eng aus,[6] dass eine Funktions-Provenance[7] für die Umsetzung desselben nicht angemessen wäre.

[5] Bose/Frew 2005; Freire et al. 2008; Moreau/Groth et al. 2008.
[6] Vgl. Kapitel 3.7.6
[7] Cheney/Chiticariu/Tan 2009; vgl. Kapitel 5.1.

11.2 Konsequenzen einer unvollständigen oder fehlerhaften Auskunft

Eine effektiver Grundrechtsschutz setzt wirksame Sanktionen voraus[8] Erteilt eine verantwortliche Stelle, sei es mit oder ohne Unterstützung eines Datenschutzauskunftssystems, eine unvollständige oder fehlerhafte Auskunft, dann handelt sie nach § 43 Abs. 1 Nr. 8a BDSG ordnungswidrig und kann nach § 43 Abs. 3 S. 1 BDSG in jedem einzelnen Fall mit einer Geldbuße von bis zu 50 000 € belegt werden. Ein wirtschaftlicher Vorteil, der nach § 43 Abs. 3 S. 2 u. 3 BDSG einen höheren Betrag rechtfertigen würde, ist bei einer fehlerhaften Auskunft im Regelfall nicht anzunehmen.

Mit Einführung der DSGVO wird die mögliche Höhe einer Geldbuße deutlich nach oben angepasst. Die nach Art. 58 Abs. 2 Lit. i DSGVO zuständige Aufsichtsbehörde kann nach Art. 83 Abs. 5 Lit. b DSGVO eine Geldbuße von bis zu 20 Mio. Euro oder bis zu 4 % des weltweit erzielten Jahresumsatzes, je nachdem welcher Betrag höher liegt, verhängen.

Damit ist allerdings noch nicht das Durchsetzungsdefizit des Datenschutzes behoben. Die personelle Ausstattung der Aufsichtsbehörden lässt keine Vielzahl von Bußgeldverfahren erwarten.[9] Eine Alternative könnten wettbewerbsrechtliche und verbraucherrechtliche Sanktionen, insbesondere ein im Rahmen einer Abmahnung nach § 12 Abs. 1 S. 1 UWG vertragsstrafenbewehrt durchgesetzter Anspruch auf Beseitigung und Unterlassung, sein.

Solche Sanktionen könnten zunächst nur für die Verarbeitung von Verbraucherdaten[10] durch Unternehmer[11] wirksam sein. Des Weiteren ist die grundsätzliche Abmahnbarkeit von Verstößen gegen das Datenschutzrecht auf Grundlage von § 4 Nr. 11 UWG a. F.[12] und § 3a UWG n. F.[13] in der Rechtsprechung umstritten. Fraglich ist, ob eine unvollständige oder fehlerhafte Auskunft eine gesetzlichen Vorschrift ist, die „auch dazu bestimmt ist, im Interesse der Marktteilnehmer das Marktverhalten zu regeln“.[14] Marktverhalten ist jede Tätigkeit auf dem Markt, durch die ein Unternehmer auf Verbraucher, sonstigen Marktteilnehmer oder Mitbewerber einwirkt.[15] Ein Verstoß gegen solch eine Vorschrift ist nach § 3a UWG eine unlautere Wettbewerbshandlung, soweit der „Verstoß geeignet ist, die Interessen von Verbrauchern, sonstigen Marktteilnehmern oder Mitbewerbern spürbar

[8] BVerfGE 125, 260 (339).
[9] Schulzki-Haddouti 2016, 114 ff.
[10] Verbraucherbegriff des § 13 BGB.
[11] Definiert in § 14 BGB.
[12] Unlauter handelt insbesondere, wer „11. einer gesetzlichen Vorschrift zuwiderhandelt, die auch dazu bestimmt ist, im Interesse der Marktteilnehmer das Marktverhalten zu regeln.“
[13] Vorschrift neugefasst durch das zweite Gesetz zur Änderung des Gesetzes gegen den unlauteren Wettbewerb vom 02.12.2015 (BGBl. I 2015 Nr. 49 S. 2158), in Kraft getreten am 10.12.2015, http://dipbt.bundestag.de/extrakt/ba/WP18/647/64799.html. in Klarstellung der Umsetzung der Richtlinie 2005/29/EG.
[14] Gleichlautend in § 4 Nr. 11 UWG a. F. und § 3a UWG n. F.
[15] KG Berlin, MMR 2011, 464 (465).

zu beeinträchtigen."[16]

Die ablehnende Meinung vertritt, dass die Verarbeitung personenbezogener Daten eines Verbrauchers durch einen Unternehmer sowie die daran anknüpfenden Pflichten des Unternehmers dessen Marktauftritt nicht unmittelbar betreffen.[17] Beispielsweise gelten die Informationspflichten des § 13 Abs. 1 TMG einem Verhalten, das dem Marktverhalten vorausgeht.[18] Sie könnten nur als Marktverhaltensvorschrift angesehen werden, wenn ihnen eine sekundäre Schutzfunktion im Hinblick auf die Wettbewerber innewohnen würde.[19] Der Gesetzgeber habe jedoch bei der Gesetzgebung keinen überindividuellen wettbewerblichen Schutzgedanken verfolgt. Der freie Wettbewerb sei in der Begründung zur Vorschrift[20] nur im Sinne einer Rechtfertigung der Einschränkung von Persönlichkeitsrechten der Nutzer berücksichtigt worden.[21] Da sich datenschutzbezogene im Gegensatz zu geschäftsbezogenen Informationspflichten nicht auf das kommerzielle Verhalten eines Verbrauchers auswirken, seien auch die Verbraucherinteressen nicht spürbar beeinträchtigt.[22] Der gleiche Schluss würde sich für Auskunftspflichten aufdrängen.

Die zustimmende Meinung sieht marktbezogene Datenschutzregeln als Marktverhaltensregeln für Unternehmer. Sie bezieht sich dabei auf die Begründung der DSRL. Die DSRL soll danach den grenzüberschreitenden Verkehr personenbezogener Daten auch im Hinblick auf einen fairen Wettbewerb regeln.[23] Die Vorschriften dienen demgemäß nicht nur dem Schutz der Persönlichkeitsrechte der Betroffenen, sondern auch dem Schutz der Interessen der Mitbewerber an gleichen Wettbewerbsbedingungen.[24] Aufklärungspflichten tragen außerdem zum Schutz der Verbraucherinteressen bei der Marktteilnahme bei und beeinflussen die Entscheidungen und das kommerzielle Verhalten der Verbraucher.[25] Eine Vorschrift zum Schutz anderer Rechtsgüter eines Verbrauchers ist eine Marktverhaltensvorschrift, wenn das geschützte Interesse durch die Marktteilnahme berührt wird.[26] Dem ist zu folgen. Die zunehmende wirtschaftliche Bedeutung der Erhebung und Verarbeitung personenbezogener Daten beeinflusst das Verständnis des Datenschutzrechts als Marktverhaltensregel in positivem Sinne.[27]

Allerdings ist der Auskunftsanspruch, im Gegensatz zu den Informationspflichten, zu-

[16] Das Prinzip der Spürbarkeit war schon zuvor im UWG enthalten und wird nun explizit an dieser Stelle erwähnt. – Insofern ändert sich nichts ggü. § 4 Nr. 11 UWG a. F.
[17] KG Berlin, MMR 2011, 464 (465).
[18] KG Berlin, a. a. O.
[19] KG Berlin, a. a. O.
[20] BT-Drs. 13/7385, S. 21 zum TDDSG.
[21] KG Berlin, a. a. O.
[22] LG Frankfurt, BeckRS 2014, 22875; ebenso OLG München, MMR 2012, 317.
[23] U. a. ErwGr 7 und 8.
[24] OLG Hamburg, K&R 2013, 601.
[25] OLG Hamburg, a. a. O. und ebenso LG Köln, Beschluss vom 26.11.2015, Az. 33 O 230/15; LG Hamburg, Beschluss vom 07.01.2016, Az. 315 O 550/15; LG Hamburg, Beschluss vom 13.03.2016 Az. 312 O 127/16.
[26] OLG Karlsruhe, NJW 2012, 3312 (3314).
[27] LG Düsseldorf, MMR 2016, 328 (330).

mindest der Verbraucherentscheidung zum Aufbau einer Kundenbeziehung nicht vorgelagert. Eine Auskunft soll ferner, im Gegensatz zu beispielsweise einem Werbeschreiben,[28] nicht unmittelbar das Verhalten des Verbrauchers im Markt beeinflussen. Der Unternehmer erteilt die Auskunft nicht initiativ, um sich einen wettbewerbsrechtlichen Vorteil zu verschaffen, sondern responsiv, in Reaktion auf die Anfrage des Verbrauchers. Allenfalls in den Fällen, in denen der Unternehmer den Umgang mit personenbezogenen Daten aus Imagegründen systematisch fehlerhaft darstellt, kann von einer mittelbaren marktbezogenen Beeinflussung des Verbraucherverhaltens ausgegangen werden.

Augenscheinlich eindeutiger stellt sich die Situation im Hinblick auf das UKlaG dar. Mit dem Gesetz zur Verbesserung der zivilrechtlichen Durchsetzung von verbraucherschützenden Vorschriften des Datenschutzrechts vom 17.02.2016[29] wurden die Vorschriften des BDSG in den Kanon der Verbraucherschutzgesetze aufgenommen,[30] die die Zulässigkeit der Erhebung, Verarbeitung oder Nutzung personenbezogener Daten eines Verbrauchers durch einen Unternehmer regeln. Bei einer Zuwiderhandlung gegen diese Vorschriften kann ein Unternehmer nach § 2 Abs. 1 UKlaG auf Unterlassung und Beseitigung in Anspruch genommen werden. Beseitigung meint die Maßnahmen gemäß der datenschutzrechtlichen Vorschriften auf Löschung, Sperrung und Berichtigung wie in § 35 BDSG zu finden.[31]

Fraglich ist, ob der Auskunftsanspruch eine Regelung ist, die in den Anwendungsbereich des UKlaG fällt. Nach der Gesetzesbegründung werden alle datenschutzrechtlichen Vorschriften erfasst, „die Unternehmer *für* eine zulässige Erhebung, Verarbeitung oder Nutzung von Verbraucherdaten [...] beachten müssen."[32] Eine Erhebung, Verarbeitung und Nutzung personenbezogener Daten ist nach der Maßgabe von § 4 Abs. 1 BDSG zulässig. Sie wird nicht dadurch unzulässig, dass eine Auskunft nach § 34 BDSG fehlerhaft oder unvollständig ist. Die Auskunft ist nicht kausal für die Datenpreisgabe.[33] § 34 BDSG ist eine Vorschrift, die Unternehmer *bei* der zulässigen Verarbeitung von Verbraucherdaten beachten müssen. Somit ist im Hinblick auf § 34 BDSG kein Unterlassungsanspruch gegen den Rechtsverstoß einer unvollständigen Auskunft gegeben.

Ungeachtet all dessen eignet sich die Vollständigkeit und Korrektheit der Auskunft kaum für eine Abmahnung. Erstens fallen die Anspruchsberechtigten der Auskunft (Betroffene), die die Vollständigkeit und Korrektheit überprüfen könnten, und die Anspruchsberechtigten der Unterlassung[34] auseinander. Zweitens ist eine Auskunft individuell und geht nicht einer Vielzahl von Verbrauchern in gleichartiger oder zumindest vergleichbarer

[28] OLG Stuttgart, MMR 2007, 437.
[29] BGBl. I 2016 Nr. 8 S. 233, http://dipbt.bundestag.de/extrakt/ba/WP18/651/65144.html.
[30] Genauer: In § 2 Abs. 2 S. 1 Nr. 11 UKlaG.
[31] BT-Drs. 18/6916, 8.
[32] BT-Drs. 18/4631, 23.
[33] Im Gegensatz zu den Hinweisen gemäß § 4a Abs. 1 S. 1 BDSG.
[34] Nach § 3 Abs. 1 S. 1 UKlaG Verbraucherverbände, Verbände zur Förderung gewerblicher oder selbständiger beruflicher Interessen, Industrie- und Handelskammern und die Handwerkskammern.

Weise zu, wie dies beispielsweise für Allgemeine Geschäftsbedingungen (AGB) der Fall ist. Es ist somit nur schwer möglich, eine Unterlassungsaufforderung mit hinreichender Präzision zu formulieren.

11.3 Zwischenfazit

Das in dieser Arbeit vorgestellte Datenschutzauskunftssystem erfüllt alle rechtlichen Kriterien des Auskunftsanspruchs und damit das Konstruktionsziel $\mathfrak{K}6$. Der Einsatz eines Datenschutzauskunftssystems ist unumgänglich, um eine Auskunft in modernen, vernetzten IT-Systemen in vollem Umfang erteilen zu können. Durch die Wahlmöglichkeit des Betroffenen zwischen Unverkettbarkeit und Transparenz, anhand einer in die Auskunftsplattform integrierten Metrik, ist der Einsatz eines Datenschutzauskunftssystems verhältnismäßig.

Ob in Zukunft ein Handlungsdruck auf verantwortliche Stellen entsteht, Auskunftsansprüche vollumfänglich zu befriedigen, hängt von der Durchsetzung des Anspruchs ab. Die Aufsichtsbehörden bekommen mit der DSGVO stärkere Sanktionsmöglichkeiten in die Hand. Ihre personelle Ausstattung und ihr Selbstbild als kooperative Institution ändert sich dadurch nicht.

Durch Wettbewerbs- und Verbraucherrecht ist keine Verbesserung der Betroffenenrechte zu erwarten. Der Auskunftsanspruch ist keine Voraussetzung für eine zulässige Erhebung, Verarbeitung oder Nutzung personenbezogener Daten. Demnach können Unternehmen weder nach UWG noch nach UKlaG abgemahnt werden.

12 Nutzerrezeption des Datenschutzauskunftssystems

Unabhängig von seiner technischen und rechtlichen Güte ist das Datenschutzauskunftssystem nur dann sinnvoll, wenn es von den potentiellen Nutzern des Systems, den Betroffenen der Datenverarbeitung, erfolgreich eingesetzt werden kann. Deshalb wurde die Auskunftsplattform des Datenschutzauskunftssystems, *PrivacyInsight*,[1] in einer Nutzerstudie auf ihre Gebrauchstauglichkeit hin überprüft.[2]

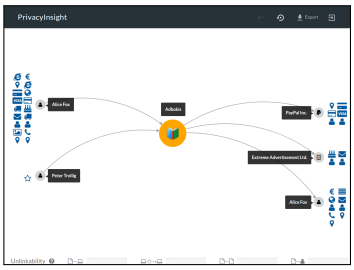
Abschnitt 12.1 stellt den Aufbau und Ablauf der Studie und die Zusammensetzung der für die Studie ausgewählten Probanden vor. Der Vergleich von *PrivacyInsight* mit bestehenden Ansätzen zur Auskunft wird in Abschnitt 12.2 vorgestellt. Die Ergebnisse der Evaluation des erweiterten Funktionsumfangs von *PrivacyInsight* findet sich in Abschnitt 12.3. Ergänzend wurden in der Studie auch Aspekte der Unverkettbarkeitsmetrik aus Kapitel 9 und 10 abgefragt. Die Einschätzung der Probanden wird in Abschnitt 12.4 wiedergegeben. Den Abschluss in Abschnitt 12.5 bildet die Analyse allgemeiner Fragen zur Erwartungshaltung der Betroffenen.

12.1 Struktur der Studie

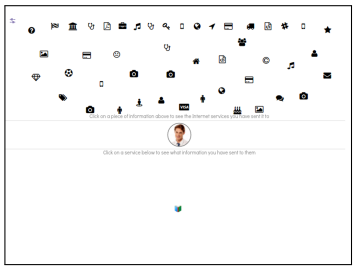
Mit der vorliegenden Nutzerstudie wurden drei Ziele verfolgt. Erstens sollte überprüft werden, ob Informationen über die Verarbeitung personenbezogener Daten mit *PrivacyInsight* genauso effizient und effektiv gefunden werden können, wie bei bestehenden Auskunftsverfahren. Probanden wurden dazu Aufgaben gestellt. Die Quote ihrer erfolgreichen Lösungen sowie ihr Zeitbedarf wurden gemessen. Zweitens sollte evaluiert werden, wie die Gebrauchstauglichkeit von *PrivacyInsight*, auch im Vergleich mit den bereits bestehenden Verfahren, durch Anwender empfunden wird. Mit Hilfe zweier etablierter Usability-Fragebögen wurde deshalb die subjektive Einschätzung der Probanden abgefragt. Drittens war die Frage zu beantworten, ob die zusätzlichen Funktionen von *PrivacyInsight*, insbesondere die Metrik für Unverkettbarkeit, von den Anwendern als nutzbringend empfunden werden. Dafür wurde ein zusätzlicher, auf die Studie abgestimmter Fragebogen hinzugezogen.

[1] Zum Funktionsumfang, siehe Kapitel 10.3

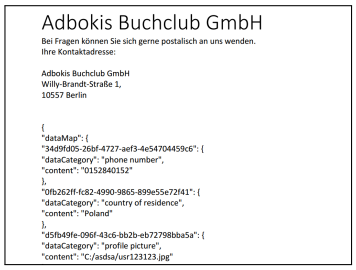
[2] Die Studie wurde bereits in Bier/Kühne/Beyerer 2016 veröffentlicht.



(a) PrivacyInsight



(b) GenomSynlig



(c) JSON-Dokument

Abbildung 12.1: Auskunftsverfahren im Vergleich

Als Vergleichsmaßstab für die ersten beiden Ziele der Studie wurden zwei Auskunftsverfahren gewählt: (1) Ein *JSON-Dokument* (Abbildung 12.1c), welches den Stand der Praxis für Datenschutzauskünfte repräsentiert[3] und (2) *GenomSynlig* (Abbildung 12.1b), die Auskunftsplattform von A4Cloud und das derzeit beste, in der Forschung verfügbare Privacy-Dashboard. Von *GenomSynlig* wurde die mit *PrivacyInsight* vergleichbare Ansicht „Trace View“ herangezogen.[4]

Das JSON-Dokument besteht aus 45 Seiten im PDF-Format. Am Anfang des Dokuments befindet sich ein Abschnitt mit der Firma und einer postalischen Kontaktadresse der verantwortlichen Stelle. Auf den übrigen Seiten werden die personenbezogenen Daten sowie die verarbeitenden und empfangenden Stellen aufgelistet. Die Struktur entspricht der Datenstruktur in Kapitel 6.4.

GenomSynlig bietet die Möglichkeit unterschiedliche Filter zu setzen. In der Evaluation wurde es so konfiguriert, dass nur ein Unternehmen, die AdBokis GmbH, angezeigt wurde. Dadurch wurde die Übersichtlichkeit für die Aufgabenstellung erhöht.

12.1.1 Stichprobe

An der fünftägigen Studie im Frühjahr 2016 nahmen insgesamt 31 Personen teil. Die Probanden wurden aus dem Umfeld der Karlsruher Forschungslandschaft rekrutiert. Deshalb hatten 74 % der Probanden einen akademischen Abschluss. Alle Probanden hatten zumindest das Abitur, die Fachhochschulreife oder einen Abschluss der erweiterten Oberschule. 32 % der Teilnehmer waren Studenten, 65 % waren erwerbstätig. Da Studenten bereits

[3] Beispielsweise als Downloadformat im Google Dashboard, https://www.google.com/dashboard, abgerufen am 9. Mai 2017; ähnlich komplex sind auch die Facebook-Datenbestände, http://www.europe-v-facebook.org/DE/Datenbestand/datenbestand.html, abgerufen am 9. Mai 2017; die papierenen Auskünfte deutscher Unternehmen (siehe Kapitel 1.1.1) sind meist tabellarisch strukturiert, aber keineswegs deutlich übersichtlicher.

[4] http://hci.cse.kau.se:8000/Datatrack_views/datatrack-traceview.html, eine kurze Einführung zum allgemeinen Funktionsumfang findet sich in Kapitel 1.1.2.

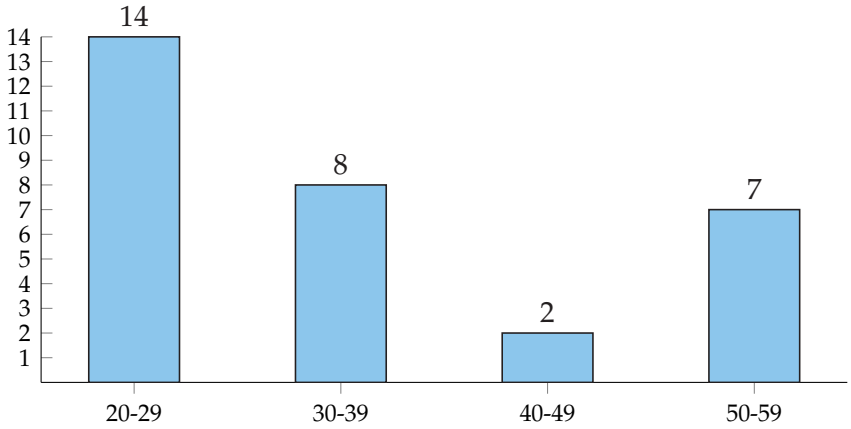


Abbildung 12.2: Altersverteilung der Stichprobe

einen Bachelorabschluss haben können, sind sie nicht äquivalent mit den Probanden ohne akademischen Abschluss.

39 % der Teilnehmer waren weiblich, 61 % männlich. 65 % der Teilnehmer kamen aus einer Großstadt oder ihrem Einzugsbereich. Die übrigen Teilnehmer verteilten sich relativ gleichmäßig auf Städte, Kleinstädte und den ländlichen Bereich.

Die Teilnehmer besaßen ausgeprägtes technisches Hintergrundwissen. 68 % der Teilnehmer können technische Probleme fast immer selbst lösen, 19 % zumindest manchmal. 29 % der Teilnehmer haben bereits selbst eine Webanwendung entwickelt. Allerdings hatten nur 19 % der Teilnehmer in der Vergangenheit bereits selbst ein oder mehrmals ein Auskunftsgesuch nach § 34 BDSG gestellt.

Im Vergleich zur Gesamtbevölkerung ist die Stichprobe überdurchschnittlich gebildet und technisch affin. Außerdem ist die Quote männlicher Teilnehmer höher als die weiblicher Teilnehmer. Auch die Altersverteilung ist nicht repräsentativ. Der Anteil der Internetnutzer an der deutschen Gesamtbevölkerung, die 45 Jahre oder älter sind, liegt bei ca. 46 %[5] während der Anteil der über 45 jährigen in der Stichprobe bei höchstens 29 % (9 Personen, Abbildung 12.2) liegt.

Auf der anderen Seite entspricht die Gesamtbevölkerung auch nicht den ersten Nutzern einer neuen Datenschutztechnologie. Erstnutzer für Datenschutztechnologien sind überwiegend männlich, jung und gebildet.[6] Insofern sind die Ergebnisse der Studie mit der gewählten Stichprobe ein guter Indikator für die zukünftige Akzeptanz einer Auskunftsplattform. Die Gebrauchstauglichkeit wird anhand einer Personengruppe bestimmt, die das System tatsächlich verwenden würde.

[5] Statistisches Bundesamt 2015, 206.
[6] Spiekermann 2005.

12.1.2 Versuchsaufbau und Ablauf

Die Nutzerstudie fand unter Laborbedingungen statt. Jeder Teilnehmer bearbeitete die Aufgaben und Fragebögen alleine und unter Beobachtung des Studienleiters. Im Labor wurden zwei getrennte Arbeitsplätze genutzt.[7] Am einen Arbeitsplatz wurden die gestellten Aufgaben mit den drei elektronischen Auskunftvarianten bearbeitet, am anderen Arbeitsplatz wurden die schriftlichen Fragebögen ausgefüllt. Diese Konfiguration erlaubte eine Vorbereitung der Aufgaben während der Bearbeitung der Fragebögen durch den Probanden.

Nach einer organisatorischen Einführung wurde den Probanden das Szenario erläutert. Es entsprach im Großen und Ganzen dem Minimalbeispiel aus Kapitel 1.6. Für *GenomSynlig* wurden leichte Anpassungen an den Rollen vorgenommen. Der Kunde hieß bei Tests mit diesem System, dem voreingestellten Namen im online verfügbaren Prototypen entsprechend, Bob Bobsson. Die genaue Szenariobeschreibung findet sich im Anhang F.1.

Der Versuch bestand organisatorisch aus zwei Aufgabenteilen. Zwischen und nach diesen Teilen waren die Fragebögen auszufüllen.

Im ersten Teil wurden die drei Auskunftsverfahren *PrivacyInsight*, *GenomSynlig* und JSON miteinander verglichen. Um einen Einfluss der Reihenfolge der Systeme auf die Ergebnisse auszuschließen, wurden die Auskunftsverfahren in allen 6 möglichen Permutationen getestet. Jedem Probanden wurde eine Permutation zugewiesen. So wurde jede Permutation mindestens 5 Mal getestet. Zumindest im ersten Teil war den Probanden dadurch nicht klar, welches System das im Rahmen dieser Arbeit entwickelte war.

Den Probanden wurden für jedes System 5 Aufgaben gestellt,[8] die innerhalb von maximal 2 Minuten zu bearbeiten waren. Die Bearbeitungszeit und der Lösungserfolg wurden protokolliert. Wurde die Aufgabe nicht innerhalb von 2 Minuten abgearbeitet, galt sie als nicht gelöst. Nach dem Ende der Bearbeitung aller Aufgaben hatte der Proband jeweils noch eine Minute Zeit sich mit dem System vertraut zu machen. Anschließend wurde er angewiesen, einen Usability-Fragebogen auszufüllen.

Im zweiten Teil wurden die besonderen Funktionen von *PrivacyInsight* getestet. Der zweite Teil begann mit einer Einführung in die Funktionen und die Benutzung von *PrivacyInsight* durch den Studienleiter. Anschließend hatte der Proband 10 Aufgaben zu bearbeiten,[9] die über den Funktionsumfang von *GenomSynlig* und JSON hinausgehen. Dem Probanden wurden dadurch die Möglichkeiten von *PrivacyInsight* demonstriert. Für jede Aufgabe waren wieder 2 Minuten vorgesehen. Die Ergebnisse wurden protokolliert. Nach Bearbeitung aller Aufgaben hatte der Proband nochmals 2 Minuten Zeit, um sich mit *PrivacyInsight* vertraut zu machen. Anschließend waren zwei weitere Usability-Fragebögen auszufüllen.

[7] Blickmesslabor des Fraunhofer IOSB in Karlsruhe.
[8] Siehe Abbildung F.3 im Anhang.
[9] Siehe Abbildungen F.5 und F.6 im Anhang.

Ein Usability-Fragebogen war der kompakte *System Usability Scale (SUS)*.[10] Der aus 10 Items bestehende SUS ist eine technologieunabhängige Likert-Skala. Das Antwortformat ist auf 5 Stufen von 1 bis 5 festgelegt. Die daraus berechnete Skala kann Werte zwischen 0 und 100 annehmen. Diese Werte lassen sich in ein Bewertungssystem von „schlechtestmöglich" bis „bestmöglich" übersetzten.[11] Während der SUS ursprünglich eindimensional war, stellten Lewis und Sauro fest, dass „Erlernbarkeit" als zweiter Faktor abgeleitet werden kann.[12] Der SUS wurde im ersten Teil dazu verwendet, um die subjektive Gebrauchstauglichkeit der drei Systeme zur Auskunftserteilung miteinander zu vergleichen. Der SUS im zweiten Teil sollte, im Vergleich mit den Ergebnissen aus dem ersten Teil, zeigen, wie sich die Wahrnehmung der Nutzer durch die verbesserte Kenntnis von *PrivacyInsight* verändert hat.

Der in der Literatur vorgestellte PET-USES-Fragebogen[13] wurde nicht verwendet, da viele Fragen von PET-USES nicht zu Auskunftsplattformen passen. Der erste Teil von PET-USES hat eine hohe Übereinstimmung mit dem SUS.

Der zweite, im zweiten Teil verwendete, Fragebogen war der umfangreichere *User Experience Questionnaire (UEQ)*.[14] Der UEQ besteht aus 26 Items mit einem 7-stufigen Antwortformat semantischer Differentiale.[15] Mit dem UEQ wird der subjektive Eindruck eines getesteten Systems auf einen Probanden abgefragt.[16] Die 6 Skalen des UEQ sind Attraktivität, Effizienz, Durchschaubarkeit, Steuerbarkeit, Stimulation und Originalität. Anhand dieser Skalen wurde die Nutzerrezeption von *PrivacyInsight* genauer analysiert.

Zum Abschluss der Studie wurden jedem Teilnehmer noch ein demographischer und ein ergänzender Fragebogen vorgelegt. Auf die demographischen Angaben wurde im vorhergehenden Abschnitt eingegangen. Die Antworten auf die ergänzenden Fragen werden in den Abschnitten 12.4 und 12.5 besprochen.

Der genaue Ablauf eines Probandendurchlaufs findet sich in Anhang F.1. Die gestellten Aufgaben finden sich im Anhang F.2, die verwendeten Fragebögen im Anhang F.3.

12.2 Vergleich von *PrivacyInsight* mit alternativen Auskunftsverfahren

Im ersten Teil der Studie haben die 5 gestellten Aufgaben folgende Szenarien und Fragestellungen abgedeckt: (1) Wurde ein personenbezogenes Datum einer bestimmten Kategorie von der verantwortlichen Stelle erhoben? (2) Welches personenbezogene Datum wird

[10] Brooke 1996.
[11] Bangor/Kortum/Miller 2009.
[12] Lewis/Sauro 2009.
[13] Wästlund/Wolkerstorfer/Köffel 2010.
[14] Laugwitz/Held/Schrepp 2008.
[15] Semantische Differentiale sind begrifflich entgegengesetzte Eigenschaftspaare wie „schnell - langsam".
[16] Rauschenberger/Thomaschewski/Schrepp 2013.

Aufgabe	1	2	3	4	5
PrivacyInsight	51,29	38,65	55,45	24,97	65,52
GenomSynlig	60,13	42,74	57,61	74,52	70,71
JSON	38,84	23,90	52,32	50,29	85,23

Tabelle 12.1: Mittlerer Zeitbedarf zur Aufgabenbearbeitung in Sekunden (Maximum: 120 s) im ersten Teil

gespeichert (konkreter Inhalt)? (3) Ist es erkenntlich, wenn mehrere Daten einer Kategorie gespeichert wurden? (4) Wie kann der Betroffene die Löschung eines personenbezogenen Datums beantragen? (5) Wie kann der Fluss eines personenbezogenen Datums vom Betroffenen zur verantwortlichen Stelle dargestellt werden? Alle Aufgaben sind mit allen drei Auskunftssystemen, zumindest auf Umwegen, lösbar.

Aufgabenbewältigung Die erste Aufgabe wurde nur von 35 % der Probanden mit *GenomSynlig* richtig gelöst, wobei 7 Teilnehmer am Ablauf der Zeit scheiterten. Mit *PrivacyInsight* waren bereits 71 % der Probanden in der Lage die Aufgabe zu lösen, mit JSON waren es sogar 90 %. Für den Zeitbedarf ergab sich die gleiche Reihung (vgl. Tabelle 12.1).

In der zweiten Aufgabe schnitt *GenomSynlig* mit einer Erfolgsquote von 48 % mit Abstand am schlechtesten ab. *PrivacyInsight* und JSON waren mit einer Erfolgsquote von 90 % bzw. 94 % fast gleich auf. Diejenigen Probanden, die mit *PrivacyInsight* innerhalb der vorgegebenen Zeit mit der Bearbeitung der Aufgabe fertig wurden, lösten die Aufgabe alle richtig. Die Antworten bei *GenomSynlig* kamen zwar fast genauso schnell, waren aber häufig falsch. Die Probanden wurden von der Darstellung in die Irre geführt. Personenbezogene Daten konnten nicht klar einer verantwortlichen Stelle zugeordnet werden.

Die dritte Aufgabe wurde für alle Systeme ähnlich schnell bearbeitet, allerdings mit stark unterschiedlicher Qualität der Ergebnisse. 68 % der Probanden fanden mit *PrivacyInsight* die richtige Lösung während es mit *GenomSynlig* und JSON nur 52 % bzw. 39 % waren. Ähnlich wie bei der zweiten Aufgabe war die Zuordnung personenbezogener Daten in *PrivacyInsight* besser erkennbar.

Die vierte Aufgabe zielte auf die Integration der Auskunft und der weiteren Betroffenenrechte ab. Diese ist mit *PrivacyInsight* deutlich am besten gelöst. Nur ein Proband war nicht in der Lage, die Aufgabe erfolgreich zu bearbeiten. Er scheiterte am vorgegebenen zeitlichen Rahmen. Alle anderen Probanden fanden in kurzer Zeit den richtigen Button. Ganz anders bei *GenomSynlig* und JSON. Bei diesen Ansätzen lag die Erfolgsquote nur bei 45 % bzw. 64 %. *GenomSynlig* bietet zwar auch einen Löschbutton, dieser ist jedoch sehr versteckt platziert. Bei JSON mussten die Probanden auf die Idee kommen, eine postalische Anfrage zu stellen.

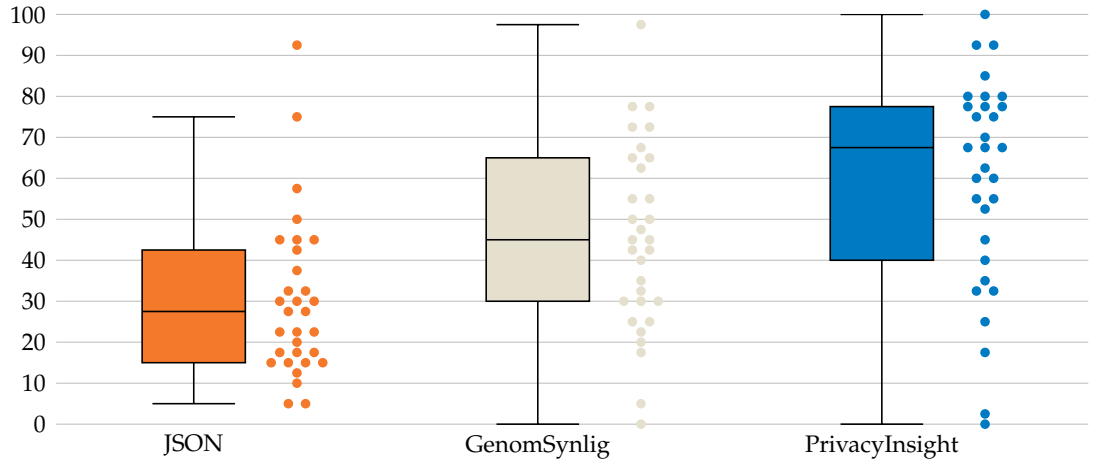


Abbildung 12.3: SUS-Score der drei Vergleichsverfahren

Die letzte Aufgabe wandte sich den Datenflüssen zu. Bei JSON sind diese anhand der Datenkennungen manuell nachvollziehbar. Die letzte war mit Abstand die schwierigste Aufgabe. Mit *PrivacyInsight* waren immerhin 58 % der Probanden in der Lage, einen Datenfluss hervorzuheben. Mit *GenomSynlig* und JSON waren es nur 35 % bzw. 16 %.

Mit *GenomSynlig* hatten die Probanden über die ersten vier Aufgaben hinweg den größten Zeitbedarf. Bei der letzten Aufgabe stand JSON etwas schlechter da, da eine nicht-interaktive Darstellung für das Aufzeigen eines Datenflusses nicht geeignet ist. Bei den ersten drei Aufgaben war JSON am schnellsten zu verwenden, bei den anderen beiden war *PrivacyInsight* die am schnellsten zu bedienende Oberfläche. Der Zeitvorteil von JSON lag darin begründet, dass es bei den ersten Aufgaben möglich war, die Textsuche im PDF zu verwenden. Lösungsquote und Geschwindigkeit zusammengenommen brachte *PrivacyInsight* für die Probanden die besten Ergebnisse.

Fragebögen Die Ergebnisse des SUS sind in Abbildung 12.3 zusammengefasst. Alle drei Systeme lagen nicht im oberen Bereich der Skala. Dabei ist zu berücksichtigen, dass die Gebrauchstauglichkeit in einer Testsituation unter Stress erfahren wurde. Außerdem handelt es sich bei Auskunftsplattformen um eine völlig neue Anwendungsgattung.

Trotz seiner ordentlichen Performance in den ersten drei Aufgaben wurde JSON im Median nur mit 27,5 Punkten bewertet. *GenomSynlig* erreichte einen Median von 45 Punkten. Auf der SUS-Skala steht dies für eine eine geringe bis mittelmäßige Gebrauchstauglichkeit. *PrivacyInsight* lag im Median bei 67,5 Punkten. Dies ist eine mittlere bis gute, aber weder eine exzellente noch eine bestmögliche Bewertung. Die Erlernbarkeit lag bei *PrivacyInsight* und *GenomSynlig* in etwa auf dem gleichen Niveau wie die Gesamtbewertung.

Die Streuung der Bewertungen war bei allen drei Verfahren groß. Die Permutation

Aufgabe	1	2	3	4	5
Erfolgsquote	70,97 %	61,29 %	83,87 %	32,26 %	80,65 %
Mittlerer Zeitbedarf	44,90 s	72,09 s	40,70 s	54,81 s	45,13 s

Aufgabe	6	7	8	9	10
Erfolgsquote	83,87 %	93,55 %	93,55 %	93,55 %	100,00 %
Mittlerer Zeitbedarf	43,45 s	17,03 s	16,57 s	22,06 s	3,06 s

Tabelle 12.2: Erfolgsquote und mittlerer Zeitbedarf zur Aufgabenbearbeitung im zweiten Teil (nur *PrivacyInsight*)

der Reihenfolge der Verfahren hatte keinen signifikanten Einfluss auf die Bewertungen. Gemeinsam mit der Aufgabenbewältigung kann dennoch *PrivacyInsight* als das klar überlegene System identifiziert werden.

12.3 Nutzerrezeption des erweiterten Funktionsumfangs von *PrivacyInsight*

Im zweiten Teil wurden die Probanden ausschließlich mit *PrivacyInsight* konfrontiert. Die 10 Aufgaben drehten sich rund um den Funktionsumfang von *PrivacyInsight*, der über *GenomSynlig* und eine herkömmliche, textuelle Auskunft hinausgeht.

Die ersten beiden Aufgaben befassen sich mit der Übermittlung personenbezogener Daten. Die dritte Aufgabe fragt nach dem Zweck der Speicherung. Die Aufgaben 4 und 6 nehmen den gesamten Datenfluss in den Blick. Die fünfte Aufgabe setzt sich mit der Kategorie der gespeicherten Daten auseinander. Die letzten vier Aufgaben umfassen den Datenexport und -download, einen Antrag auf Berichtigung und abschließend das Zurücksetzen der Darstellung. Der genaue Aufgabentext findet sich in Anhang F.2.

Aufgabenbewältigung Die Erfolgsquote war über alle Aufgaben hinweg sehr hoch (Tabelle 12.2). Die letzten vier Aufgaben konnten von über 90 % der Probanden gelöst werden. Der Zeitaufwand lag im Mittel bei weniger als 25 s und war damit weitaus niedriger als bei den Aufgaben im ersten Teil. Aufgrund der vorherigen Einweisung konnten alle Probanden die letzte Aufgabe in unter 20 Sekunden lösen.

Einzig Aufgabe 4 konnte von weniger als der Hälfte der Probanden gelöst werden. Nur 32 % der Probanden waren in der Lage, zu identifizieren, durch wie viele Abteilungen ein personenbezogenes Datum floss. Zu erkennen, wann die Hierarchieebene der Abteilungen in der Informationsflussvisualisierung geöffnet war, war schwer.

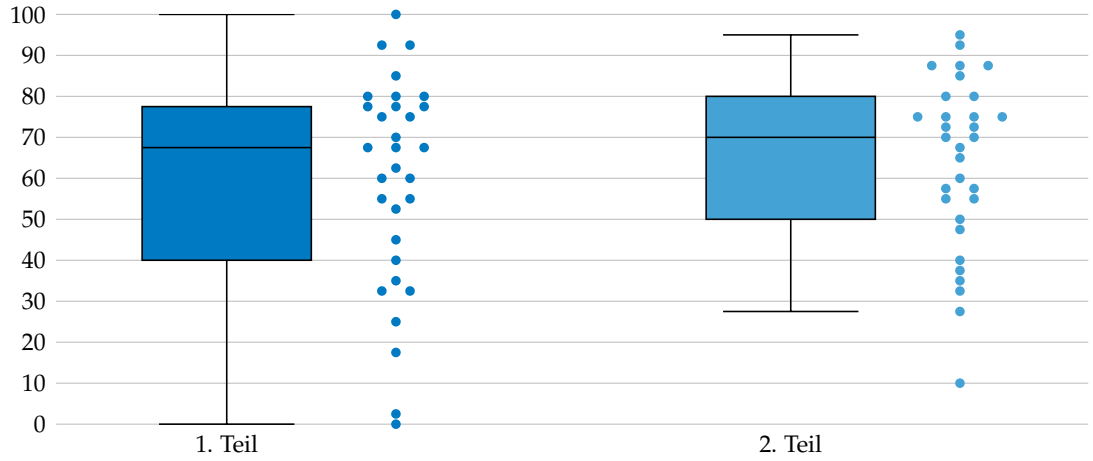


Abbildung 12.4: SUS-Score von *PrivacyInsight* im ersten und zweiten Teil

Es war allerdings nicht so, dass die Probanden das Expandieren von Knoten und das Hineinnavigieren in tiefere Ebenen grundsätzlich nicht verstanden hätten. Die Mehrzahl der Probanden war nach den in der Studie gemachten Beobachtungen in der Lage, Knoten zu öffnen und Untereinheiten zu identifizieren.

Fragebögen Der wesentliche Unterschied zwischen den SUS-Ergebnissen von *PrivacyInsight* im ersten und im zweiten Teil ist eine starke Reduktion der Streuung (vgl. Abbildung 12.4). Der Interquartilsabstand beträgt nur noch 30 Punkte (zuvor: 37,5 Punkte). Nur noch ein Proband bewertet die Gebrauchstauglichkeit von *PrivacyInsight* mit weniger als 27,5 Punkten (zuvor: 4 Probanden). Der Median bleibt dabei fast unverändert und steigt nur leicht auf 70 Punkte. Die Probanden sind sich also nach wiederholter Verwendung von *PrivacyInsight* einiger über dessen Bewertung. Wissen hat eine konsolidierende Wirkung.

Aus dem UEQ werden für alle Skalen Werte im Bereich von -3 bis +3 berechnet. Gemäß dem Auswerteschema werden Werte von -3 bis -0,8 als negativ, Werte von -0,8 bis +0,8 als neutral und Werte zwischen +0,8 und +3 als positiv angesehen. Probanden vermeiden normalerweise extreme Bewertungen, weshalb sich der Wertebereich realstischer meist zwischen -2 und +2 bewegt. Mit den Auswertetools des UEQ werden Vergleichswerte etablierter Systeme mitgeliefert. Da *PrivacyInsight* kein Produktivsystem ist, ist ein direkter Vergleich mit diesem Datensatz mit Vorsicht zu interpretieren.

PrivacyInsight erreichte im Mittel einen Attraktivitätswert von 1,25 – ein überdurchschnittlicher Wert. Die Effizienz erreichte einen durchschnittlichen Wert von 1,27.

Die Durchschaubarkeit wurde mit Abstand am schlechtesten bewertet. Sie erreichte nur einen mittleren Wert von 0,49. Damit befindet sie sich zwar noch im neutralen Bereich, jedoch unter den schlechtesten 25 % der Systeme im Hinblick auf den Vergleichsdatensatz.

PrivacyInsight wird von den Probanden als verwirrend und kompliziert wahrgenommen. Da die Darstellungsform für alle Probanden neu war, ist dies nicht verwunderlich.

Der Durchschaubarkeit folgt die Steuerbarkeit. Sie wird mit einem mittleren Wert von 0,87 als durchschnittlich angesehen.

Demgegenüber geben Stimulation und Originalität einen positiven Ausblick. Sie schneiden mit mittleren Werten von 1,26 und 1,5 überdurchschnittlich gut ab. *PrivacyInsight* wird als originell und neuartig beschrieben.

Die Permutation der Reihenfolge der Systeme im ersten Teil hat keinen messbaren Einfluss auf die Bewertungen im zweiten Teil.

Den Probanden wurden noch zusätzliche Fragen gestellt, um ihre Einschätzung des Nutzens von *PrivacyInsight* abzufragen. Da diese Fragen am Ende gestellt wurden und den meisten Probanden klar gewesen sein dürfte, dass *PrivacyInsight* die getestete Neuentwicklung ist, ist damit zu rechnen, dass sie nicht unvoreingenommen beantwortet wurden.

Dennoch ist erfreulich, dass, bis auf eine Person, alle Probanden für die zusätzlich in *PrivacyInsight* verfügbaren Informationen eine höhere Komplexität hingenommen haben. Die Integration der Datenschutzauskunft mit den übrigen Betroffenenrechten auf Löschung, Sperrung und Berichtigung hielten alle Probanden für vollkommen oder nahezu vollkommen sinnvoll.

12.4 Verständlichkeit und Nutzen der Unverkettbarkeitsmetriken

Im Rahmen der Nutzerevaluation wurde auch die Einstellung der Probanden zur Unverkettbarkeitsmetrik abgefragt. Dazu wurde ihnen zuerst der in Anhang F.3 abgedruckte Erläuterungstext vorgelegt. Der Erläuterungstext ist identisch mit dem in *PrivacyInsight* hinterlegten. Die Probanden konnten gleichzeitig einen Blick auf die reale Einbindung der Metrik in Form von Zustandsbalken werfen.

Von den 31 Probanden waren 21 der Meinung, das Konzept der Unverkettbarkeitsmetrik verstanden zu haben. 7 Probanden waren sich aufgrund der gegebenen Kurzbeschreibung nicht sicher. Dass ein Proband der Auffassung ist, die Metrik verstanden zu haben, heißt noch nicht, dass dies tatsächlich der Fall ist. In der Akzeptanz von Datenverarbeitungsverfahren spielt allerdings das wahrgenommene Verhalten eines Systems eine viel größere Rolle als das tatsächliche.

Nichtsdestotrotz wurde versucht, in Interviews nach der Bearbeitung der Fragebögen abzuklopfen, ob die Metrik tatsächlich verstanden wurde. Die Probanden wurden aufgefordert, den Studienleitern im Rahmen eines finalen Feedbacks die Bedeutung der Metrik so zu erklären, wie sie sie verstanden haben. Dabei hat sich der Eindruck ergeben, dass subjektive und objektive Wahrnehmung übereinstimmen. Verallgemeinerbar ist dieser unsystematische Eindruck allerdings nicht.

Von den 21 Probanden, die die Metrik verstanden hatten, hielten 6 die Metrik nicht für hilfreich, für 9 war sie eine nützliche Entscheidungsgrundlage für ihre Opt-out-Möglichkeit. Insgesamt ein zufriedenstellendes, wenn auch nicht gutes, Ergebnis für eine bis dato unbekannte Entscheidungshilfe.

Die Probanden wurden mit Unlinkability-Werten von 76 %, 80 %, 86 % und 89 % (Speicher- und Verarbeitungsrelation, Datenflussrelation, Verknüpfungsrelation und Identifikationsrelation) konfrontiert. Unter dieser Maßgabe wollte kein Proband direkt auf das Datenschutzauskunftssystem verzichten. 8 Probanden waren allerdings unentschieden.

Ein Schwellwert, ab dem eine überwiegende Anzahl der Betroffenen auf das Datenschutzauskunftssystem verzichten würde, wurde nicht ermittelt. Dies wäre eine interessante Aufgabe für eine zukünftige Nutzerbefragung.

12.5 Erwartungen der Betroffenen

Neben den demographischen Daten wurde am Ende der Studie abgefragt, wen die Probanden in der Verantwortung für den Schutz personenbezogener Daten sehen. Im Verhältnis zwischen Staat und Bürger sowie Staat und Unternehmen waren sich die Probanden uneins. Im Mittel wurden beide Akteure in gleichem Maße in die Pflicht genommen, wobei die Streuung so stark war, dass von keinem eindeutigen Bild gesprochen werden kann. Anders sah die Situation im Verhältnis zwischen Unternehmen und ihren Kunden aus. Nur 3 Probanden (10 %) sahen eher den Kunden in der Pflicht. Weitere zwei Probanden (6 %) sahen die Verantwortung in gleichem Maße bei Kunden und Unternehmen. Alle anderen (84 %) sahen vor allem das Unternehmen in der Pflicht, 9 Probanden (29 %) sogar ausschließlich.

Daraus ergibt sich ein Widerspruch zwischen rechtlicher und tatsächlicher Sicht auf den Datenschutz. Während sich Bürger und Unternehmen als Private zunächst gleichberechtigt gegenüberstehen, ist der Staat Grundrechtsverpflichteter und muss daher für das Recht auf informationelle Selbstbestimmung einstehen.

Die Probanden sehen dagegen das primäre Machtgefälle zwischen Kunden und Unternehmen. Die Datenindustrie wird auch nach Snowden[17] als die größere Gefahr gesehen.

Eine Umfrage von Statista bestätigt die Wahrnehmung.[18] Während 2009 noch 44 % der Befragten den Staat als den für den Datenschutz im Internet Verantwortlichen benennen, sind es 2014 nur noch 15 %. Auf der anderen Seite versiebenfacht sich die Zahl der Befragten, die Anbieter von Online-Diensten und Hersteller von Hard- und Software in der Pflicht sehen, auf 22 %. Die Fragestellung ist eine etwas andere, doch das Ergebnis zeigt in die gleiche Richtung.

[17] Preneel 2015.

[18] http://de.statista.com/statistik/daten/studie/243353/umfrage/verantwortung-fuer-den-datenschutz-im-internet, abgerufen am 9. Mai 2017.

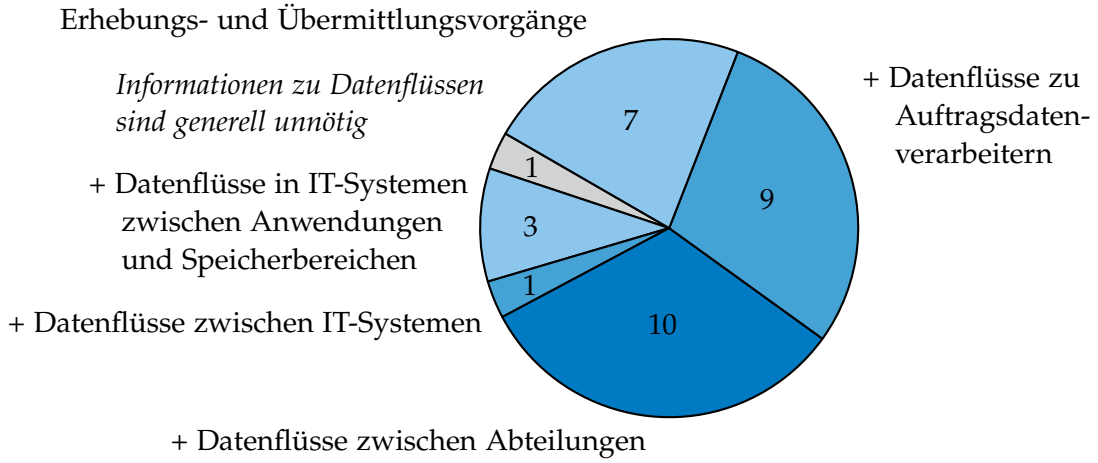


Abbildung 12.5: Erwarteter Detailgrad der Informationen über Datenflüsse im Rahmen einer Auskunft (im Uhrzeigersinn)

Welche Schlussfolgerungen ergeben sich daraus für das Recht auf Auskunft? Eine weitere Differenzierung zwischen Unternehmen und Staat wurde nicht vorgenommen. Teil des Fragebogens war jedoch die Frage, welche Informationstiefe die Probanden im Rahmen einer Auskunft nach § 34 BDSG erwarten würden. Wie in Abbildung 12.5 abzulesen, wären 87 % der Probanden (26 Personen) damit zufrieden, wenn ihnen Informationen zu Erhebungs- und Übermittlungsvorgängen, Datenflüsse zu Auftragsdatenverarbeitern und Datenflüsse zwischen Abteilungen mitgeteilt würden. Dies entspricht in etwa den rechtlichen Vorgaben. Nutzererwartungen und Datenschutzrecht sind in Deckung.

12.6 Zwischenfazit

Die Auskunftsplattform *PrivacyInsight* hat sich im Rahmen der Studie als überlegen gegenüber bereits existierenden Auskunftsverfahren aus Literatur und Praxis herausgestellt. Der Funktionsumfang des Datenschutzauskunftssystems wurde gut angenommen und positiv bewertet. *PrivacyInsight* erfüllt die in Hypothese $\mathfrak{H}5$ formulierten Hoffnungen.

Die Probanden als typische Nutzer der ersten Stunde halten die Integration von Auskunft und weiteren Betroffenenrechten für den richtigen Weg. Die Unverkettbarkeitsmetrik wird von manchen als hilfreiche Information, von machen noch mit einer gewissen Ambivalenz betrachtet. Inwiefern von einem Opt-out Gebrauch gemacht würde, ist noch nicht absehbar.

13 Zusammenfassung und Ausblick

Die vorliegende Arbeit liefert eine interdisziplinäre Gesamtbetrachtung zur Datenschutzauskunft in drei Themenfeldern. Sie stellt vor, wie sich die Umsetzung des Rechts auf Auskunft im Hinblick auf den technologischen Fortschritt entwickeln könnte.

Herausforderung ist der informationelle Kontrollverlust des modernen Menschen. Er wird durch drei Trends geprägt.[1] *Erstens* werden Daten an immer mehr Stellen durch immer neue Mechanismen und Sensoren erhoben. Bei jedem Besuch einer Website, mit jeder E-Mail und bei jeder Bewegung durch den öffentlichen Raum unter der Beobachtung von Kameras[2] und Smartphones werden neue personenbezogene Daten erhoben. *Zweitens* ist der Austausch und die Speicherung personenbezogener Daten so leicht möglich wie noch nie. In der Cloud ist nahezu unbegrenzter Speicher für jedermann verfügbar. Die Bandbreite der Übertragungskanäle erlaubt es, von überall jederzeit auf alle gespeicherten Daten zuzugreifen. *Drittens* erhöhen sich die Möglichkeiten der Auswertung personenbezogener Daten signifikant. Durch die Anwendung von Big-Data-Techniken können aus großen Datenschätzen neue Informationen abgeleitet werden. Methoden der künstlichen Intelligenz erlauben die Erkennung komplexer Muster.

In allen drei Dimensionen muss sich auch die Erbringung der Datenschutzauskunft weiterentwickeln, um den prognostizierten Kontrollverlust zu verhindern. Personenbezogene Daten müssen bereits zum Erhebungszeitpunkt identifiziert und der späteren Auskunft zugänglich gemacht werden. Die Erfassungsmechanismen müssen in allen datenerhebenden Systemen integriert sein. Die Speicherung der Historie eines personenbezogenen Datums, der Provenance, muss so organisiert sein, dass eine Auskunft bei Bedarf unverzüglich generiert werden kann. Der Zugriff des Betroffenen auf die Auskunft muss unkompliziert, schnell, benutzerfreundlich und nachvollziehbar möglich sein. All dies lässt sich nur durch ein automatisiertes Datenschutzauskunftssystem bewerkstelligen. Dies ist das erste Themenfeld der vorliegenden Arbeit.

Gleichzeitig entsteht ein Profilbildungspotential durch das Auskunftsverfahren selbst. Die Architektur eines Datenschutzauskunftssystems muss so entworfen werden, dass die Verkettung personenbezogener Daten nur für die Auskunft stattfindet. Der Betroffene muss in die Lage versetzt werden, zu entscheiden, welchen Kompromiss zwischen Transparenz und Unverkettbarkeit er einzugehen bereit ist. Damit ist das zweite Themenfeld

[1] Seemann 2016, 129 f.

[2] Nicht mehr nur als stationäre (intelligente) Videoüberwachung, sondern auch durch Drohnen, Dashcams, Wearables à la Google Glass und Smartphones mobil eingesetzt.

umrissen.

Die technologische Entwicklung wird von einem Umbruch der rechtlichen Grundlagen des Datenschutzes begleitet. Der zentrale Faktor ist die Europäisierung des Datenschutzes. Nationale Regelungen werden durch die Datenschutz-Grundverordnung abgelöst. Dadurch entstehen neue Betroffenenrechte wie das Recht auf Datenübertragbarkeit. Dieses Recht ist ein Beispiel für die Konvergenz des Wettbewerbs- und Datenschutzrechts auf nationaler und europäischer Ebene. Die Einordnung aller rechtlichen Aspekte ist das dritte Themenfeld.

Die vorliegende Arbeit trägt mit den im folgenden Abschnitt zusammengefassten Ergebnissen zur Klärung und Bewältigung der aufgeworfenen Problemstellungen bei.

13.1 Zusammenfassung der Ergebnisse

Gemäß der Rechtsprechung des Bundesverfassungsgerichts muss das Datenschutz-Schutzziel der Transparenz nicht zwingend durch ein Recht auf Auskunft umgesetzt werden (↯ §1). Die Datenschutzauskunft ist eine Umsetzungsvariante des Rechts auf Kenntnisnahme, verankert im Recht auf informationelle Selbstbestimmung des Art. 2 I i. V. m. Art. 1 I GG. Der Gesetzgeber hat sich hingegen für ein Recht auf Auskunft entschieden. In den §§ 19, 34 BDSG legt er für öffentliche und nicht-öffentliche Stellen nahezu inhaltsgleiche Anforderungen fest. Anders stellt sich die Situation europarechtlich dar. Das Recht auf Auskunft ist in Art. 8 Abs. 2 Satz 2 GRCh explizit kodiert (§1). Folgerichtig findet sich auch in der DSGVO ein Recht auf Auskunft.

Die DSGVO ändert nichts daran, dass das Recht auf Auskunft jedem Betroffenen ohne weitere Voraussetzungen zusteht. Sowohl nach den bisherigen Regelungen als auch nach denen der DSGVO ist es möglich, eine Auskunft auf einer elektronischen Plattform zu erteilen, soweit diese angemessen gesichert ist und eine Downloadmöglichkeit bietet. Im Rahmen des Rechts auf Datenübertragbarkeit legt die DSGVO fest, dass solch eine Downloadmöglichkeit die stellenübergreifende Nutzung personenbezogener Daten ermöglichen soll. Der Betroffene soll selbst in der Lage sein, zu entscheiden, welche verantwortliche Stelle welche Datenverarbeitung für ihn übernimmt. Datenschutz hat insofern eine marktgestaltende Wirkung. Verbraucherschutzrecht und Datenschutzrecht verbinden sich.

Die Auskunft umfasst die gespeicherten personenbezogenen Daten, den Zweck der Speicherung, den Speicherort sowie Herkunft und Empfänger. Aufgrund des, durch die DSGVO bestätigten, weiten Empfängerbegriffs ist eine durchgängige Nachvollziehbarkeit der Flüsse personenbezogener Daten erforderlich (§2). Es sind alle Stellen zu nennen, die personenbezogene Daten zu einem dezidiert unterschiedlichen Zweck verarbeiten sowie der zwischen ihnen stattfindende Datenaustausch.

Diesen Anforderungen wird die Praxis gegenwärtig nicht gerecht. Die durchgeführte

Studie zeigt, dass selbst grundlegende Aspekte, wie der Zweck der Speicherung, nicht von allen Unternehmen in ihre Auskunft mitaufgenommen werden. Eine Nachvollziehbarkeit der stattfinden Datenflüsse ist in keinem Fall gegeben.

Solange Unternehmen keine Sanktionen fürchten müssen ist fraglich, wie schnell flächendeckende Rechtskonformität hergestellt wird. Die DSGVO gibt den Aufsichtsbehörden neue Instrumente in die Hand. Der monetäre Sanktionsrahmen wird empfindlich erhöht. Der nationale Gesetzgeber versucht gleichzeitig neue Sanktionsmöglichkeiten auf wettbewerbsrechtlicher Ebene zu schaffen. Die Neufassung des UKlaG soll die Abmahnung von Datenschutzverstößen ermöglichen. Der Auskunftsanspruch ist davon jedoch nicht erfasst.

Das entworfene Ableitungsmodell EVAL unterstützt die Konzeption eines Datenschutzauskunftssystems. Es erlaubt den systematischen Übergang von der rechtlichen Analyse zur technischen Realisation (ℜ1). Zu diesem Zweck führt es die analytischen Ebenen der Datenschutz-Schutzziele, der datenschutzrechtlichen Anforderungen und Kriterien sowie der technischen Anforderungen ein. Für den Auskunftsanspruch sind die einzelnen Elemente dieser Ebenen vollständig beschrieben. Das in dieser Arbeit vorgestellte Datenschutzauskunftssystem erfüllt diese Anforderungen (ℜ6).

Das Datenschutzauskunftssystem trägt in den oben genannten drei Dimensionen dazu bei, dass verantwortliche Stellen die Erteilung einer Auskunft besser umsetzen können. Grundlage ist die vorgestellte Architektur, die Erfassung, Verarbeitung, Speicherung und Auswertung des Umgangs mit personenbezogenen Daten, die Personal-Data-Provenance, integriert. Die Verbindung aus Provenance-Tracking und Usage-Control sorgt dafür, dass die Auskunft und die weiteren Betroffenenrechte auf Löschung, Sperrung und Berichtigung gemeinsam realisiert werden (ℜ2). Ist die Provenance eines Datums bekannt, ist auch bekannt, wo UC-Policies die übrigen Betroffenenrechte durchsetzen müssen.

Die Ereignisse der Datenverarbeitung von einmal erkannten personenbezogene Daten werden von einem PEP abgefangen. Die Ereignisse werden mit Hilfe einer generischen Semantik beschrieben. So können die Auswirkung eines Ereignisses auf das Informationsflussmodell für neu hinzukommende Applikationen direkt bestimmt werden. Die Personal-Data-Provenance speist sich direkt aus dem, um den Betroffenen erweiterten, Informationsflussmodell. Das Data-Provenance-Modell dieser Arbeit ist, im Gegensatz zu beispielsweise dem aus A4Cloud, datenzentriert. Es werden keine Verarbeitungsprotokolle, sondern Strukturen für jedes einzelne personenbezogene Datum gespeichert. Die Personal-Data-Provenance hat eine Baumstruktur mit den Repräsentationen der einzelnen Datenverwendungen als Knoten und den temporären und räumlichen Beziehungen als Kanten. Unterschiedliche Repräsentationstypen erlauben die Erfassung der Provenance gemäß der datenschutzrechtlichen Anforderungen (ℜ3). Die Präzision des Trackings und damit der Provenance ist stark davon abhängig, ob PEPs für alle datenverarbeitenden Systeme vorhanden sind. Noch nicht abbildbar ist eine Vermischung personenbezogener

Daten in einem gemeinsamen Container, so dass eine anschließende Trennung wieder möglich ist ($\lightning \mathfrak{K}3$).

Würden alle technisch erfassbaren Verarbeitungszustände in der Personal-Data-Provenance gespeichert, würde dies dem Grundsatz der Datenminimierung zuwiderlaufen. Mit Hilfe von Lösch- und Abstraktionsregeln kann erreicht werden, dass nur genau die Provenance gespeichert wird, die für eine Auskunft erforderlich ist. Tests zeigen, dass diese Regeln potentiell zu einer besseren Speichereffizienz führen ($\mathfrak{H}3$). Der Speicherbedarf wächst sowohl im Verhältnis zu den Ereignissen als auch zu der Anzahl der personenbezogenen Daten maximal linear. Das System ist auch bei hoher Ereignisfrequenz und Datendichte noch benutzbar. Der Overhead der Provenance ist kaum spürbar. Allerdings können die verwendeten PEPs die datenverarbeitenden Systeme merklich bremsen.

Die Personal-Data-Provenance ist im Sinne einer informationellen Gewaltenteilung dezentral auf den einzelnen datenverarbeitenden Systemen gespeichert. Finden systemübergreifende Informationsflüsse statt, werden die dazugehörigen Ereignisse anhand einer systemübergreifenden Semantik interpretiert. Dazu wurden in dieser Arbeit die Ideen von Lovat[3] und Kelbert[4] erweitert. Scopes sind nun auch zwischen Systemen möglich. Anwendungen können dynamisch gekoppelt werden.

Erst wenn der Betroffene eine Auskunft anfordert, wird die Personal-Data-Provenance aggregiert. Dazu wird der Provenance-Baum traversiert und anhand der Repräsentationen, die eine Datenweitergabe beschreiben, zusammengefügt ($\mathfrak{K}4$). Die Auskunftsplattform *PrivacyInsight* ist eine interaktive Webapplikation. Sie erlaubt es dem Betroffenen, Flüsse personenbezogener Daten in anpassbarer Informationstiefe nachzuverfolgen. Einblicke in die gespeicherten personenbezogenen Daten sind genauso möglich, wie die Wahrnehmung der Betroffenenrechte auf Löschung, Sperrung und Berichtigung. Eine Nutzerstudie zeigt, dass *PrivacyInsight* der bisher besten Auskunftsplattform, *GenomSynlig*,[5] in allen Belangen überlegen ist. Mit einer kurzen Einführung war allen Probanden auch der erweiterte Funktionsumfang von *PrivacyInsight* gut zugänglich ($\mathfrak{H}5$). Der Umgang mit personenbezogenen Daten wird für die Betroffenen nachvollziehbar.

Ein Datenschutzauskunftssystem hat Auswirkungen auf die Profilbildungsfähigkeit von Stellen im auskunftspflichtigen Unternehmen. Eine Metrik für Unverkettbarkeit macht diese sichtbar. Diese Arbeit geht dabei in mehrfacher Hinsicht über den Stand der Forschung hinaus. Erstens wurde die Definition eines Grades von Unverkettbarkeit so verallgemeinert, dass sie für beliebige Angreifer, Entitäten, Systeme und Verkettungsrelationen instanziierbar ist. Zweitens wurde die Metrik für das Szenario eines Datenschutzauskunftssystems in 4 Varianten instanziiert ($\mathfrak{K}5$). Drittens wurde aufgezeigt, wie die Metrik konkret berechnet werden kann. Dazu wurden sowohl mathematisch-analytische Über-

[3] Lovat 2015.
[4] Kelbert/Pretschner 2015.
[5] Angulo et al. 2015.

legungen angestellt als auch eine neue Heuristik entwickelt. Die Berechenbarkeit wurde anhand einer Fallstudie belegt. Der Aufwand ist jedoch stark von der Größe der Datensätze abhängig.

Zuletzt wurde untersucht, welche Konsequenzen aus einer Metrik für Unverkettbarkeit gezogen werden können. Es ist schwierig, verschiedene Systemarchitekturen miteinander zu vergleichen. Die Metrik gibt nur Auskunft über die Unverkettbarkeit in einer bestimmten Situation zu einem bestimmten Zeitpunkt. Nur in Extremfällen können aus Szenarien Schlussfolgerungen für ganze Systemarchitekturen abgeleitet werden. Die Architektur *Insynd*[6] ist der Architektur dieser Arbeit in manchen Metriken überlegen, sofern man in Kauf nimmt, dass jeder Betroffene vor einer etwaigen Erhebung personenbezogener Daten in einen Schlüsselinitialisierungsprozess eingebunden werden muss.

Gut geeignet ist die Metrik hingegen, um den Betroffenen zu einem festen Zeitpunkt über seine Unverkettbarkeitssituation zu informieren. Fraglich ist in solch einem Zusammenhang, wie gut ein Laie eine komplexe Metrik interpretieren kann. Die durchgeführte Nutzerstudie zeigt, dass das Konzept der Metrik für einen Großteil der möglichen technikaffinen Erstverwender verständlich ist. Aussagen über die Bedeutung bestimmter Werte lassen sich daraus noch nicht ableiten (↯ §4).

Die Metrik wurde in *PrivacyInsight* eingebunden und mit einer Opt-out-Möglichkeit für das Provenance-Tracking verknüpft (§4). Rechtlich ist so eine Möglichkeit umstritten, jedoch vertretbar. Das Recht auf Auskunft ist zwar nach dem Gesetzeswortlaut unabdingbar. Aus verfassungs- und europarechtlichen Gründen kann dies jedoch für den Sonderfall eines Datenschutzauskunftssystems nicht uneingeschränkt gelten. Die befragten Probanden würden jedenfalls auf solch eine Möglichkeit nur ungern verzichten. Ob sie sie letztendlich wahrnehmen, ist eine andere Sache.

13.2 Abgrenzung

13.2.1 Verwandte Arbeiten

Ableitung datenschutzrechtlicher Anforderungen Kiyavitskaya, Krausová und Zannone[7] begründen, warum die Ableitung rechtlicher Anforderungen schwierig ist. Die von ihnen angesprochenen Punkte wie die Mehrdeutigkeit des Rechts und die unterschiedliche Methodik in Rechtswissenschaft und Informatik werden von dieser Arbeit adressiert.

Hammer, Pordesch und Roßnagel,[8] Schwenke,[9] Schulz[10] und Kahlert[11] verwenden die

[6] Pulls/Peeters 2015.
[7] Kiyavitskaya/Krausová/Zannone 2008.
[8] Hammer/Pordesch/Roßnagel 1993.
[9] Schwenke 2006.
[10] Schulz et al. 2011.
[11] Kahlert, DuD 2014, 86.

Methode KORA zur Konkretisierung rechtlicher Anforderungen an IT-Systeme. Bei allen werden die Schwächen der Methode KORA deutlich. Die von ihnen als „rechtliche Kriterien“ bezeichneten Begriffe sind nur allgemeine Zielvorgaben. Die eigentliche Auslegung des Rechts durch sprachliche, logische, systematische, historische oder teleologische Methoden findet zum Großteil erst auf der Ebene der „technischen Gestaltungsziele“ statt. Dort sollte allerdings schon der Wechsel zur technischen Sprache stattgefunden haben. Echte, detaillierte technische Anforderungen werden nicht erreicht. Das in dieser Arbeit neu eingeführte EVAL-Modell führt zusätzliche Ableitungsschritte ein und präzisiert die Abgrenzung zwischen diesen. Darüber hinaus findet die Ableitung bis hin zur Implementierung statt.

Unter vielen diskutieren Spiekermann und Cranor[12], Langheinrich[13] und Hansen[14] die bei der Entwicklung datenschutzfreundlicher Systeme zu verfolgenden grundsätzlichen Herausforderungen und Ziele. Aus diesen leiten sie ad hoc technische Anforderungen ab. Sie tun dies jedoch nicht durch eine verallgemeinerbare systematische Ableitung wie in EVAL. Ihr Beitrag liegt vielmehr darin, die Bedeutung und Anwendung von Datenschutz-Schutzzielen für Entwickler greifbar zu machen.

Probst[15] listet generische Maßnahmen zur Adressierung der Datenschutz-Schutzziele auf. Da sie nicht als Anforderungen an ein konkretes System entworfen wurden, pendeln sie irgendwo zwischen den rechtlichen Anforderungen und den rechtlichen Kriterien von EVAL. Ist für ein zu entwickelndes System aus zeitlichen oder finanziellen Gründen keine vollständige EVAL-Analyse möglich, ist diese Art der Auflistung die nächstbeste Alternative.

Aldeco-Pérez und Moreau[16] schlagen vor, Provenance für die Auditierung der Verwendung personenbezogener Daten heranzuziehen. Sie leiten aus dem UK Data Protection Act sieben Prinzipien für solch eine Provenance ab. Ergänzend schlagen sie eine generische Architektur für die Auditierung mit Hilfe einer Personal-Data-Provenance vor. Darüber hinaus wird ausgeführt, wie geeignete Queries auszusehen haben. Demgegenüber nimmt diese Arbeit eine weit umfangreichere Ableitung von Anforderungen an eine Provenance für den Auskunftsanspruch vor. Die Architektur dieser Arbeit verbindet Usage-Control und Provenance und wurde prototypisch implementiert. Da die Art der Abfrage einer Personal-Data-Provenance zum Zweck der Auskunft vordefiniert ist, sind Queries nicht Teil dieser Arbeit.

[12] Spiekermann/Lorrie Faith Cranor 2009.
[13] Langheinrich 2001.
[14] Hansen 2011.
[15] Probst, DuD 2012, 439.
[16] Aldeco Pérez/Moreau 2008.

Informationsfluss- und Provenance-Modell Lovat, Oudinet und Pretschner[17] verwenden einen ähnlichen Graphen je Datum, wie er in dieser Arbeit im Provenance-Modell Verwendung findet, um die gegenwärtige „Menge“ eines sensitiven Datums in einem Container zu bestimmen. Dagegen hinterlegt diese Arbeit die Historie der Informationsflüsse in einer Provenance-Datenstruktur.

Die Ideen von Lovat[18] zu strukturierten Informationsflüssen sind eine Ergänzung der Modelle in dieser Arbeit. Eine zukünftige Einbindung, um die Struktur der Provenance zwischen unterschiedlichen Verarbeitungskomponenten zu erhalten, ist denkbar und sinnvoll. Das Modell für schichtenübergreifende Informationsflüsse von Lovat[19] wird in dieser Arbeit aufgenommen und auf systemübergreifende Informationsflüsse verallgemeinert. Ergänzend wird es einer Beschreibung in der Informationsflusssemantik zugänglich gemacht.

Die Abgrenzung des systemübergreifenden Informationsfluss- und Provenance-Trackings gegenüber den Arbeiten von Kelbert[20] wurde bereits in Kapitel 8.1 vorgenommen. Zusammenfassend ermöglicht der Ansatz von Kelbert im Gegensatz zu dieser Arbeit die systemübergreifende Durchsetzung von UC-Policies. Die Konzepte dieser Arbeit lassen dafür das Tracking von Informationsflüssen aus Anwendungen heraus zu. Die Enden einer Verbindung können anhand beliebiger Identifikatoren zugeordnet werden, die in einer Informationsflusssemantik hinterlegt sind. Darüber hinaus wird die Provenance der Datenflüsse aufgezeichnet.

Des Weiteren existieren bereits Konzepte zur Propagierung von Taintmarken über Systemgrenzen hinweg. NEON[21] ist ein Monitor für virtuelle Maschinen und unterstützt Taintmarken zum Tracken von Informationsflüssen auf Byteebene. Diese Art der Repräsentation von Taintmarken limitiert die mögliche Anzahl getrackter Daten je Byte auf 32. Dagegen erlaubt es das in dieser Arbeit verwendete Informationsflussmodell theoretisch unbegrenzt viele Daten zu tracken. Tainting auf Netzwerkebene findet zwar nicht auf Byteebene, sondern pro Netzwerkpaket mit bis zu 256 Daten pro Paket statt. Die einzelnen, im Netzwerkpaket transportierten Bytes lassen sich auf Empfängerseite allerdings nicht mehr unterscheiden. SeeC[22] speichert je Prozess eines IT-Systems Taintmarken in einem Schattenspeicher und erzeugt für jeden Socket eine *write queue* für Taintmarken zu Datenflüssen über TCP/IP. Bei SeeC ist die mögliche Anzahl getrackter Daten je Byte ebenfalls auf 32 beschränkt. Generell gilt für die Propagierung von Taintmarken, dass sich daraus nicht die in dieser Arbeit erforderliche Provenance ergibt. Es wird nur die Zuordnung der Daten im aktuellen Systemzustand gespeichert.

[17] Lovat/Oudinet/Pretschner 2014.
[18] Lovat/Kelbert 2014.
[19] Lovat 2015.
[20] Kelbert/Pretschner 2013; Kelbert/Pretschner 2014; Kelbert/Pretschner 2015.
[21] Q. Zhang et al. 2010.
[22] Kim et al. 2009.

GARM[23] erlaubt mit Hilfe von Application-Rewriting ein Provenance-Tracking über Systemschichten hinweg. Insofern ist es eine hartcodierte Lösung für bestimmte Abstraktionsschichten. GARM differenziert nicht zwischen Daten und Containern und lässt kein systemübergreifendes Proveance-Tracking zu, wie es diese Arbeit bietet.

Muniswamy-Reddy et al.[24] integrieren in ihrem Framework Provenance-Collection und -Storage über Systemschichten hinweg. Das Framework baut auf dem Provenance-Dateisystem PASS[25] auf. Es ist keine generische Lösung, wie die in dieser Arbeit vorgestellte, sondern ist spezifisch auf die drei Schichten Dateisystem, Browser und Workflow-Engine zugeschnitten. Insbesondere beruht es auf der Betriebssystemintegration von PASS, während der Ansatz dieser Arbeit auch auf Anwendungsebene funktioniert.

Herkenhöhner et al.[26] stellen eine Architektur und ein Nachrichtenformat vor, um die für das Recht auf Auskunft erforderliche Provenance in organisationsübergreifenden Webservices automatisiert zu sammeln und dem Betroffenen zur Verfügung zu stellen. Sie stellen ausschließlich ein Nachrichtenformat für verteilte Data-Provenance-Systeme vor, entwerfen und implementieren jedoch kein solches System. Diese Arbeit geht jedoch alle Schritte, von der Anforderungsanalyse bis zur Implementierung. Nachrichtenformate wurden in dieser Arbeit nicht auf der Ebene von Webservices, sondern implementierungsnah definiert.

Kapitel 5.1 listet eine Reihe weiterer Provenance-Systeme auf, die spezifisch für bestimmte Abstraktionsschichten sind und sich an einen speziellen Anwendungsfall wie das wissenschaftliche Rechnen richten. All diese Systeme sind nicht geeignet, um zielgerichtet die Provenance personenbezogener Daten gemäß der datenschutzrechtlichen Anforderungen zu sammeln. Diese Arbeit stellt einen generischen Ansatz vor, der auf beliebigen Abstraktionsschichten funktioniert und richtet die Art der Provenance speziell auf den Anwendungsfall der Datenschutzauskunft aus.

Pulls et al.[27] stellen das kryptographische Schema *Insynd* vor. Es ermöglicht, Personal-Data-Provenance zu sammeln, ohne die Verknüpfungen zwischen den einzelnen Logs zu offenbaren. Wie in Kapitel 9.10.3 diskutiert, kann mit diesem Schema in gewissen Fällen der durch den Einsatz eines Datenschutzauskunftssystems entstehende Grad der Unverkettbarkeit verbessert werden. Der Ansatz hat allerdings den Nachteil, dass der Betroffene in den Setup-Prozess des Logging-Schemas involviert werden muss. Geklärt wird außerdem nicht, wie sich die Personal-Data-Provenance zusammensetzt und wie sie aggregiert werden kann.

Die in Kapitel 5.1 gelisteten Arbeiten zur Sicherheit von Data-Provenance sind orthogonal zu dieser Arbeit. Die Sicherheit von Data Provenance wird gemäß der Annahmen in

[23] Demsky 2009; Demsky 2011.
[24] Muniswamy-Reddy et al. 2009.
[25] Holland et al. 2008.
[26] Herkenhöner et al. 2010.
[27] Pulls/Peeters/Wouters 2013; Pulls/Peeters 2015.

Kapitel 5.3.3 vorausgesetzt. Diese Arbeit liefert zu diesem Themenkomplex keine eigenen Beiträge.

Ebenso liefert diese Arbeit keine eigenen Beiträge zu Usage-Control. Die in Kapitel 5.2 vorgestellten Forschungsergebnisse werden, aus den im selbigen Kapitel geschilderten Gründen, im Datenschutzauskunftssystem verwendet.

Unverkettbarkeit Bestehende Arbeiten zur Bestimmung des Begriffs der Unverkettbarkeit wurden bereits in Kapitel 9.1.1 vorgestellt. Gegenüber den dort erwähnten Definitionen ist die in dieser Arbeit verwendete Definition genauer. Sie benennt den Angreifer und die verwendeten Relationen für das „zusammenführen" personenbezogener Daten, Betroffener und datenverarbeitender Systeme.

Die in Kapitel 9.1.2 vorgestellten Modelle zur Relationenunterscheidbarkeit haben alle den Nachteil, dass sie nach dem „all-or-nothing"-Ansatz arbeiten. In einer Situation, in der ein gewisser Wissenszuwachs unabdingbar ist und in Kauf genommen wird, wie bei Auskunftssystemen, ist solch ein Ansatz jedoch nicht hilfreich. Die Metrik würde jederzeit trivial messen, dass die Unverkettbarkeit nicht gewahrt bleibt. Deshalb untersucht diese Arbeit relative Ansätze im Sinne entropiebasierter Unverkettbarkeitsmetriken.

Eine erste entropiebasierte Metrik für Anonymität wird von Serjantov und Danezis[28] vorgestellt. Sie überführen das klassische „Anonymity Set" auf ein nach den Wahrscheinlichkeiten der einzelnen Elemente der Menge gewichtetes Maß. Diaz et al.[29] ergänzen die Normierung des Anonymitätsmaßes. Diese Arbeit verwendet dagegen eine entropiebasierte Metrik zur Messung des Grads der Unverkettbarkeit. Dazu kann auf beliebige Relationen zurückgegriffen werden. Die Metriken für Anonymität sind ein Spezialfall davon und können von der in dieser Arbeit vorgestellten Metrik mit abgebildet werden.

Die Arbeiten von Steinbrecher und Köpsel[30] übertragen den informationstheoretischen Ansatz auf Unverkettbarkeit. Der Ansatz wird von Franz et al.[31] aufgegriffen. Beide beschränken die Metrik auf Äquivalenzrelationen. Franz et al. diskutieren zusätzlich bestimmte Sonderfälle von Beobachtungen, die ein Angreifer machen kann. Beispiele sind die Anzahl der Äquivalenzklassen, die Kardinalität der Äquivalenzklassen und die Aufdeckung der Verkettungsrelation für eine Teilmenge der betrachteten Entitäten. Fischer, Katzenbeisser und Eckert[32] beschränken sich ebenfalls auf Äquivalenzrelationen. Sie definieren ergänzend eine innere und äußere Struktur von Unverkettbarkeit. Diese Arbeit betrachtet beliebige Relationen. Eine Beschränkung auf Äquivalenzrelationen würde die Modellierung einer Beziehung zwischen personenbezogenen Daten und Betroffenen nicht zulassen.

[28] Serjantov/Danezis 2003.
[29] Diaz et al. 2003.
[30] Steinbrecher/Köpsell 2003.
[31] Franz/Meyer/Pashalidis 2007.
[32] Fischer/Katzenbeisser/Eckert 2008.

Pashalidis[33] verallgemeinert entropiebasierte Metriken von Äquivalenzrelationen auf alle homogenen Relationen. Je nach Art der Relation bezeichnet er die Metrik unterschiedlich und führt Definitionen wie „Fairness“ und „Undurchsichtigkeit“ ein. Diese Arbeit beschränkt sich auf Unverkettbarkeitsmetriken und betrachtet dabei beliebige Relationen. Darüber hinaus wird in dieser Arbeit das Hintergrundwissen des Angreifers explizit modelliert. Statt eines Maximum-Entropie-Priors wird ein angreiferspezifischer Prior bestimmt.

Benutzeroberflächen zur automatisierten Auskunftserteilung Bestehende Ansätze zur Visualisierung der Verarbeitung personenbezogener Daten wurden bereits in Kapitel 1.1.2 vorgestellt.

Das Data Disclosure Log von Kolter et al.[34] ist im Gegensatz zum in dieser Arbeit entwickelten *PrivacyInsight* keine Auskunftsplattform, sondern greift nur auf die auf Betroffenenseite bereits vorhandenen Informationen zu. Gleiches gilt für Data Track.[35] Beide Tools sind darüber hinaus nicht webtauglich, sondern konventionelle Desktopanwendungen.

Translucene Map von Kani-Zabihi und Helmhout[36] stellt Datenverarbeitungsvorgänge ausschließlich auf Grundlage des Verfahrensverzeichnisses dar.

GenomSynlig[37] ist das einzige in der Literatur existierende Tool zur interaktiven elektronischen Auskunftserteilung. Es liefert allerdings nur Informationen, wenn der Betroffene die Quelle der personenbezogenen Daten ist. Vom Informationsgehalt her geht *PrivacyInsight* unter anderem in der Darstellung interner Datenflüsse und der Visualisierung aller Empfänger über GenomSynlig hinaus. Die Nutzerstudie des Kapitels 12 hat außerdem gezeigt, dass *PrivacyInsight* besser bedienbar ist.

13.2.2 Grenzen der Arbeit

Diese Arbeit beschränkt sich auf deutsches und europäisches Datenschutzrecht. Sie berücksichtigt die Rechtslage nach dem BDSG umfassend. Das dem im Grundsatz entsprechende Auskunftsrecht der DSGVO wird themenbezogen mitbetrachtet, jedoch bis auf das Recht auf Datenübertragbarkeit nicht vertieft behandelt.[38]

Die angenommenen Voraussetzungen für die zuverlässige Erbringung einer automatisierten Auskunft werden in Kapitel 5.3.3 erörtert. Insbesondere die Annahme, dass das Datenschutzauskunftssystem in allen datenverarbeitenden IT-Systemen korrekt aufgesetzt

[33] Pashalidis 2008.
[34] Kolter/Netter/Pernul 2010.
[35] Wästlund/Fischer-Hübner et al. 2010.
[36] Kani-Zabihi/Helmhout 2012.
[37] Fischer-Hübner/Angulo/Pulls 2014; Angulo et al. 2015.
[38] Vgl. die Ausführungen in Kapitel 2.2.2.

sein muss, begrenzt die Nutzbarkeit der Ergebnisse dieser Arbeit. Aufgrund der Annahme ist eine automatisierte Auskunft beim Rückgriff auf Cloud-Auftragsdatenverarbeiter schwierig. Die verantwortliche Stelle müsste den Cloud-Anbieter dazu verpflichten, seine verwendete Software anzupassen. Insbesondere bei großen, internationalen Anbietern ist dies kaum zu erreichen. Durch die DSGVO und Standardisierungsbemühungen könnte dieses Problem möglicherweise überwunden werden.

Ein noch nicht vollständig gelöstes Problem des Informationsflusstrackings ist *Label Creep*. Kann bei einem Ereignis nicht festgestellt werden, ob ein personenbezogenes Datum geflossen ist, muss konservativ angenommen werden, dass ein Fluss stattgefunden hat. Dadurch werden mehr Container als Speicherorte personenbezogener Daten aufgefasst als tatsächlich personenbezogene Daten enthalten. Ein Ansatz wie quantitatives Informationsflusstracking[39] kann nicht verfolgt werden. Es ist nicht möglich eine Informationsmenge zu definieren, ab der personenbezogene Daten nicht mehr sensibel oder vorhanden sind.

Die in Kapitel 7.2 vorgestellten Abstraktionsregeln sind containerzentriert und nicht datenzentriert. Insofern ist es im Rahmen des Modells nicht möglich, für unterschiedlich sensible Daten den Detailgrad des Trackings zu variieren. Auskunftsrechtlich sind besonders sensible personenbezogene Daten zwar nicht anders zu betrachten als andere Daten. Anforderungen aus anderen Bereichen könnten jedoch möglicherweise zu einem anderen Ergebnis führen.

Eine Einschränkung des im Kapitel 7.3 beobachten Vorteils von Abstraktionsregeln beim Speicherverbrauch ergibt sich daraus, dass Abstraktionsregeln nur wirksam sein können, wenn ein hoher Abstraktionsgrad der Provenance akzeptabel ist. Werden einzelne Verarbeitungsschritte zu vollständig unterschiedlichen Zwecken durchgeführt, müssen diese Schritte in der Provenance erkennbar sein. Ein höherer Speicherbedarf ist dann die Folge.

Die Frage, wie unstrukturierte personenbezogene Daten zum Erhebungszeitpunkt erkannt und mit Provenance-Policies annotiert werden können, ist bisher ungelöst. Ihre Beantwortung ist Voraussetzung für eine vollständige automatisierte Auskunftserteilung. Erste Ideen existieren für die Detektion personenbezogener Daten in E-Mails.[40]

Ein Bereich, in dem es besonders schwer ist Daten einem Betroffenen zuzuordnen und so eine Auskunft zu ermöglichen, ist die Videoüberwachung. Zumindest in abgegrenzten, nicht-öffentlichen Bereichen könnten intelligente Systeme zukünftig eine Lösung bieten. Mitarbeiter könnten sich zu Arbeitsbeginn gegenüber einer Kamera ausweisen und dann im Nachhinein die über sie aufgezeichneten Bilder einsehen.

Nur unvollständig gelöst ist die Frage nach Integrität und Verfügbarkeit bei gleichzeitiger Vertraulichkeit der Provenance. Die Arbeiten von Pulls bieten,[41] mit den oben

[39] Lovat/Oudinet/Pretschner 2014.
[40] Bier/Prior 2014.
[41] Pulls/Peeters 2015.

sowie im Kapitel 9.10.3 erwähnten Einschränkungen, eine Alternative zum hier verfolgten Konzept der dezentralen Speicherung.

Die dem für Unverkettbarkeit verwendeten Angreifermodell zugrundeliegenden Annahmen werden in Kapitel 9.6 erläutert. Eine wesentliche Annahme ist die Unabhängigkeit der Flüsse zweiter Daten. Durch diese konservative Annahme wird ein möglicher Teil des Hintergrundwissens des Angreifers ausgeblendet. Es ist eine Vereinfachung, die so lange vorzuziehen ist, so lange tatsächliche Abhängigkeiten unbekannt sind.

Die vorgestellte Metrik für den Grad der Unverkettbarkeit basiert auf einer rückschauenden Betrachtung. Für die Einbindung der Metrik bei der Erhebung personenbezogener Daten ist eine clientseitige Vorausberechnung notwendig. Die dafür erforderlichen Verfahren und Vorhersagemodelle sind Aufgabe zukünftiger Arbeiten. Der Betroffene könnte in der Konsequenz a priori entscheiden, ob er seine personenbezogenen Daten teilt und ob er ein Datenschutzauskunftssystem nutzen möchte.

Die Nutzerstudie des Kapitels 12 ermöglicht keine Aussage darüber, ob die Unverkettbarkeitsmetrik tatsächlich verstanden wurde und ob sie tatsächlich die Entscheidung des Betroffenen beeinflusst. Festgestellt wurde nur ein subjektiver Eindruck der Testpersonen.

13.3 Ausblick

Mit dem vorliegenden Datenschutzauskunftssystem ist ein großer Schritt hin zu einer, sowohl für verantwortliche Stellen als auch für Betroffene, besseren Umsetzbarkeit des Rechts auf Auskunft getan. Eine elektronische Auskunft entlastet die Unternehmen personell und gibt den Betroffenen gleichzeitig die Gewissheit, jederzeit in die Verarbeitung ihrer personenbezogenen Daten Einblick nehmen zu können. Eine elektronische Auskunft ist zuverlässiger als die bisherigen manuellen Verfahren und gibt dem Betroffenen die Gelegenheit, im Anschluss an die Auskunft, unmittelbar seine weiteren Rechte wahrzunehmen. *PrivacyInsight* hat bisher von allen Seiten außerordentlich positive Resonanz bekommen.

Ein für die Zukunft wichtiges Handlungsfeld ist die Verknüpfung unterschiedlicher datenverarbeitender Stellen. Die DSGVO hat dafür die Voraussetzungen geschaffen. Gemeinsam für die Verarbeitung Verantwortliche sind vorgesehen. Auftragsdatenverarbeiter sind schon seit jeher in den Verantwortungsbereich einer verantwortlichen Stelle miteinbezogen. Insofern ist es von höchstem Interesse, Schnittstellen für eine stellenübergreifende Auskunft zu schaffen. Nur so kann dem Betroffenen langfristig eine nahtlose Nachvollziehbarkeit der Datenverarbeitungsvorgänge geboten werden.

So vorausschauend die DSGVO auch ist, sie ist nicht in jeder Hinsicht konsequent. Sie führt zwar mit dem Recht auf Datenübertragbarkeit eine Regelung für den Austausch personenbezogener Daten ein, der Verordnungsgeber hat sich jedoch nicht durchringen können auch die notwendigen technischen Anforderungen zu stellen. ErwGr 68 stellt fest,

dass es mit dem Recht auf Datenübertragbarkeit keine Pflicht zu kompatiblen IT-Systemen gibt. Eine so ausgestaltete Regelung ist stumpf.

Die Abstimmung zwischen den einzelnen Transparenzmaßnahmen könnte auch optimaler gestaltet sein. Unterrichtung, Benachrichtigung und Auskunft sind voneinander fast unabhängige Verpflichtungen. Es wäre in der Praxis sinnvoller, wenn die verantwortliche Stelle die passende Transparenzmaßnahme selbst wählen könnte. Wird der Betroffene bei einer einmaligen Datenverarbeitung vollständig unterrichtet, macht ein Auskunftsanspruch keinen Sinn mehr. Werden, wie bei Big Data in der Cloud, Daten vielfältigst verarbeitet und geteilt, macht es möglicherweise Sinn, den Betroffenen nicht über jede Übermittlung zu benachrichtigen, sondern ihm stattdessen eine einfache, stellenübergreifende Auskunftsmöglichkeit zur Verfügung zu stellen.

Abschließend noch ein Satz zur Rechtsdurchsetzung. Die härteren Sanktionen der DSGVO wirken abschreckend. Solange die Aufsichtsbehörden personell schlecht aufgestellt sind,[42] ist der vom nationalen Gesetzgeber eingeschlagene Weg jedoch der effektivere. Die Regelungen des UKlaG sollten so präzisiert werden, dass sie auch die Auskunft und die übrigen Betroffenenrechte umfassen.

[42] Nebenbei bemerkt sind stärkere und von jeder Kontrolle unabhängige Aufsichtsbehörden auch nicht der Traum eines jeden liberalen Staatstheoretikers.

Appendix

A Studie zum Status Quo des Auskunftsrechts in der Praxis

Die eigens erstellte,[1] Studie aus den Jahren 2015/16 verfolgte einen dualen, erst quantitativen und anschließend qualitativen, Ansatz. Die zunächst ausgewählte Stichprobe war weitaus größer und umfasste 612 Webseiten. Die Registrierung der entsprechenden Anzahl echter Personen und ihrer E-Mailadressen hätte einen enormen Aufwand bedeutet. Erfreulicherweise verpflichtet § 13 Abs. 8 TMG die Betreiber von Telemediendiensten, eine pseudonyme Nutzung der Dienste zu ermöglichen. Deshalb konnte auf künstlich erzeugte Persönlichkeitsprofile zurückgegriffen werden. Die umfangreichen Profile, die Namen, postalische Adresse, Alter, Geschlecht und Interessen umfassten, wurden mit dem *Fake Name Generator* erzeugt.[2] Für jedes Profil wurde nach dem Schema vorname.nachname@domain.de eine neue E-Mailadresse auf einem eigens dafür eingerichteten E-Mailserver erstellt. Die verwendete Domain wurde extra für die Studie registriert. Sie war somit noch unbekannt und es war kein zufällig an diese Domain adressierter Spam zu erwarten. Je Webseite wurde ein Profil registriert, wobei 151 Benutzeraccounts erstellt und 460 Newsletterregistrierungen vorgenommen wurden. Es wurden so viele personenbezogene Daten wie möglich angegeben, um den Datensatz so attraktiv wie möglich zu machen.

Die repräsentative Stichprobe für die Studie wurde durch die Klassifikation der Wirtschaftszweige WZ 2008 des Statistischen Bundesamtes bestimmt.[3] WZ 2008 basiert auf NACE[4] gemäß der Richtlinie 29/2002/EG und UN ISIC.[5] Die Klassifikation ist hierarchisch strukturiert und besteht aus 21 Abschnitten, 88 Abteilungen, 272 Gruppen und 615 Klassen. Bei der Studie wurden nur Abschnitte mit einer Onlinepräsenz von mindestens 60 % gemäß den Angaben des statistischen Bundesamtes berücksichtigt.[6] Außerdem wurde Wert auf eine heterogene Zusammensetzung der Stichprobe gelegt. Deshalb wurden

[1] Bier/Kömpf/Beyerer 2017.

[2] http://www.fakenamegenerator.com.

[3] https://www.destatis.de/DE/Methoden/Klassifikationen/GueterWirtschaftklassifikationen/klassifikationenwz2008.pdf?__blob=publicationFile, abgerufen am 9. Mai 2017.

[4] Statistische Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft, *franz. Nomenclature statistique des activités économiques dans la Communauté européenne*.

[5] United Nations International Standard Industrial Classification of all Economic Activities.

[6] https://www.destatis.de/DE/ZahlenFakten/GesamtwirtschaftUmwelt/UnternehmenHandwerk/IKTUnternehmen/Tabellen/04_AnteilUnternehmenInternetzugang_Website_IKT_Unternehmen.html, abgerufen am 9. Mai 2017.

Abteilungen ausgeschlossen, in denen keine Unternehmen unterschiedlicher Größe gemäß der KMU-Kriterien (siehe Tabelle A.1), definiert in der Empfehlung der EU-Kommission 2003/361/EG, gefunden werden konnten. Die Angaben zu einzelnen Unternehmen wurden, wo möglich, dem Bundesanzeiger entnommen.[7] Aus praktischen Gründen wurden

Unternehmenskategorie	Mitarbeiter	Jahresumsatz	Jahresbilanzsumme
Mittlere Unternehmen	< 250	$\leq$ €50 m	$\leq$ €43 m
Kleine Unternehmen	< 50	$\leq$ €10 m	$\leq$ €10 m
Kleinstunternehmen	< 10	$\leq$ €2 m	$\leq$ €2 m

Tabelle A.1: Definition der Kleinstunternehmen sowie der kleinen und mittleren Unternehmen (KMU-Definition)[8]

weiterhin all jene Abteilungen ausgeschlossen, in denen die meisten Unternehmenswebseiten keine Möglichkeit zur Registrierung einer E-Mailadresse zuließen. Dieses Kriterium wurden anhand zufälliger Stichproben der Suchmaschinentreffer bei Eingabe der Abteilung als Suchbegriff überprüft. Von jedem Abschnitt wurden die Gruppen für die Studie ausgewählt, die die oben genannten Kriterien am besten erfüllen. Insgesamt bestand die Stichprobe aus 612 Unternehmen in 17 Gruppen, d. h. aus 36 Unternehmen je Gruppe.[9]

Die Studie wurde unter der Hypothese erstellt, dass ein gewisser Teil der Unternehmen personenbezogene Daten an Dritte weitergibt.[10] Die Nutzung der weitergegebenen personenbezogenen Daten, genauer gesagt der E-Mailadressen, durch Dritte könnte dann anhand auf dem Mailserver eingehender E-Mails erkannt werden. Anschließend könnte anhand der individuellen E-Mailadresse die Übermittlung der Daten zu dem Unternehmen zurückverfolgt werden, bei dem sie ursprünglich registriert wurde. Auf dieser Grundlage wäre es möglich, Auskunftsersuchen bei den ursprünglichen Unternehmen zu stellen und zu überprüfen, ob die Datenübermittlungen wahrheitsgemäß angegeben würden.

Der Mailserver wurde zunächst für 102 Tage im Frühling und Sommer 2015 überwacht und anschließend noch für weitere 6 Monate unter Beobachtung gestellt. Im ersten Zeitraum gingen nur unter fünf Adressen E-Mails von Domains ein, die nicht dem Unternehmen zugeordnet werden konnten, bei dem die Adresse ursprünglich registriert wurde. Im zweiten Beobachtungszeitraum hat sich daran nichts mehr geändert. In einem

[7] http://www.bundesanzeiger.de.

[8] Empfehlung der Kommission vom 6. Mai 2003 betreffend die Definition der Kleinstunternehmen sowie der kleinen und mittleren Unternehmen, 2003/361/EG.

[9] Die Ausgewählten Gruppen waren, diejenigen mit dem WZ 2008 Code 11.0, 18.2, 45.1, 47.9, 55.1, 56.1, 58.1, 58.2, 59.1, 63.1, 63.9, 79.1, 86.9, 90.0, 92.0, 93.1 und 96.0.

[10] Gestützt durch Skandale (Zeit Online, http://www.zeit.de/online/2008/37/datenschutz-ergebnis, abgerufen am 9. Mai 2017) und die Erkenntnis, dass die Verpflichtung der Mitarbeiter auf das Datengeheimnis nach § 5 S. 2 BDSG in vielen Fällen nicht erfolgt (Ehmann in: Simitis, BDSG 2014, § 5 Rn. 30).

der Fälle konnte aufgeklärt werden, dass die Domain doch zum ursprünglichen Unternehmen gehört. Von den verbleibenden vier Fällen antworteten zwei Unternehmen, eine Online-Lotterie und ein Escort-Service, nicht auf das Auskunftsersuchen. Ein Unternehmen, ebenfalls eine Online-Lotterie, antwortete, dass es die Anfrage nicht verstünde, und ein Unternehmen, ein Online-Magazin, wies den Verdacht zurück, dass es Daten weitergegeben habe. Bei letzterem Unternehmen lässt die Art der eingegangenen E-Mails darauf schließen, dass Spammer an die registrierte Adresse gelangt sind. Von den verdächtigen Unternehmen waren drei nicht-deutscher Herkunft, während sich in der Stichprobe insgesamt nur etwa 12 % nicht-deutsche Unternehmen befanden.

Um die qualitative Analyse auf eine breitere Basis zu stellen, wurden zusätzlich Auskunftsersuchen an Unternehmen versendet, bei denen die beteiligten Forscher bereits seit längerem registriert waren. Insgesamt kamen so 40 Auskunftsersuchen zusammen. Die Auskunftsersuchen wurden zunächst per E-Mail gestellt.[11] Wo nötig, insbesondere im Bankensektor, wurde eine postalische Anfrage nachgereicht. Hinweise auf die Art der personenbezogenen Daten wurden auf Nachfrage bereitwillig gegeben.

In der erweiterten Stichprobe befanden sich sechs international agierende US-amerikanische Unternehmen, die auch auf dem deutschen Markt stark vertreten sind. Bei zwei Unternehmen konnte die Herkunft nicht zweifelsfrei geklärt werden. Die übrigen 32 Unternehmen stammen aus Deutschland, darunter sechs DAX-Konzerne und wesentliche, deutschlandweit vertretene Internetzugangs-und Mobilfunkanbieter. Versicherungen, Banken, Handel, Logistik, Auskunfteien und viele weitere Branchen waren ebenso vertreten. Die Stichprobe ist nicht repräsentativ, bietet aber einen guten Querschnitt durch die deutsche Unternehmenslandschaft.

Die qualitative Prüfung der Auskünfte erfolgte entlang der zwölf in Tabelle A.3 aufgelisteten Anforderungen. Die Anforderungen wurden aus den Vorgaben des Datenschutzrechtes abgeleitet. Details dazu finden sich im Kapitel 3 dieser Arbeit. Querreferenzen sind in der Tabelle mitangegeben.

Die Anforderungen teilen sich in drei formale (# 1-3) und neun inhaltliche (# 4-12) Anforderungen auf. Anforderung 1 wurde als erfüllt gewertet, wenn auf der Webseite des Unternehmens eine geeignete E-Mailadresse oder Postadresse angegeben oder ein Webformular verfügbar war. Die Anforderung wurde als teilweise erfüllt angesehen, falls es möglich war, das Unternehmen auf irgendeinem anderen Weg zu kontaktieren. Anforderung 2 wurde als erfüllt angesehen, falls die Anfrage innerhalb von 28 Tagen (4 Wochen) beantwortet wurde. Unternehmen, deren Antworten als teilweise erfüllt angesehen wurden, hatten zwar zum Ende des Zeitraums geantwortet, jedoch nur auf mehrfache Nachfrage. Unternehmen, die die Anforderung nicht erfüllt haben, hatten auch nach Ablauf der doppelten Zeit noch nicht geantwortet. Im Durchschnitt wurden Auskunfts-

[11] Wortlaut: Sehr geehrte Damen und Herren, hiermit mache ich von meinem datenschutzrechtlichen Auskunftsrecht Gebrauch. Bitte übersenden Sie mir die vollständige Auskunft per E-Mail oder postalisch an [...]. Mit freundlichen Grüßen [...].

#	Anforderung	Kapitel
1	Das Recht auf Auskunft muss über die üblichen Kommunikationskanäle wahrnehmbar sein.	3.3
2	Das Auskunftsersuchen muss innerhalb einer angemessenen Frist beantwortet werden.	3.5
3	Die Auskunft muss für gewöhnliche Bürger verständlich sein.	3.5
4	Alle von der verantwortlichen Stelle gespeicherten personenbezogenen Daten müssen in nicht-abstrakter Weise aufgeführt werden.	3.7.1
5	Der Ort der Speicherung (Datei, Datenbank, Stelle, Abteilung,...) aller personenbezogenen Daten muss mitgeteilt werden.	3.7.1
6	Die Herkunft aller personenbezogenen Daten muss erwähnt werden.	3.7.3
7	Jede Übermittlung personenbezogener Daten muss angegeben werden.	3.7.2
8	Alle Empfänger oder Kategorien von Empfängern personenbezogener Daten müssen erwähnt werden.	3.7.2
9	Die übermittelten personenbezogenen Daten müssen aufgelistet werden.	3.7.2
10	Der interne Fluss personenbezogener Daten muss auf nachvollziehbare Weise beauskunftet werden.	3.7.2
11	Der Zweck der Verarbeitung muss für jedes personenbezogene Datum angegeben werden.	3.7.5
12	Der Zweck der Erhebung eines personenbezogenen Datums muss in jedem Verarbeitungsschritt nachvollziehbar sein.	3.7.5

Tabelle A.3: Allgemeine Anforderungen an Datenschutzauskünfte

anfragen nach etwa 8 Tagen beantwortet. Anforderung 3 wurde als erfüllt angesehen, falls keine unverständliche Codierung der Daten verwendet wurde. Die Anforderung wurde als teilweise erfüllt gewertet, falls nur vereinzelte Angaben unverständlich waren. Die Anforderungen 4-12 wurden als erfüllt betrachtet, wenn die jeweiligen Angaben vollständig waren. Sie wurden als teilweise erfüllt gewertet, wenn jeweils zumindest ein überwiegender Anteil der erforderlichen Angaben vorhanden war. Die Vollständigkeit wurde großzügig interpretiert.

Abbildung A.1 zeigt den Erfüllungsgrad der Anforderungen durch die erteilten Auskünfte. Die Situation bei den formalen Kriterien liegt im akzeptablen Bereich. Bis auf ein Unternehmen konnten alle auf irgendeine Art und Weise kontaktiert werden. 27 Unternehmen (67,5 %) beantworteten die Auskunftsanfrage in weniger als 15 Tagen. Nur sechs Unternehmen stellten auch nach 6 Monaten noch keine Auskunft zur Verfügung. Vier davon waren die verdächtigen Unternehmen aus dem ersten Teil der Studie. Alle

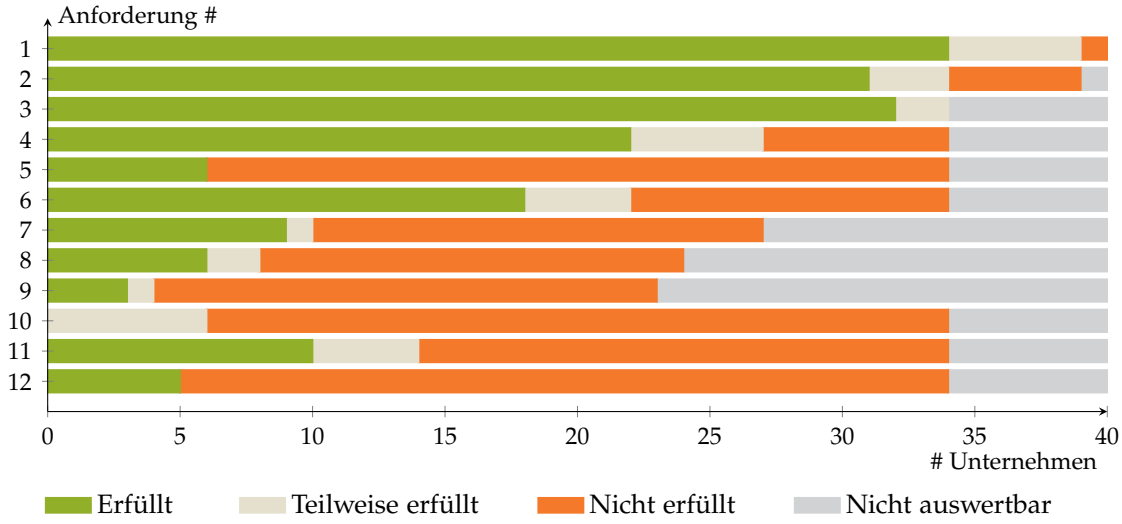


Abbildung A.1: Erfüllungsgrad der Anforderungen durch die erteilten Auskünfte

Auskünfte bis auf zwei waren klar strukturiert und verständlich formuliert.

Die Lage bei den inhaltlichen Kriterien fällt deutlich schlechter aus. Insbesondere Angaben zu Speicherort, Empfängern, Datenflüssen und dem Zweck der Speicherung sind unvollständig. Der Speicherort wurde meist dann sichtbar, wenn Screenshots von Datenbankanwendungen als Teil der Auskunft mitgeliefert wurden. Dies war immerhin bei sechs Unternehmen der Fall. Die internen Flüsse personenbezogener Daten wurden in keinem Fall voll zufriedenstellend wiedergegeben. Nur wenige Auskünfte beinhalten überhaupt Angaben dazu. Positive Ausnahme war ein deutscher Versand- und Einzelhändler, der grundlegende Informationen zu den an der Datenverarbeitung beteiligten Abteilungen mitlieferte. Er stand auch insgesamt mit 11 erfüllten oder teilweise erfüllten Anforderungen an der Spitze.

B Ableitung datenschutzrechtlicher Anforderungen und Kriterien sowie technischer Anforderungen an Datenschutzauskunftssysteme

In den nachfolgenden Abschnitten wird die Ableitung der einzelnen datenschutzrechtlichen Anforderungen und Kriterien sowie der technischer Anforderungen an Datenschutzauskunftssysteme gemäß EVAL (Kapitel 4.1.2) im Detail nachvollzogen. Für jeden Teilschritt des Models wird dargelegt, inwiefern die abgeleiteten Anforderungen und Kriterien vollständig sind. Außerdem werden Argumente für ihre jeweilige Korrektheit angeführt.

B.1 Ableitung datenschutzrechtlicher Anforderungen

Als einfachgesetzliche Konkretisierung der rechtlichen Ziele lassen sich die rechtlichen Anforderungen aus den maßgeblichen Gedanken der Gesetzgebung herleiten. Wesentliche Rechtsgrundlage für den Datenschutz in Deutschland ist das BDSG. Deshalb bezieht sich die Analyse auch schwerpunktmäßig auf dieses Gesetz.

An den Stellen, an denen es sinnvoll erscheint wird darüber hinaus auch die gegenwärtige und zukünftige europäische Gesetzgebung, insbesondere die DSGVO, mit in den Blick genommen. So wird vermieden, dass wichtige zukünftige oder internationale Anforderungen außer Acht gelassen werden.

An wenigen Punkten werden auch Regelungsvorgaben des Bundesverfassungsgerichts mit in den Blick genommen, die zwar nicht den Weg in die einfachen Gesetze gefunden haben, aber in der Gesamtschau der Grundrechte dennoch auch für den Datenschutz zu berücksichtigen sind.

Die Analyse der Vollständigkeit der Anforderungen bezieht sich nur auf das Bundesdatenschutzgesetz. Rechtliche Anforderungen sind nicht statisch. Ebenso wie die Gesetze selbst unterliegen auch sie einem ständigen Wandel. Deshalb kann auch niemals eine absolute Vollständigkeitsaussage getroffen werden.

B.1.1 Synthese - Korrektheit der Anforderungen

Eine datenschutzrechtliche Anforderung ist dann korrekt, wenn sie sich auf eine rechtliche Grundlage stützen kann, diese Rechtsgrundlage zutreffend wiedergegeben ist und die Anforderung klar gegenüber anderen Anforderungen abgegrenzt ist. Die Tabelle B.1 gibt einen Überblick über die einfachgesetzliche Verankerung der datenschutzrechtlichen Anforderungen. In den nachfolgenden Abschnitten wird für jede Anforderung behandelt, inwiefern sie korrekt hergeleitet wurde.

Datenschutzrechtliche Anforderung	Einfachgesetzliche Verankerung
Datenvermeidung	§ 3a S. 1 BDSG, Art. 5 Abs. 1 Lit. c DSGVO
Datensparsamkeit	§ 3a S. 1 BDSG, Art. 5 Abs. 1 Lit. c u. Art. 32 Abs. 1 Lit. a DSGVO
Anonyme und pseudonyme Dienste	§ 13 Abs. 6 u. 7 TMG
Datenschutzfreundliche Grundeinstellungen	Art. 25 Abs. 2 DSGVO
Datengeheimnis	§ 5 BDSG
Zutritts- und Zugangskontrolle	Anlage zu § 9 S. 1 BDSG Nr. 1 u. 2, Art. 5 Abs. 1 Lit. f DSGVO
Zugriffs- und Weitergabekontrolle	Anlage zu § 9 S. 1 BDSG Nr. 3 u. 4, Art. 32 Abs. 1 Lit. b DSGVO
Verfügbarkeitskontrolle	Anlage zu § 9 S. 1 BDSG Nr. 7, Art. 32 Abs. 1 Lit. c DSGVO
Datenkorrektheit	Art. 6 Lit. d DSRL, Art. 5 Abs. 1 Lit. d DSGVO
Revisionssichere Protokollierung (Eingabekontrolle)	Anlage zu § 9 S. 1 BDSG Nr. 5, Art. 5 Abs. 1 Lit. f DSGVO
Recht auf Datenübertragbarkeit	Art. 20 DSGVO
Prinzip der Verantwortlichkeit	Definitorisch in § 3 Abs. 7 BDSG, Art. 5 Abs. 2 DSGVO
Prinzip der nicht-automatisierten Einzelentscheidung	§ 6a BDSG, , Art. 22 DSGVO
Einwilligung	§ 4a BDSG, Art. 7 DSGVO
Einheitliche Ansprechpartner	Art. 26 DSGVO
Datenschutzkontrolle durch unabhängige BfD	§ 4f, §§ 22 ff. BDSG, Art. 39 DSGVO
Meldepflicht	§ 4d Abs. 1 BDSG, Art. 36 DSGVO
Aufklärungspflichten	u. a. § 4a Abs. 1 S. 2 BDSG
Hinweis- und Kennzeichnungspflichten	u. a. § 6b Abs. 2 BDSG
Unterrichtungspflichten	§ 4 Abs. 3 S. 1 BDSG, Art. 13 DSGVO

Benachrichtigungspflichten	§§ 33, 19a BDSG, Art. 14 DSGVO
Auskunftsanspruch	§§ 19, 34 BDSG, Art. 15 DSGVO
Vorabkontrolle und Datenschutzfolgenabschätzung	§ 4d Abs. 5 BDSG, Art. 35 DSGVO
Datenschutzaudit	§ 9a BDSG, , Art. 41, 42 DSGVO
Informationspflichten (bei DS-Verletzungen)	§ 42a BDSG
Führung eines Verfahrensverzeichnisses	§ 4g Abs. 2 i. V. m. § 4e S. 1 BDSG
Recht auf Berichtigung	§§ 20 Abs. 1, 35 Abs. 1 BDSG, Art. 16 DSGVO
Recht auf Löschung	§§ 20 Abs. 2, 35 Abs. 2 BDSG, Art. 17 DSGVO
Recht auf Sperrung	§§ 20 Abs. 3, 35 Abs. 3 BDSG, Art. 18 DSGVO
Widerspruchsrecht	§ 20 Abs. 5 BDSG, Art. 21 DSGVO
Zweckbestimmung	Über das ges. BDSG verteilt; u. a. § 28 Abs. 1 S. 2 BDSG
Zweckbindung und Zwecktrennung	Über das ges. BDSG verteilt, technisch in der Anlage zu § 9 S. 1 BDSG Nr. 8, Art. 5 Abs. 1 Lit. b DSGVO
Organisatorische und technische Gewaltenteilung	BVerfG, 18.12.1987, 1 BvR 962/87 = NJW 1988, 959
Grundsatz der Direkterhebung	§ 4 Abs. 2 S. 1 BDSG
Verbot mit Erlaubnisvorbehalt	§ 4 Abs. 1 BDSG, Art. 6 DSGVO
Verhältnismäßigkeitsprinzip	u. a. § 4 Abs. 2 S. 2 Nr. 2 Lit. a BDSG

Tabelle B.1: Tabellarische Übersicht der Herkunft datenschutzrechtlicher Anforderungen

Datenvermeidung Die Datenvermeidung findet sich wörtlich in § 3a S. 1 BDSG. Sie fordert, dass nur Daten erhoben werden, die für den jeweiligen definierten Zweck tatsächlich benötigt werden. Die Datenvermeidung führt zum höchsten Datenschutzniveau, da nur von Daten, die tatsächlich erhoben werden, eine Gefährdung für das Recht auf informationelle Selbstbestimmung ausgehen kann.

Datensparsamkeit Die Datensparsamkeit ist, wie die Datenvermeidung, in § 3a S. 1 BDSG enthalten. Im Gegensatz zur Datenvermeidung stellt die Datensparsamkeit keine Anforderungen an die Erhebung, sondern an die Verarbeitung und Nutzung personenbezogener Daten. Sie gibt vor, dass für jeden einzelnen Verarbeitungsvorgang nur diejenigen Daten verwendet werden sollen, die für den jeweiligen Zweck erforderlich sind.

In Konkretisierung wird in § 3a S. 2 BDSG erwähnt, dass dies auch bedeutet, dass

personenbezogene Daten vor dem Eingang in ein Datenverarbeitungverfahren soweit möglich zu anonymisieren oder zu pseudonymisieren sind.

Anonyme und pseudonyme Dienste Der Vollständigkeit halber ist die Anforderung der Bereitstellung anonymer und pseudonymer Dienste aufgeführt. Diese ist keine Anforderung des BDSG, sondern eine des § 13 Abs. 6 u. 7 TMG. Sie bezieht sich auf eine Kommunikationsbeziehung zwischen einem (potentiell) Betroffenen und der verantwortlichen Stelle und wirkt bereits zum Erhebungszeitpunkt. Insofern handelt es sich um eine für Telemedien spezifische Regelung der Datenvermeidung.

Datenschutzfreundliche Grundeinstellungen Datenschutzfreundliche Grundeinstellungen sind im BDSG nicht explizit als Anforderung enthalten, entsprechen jedoch bereits heute dem Geist der Datenvermeidung und der Zweckbindung. Explizit aufgeführt sind sie in Art. 25 Abs. 2 DSGVO. Datenschutzfreundliche Grundeinstellungen müssen im Hinblick auf die Menge der erhobenen personenbezogenen Daten, den Umfang ihrer Verarbeitung, ihre Speicherdauer und ihre Zugänglichkeit für Dritte gewählt werden.

Datenschutzfreundliche Grundeinstellungen beziehen sich nicht ausschließlich auf die Verarbeitungsvorgänge bei der verantwortlichen Stelle. Als Teil des Datenschutzes durch Technikgestaltung sind sie beim Entwurf von Software und und beim Design von Benutzeroberflächen mitzuberücksichtigen (ErwGr 78 DSGVO). Datenschutzfreundliche Grundeinstellungen stehen in Beziehung zur Einwilligung indem sie die Zahl der explizit erforderlichen Einwilligungen erhöhen.

Datengeheimnis Das Datengeheimnis des § 5 BDSG ist eine organisatorische Anforderung. Es richtet sich an das mit der Datenverarbeitung betraute Personal und statuiert ein unmittelbares, persönliches, gesetzliches Verbot der zweckfremden Verwendung personenbezogener Daten. Die formelle Verpflichtung auf das Datengeheimnis führt zu einer rechtlich stärkeren Bindung mit entsprechenden Sanktionen.

Zutritts- und Zugangskontrolle Die Forderung nach Zutritts- und Zugangskontrolle vereint die physische Absicherung von Datenverabeitungsanlagen. Sie ist in der Anlage zu § 9 S. 1 BDSG Nr. 1 u. 2 verankert. Die Zutrittskontrolle verhindert den physischen Zugang zu Räumlichkeiten, in denen Datenverarbeitungsanlagen stehen. Zugangskontrolle verhindert die Nutzung von Datenverabeitungssystemen durch Unbefugte mit hard- und softwaretechnischen Mitteln.

Zugriffs- und Weitergabekontrolle Die Zugriffs- und Weitergabekontrolle stellt Anforderungen an das Rechtemanagement (engl. Access Control) in einer Datenverarbeitungsanlage. Sie ist in der Anlage zu § 9 S. 1 BDSG Nr. 3 u. 4 kodifiziert. Die Zugriffskontrolle

reglementiert den Zugriff auf personenbezogene Daten im Zuge eines zweck- und rollenadäquaten Rechtemanagements. Die Weitergabekontrolle verlangt, dass personenbezogene Daten, auf die eine Person Zugriff hat, von dieser nur übermittelt oder weitergegeben werden können, wenn dies datenschutzrechtlich zulässig ist.

Verfügbarkeitskontrolle Die Verfügbarkeitskontrolle in der Anlage zu § 9 S. 1 BDSG Nr. 7 verlangt, dass personenbezogene Daten gegen zufällige Zerstörung oder Verlust geschützt werden. Sie ist der speicherseitige Aspekt des Verfügbarkeitsziels.

Datenkorrektheit Die Datenkorrektheit hat eine große inhaltliche Nähe zum Recht auf Berichtigung. Ob sie deshalb als eigene Anforderung gelten kann, ist fraglich. Erwähnt wird sie in Art. 6 Lit. d DSRL und nicht im BDSG selbst. Für die Aufnahme als eigene Anforderung spricht, dass die Datenkorrektheit, im Gegensatz zum Recht auf Berichtigung, präventiv wirken soll. Deshalb wurde sie in den Anforderungskatalog übernommen.

Revisionssichere Protokollierung (Eingabekontrolle) Die revisionssichere Protokollierung[1] oder Eingabekontrolle, wie sie in der Anlage zu § 9 S. 1 BDSG Nr. 5 genannt wird, ist Teil des Integritätsschutzes. Sie fordert die Nachvollziehbarkeit der Erhebung, Modifikation und Löschung von personenbezogenen Daten.

Recht auf Datenübertragbarkeit Das Recht auf Datenübertragbarkeit, wie es in Art. 20 DSGVO dargelegt wird, ist keine traditionelle Datenschutzanforderung. Sie geht in ihrem Charakter eher auf das Wettbewerbsrecht zurück und soll den Wechsel des digitalen Dienstleisters ermöglichen, indem Kompatibilität und Exportmöglichkeiten gefordert werden. Weitere Ausführungen zum Recht auf Datenübertragbarkeit finden sich in Kapitel 10.2.

Prinzip der Verantwortlichkeit Das Prinzip der Verantwortlichkeit manifestiert sich im BDSG in der verantwortlichen Stelle. Definitorisch findet es sich in § 3 Abs. 7 BDSG, ist aber letztendlich über das gesamte Datenschutzrecht verteilt. Verantwortlichkeit heißt, dass nur dort eine Verarbeitung personenbezogener Daten stattfinden darf, wo geklärt ist, welche Stelle die Verantwortung dafür trägt. Das Prinzip wirkt auch in die interne organisatorische Ausgestaltung des Betriebs eines Datenverarbeitungssystems hinein. Dazu zählt, dass Rechte, Befugnisse und Entscheidungsgewalt im gleichen Schritt mit der Verantwortung vergeben werden.

[1] BVerfGE 125, 260 (325 f.).

Prinzip der nicht-automatisierten Einzelentscheidung Um den Betroffenen nicht zum Objekt der Datenverarbeitung werden zu lassen und um zu verhindern, dass Maschinen Menschen kontrollieren, darf nach § 6a Abs. 1 S. 1 BDSG jede Entscheidung, die den Betroffenen erheblich beeinträchtigt, grundsätzlich nicht ausschließlich auf eine automatisierte Verarbeitung gestützt werden.

Einwilligung Neben Rechtsvorschriften kann nach § 4 Abs. 1 BDSG nur die Einwilligung nach § 4a BDSG zur Zulässigkeit eines Umgangs mit personenbezogenen Daten führen. Die Anforderung der Einwilligung mündet darin, dass für jede Datenerhebung, -verarbeitung und -nutzung eine Einwilligung beim Betroffenen einzuholen ist, soweit keine andere Rechtsgrundlage existiert.

Einheitliche Ansprechpartner Die Forderung nach einem einheitlichen Ansprechpartner (engl. Single Point of Contact – SPOC) ist nicht Teil des BDSG. Bei gemeinsam für eine Verarbeitung Verantwortlichen sieht Art. 26 Abs. 1 DSGVO vor, dass die Verantwortlichen eine Vereinbarung darüber schließen, welche Stelle welche datenschutzrechtlichen Verpflichtungen erfüllt. Eine Vereinbarung über eine einheitliche Auskunft soll den Betroffenen darin unterstützen, sein Auskunftsrecht bei dem Ansprechpartner geltend zu machen, bei dem er seine Rechte am effektivsten und effizientesten wahrnehmen kann. Die Vereinbarung ist für den Betroffenen jedoch nicht bindend (siehe Kapitel 3.3). Art. 26 Abs. 3 DSGVO sieht, wie bereits § 6 Abs. 2 S. 2 BDSG, vor, dass der Betroffene seine Rechte gegenüber jedem Beteiligten geltend machen kann. ErwGr 59 sieht gemeinsam mit Art. 12 Abs. 1 DSGVO vor, Anträge elektronisch zu stellen. Deren Beantwortung sollte nach ErwGr 63 idealerweise direkt per Zugriff auf eine elektronische Plattform erfolgen.

Datenschutzkontrolle durch unabhängige BfD Die Datenschutzkontrolle durch unabhängige BfD ist für öffentliche und nicht-öffentliche Stellen, die personenbezogene Daten verarbeiten, in § 4f BDSG festgelegt. Der aufsichtsrechtliche Gegenpart sind der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit (BfDI, §§ 22 ff. BDSG) und die Aufsichtsbehörden der Länder (§ 38 BDSG). Die Unabhängigkeit all dieser Institutionen wurde unlängst durch den EuGH gestärkt.[2]

Meldepflicht Hat eine verantwortliche Stelle keinen BfD berufen, unterliegt sie der Meldepflicht nach § 4d Abs. 1 BDSG. Sie muss, bis auf Ausnahmen, alle Verfahren der automatisierten Datenverarbeitung vor ihrer Inbetriebnahme bei der Aufsichtsbehörde bzw. dem übergeordneten Beauftragten für den Datenschutz melden.

[2] EuGH, NJW 2010, 1265 (1265).

Aufklärungspflichten Aufklärungspflichten,[3] wie sie unter anderem in § 4a Abs. 1 S. 2 BDSG für die Einwilligung festgelegt sind, sollen sicherstellen, dass dem Betroffenen die Konsequenzen seines Handelns oder des Handelns der verantwortlichen Stelle deutlich gemacht werden. Im Gegensatz zu Hinweisen und Kennzeichnungen sind sie immer direkt an einen bestimmten Betroffenen gerichtet. Die Unterrichtung folgt, anders als die Aufklärung, der Erhebung nach.

Hinweis- und Kennzeichnungspflichten, Unterrichtungspflichten, Benachrichtigungspflichten und Auskunftsanspruch Die der Erhebung personenbezogener Daten nachgelagerten Transparenzpflichten wurden bereits im Kapitel 3.1 beschrieben und gegeneinander abgegrenzt.

Vorabkontrolle und Datenschutzfolgenabschätzung Die Vorabprüfung automatisierter Verarbeitungsanlagen, die besondere Risiken und Gefährdungen für den Betroffenen in sich tragen, ist in § 4d Abs. 5 BDSG festgelegt. Die Vorabkontrolle wird vom BfD durchgeführt und ist im Allgemeinen organisatorischer Natur. Auf europäischer Ebene ist in Art. 35 DSGVO eine noch weitergehende Datenschutzfolgenabschätzung enthalten, die den gesamten Entwicklungsprozess einer Datenverarbeitungsanlage betrifft.

Datenschutzaudit Ein Datenschutzaudit als Transparenzmaßnahme ist in § 9a BDSG verankert. Kriterien für ein Datenschutzaudit sollten in ein eigenes Gesetz eingehen,[4] das aber nie verwirklicht wurde.

Informationspflichten (bei DS-Verletzungen) Jenseits der allgemeinen Transparenzanforderungen statuiert § 42a BDSG eine weitergehende Informationspflicht bei unrechtmäßiger Kenntniserlangung personenbezogener Daten durch Dritte.

Führung eines Verfahrensverzeichnisses Der BfD der verantwortlichen Stelle hat nach § 4g Abs. 2 i. V. m. § 4e S. 1 BDSG ein Verfahrensverzeichnis, im Umfang den meldepflichtigen Angaben ähnlich, zu führen und jedermann verfügbar zu machen.

Recht auf Berichtigung Nach den §§ 20 Abs. 1, 35 Abs. 1 BDSG sind diejenigen personenbezogenen Daten, die unrichtig sind, zu berichtigen. Im Regelfall geschieht dies auf Antrag des Betroffenen, nachdem er aufgrund seines Auskunftsrechts Einblick in seine personenbezogenen Daten erhalten hat. Das Recht auf Berichtigung ist somit das erste der drei universellen Interventionsrechte.

[3] Siehe auch BVerfGE 65, 1 (46).
[4] BT-Drs. 16/12011.

Recht auf Löschung Das zweite wesentliche Interventionsrecht ist das Recht auf Löschung. Demgemäß sind, entsprechend der §§ 20 Abs. 2, 35 Abs. 2 BDSG, diejenigen personenbezogenen Daten, die für einen bestimmten Zweck nicht mehr erforderlich sind, zu löschen. Die Löschung kann auf Antrag des Betroffenen erfolgen, kann aber auch intern durch wohldefinierte Löschregeln ausgelöst werden.

Recht auf Sperrung Ist eine Löschung nicht möglich, tritt an ihre Stelle die Sperrung nach §§ 20 Abs. 3, 35 Abs. 3 BDSG. Eine Sperrung bedeutet, dass die Daten so gekennzeichnet werden, dass ihre Verarbeitung und Nutzung weitestmöglich eingeschränkt werden (§ 3 Abs. 4 Nr. 4 BDSG).

Widerspruchsrecht Das Widerspruchsrecht ist eine Sonderregelung für öffentliche Stellen und entstammt dem Verwaltungsverfahrensrecht. Es ist in § 20 Abs. 5 BDSG für das Datenschutzrecht übernommen worden.

Zweckbestimmung Die Zweckbestimmung ist als Kernforderung des Datenschutzes über das gesamte BDSG verteilt. Für die Erhebung personenbezogener Daten zu eigenen Geschäftszwecken findet sich die Forderung beispielsweise in § 28 Abs. 1 S. 2 BDSG. Weitere Ausführungen finden sich im Kapitel 3.7.5.

Zweckbindung und Zwecktrennung Ebenso wie die Zweckbestimmung sind auch Zweckbindung und Zwecktrennung über das gesamte BDSG verteilt. Die Zweckbindung fordert, dass personenbezogene Daten nur zu dem Zweck verarbeitet und genutzt werden dürfen, zu dem sie erhoben wurden und der für sie dokumentiert wurde. Unter vielen findet sich diese Festlegung beispielsweise in § 28 Abs. 3 S. 7 BDSG.

Die Zwecktrennung ist Ausfluss eines wesentlichen Aspekts der Unverkettbarkeit. Personenbezogene Daten, die zu unterschiedlichen Zwecken verarbeitet, insbesondere gespeichert werden, dürfen nicht zusammengeführt werden. Mit technischem Bezug ist dies in der Anlage zu § 9 S. 1 BDSG Nr. 8 festgelegt.

Organisatorische und technische Gewaltenteilung Die organisatorische und technische Gewaltenteilung im Rahmen der informationellen Gewaltenteilung folgt aus dem Gebot der Zweckbindung und Zwecktrennung. Ihr liegt das verwaltungsrechtliche Abschottungsprinzip zugrunde. Die informationelle Gewaltenteilung ist nicht im BDSG festgelegt, sondern ergibt sich aus der Rechtsprechung des Bundesverfassungsgerichts,[5] in dessen Lichte das BDSG auszulegen ist.

Während sich Zweckbindung und Zwecktrennung auf die personenbezogenen Daten selbst beziehen, ist die informationelle Gewaltenteilung eine Forderung, die direkt an die

[5] BVerfGE 65, 1 (69); BVerfG, NJW 1988, 959 (961).

organisatorischen und technischen Einrichtungen gestellt wird. Die Zwecktrennung untersagt die Zusammenführung personenbezogener Daten. Die Gewaltenteilung verpflichtet dazu, dass Zugriffsrechte, Rollen sowie physische und logische Speicherorte nicht beliebig festgelegt werden. Sie sind entsprechend dem Zweck und nach dem Prinzip der Machtdistribution festzulegen.

Grundsatz der Direkterhebung Der Grundsatz der Direkterhebung ist in § 4 Abs. 2 S. 1 BDSG festgelegt. Er stärkt die Einflussnahme und Selbstbestimmung des Betroffenen bei der Erhebung personenbezogener Daten.

Verbot mit Erlaubnisvorbehalt Das Verbot mit Erlaubnisvorbehalt des Datenschutzes kann für nicht-öffentliche Stellen nicht aus den Grundrechten abgeleitet werden, sondern ist eine rechtspolitische Entscheidung des Gesetzgebers.[6] Diese Grundsätzliche Forderung findet sich prominent in § 4 Abs. 1 BDSG.

Verhältnismäßigkeitsprinzip Das Verhältnismäßigkeitsprinzip ist nicht aus Datenschutzgrundrechten abgeleitet, sondern Teil des Rechtsstaatsprinzips in Art. 20 Abs. 2 GG.[7] Es findet sich in den Vorschriften des BDSG unter anderem in § 4 Abs. 2 S. 2 Nr. 2 Lit. a BDSG, ist jedoch ein grundsätzliches Prinzip, das sich über das gesamte deutsche Recht und damit auch das Datenschutzrecht erstreckt.

B.1.2 Vollständigkeit in Bezug auf das BDSG

Die Vollständigkeit der Anforderungen in Bezug auf das BDSG kann nicht im formalen Sinne bewiesen werden. Dennoch soll durch eine kursive Gesamtschau auf das BDSG deutlich gemacht werden, dass dessen wesentliche Anforderungen in Tabelle B.1 enthalten sind und keine Vorschrift in der Analyse unberücksichtigt blieb.

Das BDSG ist in sechs Abschnitte unterteilt. Der erste Abschnitt (§§ 1 bis 11) umfasst allgemeine Bestimmungen und Definitionen, die sowohl für öffentliche, als auch für nicht-öffentliche Stellen gelten. Der zweite Abschnitt (§§ 12 bis 26) widmet sich der Datenverarbeitung der öffentlichen Stellen, während der dritte Abschnitt (§§ 27 bis 38a) die Datenverarbeitung nicht-öffentlicher Stellen in den Blick nimmt.

Die Abschnitte vier (§§ 39 bis 42a), fünf (§§ 43 und 44) und sechs (§§ 45 bis 48) beinhalten Sondervorschriften, Schlussvorschriften und Übergangsvorschriften. In ihnen finden sich keine allgemeingültigen Anforderungen. Sonderregularien wie die Gegendarstellung im Medienrecht (§ 41 Abs. 2) bleiben im Folgenden unberücksichtigt. Einzig die aufge-

[6] Rudolf in: Merten/Papier, HGR IV 2011, § 90 Rn. 28.
[7] BVerfGE 19, 342 (348 f.).

nommenen Informationspflichten bei unrechtmäßiger Kenntniserlangung von Daten des § 42a haben allgemeine Bedeutung.

Die Abschnitte 2 und 3 sind jeweils noch einmal in drei Unterabschnitte unterteilt. Sie befassen sich jeweils mit den eigentlichen Rechtsgrundlagen der Datenverarbeitung, den Rechten des Betroffenen und dem BfDI bzw. der Aufsichtsbehörde. Der letzte Unterabschnitt wird insgesamt durch die auch betrieblich beziehungsweise behördlich geltende Forderung nach unabhängigen Datenschutzbeauftragten miterfasst, sind ansonsten jedoch von keiner weitergehenden Relevanz für die Erhebung der datenschutzrechtlichen Anforderungen für den Einsatz von Datenverarbeitungsanlagen.

Die §§ 1 bis 3 beschreiben den Anwendungsbereich des Gesetzes und enthalten Definitionen und Begriffsbestimmungen. Aus diesen ergeben sich keine Anforderungen, sie tragen jedoch zum Verständnis sich an anderer Stelle ergebender Anforderungen bei. Die Anforderungen der §§ 3a (Datenvermeidung und Datensparsamkeit), 4 (Grundsatz der Direkterhebung, Verbot mit Erlaubnisvorbehalt und Unterrichtungspflichten sowie beispielhaft das Verhältnismäßigkeitsprinzip) sowie 4a (Einwilligung und beispielhaft die dazugehörigen Aufklärungspflichten) wurden in der Anforderungsanalyse berücksichtigt (siehe Tabelle B.1). Die Auslandsübermittlung in den § 4b und 4c steht als Spezialregelung außerhalb des Betrachtungsumfangs dieser Analyse. Sie stellen organisatorisch-rechtliche Anforderungen, die zu erfüllen sind, bevor eine Übermittlung stattfindet. Die Meldepflicht (§§ 4d und 4e) sowie die Bestellung eines BfD (§§ 4f und 4g) sind berücksichtigt. Die Vorabkontrolle und die Führung eines Verfahrensverzeichnisses als besondere Aufgaben des BfD wurden gesondert erfasst. Das Datengeheimnis des § 5 ist ebenfalls in der Liste der extrahierten Anforderungen enthalten.

§ 6, der die Rechte des Betroffenen beschreibt, greift dem zweiten Unterabschnitt der Abschnitte 2 und 3 vor. Es finden sich an dieser Stelle keine eigenen Anforderungen, sondern nur Kriterien für die Ausgestaltung der Anforderungen des zweiten Unterabschnitts. Der nachfolgende § 6a statuiert das Prinzip der nicht-automatisierten Einzelentscheidung und erweitert in Absatz 3 den Kriterienkatalog des Auskunftsanspruchs um den logischen Aufbau der automatisierten Verarbeitung. Ersteres ist berücksichtigt, letzteres findet sich auf der nächsten Ebene des EVAL-Modells wieder. Für die Videoüberwachung enthält § 6b besondere Kriterien für die einzelnen datenschutzrechtlichen Anforderungen, insbesondere die Betroffenenrechte. Eine Besonderheit ist die Hinweis- und Kennzeichnungspflicht, weshalb diese als eigene Anforderung übernommen wurde. Mobile personenbezogene Speicher- und Verarbeitungsmedien unterliegen der besonderen Unterrichtungspflicht des § 6c, die bereits durch § 4 in den Anforderungskatalog aufgenommen wurde.

Die §§ 7 und 8 beschäftigen sich mit dem Schadenersatz und stehen eher im Kontext des vierten und fünften Abschnitts. Sie enthalten keine eigenen Anforderungen, die auf technische oder organisatorische Vorgänge übertragbar wären. Demgegenüber eröffnet § 9 gleich einen ganzen Katalog von Anforderungen, die als technisch-organisatorische Maßnahmen bezeichnet werden. Sie finden sich gesammelt in einem Anhang des BDSG.

Im Detail handelt es sich um die Zutritts- und Zugangskontrolle, die Zugriffs- und Weitergabekontrolle, die Eingabekontrolle als Teil der revisionssicheren Protokollierung, die Verfügbarkeitskontrolle und die Zwecktrennung. Die Auftragskontrolle ist Teil des Verantwortlichkeitsprinzips und wurde deshalb nicht als eigene Anforderung erfasst. Der § 9a bringt das Datenschutzaudit mit in die Liste der Anforderungen ein. § 10 listet besondere Kriterien hinsichtlich automatisierter Abrufverfahren auf, die sich insbesondere auf die Nachprüfbarkeit im Auditverfahren beziehen. § 11 ergänzt eine Reihe von Dokumentationspflichten und Verfahrensanweisungen für die Auftragsdatenverarbeitung. Diese wurden bei der durchgeführten Analyse explizit außen vor gelassen.

Ab dem § 12 folgen die Abschnitte 2 und 3. In ihrem ersten Unterabschnitt finden sich jeweils detaillierte Regelungen für die Zulässigkeit der Datenverarbeitung, die Zweckbestimmung und die Zweckbindung. Diese Regelungen spiegeln sich im anwendungsspezifischen Kriterienkatalog wider. Einzelne Anforderungen, die über das bisher erwähnte hinausgehen, lassen sich jedoch nicht lokalisieren. Die Vielfalt der Regelungen beruht vor allem auf dem Prinzip des Verbots mit Erlaubnisvorbehalt, das die Auflistung aller Erlaubnistatbestände erforderlich macht.

Wesentliche Anforderungen finden sich im zweiten Unterabschnitt mit den Rechten des Betroffenen. Dies sind die Benachrichtigung (§§ 33, 19a), der Auskunftsanspruch (§§ 19, 34), das Recht auf Berichtigung (§§ 20 Abs. 1, 35 Abs. 1), das Recht auf Löschung (§§ 20 Abs. 2, 35 Abs. 2) und das Recht auf Sperrung (§§ 20 Abs. 3, 35 Abs. 3). Ergänzend findet sich im zweiten Abschnitt auch noch das Widerspruchsrecht (§ 20 Abs. 5).

Dieser Gesamtüberblick über das BDSG gibt einen Anhaltspunkt für die Vollständigkeit der erhobenen Anforderungen.

B.2 Ableitung datenschutzrechtlicher Kriterien

Die Ableitung datenschutzrechtlicher Kriterien wurde im wesentlichen in Kapitel 3 behandelt. Deshalb finden sich in diesem Abschnitt vor allem Ergänzungen, die über das Kernthema des Auskunftsrechts hinausgehen.

B.2.1 Korrektheit der datenschutzrechtlichen Kriterien

Die datenschutzrechtlichen Kriterien sind korrekt, wenn sie methodisch sauber aus den Rechtsgrundlagen und ihrer Interpretation abgeleitet wurden. Kapitel 3 sollte dieser Anforderung genügen. In der Auflistung der Kriterien in Abschnitt 4.3.1 wird auf die jeweils einschlägigen Argumentationsstränge verwiesen.

Für die ergänzenden Kriterien des Abschnitts 4.3.2 wurde nicht überall eine so umfangreiche Diskussion unternommen. Sie gehen über das Kernthema dieser Arbeit hinaus. Dennoch sollen im Folgenden Hinweise gegeben werden, weshalb die dargestellten Kriterien dennoch korrekt sind.

Datenvermeidung Die Kriterien 40 und 41 übertragen das allgemeine Prinzip der Datenvermeidung auf die Protokolldaten und die zugrundeliegenden Basisdaten. Die Kriterien 43 und 44 konkretisieren, dass bei der Protokollierung keine Daten erhoben werden dürfen, die zur Überwachung von Beschäftigten dienen könnten. Sie greifen Kriterien auf, die sich aus der Verbindung des § 32 Abs. 1 BDSG mit dem Prinzip der Datenvermeidung ergeben.

Datensparsamkeit Die Kriterien 45, 46 und 47 regulieren die frühzeitige Löschung unterschiedlicher Protokolldaten nach dem Prinzip der Datensparsamkeit. Sie stehen in Beziehung zu den §§ 20 Abs. 2 und 35 Abs. 2 BDSG.

Zugriffs-, Zugangs- und Weitergabekontrolle Ein Zugriff Dritter auf personenbezogene Protokolldaten ist nach § 3 Abs. Nr. 3 BDSG eine Übermittlung. Bei einem automatisierten Auskunftsverfahren wäre darüber hinaus auch noch § 10 BDSG zu berücksichtigen. In jedem Fall untersagt § 34 Abs. 5 BDSG Übermittlungszwecke und ohne zulässigen Zweck kann es nach § 4 Abs. 1 BDSG auch keine zulässige Übermittlung geben. Dies ist in Kriterium 52 festgehalten. Kriterium 53 führt mit dem 4-Augen-Prinzip ein bewährtes Mittel der datenschutzorganisatorischen Kontrolle ein.

Verfügbarkeitskontrolle Das Kriterium 57 ergibt sich wie die Kriterien 56 und 60 aus dem Anspruch einer vollständigen und korrekten Auskunft (siehe Abschnitt 3.5). Das Kriterium präzisiert diesen Anspruch für das vorgelagerte Rechtemanagement.

Revisionssichere Protokollierung Die Kriterien 58 und 59 sichern die Integrität der Protokolldaten zu. Auch sie ergeben sich indirekt aus dem Anspruch einer vollständigen und korrekten Auskunft.

Einheitliche Ansprechpartner Kriterium 62 geht auf Art. 12 i. V. m. Art. 26 DSGVO unter Berücksichtigung der Erwägungsgründe 59 und 63 zurück. Nach heutiger Rechtslage sind verbundene verantwortliche Stellen und daraus abgeleitete gemeinsame Ansprechpartner noch nicht möglich, eine zentrale Plattform noch nicht erforderlich.

Zweckbindung und Zwecktrennung Das Kriterium 65 geht wie Kriterium 67 fast wortwörtlich auf § 34 Abs. 5 BDSG zurück. Kriterium 66 ergibt sich aus § 6 Abs. 3 i. V. m. § 31 BDSG. Die Kriterien 68 und 69 konkretisieren die Zwecktrennung für die Speicherung und die Verarbeitungsprozesse und gehen auf die Anlage zu § 9 S. 1 BDSG Nr. 8 zurück. Protokolldaten und Basisdaten haben immer einen unterschiedlichen Zweck und sind deshalb auch in jedem Fall zu trennen.

Organisatorische und technische Gewaltenteilung Die organisatorische und technische Gewaltenteilung findet sich bisher noch kaum in expliziten Regelungsvorgaben des Bundesdatenschutzgesetzes. Sie ergibt sich jedoch indirekt aus anderen Anforderungen und Kriterien. Das Kriterium 70 ergibt sich aus Datensparsamkeit und informationeller Gewaltenteilung. Es lässt sich auf die Beschränkung der Nutzung auf die Auskunftszwecke in § 28 Abs. 1 i. V. m. § 34 Abs. 5 BDSG zurückführen. Kriterium 71 ist eine weitergehende Konsequenz und technische Konkretisierung des Kriteriums 70. Daten, die nicht übermittelt werden dürfen, da dies nicht zwingend erforderlich ist, müssen lokal beziehungsweise dezentral gespeichert und verarbeitet werden. Die Kriterien 72 und 73 sind eine explizite Formulierung des Unverkettbarkeitsprinzips, angewendet auf die Situation in einem Datenschutzauskunftssystem.

B.2.2 Vollständigkeit der datenschutzrechtlichen Kriterien

Die datenschutzrechtlichen Kriterien erheben keinen allgemeinen Anspruch auf Vollständigkeit. Einzig die aus dem Auskunftsanspruch abgeleiteten Kriterien sollten in der Gesamtschau alle wesentlichen Aspekte abdecken. Die Diskussion des Auskunftsanspruchs im Kapitel 3 ist dahingehend abschließend. Der Vergleich mit der Kommentarliteratur lässt keine Lücken in den anwendungsunabhängigen Erwägungen erkennen. Die für ein Datenschutzauskunftssystem spezifischen Kriterien sind neu und bisher ohne Präzedenzfall. Eine Vollständigkeit lässt sich deshalb nicht belegen.

B.3 Ableitung technischer Anforderungen

Die technischen Anforderungen leiten sich aus den rechtlichen Kriterien ab. Sie stehen in einem N:M-Verhältnis, weshalb manche rechtlichen Kriterien von mehreren technischen Anforderungen gemeinsam umgesetzt werden (siehe Tabelle 4.2) und andererseits auch bestimmte Anforderungen eine Reihe rechtlicher Kriterien vereinen (siehe Tabelle B.2).

Die Korrektheit einer technischen Anforderung ergibt sich, wenn sie sich vollständig durch die zugrundeliegenden rechtlichen Kriterien erklären lässt. Gehen technische Anforderungen über rechtliche Kriterien hinaus, muss dies technisch begründet sein und darf nicht im Widerspruch zu rechtlichen Kriterien stehen.

Die Menge der technischen Anforderungen ist vollständig, wenn alle rechtlichen Kriterien von ihnen vollständig erfasst werden.

B.3.1 Korrektheit der technischen Anforderungen

Die Korrektheit von Anforderungen, die sich aus genau einem Kriterium ergeben, lässt sich durch einen direkten Vergleich nachvollziehen. Hauptunterschied zwischen den recht-

Technische Anforderung	Rechtliche Kriterien
5	40, 41, 43, 44
6	1, 58
7	1, 2, 3, 13, 24, 31, 64
9	1, 15, 19, 21, 28, 29, 31
10	1, 20, 28, 29, 42
11	1, 14, 17, 28, 29, 32
12	1, 12, 28, 29, 32, 44
15	1, 8, 11, 27, 28, 29
19	30, 68, 69
23	6, 57
25	6, 25, 26
26	6, 22, 23, 25
28	6, 46
29	6, 47
39	57, 59
40	65, 66, 67
55	2, 56, 60
58	18, 62

Tabelle B.2: Technische Anforderungen, die mehrere rechtliche Kriterien umsetzen

lichen Kriterien und den technischen Anforderungen ist bei einem solchen 1:1-Verhältnis der technische Bezug.

Beispielhaft sei dies an der technischen Anforderung 1 dargestellt. Sie fordert die Einbettung von Event-Listenern in allen Systemkomponenten, die personenbezogene Daten verarbeiten. Diese technische Anforderung lässt sich auf das Kriterium 1 zurückführen. Dieses fordert eine Protokollierung aller Erhebungs-, Verarbeitungs-, Nutzungs- und Übermittlungsvorgänge. Ohne Event-Listener in allen Systemkomponenten ist dieses Kriterium nicht zu erreichen, da sonst bestimmte Vorgänge nicht erfasst werden können. Deshalb ist die Anforderung 1 korrekt aus dem Kriterium 1 abgeleitet.

Weitaus interessanter sind diejenigen Anforderungen, die mehrere Kriterien umsetzen (siehe Tabelle B.2). Da sich ihre Korrektheit aus dem vollständigen Enthaltensein in der Vereinigung der Kriterien ergibt, ist eine genauere Analyse notwendig. Diese wird im Folgenden vorgenommen.

Anforderung 5 Die technische Anforderung 5 vereint die Kriterien der Datenvermeidung 40, 41, 43 und 44. Während bei den rechtlichen Kriterien einzelne Datenarten aufgelistet

werden, die nicht erhoben werden dürfen, geht die technische Anforderung den umgekehrten Weg. Sie schreibt fest, dass keine Informationen protokolliert werden dürfen, deren Protokollierung nicht speziell in einer anderen Anforderung vorgeschrieben ist. Da keine der technischen Anforderungen eine Protokollierung der in den Kriterien 40 bis 44 genannten Datenarten vorsieht und Kriterium 40 alle Informationen umfasst, die nicht Teil des Auskunftsanspruchs sind, ist die Anforderung korrekt umgesetzt.

Anforderung 6 Das Kriterium 58 ist fast 1:1 in die Anforderung 6 übergegangen. Dem Kriterium der Nachvollziehbarkeit wird durch den Verweis auf die nachfolgenden Anforderungen Genüge getan. Umgekehrt wurde Anforderung 6 dahingehend erweitert, dass jeder Umgang mit personenbezogenen Daten entsprechend der nachfolgenden Anforderungen zu protokollieren ist. Dies stützt sich auf Kriterium 1.

Anforderung 7 Die technische Anforderung 7 besteht aus mehreren Datenarten, die gemeinsam den Umfang des Protokolls zur Erhebung personenbezogener Daten festlegen. Dass ein solches Protokoll grundsätzlich zu erstellen ist, ergibt sich für einen Spezialfall aus Kriterium 24. In den anderen Fällen ist die Erstellung des Protokolls eine technische Vereinfachung, um gemäß der Kriterien 2 und 3 die Beauskunftung der nach Kriterium 13 gespeicherten Herkunftsinformationen im Nachhinein zu ermöglichen. Sind die Herkunftsdaten nicht anderweitig gespeichert, würde die technische Anforderung 7 die rechtlichen Kriterien übererfüllen. Dies widerspricht jedoch keiner der übrigen Kriterien, insbesondere nicht Kriterium 40, da eine vollständige Auskunft im Interesse des Betroffenen liegt. Für den Inhalt des Protokolls ergibt sich im Einzelnen die Angabe der Quelle aus Kriterium 13, die Angabe des Zeitpunktes aus Kriterium 31, die Angabe der Kategorie als erweiternder Analogieschluss aus Kriterium 16, die Angabe der Organisationseinheit und des IT-Systems aus Kriterium 2 in Verbindung mit Kriterium 1, die Angabe des Zwecks aus Kriterium 64.

Anforderung 9 Die technische Anforderung zum Umfang des Protokolls zur Übermittlung personenbezogener Daten ergibt sich aus ganz ähnlichen Kriterien wie bei der Erhebung. Allerdings gibt es keine Übererfüllung in der technischen Umsetzung, da die Protokollierung der Empfänger in Kriterium 15 immer gefordert ist. Das Kriterium 19 fasst bis auf den Zweck alle inhaltlichen Punkte des Protokolls zusammen. Die Kriterien 28 und 29 ergänzen diesen. Die übrigen Kriterien geben der Anforderung ihren Kontext.

Anforderung 10 Das Protokoll zur Veröffentlichung personenbezogener Daten setzt sich aus den inhaltlichen Punkten des Kriteriums 20 zusammen. Es fordert die Protokollierung von Beginn und Ende der Veröffentlichung. Die Aufnahme von Organisationseinheit und IT-System sind auf das allgemeine Prinzip der durchgängigen Historie von Kriterium 1

zurückzuführen. Kriterium 42 wirkt explizit einschränkend und verhindert, im Gegensatz zur Anforderung 9, die Aufnahme der Gegenstelle ins Protokoll. Die Kriterien 28 und 29 ergänzen die Angabe des Zwecks.

Anforderung 11 Analog der obigen Anforderungen ergibt sich die Protokollierung von Sender und Empfänger bei der internen Weitergabe aus den Kriterien 1, 14 und 17. Der Zeitpunkt der Weitergabe wurde mit aufgenommen, ist jedoch gemäß Kriterium 32 nur im Einzelfall auskunftspflichtig. Die Protokollierung des Zwecks ergibt sich aus den Kriterien 28 und 29.

Anforderung 12 Die grundsätzliche Existenz der Anforderung und die Protokollierung von Organisationseinheit und IT-System ergeben sich analog zu den übrigen Protokollanforderungen aus Kriterium 1. Als Erweiterung wurde aus Kriterium 12 die Speicherung von Name und Typ der Anwendung für gewisse Spezialfälle übernommen. Die Speicherung der Zwecke der Verarbeitung ist in den Kriterien 28 und 29 festgeschrieben. Dagegen wird die Speicherung des Zeitpunkts einer Verarbeitung durch Kriterium 32 optional gestellt und durch Kriterium 44 letztendlich untersagt.

Anforderung 15 Die Anforderung, ein Protokoll über die gespeicherten Daten zu erstellen, ergibt sich aus der Kombination des speziellen Kriteriums 8 mit dem allgemeinen Kriterium 1, welches wiederum die Grundlage für die Speicherung von Organisationseinheit und IT-System ist. Die Speicherung des genauen Dateipfads und der Metadaten gründet in dem erweiternden Kriterium 11. Die Protokollierung des Zwecks ist in den Kriterien 27, 28 und 29 festgeschrieben.

Anforderung 19 Die getrennte Speicherung und Verarbeitung der Protokolldaten finden sich, je nach Protokolltyp, in unterschiedlichen Kriterien. Für den Zweck gibt Kriterium 30 die unabhängige Speicherung vor. Für die allgemeine Datenhaltung und Datenverarbeitung der Protokolldaten stützen die Kriterien 68 und 69 die Anforderung.

Anforderung 23 Die Grundanforderung, Protokolldaten zu speichern, solange es keine Löschregel gibt, findet ihren ersten Bezugspunkt in der Speicherdauer der Basisdaten gemäß Kriterium 6. Die Löschregel als Löschberechtigung ist dem Kriterium 57 entnommen.

Anforderung 25 Die Löschregel für die Erhebungsprotokolle ist in Anforderung 25 verankert. Aufgrund von Kriterium 26 ist die Speicherdauer der Erhebungsprotokolle an die Speicherdauer der Übermittlungsprotokolle gebunden. Indirekt ist die Speicherdauer damit auch an die in Kriterium 6 beschriebene Speicherdauer der Basisdaten gebunden.

Kriterium 25 wird durch die Bindung an die Übermittlungsprotokolle miterfüllt, soweit deren Löschregel korrekt ist.

Anforderung 26 Die Löschregel in Anforderung 26 stützt sich in ihren einzelnen Spiegelpunkten mehr oder weniger wortwörtlich auf die Kriterien 22, 23, 25 und 6. Die Kriterien stellen zum Teil nur Forderungen für bestimmte Spezialfälle auf. Die Anforderung 26 übererfüllt durch ihre Verallgemeinerung die Kriterien, verletzt dabei jedoch keine anderen Kriterien oder grundlegenden Prinzipien.

Anforderung 28 Anforderung 28 ergibt sich wortwörtlich aus Kriterium 46. Das Kriterium 6 wird implizit miterfüllt, da nach Abschluss der Verarbeitung (Terminierung des Prozesses), der Prozess und damit auch die Basisdaten nicht mehr existieren.

Anforderung 29 Das Kriterium 47 konkretisiert Kriterium 6 für die Speicherung. Beide münden gemeinsam in der, in Anforderung 29 festgelegten, Bindung des Speicherprotokolls an die Basisdaten.

Anforderung 39 Das Verbot von Schreibzugriffen auf die Protokolldaten von Prozessen außerhalb des Datenschutzauskunftssystems ist eine starke technische Spezialisierung der Kriterien 57 und 59, die die Integrität der Protokolldaten in allgemeiner Weise fordern.

Anforderung 40 Anforderung 40 definiert eine auf den Betroffenen beschränkte feingranulare Regel für den lesenden Zugriff, die die Kriterien 65, 66 und 67 umsetzt. Die Zugriffsregel gilt gemäß Kriterium 66 für die übermittelten Auskunftsdaten und gemäß Kriterium 67 für die Protokolldaten selbst.

Anforderung 55 Die Anforderung an die Vollständigkeit der Protokollinformationen ergibt sich direkt aus Kriterium 2. Dass die Auskunft den Protokollinformationen zu entsprechen hat bestimmt sich nach Kriterium 56. Die Korrektheit der Auskunft ist Kriterium 60 entnommen.

Anforderung 58 Die Standardisierung des Interfaces eines Datenschutzauskunftssystems ist die Voraussetzung für die stellenübergreifende Protokollierung nach Kriterium 18. Eine zentrale Plattform gemäß Kriterium 62 ist auch auf einheitliche Interfaces angewiesen.

B.3.2 Vollständigkeit der technischen Anforderungen

Die technischen Anforderungen sind nur in Bezug auf die funktionalen Anforderungen (Anforderungen 1 bis 29), die sich aus den Kriterien des Auskunftsanspruchs ergeben

(Kriterium 1 bis 39), vollständig. Da bereits die Kriterien, die über den Auskunftsanspruch hinaus gehen, nicht vollständig sind, können es auch die diesbezüglichen technischen Anforderungen nicht sein.

Die nicht-funktionalen technischen Anforderungen sind nicht erschöpfend, da sich beispielsweise aus Kriterium 38 eine lange Liste ergänzender nicht-funktionaler Anforderungen ergibt. Das Kriterium eines einfachen und flexiblen Zugangs erfordert die genaue und umfangreiche Formulierung von Usability-Anforderungen.

Die Kriterien 4, 10, 16 und 32 weichen die in anderen Kriterien enthaltenen Protokollpflichten für bestimmte Spezialfälle auf. Sie wurden in den technischen Anforderungen zum Teil implizit berücksichtigt, erfordern jedoch keine vollständige Umsetzung in den Anforderungen. Kriterium 4 ist aber insofern wichtig, als dass es eine fortlaufende rekursive Protokollierung des Datenschutzauskunftssystems selbst und die damit einhergehenden technischen und praktischen Probleme vermeidet.

Die für die Vollständigkeitsprüfung relevante verkürzte Liste findet sich in Tabelle B.3.

Die vollständige Umsetzung der Kriterien, die nur in einer Anforderung resultieren, ist direkt nachvollziehbar und bedarf wie bei der Korrektheit keiner weiteren Erläuterung. Die komplexen Kriterien werden in den folgenden Absätzen behandelt.

Kriterium 1 Die Anforderungen 1, 2, 6, 7, 9, 10, 11, 12 und 15 umfassen alle Teilschritte im Lebenszyklus eines Datums. Sie beinhalten jeweils den Speicherort. Über Anforderung 16 ist sichergestellt, dass die Historie keine Lücken aufweist. Werden alle Anforderungen erfüllt, ergibt sich eine durchgängige Historie. Die Anforderungen sind demgemäß vollständig im Bezug auf Kriterium 1.

Kriterium 3 Gemäß Kriterium 3 müssen die Anforderungen sicherstellen, dass der Personenbezug personenbezogener Daten im Zuge der Auskunft herstellbar ist. Um dies zu erreichen, wird in Anforderung 7 die Quelle der personenbezogenen Daten genau referenziert. Das entstehende Protokoll wird gemäß Anforderung 8 mit einem eindeutigen Identifikator des Betroffenen verknüpft. Dadurch kann das Kriterium insgesamt erfüllt werden.

Kriterium 6 Das Kriterium 6 ist die Grundvorgabe für die Speicherfrist der Protokolldaten. Die Bindung an die Basisdaten wird für die Prozessverarbeitung und die Speicherung von den Anforderungen 28 und 29 umgesetzt. Alle übrigen Löschregeln (Anforderungen 23, 25, 26 und 27) sind an die beiden erstgenannten Löschregeln zeitlich gebunden und entsprechen so auch Kriterium 6. Da die Löschregeln alle Protokollarten behandeln, ist Kriterium 6 insgesamt vollständig umgesetzt.

Kriterium 25 Dieses Kriterium ist eine Spezialregelung für die Löschung der Protokolldaten bei der Übermittlung zu Werbezwecken. Sie wird für die Übermittlungsprotokolle (Empfänger) in Anforderung 26 direkt übernommen. Die Herkunftsinformationen finden sich in den Erhebungsprotokollen nach Anforderung 25. Sie sind in ihrer Speicherdauer an Anforderung 26 gebunden. Dadurch wird Kriterium 25 insgesamt vollständig umgesetzt.

Kriterien 28 und 29 Nach diesen Kriterien müssen alle Zwecke der Verwendung personenbezogener Daten protokolliert werden. Für die Übermittlung und die Veröffentlichung findet sich dies in den Anforderungen 9 und 10. Für die Speicherung und Verarbeitung ist dies in den Anforderungen 12 und 15 festgelegt. Die Aufnahme des Zwecks in Weitergabeprotokolle findet sich in Anforderung 27. Um alle Zwecke beauskunften zur können, muss ihre Abfolge erhalten bleiben. Dies ist in Anforderung 27 enthalten. Somit sind die beiden Kriterien vollständig umgesetzt.

Kriterium 31 Der Zeitpunkt der Erhebung und Übermittlung muss nach diesem Kriterium in der Auskunft mitenthalten sein. Für die ordnungsgemäße Protokollierung stehen die Anforderungen 7 und 9. Die Übereinstimmung von Protokoll und Auskunft wird für dieses Kriterium wie auch für alle vorhergehenden Kriterien von Anforderung 55 sichergestellt. Daher ist Kriterium 31 vollständig von den Anforderungen abgedeckt.

Rechtliche Kriterien	Technische Anforderungen
1	1, 2, 6, 7, 9, 10, 11, 12, 15, 16
3	7, 8
6	23, 25, 26, 27, 28, 29
7	24
8	15
9	17
11	15
12	12
13	7
14	11
15	9
19	9
17	11
20	10
21	9
22	26
23	26
24	7
25	25, 26
26	25
27	15
28	9, 10, 11, 12, 15
29	9, 10, 11, 12, 15
30	19
31	7, 9
33	13
34	14

Tabelle B.3: Ableitung rein funktionaler technischer Anforderungen aus den rechtlichen Kriterien des Auskunftsanspruchs

C Bezeichner und Namensräume

Um die Elemente der Datenstrukturen und die Komponenten der gesamten UC- und Provenance-Infrastruktur klar unterscheidbar zu machen, müssen sie mit einem global eindeutigen Kennzeichen versehen werden. Für solche Zwecke wurde der Universally Unique Identifier (UUID) standardisiert und etabliert.[1]

Neben den Instanziierungen der Komponenten der UC- und Provenance-Architektur[2] ist bei folgenden Datentypen Eindeutigkeit erforderlich:

- Container
- Datum
- Repräsentation
- Betroffener

Daneben sind Policies und (Abstraktions-)Regeln eindeutig zu kennzeichnen.

In den meisten Fällen finden pseudo-randomisierte Identifikatoren nach UUID Version 4 Anwendung. Einzig der Identifikator für ein und denselben Container muss zu unterschiedlichen Zeitpunkten von zum Teil unterschiedlichen Komponenten unabhängig generiert werden. Ist solch ein Fall abzusehen, werden namensbasierte Container-UUIDs gemäß UUID Version 3 angelegt. Als Basis-Namensraum wird dabei auf den UCN-Namensraum zurückgegriffen (siehe unten). Auf der Ebene eines einzelnen Systems wird ein Container durch die ihn verantwortende Abstraktionsschicht, seinen Typ und seinen individuellen Namen gekennzeichnet.

Usage-Control-Domänen Usage-Control stützt sich mit seinem Unternehmensfokus auf einen zentralen, hierarchischen Ansatz. Für jede Organisation, die Usage-Control und Provenance-Tracking einsetzt, wird eine der Organisation angepasste Domänenhierarchie aufgespannt. Dieser Hierarchie folgen das Policy-Management sowie die Verwaltung der UC- und Provenance-Komponenten. Domänen sind eindeutige Namen für technische Verantwortlichkeitsstrukturen im Unternehmen. Sie erfordern einen sauber definierten Namensraum.

[1] http://tools.ietf.org/html/rfc4122.

[2] PEP, PXP, PDP, PRP, PIP, PMP, ProSP, ProCP und ProDP.

Der UCN-Namensraum Für die Definition des Namensraums wird auf das URN-Schema zurückgegriffen.[3] Uniform Resource Names (URNs) stellen als Teil der Uniform Resource Identifier (URI) einen dauerhaften, ortsunabhängigen Bezeichner für eine Ressource zur Verfügung.[4]

Der Namensraum (Namespace Identifier (NID)) für eine UC-Domäne lautet „ucn“ für „Usage Control Node“.

Die UCN-Syntax sieht wie folgt aus:

```
<URN>          ::= "urn:ucn:" <Organization>

<Organization> ::= 1*<let-num-hyp> |
                   1*<let-num-hyp> ":" <OrgUnit> |
                   1*<let-num-hyp> ":" <SystemType> |
                   1*<let-num-hyp> ":" <IT-System>

<OrgUnit>      ::= 1*<let-num-hyp> |
                   1*<let-num-hyp> ":" <OrgUnit> |
                   1*<let-num-hyp> ":" <SystemType> |
                   1*<let-num-hyp> ":" <IT-System>

<SystemType>   ::= 1*<let-num-hyp> | 1*<let-num-hyp> ":"
                   <SystemType> | 1*<let-num-hyp> ":" <IT-System>

<IT-System>       ::= 1*<let-num-hyp>
```

Wobei <let-num-hyp> dem „Namespace Specific String Syntax“ aus RFC 2141 Abschnitt 2.2 entspricht:[5]

```
<let-num-hyp> ::= <upper> | <lower> | <number> | "-"

<upper>       ::= "A" | "B" | "C" | "D" | "E" | "F" | "G" | "H" |
                  "I" | "J" | "K" | "L" | "M" | "N" | "O" | "P" |
                  "Q" | "R" | "S" | "T" | "U" | "V" | "W" | "X" |
                  "Y" | "Z"

<lower>       ::= "a" | "b" | "c" | "d" | "e" | "f" | "g" | "h" |
                  "i" | "j" | "k" | "l" | "m" | "n" | "o" | "p" |
```

[3] http://tools.ietf.org/html/rfc2141.
[4] http://tools.ietf.org/html/rfc3986.
[5] http://tools.ietf.org/html/rfc2141.

```
                    "q" | "r" | "s" | "t" | "u" | "v" | "w" | "x" |
                    "y" | "z"

<number>        ::= "0" | "1" | "2" | "3" | "4" | "5" | "6" | "7" |
                    "8" | "9"
```

Der Namensraum ist hierarchisch (Baumstruktur). Jeder Teilpfad mit dem eine UC-Domäne beginnt, ist immer selbst auch eine gültige UC-Domäne und der ursprünglichen UC-Domäne übergeordnet. Die unterste Ebene des Namensraums wird als lokale Domäne bezeichnet.

Beispiel. *Ein gültiger Domänenname für ein Android-Testgerät der Vertriebsabteilung von AdBokis wäre* `urn:ucn:adbokis:sales:android-mobile-device:nexus10-test1`.

D XML-Schemata und Event-Deklarationen

D.1 XML-Schema der Informationsflusssemantik

Das folgende Listing zeigt die XML-Schema-Definition der Informationsflusssemantik.

```
<?xml version="1.0" encoding="UTF−8"?>
<xs:schema
  xmlns:xs="http://www.w3.org/2001/XMLSchema"
  elementFormDefault="qualified">

  <xs:element name="ifsemantics">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="params"/>
        <xs:element ref="actions"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>

  <xs:element name="params">
    <xs:complexType>
      <xs:sequence>
        <xs:element maxOccurs="unbounded" ref="param"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>

  <xs:element name="param">
    <xs:complexType>
      <xs:attribute ref="name" use="required"/>
      <xs:attribute ref="type" use="required"/>
    </xs:complexType>
  </xs:element>

  <xs:element name="actions">
    <xs:complexType>
      <xs:sequence>
        <xs:element maxOccurs="unbounded" ref="action"/>
```

```
      </xs:sequence>
    </xs:complexType>
  </xs:element>

  <xs:element name="action">
    <xs:complexType>
      <xs:sequence>
        <xs:element minOccurs="0" maxOccurs="unbounded" ref="scope"/>
        <xs:element maxOccurs="unbounded" ref="operation"/>
      </xs:sequence>
      <xs:attribute ref="name" use="required"/>
    </xs:complexType>
  </xs:element>

  <xs:element name="scope">
    <xs:complexType>
      <xs:simpleContent>
        <xs:extension base="xs:string">
          <xs:attribute ref="behavior" default="INTRA"/>
          <xs:attribute ref="delimiter" default="NONE"/>
          <xs:attribute ref="interSystem" default="FALSE"/>
        </xs:extension>
      </xs:simpleContent>
    </xs:complexType>
  </xs:element>

  <xs:element name="operation">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="left"/>
        <xs:element ref="right"/>
      </xs:sequence>
      <xs:attribute ref="name" use="required"/>
    </xs:complexType>
  </xs:element>

  <xs:element name="left">
    <xs:complexType>
      <xs:sequence>
        <xs:element minOccurs="1" maxOccurs="1" ref="operand"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>

  <xs:element name="right">
```

```
    <xs:complexType>
      <xs:sequence>
        <xs:element minOccurs="0" maxOccurs="unbounded" ref="operand"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>

  <xs:element name="operand" type="xs:NCName"/>
  <xs:attribute name="type" type="xs:NCName"/>
  <xs:attribute name="name" type="xs:NCName"/>
  <xs:attribute name="behavior" type="behaviors"/>
  <xs:attribute name="delimiter" type="delimiters"/>
  <xs:attribute name="interSystem" type="boolean"/>

  <xs:simpleType name="behaviors">
    <xs:restriction base="xs:string">
      <xs:enumeration value="IN"/>
      <xs:enumeration value="OUT"/>
      <xs:enumeration value="INTRA"/>
    </xs:restriction>
  </xs:simpleType>

  <xs:simpleType name="delimiters">
    <xs:restriction base="xs:string">
      <xs:enumeration value="OPEN"/>
      <xs:enumeration value="CLOSE"/>
      <xs:enumeration value="NONE"/>
    </xs:restriction>
  </xs:simpleType>

  <xs:simpleType name="boolean">
    <xs:restriction base="xs:string">
      <xs:enumeration value="TRUE"/>
      <xs:enumeration value="FALSE"/>
    </xs:restriction>
  </xs:simpleType>
</xs:schema>
```

D.2 XML-Schema eines Events

Die folgende XML-Schema-Definition zeigt die Syntax eines Events.

```
<?xml version="1.0" encoding="UTF−8"?>
<schema
  xmlns = "http://www.w3.org/2001/XMLSchema"
  targetNamespace = "http://www.iosb.fhg.de/duc/1.0/event"
  xmlns:tns = "http://www.iosb.fhg.de/duc/1.0/event"
  elementFormDefault = "qualified">

  <complexType name="EventParameterType">
    <attribute name="name" type="string" use="required" />
    <attribute name="value" type="string" use="required" />
    <attribute name="type" type="tns:EventParameterDataTypes" default="string"/>
  </complexType>

  <simpleType name="EventParameterDataTypes">
    <restriction base="string">
      <enumeration value="string" />
      <enumeration value="binary" />
      <enumeration value="int" />
      <enumeration value="long" />
      <enumeration value="bool" />
      <enumeration value="stringArray" />
    </restriction>
  </simpleType>

  <attributeGroup name="SimpleEventParameterAttributes">
    <attribute name="name" type="string" use="required" />
    <attribute name="value" type="string" use="required" />
    <attribute name="type" type="tns:EventParameterDataTypes" default="string"/>
  </attributeGroup>

  <complexType name="ComplexEventParameterType">
    <choice maxOccurs="unbounded">
      <element name="complexParameter" type="tns:ComplexEventParameterType" minOccurs="0"
          ↪ maxOccurs="unbounded" />
      <element name="parameter" type="tns:EventParameterType" minOccurs="0"
          ↪ maxOccurs="unbounded" />
    </choice>
    <attribute name="name" type="string" use="required" />
  </complexType>

  <complexType name="EventType">
```

```
    <choice maxOccurs="unbounded">
      <element name="complexParameter" type="tns:ComplexEventParameterType" minOccurs="0"
          ↪ maxOccurs="unbounded" />
      <element name="parameter" type="tns:EventParameterType" minOccurs="0"
          ↪ maxOccurs="unbounded" />
    </choice>
    <attribute name="action" type="string" />
    <attribute name="timestamp" type="tns:TimestampType" />
    <attribute name="isTry" type="boolean" default="false" />
    <attribute name="signallerComponent" type="string" />
    <attribute name="subject" type="string" />
    <attribute name="target" type="string" />
  </complexType>

  <simpleType name="ParamMatchDataTypes">
    <restriction base="string">
      <pattern value="string|dataUsage|xpath|regex|binary|int|long|bool|stringArray|context"/>
    </restriction>
  </simpleType>

  <complexType name="ParamMatchType">
    <attribute name="name" type="string" use="required" />
    <attribute name="value" type="string" use="required" />
    <attribute name="type" type="tns:ParamMatchDataTypes" default="string"/>
    <attribute name="negate" type="boolean" use="optional" default="false" />
  </complexType>

  <complexType name="EventMatchingOperatorType">
    <sequence>
      <element name="paramMatch" type="tns:ParamMatchType" minOccurs="0"
          ↪ maxOccurs="unbounded" />
    </sequence>
    <attribute name="action" type="string"/>
    <attribute name="class" type="tns:ActionClassType"/>
    <attribute name="isTry" type="boolean" default="false"/>
  </complexType>

  <simpleType name="ActionClassType">
    <restriction base="string">
      <enumeration value="USAGE"/>
      <enumeration value="SIGNALLING"/>
      <enumeration value="OTHER"/>
    </restriction>
  </simpleType>
```

```
  <simpleType name="TimestampType">
    <restriction base="dateTime"/>
  </simpleType>

  <element name="event" type="tns:EventType"/>
</schema>
```

D.3 Erweiterung von Event-Deklarationen

Um verschachtelte Parameter in Events zuzulassen, muss die OSL-Syntax von Hilty et al. leicht erweitert werden.[1] Die notwendige Erweiterung wird im Folgenden in der auf Z basierenden Syntax von OSL formuliert. In der ursprünglichen Form besteht ein Event aus einem Eventnamen und Parametern, die durch eine partielle Funktion ($\nrightarrow$) von Parameternamen auf Parameterwerte repräsentiert werden. Für verschachtelte Parameter wird die partielle Funktion $Params$ wie folgt definiert:

$$\begin{aligned} &[EventName, ParamName, ParamValue] \\ &EventClass == \{usage, signalling, other\} \\ &getclass : EventName \rightarrow EventClass \\ &Params : ParamName \nrightarrow (Params \cup ParamValue) \\ &Event == EventName \times Params \end{aligned} \tag{D.1}$$

Eine Eventdeklaration spezifiziert die Events, die in einem konkreten System beobachtet werden können. Die Eventdeklaration wird modifiziert wie folgt definiert:

$$\begin{aligned} &EventDecl == EventName \times EventClass \times ParamDecl \\ &ParamDecl : ParamName \nrightarrow \mathbb{P}\ (ParamDecl \cup ParamValue) \end{aligned} \tag{D.2}$$

D.4 Auszug aus der Informationsflusssemantik des Windows-PEP

```
<?xml version="1.0" encoding="UTF−8"?>
<ifsemantic>
  <params>
    <param name="process" type="CONTAINER" />
    <param name="file" type="CONTAINER" />
    <param name="process_name" type="DESIGNATOR" />
    <param name="InFileName" type="DESIGNATOR" />
  </params>
```

[1] Hilty et al. 2007.

```
<actions>
  <action name="CreateProcess">
    <operation name="NF_ADD_NAMING">
      <left>
        <operand>process_name</operand>
      </left>
      <right>
        <operand>process</operand>
      </right>
    </operation>
  </action>
  <action name="KillProcess">
    <operation name="SF_CLEAR">
      <left>
        <operand>process_name</operand>
      </left>
      <right></right>
    </operation>
    <operation name="AF_CLEAR_ALIASES">
      <left>
        <operand>process_name</operand>
      </left>
      <right></right>
    </operation>
    <operation name="NF_RM_NAMING">
      <left>
        <operand>process_name</operand>
      </left>
      <right></right>
    </operation>
  </action>
  <action name="CreateFile">
    <operation name="NF_ADD_NAMING">
      <left>
        <operand>InFileName</operand>
      </left>
      <right>
        <operand>file</operand>
      </right>
    </operation>
  </action>
  <action name="ReadFile">
    <operation name="SF_FLOW_TO_RTC">
      <left>
        <operand>process_name</operand>
```

```
        </left>
        <right>
          <operand>InFileName</operand>
        </right>
      </operation>
    </action>
    <action name="WriteFile">
      <operation name="SF_FLOW_TO_RTC">
        <left>
          <operand>InFileName</operand>
        </left>
        <right>
          <operand>process_name</operand>
        </right>
      </operation>
    </action>
  </actions>
</ifsemantic>
```

D.5 Im Beispiel zu systemübergreifenden Informationsflüssen verwendete Informationsflusssemantiken

```
<?xml version="1.0" encoding="UTF-8"?>
<ifsemantic>
  <params>
    <param name="network" type="CONTAINER" />
    <param name="network_name" type="DESIGNATOR" />
    <param name="applicationObject_name" type="DESIGNATOR" />
  </params>
  <actions>
    <action name="downloadFile_start">
      <scope behavior="OUT" delimiter="OPEN" intersystem="TRUE">URL</scope>
      <scope behavior="INTRA" />
      <operation name="NF_ADD_NAMING">
        <left>
          <operand>network_name</operand>
        </left>
        <right>
          <operand>network</operand>
        </right>
      </operation>
      <operation name="SF_FLOW">
```

```
        <left>
          <operand>network_name</operand>
        </left>
        <right>
          <operand>applicationObject_name</operand>
        </right>
      </operation>
    </action>
    <action name="downloadFile_end">
      <scope behavior="OUT" delimiter="CLOSE" intersystem="TRUE">URL</scope>
      <scope behavior="INTRA" />
      <operation name="SF_CLEAR">
        <left>
          <operand>network_name</operand>
        </left>
        <right> </right>
      </operation>
      <operation name="NF_RM_NAMING">
        <left>
          <operand>network_name</operand>
        </left>
        <right> </right>
      </operation>
  </actions>
</ifsemantic>
```

Listing D.1: Informationsflusssemantik von Shopware

```
<?xml version="1.0" encoding="UTF-8"?>
<ifsemantic>
  <params>
    <param name="network" type="CONTAINER" />
    <param name="network_name" type="DESIGNATOR" />
    <param name="applicationObject_name" type="DESIGNATOR" />
  </params>
  <actions>
    <action name="downloadFile_start">
      <scope behavior="OUT" delimiter="OPEN" intersystem="FALSE">URL</scope>
      <scope behavior="OUT" delimiter="NONE" intersystem="FALSE">URL</scope>
      <scope behavior="INTRA" />
      <operation name="NF_ADD_NAMING">
        <left>
          <operand>network_name</operand>
```

```
        </left>
        <right>
          <operand>network</operand>
        </right>
      </operation>
      <operation name="SF_FLOW">
        <left>
          <operand>applicationObject_name</operand>
        </left>
        <right>
          <operand>network_name</operand>
        </right>
      </operation>
    </action>
    <action name="downloadFile_end">
      <scope behavior="OUT" delimiter="CLOSE" intersystem="FALSE">URL</scope>
      <scope behavior="INTRA" />
      <operation name="SF_CLEAR">
        <left>
          <operand>network_name</operand>
        </left>
        <right> </right>
      </operation>
      <operation name="NF_RM_NAMING">
        <left>
          <operand>network_name</operand>
        </left>
        <right> </right>
      </operation>
    </action>
  </actions>
</ifsemantic>
```

Listing D.2: Remote-Informationsflusssemantik von Shopware

```
<?xml version="1.0" encoding="UTF-8"?>
<ifsemantic>
  <params>
    <param name="webRessource" type="CONTAINER" />
    <param name="downloadItem" type="CONTAINER" />
    <param name="downloadItemID" type="DESIGNATOR" />
    <param name="filename" type="DESIGNATOR" />
  </params>
```

```
<actions>
  <action name="onCreated">
    <scope behavior="IN" delimiter="OPEN">url</scope>
    <scope behavior="IN" delimiter="NONE">url</scope>
    <operation name="SF_FLOW">
      <left>
        <operand>downloadItem</operand>
      </left>
      <right>
        <operand>webRessource</operand>
      </right>
    </operation>
    <operation name="NF_ADD_NAMING">
      <left>
        <operand>downloadItemID</operand>
      </left>
      <right>
        <operand>downloadItem</operand>
      </right>
    </operation>
  </action>
  <action name="onChanged">
    <scope behavior="IN" delimiter="NONE">url</scope>
    <operation name="SF_FLOW">
      <left>
        <operand>downloadItemID</operand>
      </left>
      <right>
        <operand>webRessource</operand>
      </right>
    </operation>
  </action>
  <action name="onClose">
    <scope behavior="INTRA"></scope>
        <operation name="SF_FLOW">
      <left>
        <operand>filename</operand>
      </left>
      <right>
        <operand>downloadItemID</operand>
      </right>
    </operation>
    <operation name="SF_CLEAR">
      <left>
        <operand>downloadItemID</operand>
```

```
        </left>
        <right>
        </right>
      </operation>
      <operation name="NF_RM_NAMING">
        <left>
          <operand>downloadItemID</operand>
        </left>
        <right>
        </right>
      </operation>
    </action>
  </actions>
</ifsemantic>
```

Listing D.3: Informationsflusssemantik des Chrome-Download-Managers

E Beweise und Definitionen zur Unverkettbarkeit

E.1 Verkettungsrelationen

E.1.1 Relationenprodukt

Mit Hilfe des Relationenproduktes lassen sich zweistellige Relationen zu mehrstelligen Relationen verknüpfen.

Mehrstelliges Relationenprodukt Seien $R \subseteq A_1 \times \ldots \times A_j$ und $S \subseteq A_{j+1} \times \ldots \times A_n$ zwei Relationen (mit $1 \leq j < n$).

Definition E.1. *Dann ist das α-Produkt (mit $1 \leq \alpha \leq max\{j, n-j\}$) von R und S definiert als die Relation*

$$\begin{aligned} R \odot_\alpha S := \{&(a_{\alpha+1}, \ldots, a_j, a_{\alpha+j+1}, \ldots, a_n) \in A_{\alpha+1} \times \ldots \times A_j \times A_{\alpha+j+1} \times \ldots \times A_n \mid \\ &\exists a_1 \in A_1 \cap A_j \ldots \exists a_\alpha \in A_\alpha \cap A_{\alpha+j} : (a_1, \ldots, a_\alpha, a_{\alpha+1}, \ldots, a_j) \in R \wedge \\ &(a_1, \ldots, a_\alpha, a_{\alpha+j+1}, \ldots, a_n) \in S\} \end{aligned}$$

Zweistelliges Relationenprodukt Für ein zweistelliges Relationenprodukt $R \odot S$ der Relationen $R \subseteq A_1 \times A_2$ und $S \subseteq A_3 \times A_4$ werden mehrere Kurzschreibweisen verwendet. Zunächst, das (α, β)-Produkt (mit $\alpha \in \{1,2\} \wedge \beta \in \{3,4\}$):

Definition E.2.

$$R \,{}_\alpha\!\odot_\beta\, S := \begin{cases} \{(a_2, a_4) \in A_2 \times A_4 \mid \exists a \in A_1 \cap A_3 : (a, a_2) \in R \wedge (a, a_4) \in S\} & \textit{für } \alpha = 1 \wedge \beta = 3 \\ \{(a_2, a_3) \in A_2 \times A_3 \mid \exists a \in A_1 \cap A_4 : (a, a_2) \in R \wedge (a_3, a) \in S\} & \textit{für } \alpha = 1 \wedge \beta = 4 \\ \{(a_1, a_4) \in A_1 \times A_4 \mid \exists a \in A_2 \cap A_3 : (a_1, a) \in R \wedge (a, a_4) \in S\} & \textit{für } \alpha = 2 \wedge \beta = 3 \\ \{(a_1, a_3) \in A_1 \times A_3 \mid \exists a \in A_2 \cap A_4 : (a_1, a) \in R \wedge (a_3, a) \in S\} & \textit{für } \alpha = 2 \wedge \beta = 4 \end{cases}$$

Häufig kommt es vor, dass das Relationenprodukt aus zwei zweistelligen Relationen mit zwei gemeinsamen Bereichen und drei disjunkten Mengen gebildet werden soll. Der gemeinsame Bereich M wird als Mittler bezeichnet.

Definition E.3. *Das Relationenprodukt von $S \subseteq A \times M$ und $T \subseteq B \times M$ (mit $A \cap B = \emptyset \wedge A \cap M = \emptyset \wedge B \cap M = \emptyset$) wird kurz als*

$$R_M = S \odot_M T := S \,{}_2\!\odot_4\, T$$

geschrieben (analog für die Kombinationen mit S^{-1} und T^{-1}).

E.1.2 Mehrstellige Verkettungsrelationen und Anonymität

Verkettungsrelationen mit mehr als zwei Stellen werden verwendet, wenn Entitäten Mittler zur Verkettung anderer Entitäten sind. Die Verkettung von Quasi-Identifikatoren mit einem sensitiven Attribut in einer Datenbank ist solch ein Fall. Die Quasi-Identifikatoren fungieren als Mittler zwischen Subjekt und sensitivem Attribut $R \subseteq Q_1 \times \cdots \times Q_n \times A$.

Verkettungsrelationen mit Quasi-Identifikatoren Q_i und einem sensitiven Attribut A können auf zweistellige Relationen reduziert werden, wenn das n-Produkt $S \odot_n R$ aus der Verkettungsrelation zwischen Quasi-Identifikatoren und Subjektidentifikator $S \subseteq Q_1 \times \cdots \times Q_n \times ID$ und der Verkettungsrelation $R \subseteq Q_1 \times \cdots \times Q_n \times A$ gebildet wird.

Anhand dieser Überlegungen wird deutlich, dass sich Anonymität nach der hier vorgenommenen Definition als ein Sonderfall der Unverkettbarkeit betrachten lässt. Der Begriff der Anonymität findet auf Verkettungsrelationen Anwendung, die Identifikatoren als Entitäten enthalten. Solche Verkettungsrelationen werden in dieser Arbeit als Identifikationsrelationen bezeichnet.

Entsprechend der Vielfalt der Definitionen in der Literatur, werden Unverkettbarkeit und Anonymität allerdings als technisch gleich,[1] unterschiedlich[2] oder unabhängig[3] aufgefasst.

E.1.3 Äquivalenzklassen homogener Verkettungsrelationen

Aus jeder zweistelligen Verkettungsrelation $R \subseteq A \times B$ lassen sich die beiden homogenen Relationen R_A und R_B ableiten.

Definition E.4.

$$R_A := \{(b_1, b_2) \in B \times B \mid \exists a \in A : (a, b_1) \in R \wedge (a, b_2) \in R\}$$
$$R_B := \{(a_1, a_2) \in A \times A \mid \exists b \in B : (a_1, b) \in R \wedge (a_2, b) \in R\}$$

Der jeweils andere Bereich fungiert als Mittler für die Bildung der homogenen Relation.

[1] Bellare/Micciancio/Warinschi 2003.
[2] Hevia/Micciancio 2008; Pashalidis 2008; Bohli/Pashalidis 2011.
[3] Hughes/Shmatikov 2004.

Lemma E.5. *Sei $R \subseteq A \times B$ eine Verkettungsfunktion (linkstotal, rechtseindeutig). Dann ist R_B eine Äquivalenzrelation und teilt A in disjunkte Äquivalenzklassen.*

Beweis. Eine Relation ist eine Äquivalenzrelation, wenn sie reflexiv, symmetrisch und transitiv ist.
Die *Reflexivität* von R_B ergibt sich aus der Linkstotalität von R:

$$\forall a \in A \; \exists b \in B : \; (a,b) \in R \; \Rightarrow \; (a,a) \in R_B$$

Die *Symmetrie* lässt sich aus der Definition von R_B herleiten:

$$\forall a_1, a_2 \in A, (a_1, a_2) \in R_B \; \exists b \in B : \; (a_1, b) \in R \wedge (a_2, b) \in R$$
$$\Rightarrow \; (a_2, b) \in R \wedge (a_1, b) \in R \; \Rightarrow \; (a_2, a_1) \in R_B$$

Die *Transitivität* ergibt sich aus der Rechtseindeutigkeit von R:

$$\forall a_1, a_2, a_3 \in A, (a_1, a_2) \in R_B \wedge (a_2, a_3) \in R_B \; \exists b_1 \in B \; \exists b_2 \in B :$$
$$(a_1, b_1) \in R \wedge (a_2, b_1) \in R \wedge (a_2, b_2) \in R \wedge (a_3, b_2) \in R$$
und aus der Rechtseindeutigkeit von R folgt
$$\forall a \in A \; \forall b_1, b_2 \in B : (a, b_1) \in R \wedge (a, b_2) \in R \; \Rightarrow \; b_1 = b_2$$
Da nun $b_1 = b_2 \; \Rightarrow \; (a_1, b_1) \in R \wedge (a_3, b_1) \in R \; \Rightarrow \; (a_1, a_3) \in R_B$

□

Ist R bijektiv so ist auch R_A eine Äquivalenzrelation.

Beispiel. *Sofern ein personenbezogenes Datum $d \in \mathscr{D}$ genau einem Betroffenen $b \in \mathscr{B}$ zugeordnet werden kann (kein mehrfacher Personenbezug), ist die Identifikationsrelation $R^{\lessdot} \subseteq \mathscr{D} \times \mathscr{B}$ rechtseindeutig und linkstotal. $R_{\mathscr{B}} = R^{\equiv}$ ist eine Äquivalenzrelation auf der Menge aller personenbezogenen Daten.*

Beispiel. *Alle Zuordnungstabellen für (subjekteindeutige) Pseudonyme sind Funktionen. Die wahren Identitäten der Subjekte bilden eine Äquivalenzrelation auf der Menge der Pseudonyme der Tabelle.*

E.2 Additivität der Entropie

Die Entropie ist eine additive Größe. Es gilt $H(X^{\triangleright}) = H(X^{\triangleright}_{d_1}) + \cdots + H(X^{\triangleright}_{d_n})$.

Beweis.

$$
\begin{aligned}
&H(X^{\triangleright}) \\
&= -\sum_{R^{\triangleright} \in \mathcal{R}^{\triangleright}} \mathbb{P}(X^{\triangleright} = R^{\triangleright}) \log_2 \mathbb{P}(X^{\triangleright} = R^{\triangleright}) \\
&= -\sum_{R^{\triangleright}_{d_1} \in \mathcal{R}^{\triangleright}_{d_1}, \ldots, R^{\triangleright}_{d_n} \in \mathcal{R}^{\triangleright}_{d_n}} \mathbb{P}(X^{\triangleright}_{d_1} = R^{\triangleright}_{d_1}, \ldots, X^{\triangleright}_{d_n} = R^{\triangleright}_{d_n}) \log_2 \mathbb{P}(X^{\triangleright}_{d_1} = R^{\triangleright}_{d_1}, \ldots, X^{\triangleright}_{d_n} = R^{\triangleright}_{d_n}) \\
&= -\sum_{R^{\triangleright}_{d_1} \in \mathcal{R}^{\triangleright}_{d_1}, \ldots, R^{\triangleright}_{d_n} \in \mathcal{R}^{\triangleright}_{d_n}} \mathbb{P}(X^{\triangleright}_{d_1} = R^{\triangleright}_{d_1}) \cdots \mathbb{P}(X^{\triangleright}_{d_n} = R^{\triangleright}_{d_n}) \log_2 \mathbb{P}(X^{\triangleright}_{d_1} = R^{\triangleright}_{d_1}) \cdots \mathbb{P}(X^{\triangleright}_{d_n} = R^{\triangleright}_{d_n}) \\
&= -\sum_{R^{\triangleright}_{d_1} \in \mathcal{R}^{\triangleright}_{d_1}, \ldots, R^{\triangleright}_{d_n} \in \mathcal{R}^{\triangleright}_{d_n}} \mathbb{P}(X^{\triangleright}_{d_1} = R^{\triangleright}_{d_1}) \cdots \mathbb{P}(X^{\triangleright}_{d_n} = R^{\triangleright}_{d_n}) \\
&\qquad\qquad (\log_2 \mathbb{P}(X^{\triangleright}_{d_1} = R^{\triangleright}_{d_1}) + \cdots + \log_2 \mathbb{P}(X^{\triangleright}_{d_n} = R^{\triangleright}_{d_n})) \\
&= -\sum_{R^{\triangleright}_{d_1} \in \mathcal{R}^{\triangleright}_{d_1}, \ldots, R^{\triangleright}_{d_n} \in \mathcal{R}^{\triangleright}_{d_n}} \mathbb{P}(X^{\triangleright}_{d_1} = R^{\triangleright}_{d_1}) \cdots \mathbb{P}(X^{\triangleright}_{d_n} = R^{\triangleright}_{d_n}) (\log_2 \mathbb{P}(X^{\triangleright}_{d_1} = R^{\triangleright}_{d_1})) \; - \; \cdots \\
&\quad - \sum_{R^{\triangleright}_{d_1} \in \mathcal{R}^{\triangleright}_{d_1}, \ldots, R^{\triangleright}_{d_n} \in \mathcal{R}^{\triangleright}_{d_n}} \mathbb{P}(X^{\triangleright}_{d_1} = R^{\triangleright}_{d_1}) \cdots \mathbb{P}(X^{\triangleright}_{d_n} = R^{\triangleright}_{d_n}) (\log_2 \mathbb{P}(X^{\triangleright}_{d_n} = R^{\triangleright}_{d_n})) \\
&= -\sum_{R^{\triangleright}_{d_2} \in \mathcal{R}^{\triangleright}_{d_2}} \mathbb{P}(X^{\triangleright}_{d_2} = R^{\triangleright}_{d_2}) \quad \cdots \sum_{R^{\triangleright}_{d_n} \in \mathcal{R}^{\triangleright}_{d_n}} \mathbb{P}(X^{\triangleright}_{d_n} = R^{\triangleright}_{d_n}) \\
&\qquad\qquad \sum_{R^{\triangleright}_{d_1} \in \mathcal{R}^{\triangleright}_{d_1}} \mathbb{P}(X^{\triangleright}_{d_1} = R^{\triangleright}_{d_1}) (\log_2 \mathbb{P}(X^{\triangleright}_{d_1} = R^{\triangleright}_{d_1})) \; - \; \cdots \\
&\quad - \sum_{R^{\triangleright}_{d_1} \in \mathcal{R}^{\triangleright}_{d_1}} \mathbb{P}(X^{\triangleright}_{d_1} = R^{\triangleright}_{d_1}) \quad \cdots \sum_{R^{\triangleright}_{d_n} \in \mathcal{R}^{\triangleright}_{d_n}} \mathbb{P}(X^{\triangleright}_{d_n} = R^{\triangleright}_{d_n}) (\log_2 \mathbb{P}(X^{\triangleright}_{d_n} = R^{\triangleright}_{d_n})) = R^{\triangleright}_{d_n})) \\
&= -\sum_{R^{\triangleright}_{d_1} \in \mathcal{R}^{\triangleright}_{d_1}} \mathbb{P}(X^{\triangleright}_{d_1} = R^{\triangleright}_{d_1}) (\log_2 \mathbb{P}(X^{\triangleright}_{d_1} = R^{\triangleright}_{d_1})) - \cdots \\
&\qquad\qquad - \sum_{R^{\triangleright}_{d_n} \in \mathcal{R}^{\triangleright}_{d_n}} \mathbb{P}(X^{\triangleright}_{d_n} = R^{\triangleright}_{d_n}) (\log_2 \mathbb{P}(X^{\triangleright}_{d_n} = R^{\triangleright}_{d_n})) \\
&= H(X^{\triangleright}_{d_1}) + \cdots + H(X^{\triangleright}_{d_n})
\end{aligned}
$$

□

E.3 Ableitung des Provenance-Systemmodells aus dem Provenance-Datenmodell

Ein Provenance-System-Graph $\mathscr{G}^{\Sigma} = (\mathscr{S}, \mathscr{L}^{\Sigma}, col)$ ist ein kantengefärbter, gerichteter Graph mit den Knoten (Systemen) $\mathscr{S}$, den Kanten $\mathscr{L}^{\Sigma} \subseteq \mathscr{S} \times \mathscr{S}$ und der Färbung $col : \mathscr{L}^{\Sigma} \to \mathcal{P}(\mathscr{D})$.

Eine System vereint alle Repräsentationen, deren Container mit derselben Domäne bezeichnet wird. Lokale Domänen sind Bezeichner für Systeme. Die Abbildung von Repräsentationen auf Systeme ist durch die Funktion $loc : \mathscr{R} \to \mathscr{S}$ gegeben. Die Zuordnung der Systeme zu Repräsentationen durch die Funktion *loc* ist äquivalent mit der Zuordnung der Domäne zu einem Container durch die Umkehrfunktion der Benennungsfunktion f_{Dom}:

$$\begin{aligned}&\forall \rho_1 = (d_1,c_1,t_1), \rho_2 = (d_2,c_2,t_2) \in \mathscr{R} : \\ &f_{Dom}(c_1) = f_{Dom}(c_2) \longleftrightarrow loc(\rho_1) = loc(\rho_2)\end{aligned}$$

Die Kanten ergeben sich aus dem ursprünglichen Provenance-Graph $\mathscr{G}$. Sie verbinden Systeme abhängig von den Kanten $\mathscr{L}$ zwischen den Repräsentationen, die auf die Systeme abgebildet werden. Die Kantenfärbung $col : \mathscr{L}^{\Sigma} \to \mathcal{P}(\mathscr{D})$ erhält die Information, welche Daten von einem System in ein anderes geflossen sind. Diese Information ist im Provenance-Graph nicht in den Kanten, sondern in den Repräsentationen, also den Knoten, hinterlegt.

Die Kanten und die Färbung des Provenance-System-Graphen $\mathscr{L}^{\Sigma} \subseteq \mathscr{S} \times \mathscr{S}$ leiten sich wie folgt aus dem Provenance-Graphen ab:

$$\begin{aligned}&\forall \varsigma_1, \varsigma_2 \in \mathscr{S}\ \forall d \in \mathscr{D}\ \forall \rho_1 = (d,c_1,t_1), \rho_2 = (d,c_2,t_2) \in \mathscr{R} : \\ &\quad loc(\rho_1) = \varsigma_1 \wedge loc(\rho_2) = \varsigma_2 \wedge (\rho_1, \rho_2) \in \mathscr{L} \longrightarrow (\varsigma_1, \varsigma_2) \in \mathscr{L}^{\Sigma} \wedge d \in col(\varsigma_1, \varsigma_2) \\ &\forall \varsigma_1, \varsigma_2 \in \mathscr{S}\ \exists \rho_1, \rho_2 \in \mathscr{R} : \\ &\quad (\varsigma_1, \varsigma_2) \in \mathscr{L}^{\Sigma} \longrightarrow loc(\rho_1) = \varsigma_1 \wedge loc(\rho_2) = \varsigma_2 \wedge (\rho_1, \rho_2) \in \mathscr{L}\end{aligned}$$

F Ablauf, Aufgaben und Fragebögen der Nutzerstudie

In den folgenden drei Abschnitten werden die Vorgaben zum Versuchsablauf (Abschnitt F.1), die Aufgaben (Abschnitt F.2) und die Fragebögen (Abschnitt F.3) wörtlich wiedergegeben. In der Studie wurde *PrivacyInsight* als „Privacy Dashboard“ und *GenomSynlig* kurz als „Synlig“ bezeichnet.

F.1 Ablaufprotokoll

(Regieanweisungen sind blau hervorgehoben)

Tabs offen, Synlig gefiltert, PDF offen

Testkandidat sitzt am Survey-Schreibtisch.

Datenschutzerklärung aushändigen und ausfüllen lassen

Optional mündlich erläutern

Im Rahmen dieser Studie erfassen wir

1. den Zeitbedarf für die Aufgabenstellungen im Versuchsablauf,
2. die Antworten zu den Aufgaben im Versuchsablauf,
3. die Bewertungen und demografische Daten entsprechend des Versuchsfragebogens,
4. die Trackingdaten des Blickverhaltens im Versuch
5. und Kommentare des Versuchsteilnehmers zum Versuchsablauf und zum Versuchsobjekt.

Die Teilnehmerliste wird getrennt von den Versuchsdaten aufbewahrt. Da dennoch eine Zuordnung ihrer Daten über den Zeitstempel möglich wäre, gelten die Daten als pseudonym erhoben. Nach Ende der Studie, jedoch spätestens nach einem Monat, werden die Teilnehmerliste und die Zeitstempel gelöscht.

Die Daten werden nur anonym ausgewertet.

Bitte füllen Sie als erstes die Datenschutzerklärung für diesen Versuch aus.

Tobii-Aufzeichnung wird gestartet

Testkandidat an den PC bitten

Testkandidat sitzt vor dem PC, interagiert aber nicht mit diesem.

Fassen Sie bitte Maus oder Tastatur nicht an, außer Sie werden dazu aufgefordert. Bitte betätigen Sie niemals die ESC-Taste.

Ich werde Ihnen jetzt eine Einführung in das Szenario geben.

Nach § 34 Abs. 1 Satz 1 BDSG hat die verantwortliche Stelle dem Betroffenen auf Verlangen Auskunft über die zu seiner Person gespeicherten Daten, deren Herkunft, mögliche Empfänger und den Zweck der Speicherung zu erteilen. Um dem Nutzer die Möglichkeit zu bieten sein Auskunftsrecht online wahrzunehmen gibt es unterschiedliche Systeme und Darstellungsmöglichkeiten.

Im Rahmen der Evaluation nehmen Sie die Rolle des Betroffenen im Sinne des Datenschutzes ein. Sie haben auf Empfehlung eines Freundes beim Onlineshop AdBokis Buchclub GmbH eingekauft und sich dafür ein Benutzerkonto angelegt. Zwei Tage später erhalten Sie einen Newsletter per Mail der Firma Extreme Advertisement Ltd. Sie möchten jetzt wissen wie Ihre Daten an das Unternehmen gekommen sind. Da Sie von der Extreme Advertisement Ltd. keine Antwort erhalten erkundigen Sie sich bei AdBokis.

Wir schauen uns jetzt nacheinander drei unterschiedliche Darstellungsmöglichkeiten für Datenschutzauskünfte an. Es gibt 5 Aufgaben die unter Zeitnahme für jede Darstellungsmöglichkeit gelöst werden müssen. Beantworten Sie die Fragen zügig, nach den Fragen haben Sie noch Zeit sich das System genauer anzuschauen. Nach dem Lösen der Aufgaben muss jeweils noch ein kurzer Usability-Fragebogen ausgefüllt werden, bevor mit der nächsten Darstellungsmöglichkeit angefangen werden kann.

Je System wiederholen

Sie sehen auf dem Bildschirm gerade die erste/zweite/dritte Darstellungsvariante.

Die Darstellungsvariante wird von der AdBokis Buchclub GmbH / einem Dienstleister auf ihrer/seiner Webseite zur Verfügung gestellt.

In dieser Variante sind Sie die/der Betroffene Alice Fox / Bob Bobsson.

Ich werde jetzt Aufgaben vorlesen, die Sie bitte zu lösen versuchen. Ich werde Sie nach jeder Aufgabe fragen, ob Sie sie verstanden haben. Bitte beginnen Sie erst, wenn ich Ihnen sage, dass Sie jetzt beginnen dürfen und fassen Maus oder Tastatur vorher bitte nicht an. Ab dann haben Sie zwei Minuten Zeit, um die Aufgabe zu lösen. Wenn Sie eine Lösung gefunden haben, lassen Sie bitte Maus und Tastatur wieder los und teilen mir dies mit. Ich stoppe dann die Zeit. Sie haben Papier und Stift vor sich liegen und können sich Notizen machen. Haben Sie erstmal mit der Auflösung begonnen, dürfen Sie nicht mehr mit dem PC interagieren.

Haben Sie die Anweisungen verstanden?

Beginnen wir mit der ersten Aufgabe (-- Sie dürfen jetzt wieder an den PC).

Je Aufgabe wiederholen

Aufgabe vorlesen

Haben Sie die Aufgabe verstanden?

Wenn nein, wiederholen/erklären, wenn ja weiter

Dann dürfen Sie beginnen.

Sie haben jetzt noch eine Minute Zeit sich ohne Zielvorgabe mit dem System vertraut zu machen.

Minute warten lassen.

Bitte wechseln Sie den Arbeitsplatz. Wir haben für dieses System einen System Usability Survey vorbereitet.

SUS aushändigen und ausfüllen lassen

Im zweiten Abschnitt wollen wir uns jetzt eine der Darstellungsvarianten genauer anschauen. Hierfür gibt es zuerst eine kurze Einführung und dann 10 Aufgaben. Danach gibt es noch kurz die Möglichkeit sich das System frei anzuschauen und dann erneut einen kurzen Usability-Fragebogen. Abschließend werden noch generelle Fragen gestellt und demographische Daten erhoben. Bitte fassen Sie während der folgenden Erklärung die Maus oder Tastatur nicht an. Wir führen die folgenden Dinge nicht vor, Sie können Sie gerne während der folgenden Aufgaben ausprobieren.

Erklärung Interaktionsicons aushändigen

In der oberen Leiste finden Sie auf der rechten Seite allgemeine Interaktionsmöglichkeiten. Von links nach rechts sind das: die Option einen Schritt zurück zu gehen, den Graphen auf seine Ursprungszustand zurückzusetzen, Ihre Daten zu exportieren und abschließend sich abzumelden.

Zentral finden sie den Fluss Ihrer Daten durch das Unternehmen in Graphenform. Jeder Kreis stellt dabei eine Organisationseinheit oder ein IT-System dar. Orangene Kreise lassen sich durch einfaches Klicken in enthaltene Untereinheiten aufteilen und damit genauer Analysieren. Die grauen Kreise am linken Rand stellen die Informationsquellen dar, die grauen Kreise am rechten Rand die Informationssenken. Die blauen Grafiken neben Quellen und Senken stellen die erhobenen, bzw. übermittelten, personenbezogenen Daten dar - durch einfaches klicken auf die blauen Grafiken wird der zurückgelegte Pfad bzw. Datenfluss im Graphen hervorgehoben.

Beim Bewegen der Maus über Knoten oder Datensätze offenbaren diese ihren Namen und bieten mit Klick auf die orangenen Symbole die Möglichkeit weitere Details zur Datenverarbeitung zu erfahren (LUPE), einen Knoten Aufzuklappen (DOPPELPFEIL) oder in den Datensatz hineinzuschauen (AUGE) und diesen herunterzuladen (DOWNLOAD), zu bearbeiten (STIFT), zu löschen (MÜLLEIMER) und die Adbokis Buchclub GmbH zu kontaktieren (BRIEFUMSCHLAG).

Doppelklick und Rechtsklick haben keine Funktion, das gesamte Privacy Dashboard lässt sich, wie im Web üblich, mit einfachen Linksklicks bedienen.
Haben Sie noch Fragen?

Fangen wir mit den Aufgaben an.

Je Aufgabe wiederholen

Aufgabe vorlesen

Haben Sie die Aufgabe verstanden?

Wenn nein, wiederholen/erklären, wenn ja weiter

Dann dürfen Sie beginnen.

Wiederholen bis alle Aufgaben abgearbeitet sind

Sie haben jetzt noch zwei Minuten Zeit sich ohne Zielvorgabe mit dem System vertraut zu machen.

Zwei Minuten warten lassen.

Bitte wechseln Sie den Arbeitsplatz.

Wir haben auch für diese Runde einen System Usability Survey vorbereitet.

Außerdem noch 3 weitere Fragebögen, die wir Ihnen anschließend aushändigen.

SUS aushändigen und ausfüllen lassen

Als zweites den User Experience Questionnaire. Auch dieser Fragebogen bezieht sich ausschließlich auf die Darstellungsvariante, die sie zuletzt gesehen haben.

UEQ aushändigen und ausfüllen lassen

Und als drittes einige ergänzende Fragen.

„Privacy Dashboard“ in Frage 2 bezeichnet die Darstellungsvariante, die Sie zuletzt gesehen haben. „Synlig“ bezeichnet die Darstellungsvariante, in der Sie Bob Bobsson waren. In Frage 4 wählen sie bitte nur eine Antwort aus. Die oberen Antwortmöglichkeiten werden durch die darunterliegenden eingeschlossen. (Mit Ausnahme der letzten beiden Antwortmöglichkeiten)

Erweiterte Fragen aushändigen und ausfüllen lassen

Zuletzt einen Fragebogen für demografische Daten.

Demografische Daten ausfüllen lassen.

Vielen Dank für die Teilnahme an unserer Evaluation!

F.2 Aufgaben

Aufgaben im ersten Teil

1. Finden Sie heraus, ob die Adbokis Buchclub GmbH in der Vergangenheit ein Bild von Ihnen bekommen hat.
2. Finden Sie heraus, welche e-Mailadresse bei der Adbokis Buchclub GmbH von Ihnen hinterlegt ist.
3. Finden Sie heraus, welche IP-Adressen die Adbokis Buchclub GmbH mit Ihnen verknüpft.
4. Beantragen Sie die Löschung Ihrer Telefonnummer.
5. Lassen Sie sich nur den Datenfluss Ihres Vornamens zur Adbokis Buchclub GmbH anzeigen. (Bei JSON: Ein sequenzielles Anzeigen ist ausreichend)

Aufgaben im zweiten Teil

1. Finden Sie heraus, an welche Dritte die Adbokis Buchclub GmbH Ihre Daten weitergegeben hat.
2. Finden Sie heraus, von welchem Arbeitsplatzrechner der Adbokis Buchclub GmbH aus Ihre e-Mailadresse an die Extreme Advertisement Ltd. weitergegeben wurde.

3. Finden Sie heraus, für welchen Zweck die Adbokis Buchclub GmbH Ihre Telefonnummer speichert.
4. Finden Sie heraus, durch wie viele Abteilungen Ihre e-Mailadresse geflossen ist.
5. Listen Sie auf, welche Datenkategorien auf dem Archivserver über Sie gespeichert sind.
6. Lassen Sie sich nur den Datenfluss der Rechnung anzeigen.
7. Exportieren Sie Ihre Daten.
8. Downloaden Sie Ihr Profilbild.
9. Beantragen Sie eine Änderung Ihrer IBAN.
10. Setzen Sie die Darstellung auf den Ausgangszustand zurück.

F.3 Fragebögen

Der gesamte, über einen einzelnen Probanden gesammelte Datensatz bestand aus folgenden Teilen:

1. Deckblatt mit Notizen und Informationen zur verwendeten Permutation (Abbildung F.1)
2. Datenschutzerklärung (Abbildung F.2)
3. Protokoll der Bewertungen für den ersten Aufgabenteil zu *PrivacyInsight*, *GenomSynlig* und JSON (Abbildung F.3)
4. SUS für den ersten Aufgabenteil zu *PrivacyInsight*, *GenomSynlig* und JSON (Abbildung F.4)
5. Protokoll der Bewertungen für den zweiten Aufgabenteil (Abbildung F.5 und F.6)
6. SUS für den zweiten Aufgabenteil (Abbildung F.4)
7. UEQ für den zweiten Aufgabenteil (Abbildung F.7 und F.8)
8. Fragebogen mit erweiterten Fragen (Abbildung F.9 bis F.11)
9. Fragebogen für die demographischen Angaben (Abbildung F.12 und F.13)

Dazu kommen noch die Aufzeichnungen des Eye-Trackings.

In den folgenden Abbildungen sind die Originaldokumente dargestellt.

Deckblatt

Zeit: _____(YYYY-MM-DD, hh:mm)

Versuchsleiter: _____ (Name)

Permutation Block 1

☐ 1-2-3 ☐ 1-3-2 ☐ 2-1-3

☐ 2-3-1 ☐ 3-1-2 ☐ 3-2-1

1 = Privacy Dashboard
2 = Synlig
3 = JSON-PDF

Kommentare Versuchsteilnehmer

_____

_____

_____

_____

_____

Bemerkungen und Beobachtungen Versuchsleiter

_____

_____

_____

_____

_____

_____

Abbildung F.1: Deckblatt

Datenschutzerklärung

Im Rahmen der Studie „Privacy Dashboard“ werden personenbezogene Daten zum Zweck der Wissenschaft und Forschung gemäß den Bestimmungen des Bundesdatenschutzgesetzes erhoben, verarbeitet und genutzt. Die Teilnahme an der Studie ist freiwillig.

Folgende personenbezogene Daten werden erhoben:

- Teilnehmerliste
  - Name und Vorname
  - E-Mailadresse und ggf. Telefonnummer
  - Versuchszeitpunkt
- Versuchsdaten
  - Zeitbedarf für Aufgabenstellungen im Versuchsablauf
  - Antworten zu Aufgaben im Versuchsablauf
  - Bewertungen und demografische Daten entsprechend des Versuchsfragebogens
  - Trackingdaten des Blickverhaltens im Versuch
  - Kommentare des Versuchsteilnehmers zum Versuchsablauf und zum Versuchsobjekt

Personenbezogene Daten werden ausschließlich beim Versuchsteilnehmer erhoben. Alle Versuchsdaten werden pseudonym erhoben und mit einem Zeitstempel versehen. Ein Personenbezug besteht einzig über Zeitstempel und die Teilnehmerliste (Zuordnungstabelle).

Die Teilnehmerliste wird getrennt von den Versuchsdaten aufbewahrt. Nach Ende der Studie, jedoch spätestens nach einem Monat, wird die Teilnehmerliste gelöscht.

Die Versuchsdaten werden anhand des Zeitstempels zusammengeführt. Nach Ende der Studie, jedoch spätestens nach einem Monat, werden die Zeitstempel aus den Versuchsdaten gelöscht. Die Daten sind ab diesem Zeitpunkt vollständig anonymisiert. Eine darüber hinausgehende Verarbeitung und Nutzung (insbesondere die Auswertung, Weitergabe und Veröffentlichung) findet nur auf Grundlage der anonymisierten Daten statt.

Ich habe die Datenschutzerklärung gelesen und verstanden.

Ort, Datum: _____ Unterschrift:_____

Ich willige in die Erhebung, Verarbeitung und Nutzung meiner personenbezogenen Daten zum Zweck der Wissenschaft und Forschung nach den Vorgaben und im Umfang der obigen Datenschutzerklärung ein.

Ort, Datum: _____ Unterschrift:_____

Abbildung F.2: Datenschutzerklärung

Privacy Dashboard

Aufgaben

Finden Sie heraus, ob die Adbokis Buchclub GmbH in der Vergangenheit ein Bild von Ihnen bekommen hat.

Lösung: Ja

☐ gelöst ☐ nicht gelöst

Dauer: _____

Finden Sie heraus, welche e-Mailadresse bei der Adbokis Buchclub GmbH von Ihnen hinterlegt ist.

Lösung: Alice.Fox@Honigmail.de

☐ gelöst ☐ nicht gelöst

Dauer: _____

Finden Sie heraus, welche IP-Adressen die Adbokis Buchclub GmbH mit Ihnen verknüpft.

Lösung: 217.146.191.19, 31.130.202.80

☐ gelöst ☐ nicht gelöst

Dauer: _____

Beantragen Sie die Löschung Ihrer Telefonnummer.

Lösung: Klick auf Löschen Icon

☐ gelöst ☐ nicht gelöst

Dauer: _____

Lassen Sie sich nur den Datenfluss Ihres Vornamens zur Adbokis Buchclub GmbH anzeigen.

Lösung: Selektieren des Datums

☐ gelöst ☐ nicht gelöst

Dauer: _____

Abbildung F.3: Protokoll für den ersten Aufgabenteil

Privacy Dashboard

System Usability Scale

Um das System zu bewerten, füllen Sie bitte den nachfolgenden Fragebogen aus. Er besteht aus Aussagen über das System. Abstufungen in der Übereinstimmung mit diesen Aussagen sind durch Kreise dargestellt. Durch Ankreuzen eines dieser Kreise können Sie Ihre Übereinstimmung mit einer Aussage äußern.

Beispiel

Ich fand das System unnötig bunt.	○	○	○	○	⊗

Mit dieser Beurteilung sagen Sie aus, dass Sie der Aussage, das System sei unnötig bunt, gar nicht zustimmen.

Entscheiden Sie möglichst spontan. Es ist wichtig, dass Sie nicht lange über die Aussagen nachdenken, damit Ihre unmittelbare Einschätzung zum Tragen kommt.

Bitte kreuzen Sie immer eine Antwort an, auch wenn Sie bei der Einschätzung zu einer Aussage unsicher sind. Es gibt keine „richtige“ oder „falsche“ Antwort. Ihre persönliche Meinung zählt!

Bitte geben Sie nun Ihre Einschätzung des Systems ab. Kreuzen Sie nur einen Kreis pro Zeile an.

	Ich stimme zu		Ich bin neutral		Ich stimme nicht zu
Ich denke, dass ich das System gerne häufig benutzen würde.	○	○	○	○	○
Ich fand das System unnötig komplex.	○	○	○	○	○
Ich denke, das System war leicht zu benutzen.	○	○	○	○	○
Ich glaube, ich würde die Hilfe einer fachkundigen Person benötigen, um das System benutzen zu können.	○	○	○	○	○
Ich fand, die verschiedenen Funktionen in diesem System waren gut integriert.	○	○	○	○	○
Ich denke, das System enthielt zu viele Inkonsistenzen.	○	○	○	○	○
Ich kann mir vorstellen, dass die meisten Menschen den Umgang mit diesem System sehr schnell lernen.	○	○	○	○	○
Ich fand das System sehr umständlich zu benutzen.	○	○	○	○	○
Ich fühlte mich bei der Nutzung des Systems sehr sicher.	○	○	○	○	○
Ich musste eine Menge lernen, bevor ich mit dem System arbeiten konnte.	○	○	○	○	○

Abbildung F.4: System Usability Scale

Block 2

Aufgaben

Finden Sie heraus, an welche Dritte die Adbokis Buchclub GmbH Ihre Daten weitergegeben hat.

Lösung: PayPal, Extreme Advertisement

☐ gelöst ☐ nicht gelöst

Dauer: _____

Finden Sie heraus, von welchem Arbeitsplatzrechner der Adbokis Buchclub GmbH aus Ihre e-Mailadresse an die Extreme Advertisement Ltd. weitergegeben wurde.

Lösung: (Vertriebsabteilung) Workspace 23

☐ gelöst ☐ nicht gelöst

Dauer: _____

Finden Sie heraus, für welchen Zweck die Adbokis Buchclub GmbH Ihre Telefonnummer speichert.

Lösung: Kundenservice/-betreuung

☐ gelöst ☐ nicht gelöst

Dauer: _____

Finden Sie heraus, durch wie viele Abteilungen Ihre e-Mailadresse geflossen ist.

Lösung: 2 (Vertrieb, Kundenbetreuung)

☐ gelöst ☐ nicht gelöst

Dauer: _____

Listen Sie auf, welche Datenkategorien auf dem Archivserver über Sie gespeichert sind.

Lösung: Lieferadresse, Nachname, Vorname, e-Mailadresse, Lieferadresse, Rechnung, Wohnsitzstaat

☐ gelöst ☐ nicht gelöst

Dauer: _____

Lassen Sie sich nur den Datenfluss der Rechnung anzeigen.

Lösung: Selektieren des Datensatzes

☐ gelöst ☐ nicht gelöst

Dauer: _____

Seite **1** von **2**

Abbildung F.5: Protokoll für den zweiten Aufgabenteil (S. 1)

Block 2

Aufgaben

Exportieren Sie Ihre Daten.

Lösung: Klick in der Menüleiste

☐ gelöst ☐ nicht gelöst

Dauer: _____

Downloaden Sie Ihr Profilbild.

Lösung: Klick aufs Icon

☐ gelöst ☐ nicht gelöst

Dauer: _____

Beantragen Sie eine Änderung Ihrer IBAN.

Lösung: Klick aufs Icon

☐ gelöst ☐ nicht gelöst

Dauer: _____

Setzen Sie die Darstellung auf den Ausgangszustand zurück.

Lösung: Klick in der Menüleiste

☐ gelöst ☐ nicht gelöst

Dauer: _____

Seite **2** von **2**

Abbildung F.6: Protokoll für den zweiten Aufgabenteil (S. 2)

Block 2

User Experience Questionnaire

Um das System zu bewerten, füllen Sie bitte den nachfolgenden Fragebogen aus. Er besteht aus Gegensatzpaaren von Eigenschaften, die das System haben kann. Abstufungen zwischen den Gegensätzen sind durch Kreise dargestellt. Durch Ankreuzen eines dieser Kreise können Sie Ihre Zustimmung zu einem Begriff äußern.

Beispiel

attraktiv	○	⊗	○	○	○	○	○	unattraktiv

Mit dieser Beurteilung sagen Sie aus, dass Sie das System eher attraktiv als unattraktiv einschätzen.

Entscheiden Sie möglichst spontan. Es ist wichtig, dass Sie nicht lange über die Begriffe nachdenken, damit Ihre unmittelbare Einschätzung zum Tragen kommt.

Bitte kreuzen Sie immer eine Antwort an, auch wenn Sie bei der Einschätzung zu einem Begriffspaar unsicher sind oder finden, dass es nicht so gut zum System passt.

Es gibt keine „richtige“ oder „falsche“ Antwort. Ihre persönliche Meinung zählt!

Abbildung F.7: User Experience Questionnaire (S. 1)

Block 2

User Experience Questionnaire

	1	2	3	4	5	6	7		
unerfreulich	○	○	○	○	○	○	○	erfreulich	1
unverständlich	○	○	○	○	○	○	○	verständlich	2
kreativ	○	○	○	○	○	○	○	phantasielos	3
leicht zu lernen	○	○	○	○	○	○	○	schwer zu lernen	4
erfrischend	○	○	○	○	○	○	○	einschläfernd	5
langweilig	○	○	○	○	○	○	○	spannend	6
uninteressant	○	○	○	○	○	○	○	interessant	7
unberechenbar	○	○	○	○	○	○	○	voraussagbar	8
schnell	○	○	○	○	○	○	○	langsam	9
neu	○	○	○	○	○	○	○	alt	10
unbedienbar	○	○	○	○	○	○	○	bedienbar	11
gut	○	○	○	○	○	○	○	schlecht	12
kompliziert	○	○	○	○	○	○	○	einfach	13
abstoßend	○	○	○	○	○	○	○	anziehend	14
veraltet	○	○	○	○	○	○	○	modern	15
unangenehm	○	○	○	○	○	○	○	angenehm	16
vorhersagbar	○	○	○	○	○	○	○	unvorhersagbar	17
abwechslungsreich	○	○	○	○	○	○	○	eintönig	18
zuverlässig	○	○	○	○	○	○	○	unzuverlässig	19
ineffizient	○	○	○	○	○	○	○	effizient	20
übersichtlich	○	○	○	○	○	○	○	verwirrend	21
stockend	○	○	○	○	○	○	○	flüssig	22
aufgeräumt	○	○	○	○	○	○	○	überladen	23
schön	○	○	○	○	○	○	○	hässlich	24
sympathisch	○	○	○	○	○	○	○	unsympathisch	25
unauffällig	○	○	○	○	○	○	○	auffällig	26

Seite **2** von **2**

Abbildung F.8: User Experience Questionnaire (S. 2)

Block 2

Erweiterte Fragen

Füllen Sie bitte den nachfolgenden ergänzenden Fragebogen aus.

Entscheiden Sie möglichst spontan. Es ist wichtig, dass Sie nicht lange über die Aussagen nachdenken, damit Ihre unmittelbare Einschätzung zum Tragen kommt.

Bitte kreuzen Sie immer eine Antwort an, auch wenn Sie bei der Einschätzung zu einer Aussage unsicher sind. Es gibt keine „richtige“ oder „falsche“ Antwort. Ihre persönliche Meinung zählt!

Insgesamt würde ich die Nutzerfreundlichkeit des Systems „Privacy Dashboard“ wie folgt bewerten:

Schlechtest-möglich	Schrecklich	Gering	Okay	Gut	Exzellent	Best-möglich
○	○	○	○	○	○	○

Sind die im Privacy Dashboard zusätzlich verfügbaren Informationen die gesteigerte Komplexität im Vergleich zu Synlig wert?

Ganz und gar nicht	○	○	○	○	○	○	○	Vollkommen

Halten Sie die Integration von Datenschutzauskunft und weiteren Betroffenenrechten (Löschung, Berichtigung) in einer Benutzeroberfläche für sinnvoll?

Ganz und gar nicht	○	○	○	○	○	○	○	Vollkommen

Welche Informationstiefe würden Sie im Datenflussgraph für personenbezogene Daten im Rahmen einer Auskunft nach § 34 Bundesdatenschutzgesetz erwarten?

- ○ Nur Erhebungs- und Übermittlungsvorgänge
- ○ Datenflüsse zu Auftragsdatenverarbeitern
- ○ Datenflüsse zwischen Abteilungen (einschließlich Auftragsdatenverarbeitern)
- ○ Datenflüsse zwischen einzelnen IT-Systemen
- ○ Datenflüsse in IT-Systemen zwischen Applikationen und Speicherbereichen
- ○ Ich halte Informationen zu Datenflüssen grundsätzlich für unnötig
- ○ Weiß nicht

Seite **1** von **3**

Abbildung F.9: Erweiterte Fragen (S. 1)

Block 2

Erweiterte Fragen
Unlinkability

Unlinkability (Deutsch: Unverkettbarkeit) misst, wie viel besser Abteilungen in der Adbokis Buchclub GmbH dadurch, dass sie Zugriff auf die für die Auskunft gesammelten Informationen haben, über die Zusammenhänge zwischen Ihnen, Ihren Daten und den Orten der Datenverarbeitung Bescheid wissen (relative Metriken). Im Privacy Dashboard sehen sie vier verschiedene Metriken für Unlinkability:

- Ganz auf der linken Seite eine Speicher- und Verarbeitungsmetrik. Sie beschreibt, wie viel Abteilungen in der Adbokis Buchclub GmbH über die Organisationseinheiten und IT-Systeme in denen Ihre Daten verarbeitet und gespeichert werden, wissen.
- Links-mittig eine Datenflussmetrik. Sie charakterisiert, wie viel Abteilungen in der Adbokis Buchclub GmbH über die Herkunft Ihrer Daten, die Weitergabe Ihrer Daten zwischen Organisationseinheiten und IT-Systemen und die Übermittlung Ihrer Daten an andere Unternehmen wissen.
- Rechts-mittig eine Verknüpfungsmetrik. Sie kennzeichnet, wie gut Abteilungen in der Adbokis Buchclub GmbH feststellen können, ob zwei Datensätze zur gleichen Person gehören.
- Ganz rechts eine Identifikationsmetrik. Sie legt offen, wie gut Abteilungen in der Adbokis Buchclub GmbH Daten Ihnen als Person zuordnen können.
- Beispielsweise bedeutet eine Identifikationsunlinkability von 80%, dass Abteilungen in der Adbokis Buchclub GmbH 80% der Informationen durch die Auskunft nicht bekommen haben, die sie vor Installation eines Auskunftssystems noch gebraucht hätten, um jede Zuordnung offen zu legen. Auf der anderen Seite bedeutet dies aber auch, dass Sie für das Unternehmen um 20% transparenter geworden sind.

Die Angabe der Unlinkability gibt Ihnen als Betroffenem der Datenverarbeitung die Möglichkeit zwischen den Vorteilen abzuwägen, die Ihnen das Privacy Dashboard bietet (Transparenz über die Datenflüsse, Einblick in die Daten, Löschung und Berichtigung) und dem Nachteil, dass sie auch für das datenverarbeitende Unternehmen transparenter werden.

Deshalb können sie durch Klick auf den entsprechenden Button im Privacy Dashboard bestimmen, dass Sie in Zukunft auf eine Nutzung des Privacy Dashboards verzichten wollen. Eine Datensammlung zum Zweck der Auskunft findet dann nicht mehr statt.

Abbildung F.10: Erweiterte Fragen (S. 2)

Block 2

Erweiterte Fragen

Haben Sie das Konzept von Unlinkability verstanden?

- ○ Ja
- ○ Nein
- ○ Weiß nicht

Halten Sie die angezeigten Unlinkabilitymetriken für hilfreich?

- ○ Ja
- ○ Nein
- ○ Weiß nicht

Würden Sie aufgrund der gegebenen Unlinkability (76%, 80%, 86%, 89% - Erläuterung siehe oben) in Zukunft lieber auf die Nutzung eines Privacy Dashboards verzichten?

- ○ Ja
- ○ Nein
- ○ Unentschieden
- ○ Weiß nicht

Wer ist Ihrer Meinung nach für den Schutz von personenbezogenen Daten verantwortlich?

Bitte vergleichen sie paarweise.

	1	2	3	4	5	6	7	
Der Staat	○	○	○	○	○	○	○	Der Bürger
Die Unternehmen	○	○	○	○	○	○	○	Der Staat
Die Kunden	○	○	○	○	○	○	○	Die Unternehmen

Ich bin mir bei meinen Einschätzungen über alle Fragebögen hinweg...

Vollkommen unsicher	○	○	○	○	○	○	○	Sehr sicher

Seite **3** von **3**

Abbildung F.11: Erweiterte Fragen (S. 3)

Demografische Angaben

Bitte beantworten Sie uns noch einige letzte Fragen zu Ihrer Person. Die Angaben dienen rein statistischen Zwecken. Wir versichern Ihnen, dass wir alle Angaben nur in anonymisierter Form auswerten werden.

Welches Geschlecht haben Sie?

- ○ Weiblich
- ○ Männlich
- ○ Keine Angabe

Wie alt sind Sie?

- ○ 14-19 Jahre
- ○ 20-29 Jahre
- ○ 30-39 Jahre
- ○ 40-49 Jahre
- ○ 50-59 Jahre
- ○ 60-69 Jahre
- ○ 70 und älter
- ○ Keine Angabe

Welchen höchsten Schulabschluss haben Sie?

- ○ Sonderschulabschluss, Abschluss der Förderschule
- ○ Volks- oder Hauptschulabschluss / Polytechnische Oberschule 8. Klasse
- ○ Mittlere Reife (Realschulabschluss, Polytechnische Oberschule 10. Klasse)
- ○ Fachhochschulreife, Abitur, Erweiterte Oberschule
- ○ Abgeschlossenes Studium
- ○ Weiß nicht
- ○ Keine Angabe

Sollten Sie einen ausländischen Schulabschluss haben, so versuchen Sie bitte diesen einem der deutschen Bildungsabschlüsse zuzuordnen.

Was machen Sie derzeit hauptsächlich?

- ○ Erwerbstätig (auch selbstständig)
- ○ In Ausbildung, Schule, Umschulung, Studium, Wehr-/Zivildienst
- ○ Rentner, Pensionär, Vorruhestand
- ○ Hausmann/-frau, Elternzeit, Mutterschutz
- ○ Zurzeit arbeitslos
- ○ Keine Angabe

Seite **1** von **2**

Abbildung F.12: Demographische Angaben (S. 1)

Demografische Angaben

Wie würden Sie Ihre Wohngegend einordnen?

Ist sie…

- ○ Großstädtisch (> 100 000 Einwohner) und Einzugsbereich
- ○ Städtisch (> 20 000 Einwohner)
- ○ Kleinstädtisch
- ○ Ländlich geprägt
- ○ Weiß nicht
- ○ Keine Angabe

Wenn Sie technische Probleme mit beispielsweise Ihrem PC oder Handy haben, können Sie diese...

- ○ Fast immer selbst lösen
- ○ Manchmal selbst lösen
- ○ Fast niemals selbst lösen
- ○ Weiß nicht
- ○ Keine Angabe

Welche Erfahrung haben Sie mit Webanwendungen (Software, die über einen Webbrowser angezeigt und bedient wird)?

- ○ Ich habe bereits selbst Webanwendungen programmiert
- ○ Ich nutze regelmäßig Webanwendungen
- ○ Ich nutze gelegentlich Webanwendungen
- ○ Ich weiß nicht, was Webanwendungen sind
- ○ Keine Angabe

Haben Sie bereits selbst ein Auskunftsgesuch nach § 34 Bundesdatenschutzgesetz gestellt?

- ○ Schon mehrmals
- ○ Einmalig
- ○ Noch nie
- ○ Weiß nicht
- ○ Keine Angabe

Seite **2** von **2**

Abbildung F.13: Demographische Angaben (S. 2)

Glossar

Abruf

Abruf ist der Bezug personenbezogener Daten per Datenübertragung durch einen Empfänger. Die weitergebende Stelle sorgt gegebenenfalls für die notwendigen technischen Voraussetzungen, wird jedoch nicht selbst aktiv. Weitergabe an einen Empfänger und Abruf durch einen Empfänger sind in ihren Rechtsfolgen gleich.

(Datenschutz-)Auskunftsplattform

Die Benutzeroberfläche, auf der dem Betroffenen eine Auskunft präsentiert wird. Eine Auskunftsplattform hat im Regelfall interaktive Elemente zur Änderung von Datenschutzvoreinstellungen oder der Wahrnehmung von Betroffenenrechten. Das englischsprachige Synonym für eine Auskunftsplattform ist Privacy Dashboard. Die in dieser Arbeit entwickelte Auskunftsplattform heißt *PrivacyInsight*.

Auswerten

Auswerten ist das Erzeugen neuer personenbezogener Daten mit eigener Aussagekraft über den Betroffenen aus bestehenden personenbezogenen Daten, insbesondere die Berechnung von Scorewerten und die automatisierte Einzelentscheidung. Dabei können auch nicht-personenbezogene Daten in die Auswertung einfließen.

Daten

Daten sind nach einer festgelegten Syntax gebildete Zeichenketten. Daten sind nicht zwingend körperlich.

Datenschutzauskunftssystem

Ein System, das die für eine Datenschutzauskunft erforderlichen Informationen automatisiert sammelt und dem Betroffenen auf Anfrage bereitstellt.

Einsichtnahme

Einsichtnahme ist die höchstpersönliche Betrachtung personenbezogener Daten durch den Empfänger. Die Einsichtnahme muss nicht optisch, sondern kann auch akustisch oder haptisch erfolgen. Sie unterscheidet sich vom Abruf durch den Medienbruch. Eine Verarbeitung der eingesehenen personenbezogenen Daten durch

den Empfänger ist nicht ohne weiteres möglich, sondern erfordert eine vorhergehende Erfassung.

Erheben

Erheben ist nach § 3 Abs. 3 BDSG das „Beschaffen von [personenbezogenen] Daten über den Betroffenen." Die Erhebung kann beim Betroffenen selbst oder bei einem Dritten stattfinden.

Informationen

Informationen sind Daten gemeinsam mit ihrer Semantik.
Personenbezogene Daten im Sinne des Datenschutzgesetzes sind Informationen. Insofern werden Daten und Informationen vielfach synonym verwendet. Die Legaldefinition personenbezogener Daten wird in Kapitel 3.7.1 diskutiert.

Löschen

Löschen ist nach § 3 Abs. 4 S. 2 Nr. 5 BDSG das „Unkenntlichmachen gespeicherter personenbezogener Daten." Löschen ist keine Verarbeitung i. e. S., da personenbezogene Daten nicht Teil der Eingabe oder der Ausgabe des Löschvorgangs sind. Die Löschung hat keine personenbezogenen Daten zum Ergebnis.

Nutzung

Die Nutzung ist ein Auffangtatbestand im BDSG. Die Nutzung ist nach § 3 Abs. 5 BDSG „jede Verwendung personenbezogener Daten, soweit es sich nicht um Verarbeitung handelt."

Nutzungskontrolle

Nutzungskontrolle (engl. Usage Control (UC)) bezeichnet Verfahren, die steuern, unter welchen Umständen Zugriff auf Daten gewährt wird und wie nachfolgend mit den Daten umgegangen werden darf. Nutzungskontrolle kann als eine Erweiterung von Access-Control über den Zugriffszeitpunkt hinaus betrachtet werden.

Personenbezogene Daten

Personenbezogene Daten sind nach § 3 Abs. 1 BDSG „Einzelangaben über persönliche und sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person (*Betroffener*)." Der komplexe Begriff der personenbezogenen Daten wird im Detail in Kapitel 3.7.1 diskutiert.

(Ereignis-)Protokoll

Ein (Ereignis-)Protokoll ist die A-posteriori-Aufzeichnung und Dokumentation von Ereignissen in IT-Systemen und Organisationen, die personenbezogene Daten verwenden. Ein Protokoll ist vorgangsbezogen. Ein Protokoll umfasst (1) den Zeitpunkt und die Reihenfolge der Ereignisse, (2) die Ereignisse selbst und (3) deren Verursacher. Verursacher sind beteiligte Personen und andere Akteure sowie technische Systeme. Ein Protokoll kann in einer Protokolldatei (auch: Logdatei) gespeichert werden.
In Abgrenzung dazu ist ein *Kommunikationsprotokoll* die A-priori-Festlegung von Ablauf, Syntax, und Semantik eines Kommunikationsvorgangs.

(Data-)Provenance

Data-Provenance ist die dokumentierte Ableitungshistorie eines Datums, ausgehend von seiner ursprünglichen Quelle. Provenance gibt Auskunft über die Herkunft von Daten sowie über die Erzeugung von Daten aus anderen Daten. Während ein Protokoll vorgangsbezogen ist, ist Provenance datenbezogen. Die Data-Provenance für die personenbezogenen Daten eines Betroffenen wird als *Personal-Data-Provenance* bezeichnet.

Speichern

Speichern ist nach § 3 Abs. 4 S. 2 Nr. 1 BDSG das „Aufbewahren personenbezogener Daten auf einem Datenträger zum Zwecke ihrer weiteren“ Verwendung.

Sperren

Sperren ist nach § 3 Abs. 4 S. 2 Nr. 4 BDSG das „Kennzeichnen gespeicherter personenbezogener Daten, um ihre weitere Verarbeitung oder Nutzung einzuschränken.“ Sperren ist keine Verarbeitung i. e. S., da personenbezogene Daten nicht verändert, sondern nur ergänzende Metadaten hinzugefügt werden.

System

Ein System ist allgemein eine Menge von Entitäten oder Komponenten gemeinsam mit den Beziehungen zwischen diesen. Ein System entsteht dadurch, dass es operiert und sich gegenüber seiner Umgebung durch Exklusivität abgrenzt. So allgemein wie der Systembegriff ist, so vielfältig wird er auch verwendet. Im *Provenance-Systemmodell* ist ein System eine Ansammlung technischer (z. B. ein Cluster) oder organisatorischer (z. B. eine Abteilung) Entitäten, denen ein gemeinsames Wissen unterstellt wird. Interne Datenflüsse und verarbeitete Daten sind allen beteiligten Entitäten bekannt. Sofern eine Ansammlung organisatorischer Entitäten gemeint ist, entspricht der Systembegriff dem organisatorischen Stellenbegriff. Der Umfang eines

Systems legt den Verantwortungsbereich der UC- und Provenance-Komponenten fest. Im *Angreifermodell* ist ein System $\Sigma^{\mathscr{A}}$ ein Modell der Menge aller Entitäten, mit denen der Angreifer in Interaktion treten kann. Ein *IT-System* ist jedes datenverarbeitende technische Gerät, das von anderen Systemen, beispielsweise über eine IP-Adresse, adressierbar ist. Ein IT-System kann auch virtualisiert sein.

Übermitteln

Übermitteln ist nach § 3 Abs. 4 S. 2 Nr. 3 BDSG die Weitergabe personenbezogener Daten an einen Dritten oder die Einsichtnahme oder der Abruf personenbezogener Daten durch einen Dritten.

Umgang

Umgang ist die Erhebung, Verarbeitung und Nutzung personenbezogener Daten.

(Grad der) Unverkettbarkeit

Der Grad der Unverkettbarkeit ist die Unsicherheit eines Angreifers A über die wahre Verkettungsrelation in einer Menge von Kandidatenrelationen $\mathcal{R}$ im Gesamtsystem $\Sigma^{\mathscr{A}}$, nachdem das Beobachtungsereignis I eingetreten ist, im Vergleich zur Unsicherheit mit seinem vorherigen Wissensstand.

Verändern

Verändern ist nach § 3 Abs. 4 S. 2 Nr. 2 BDSG die Modifikation, das „inhaltliche Umgestalten gespeicherter personenbezogener Daten."

Verarbeitung

Nach § 3 Abs. 4 S. 1 BDSG ist Verarbeitung jedes „Speichern, Verändern, Übermitteln, Sperren und Löschen personenbezogener Daten." Die Verwendung des Begriffs „Verarbeitung" in der DSRL ist extensiver und entspricht dem Begriff des Umgangs im Bundesdatenschutzgesetz.
Der Begriff der Verarbeitung personenbezogener Daten im BDSG ist unglücklich definiert. Zum einen umfasst er nicht alles, was landläufig unter Verarbeitung verstanden wird (Auswertung, insb. die automatisierte Einzelentscheidung), zum anderen nimmt er die in ihrer Qualität anderen Tatbestände der Speicherung und Übermittlung mit auf. Dadurch erzeugt er eine begriffliche Asymmetrie. Die Übermittlung an Dritte und die Erhebung von Dritten stehen einander gleichberechtigt gegenüber. Der eine Tatbestand ist jedoch Teil der Verarbeitung, der andere nicht. Weitergabevorgänge innerhalb einer verantwortlichen Stelle sind ebenso keine Verarbeitung. Deshalb wird in den Kapiteln 4 bis 9 der sauber abgegrenzte Begriff der Verarbeitung i. e. S. verwendet.

Verarbeitung i. e. S.

Die Verarbeitung im engeren Sinne bezeichnet die Veränderung und Auswertung (algorithmische Informationsverarbeitung) von personenbezogenen Daten in Prozessen, die Speicherung der Daten zur sofortigen Weiterverarbeitung in dem einem Prozess zugewiesenen Teil des Arbeitsspeichers und dem flüchtigen Speicher (Register und Cache) eines Prozessors sowie die Aufbereitung der Daten für die Einsichtnahme durch den Benutzer über eine Benutzerschnittstelle, für die Weitergabe, für den Abruf oder für die dauerhafte Speicherung. Die Speicherung zur dauerhaften Aufbewahrung, auch im Arbeitsspeicher (RAM-Disk), ist nicht Teil der Verarbeitung i. e. S.

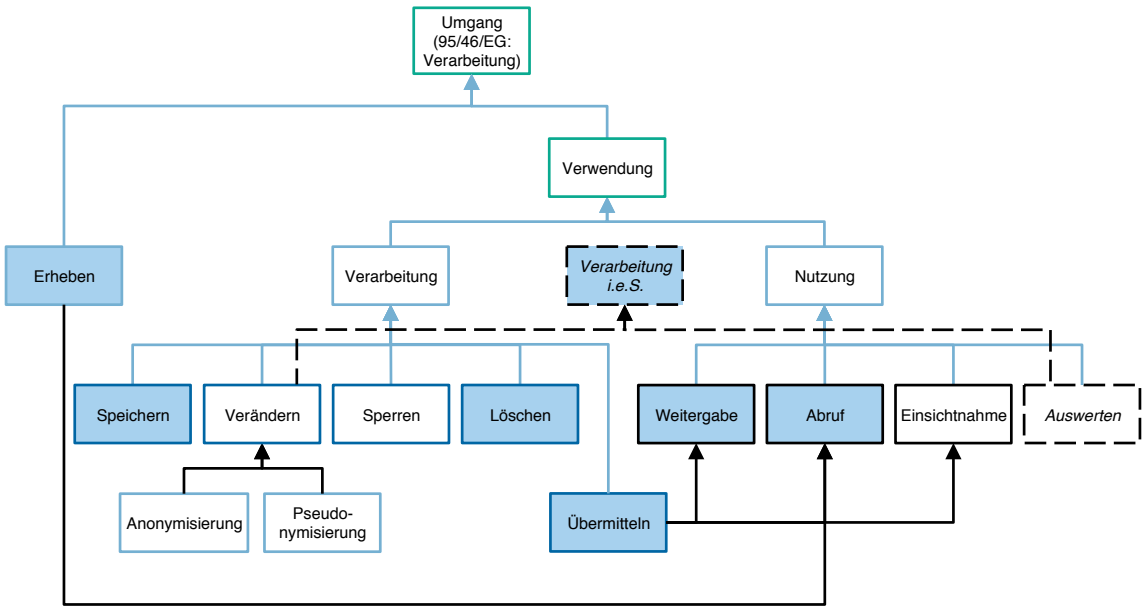


Abbildung: Begriffshierarchie Datenschutz[1]

Veröffentlichung

Eine Veröffentlichung ist das Bereithalten personenbezogener Daten zum Abruf oder zur Einsicht durch eine nicht klar abgegrenzte Personengruppe.

[1] Legende: Übergeordneter Sammelbegriff im BDSG; der Begriff wird in § 3 BDSG in einem eigenen Absatz definiert; der Begriff wird in § 3 BDSG untergeordnet definiert; der Begriff wird im BDSG verwendet; der Begriff wird im BDSG nicht verwendet; der Vorgang wird durch das implementierte Provenance-Tracking erfasst; **A** → **B** A ist im rechtlichen Sinne B; **A** ⟶ **B** A kann im technischen Sinne B sein; **A** -→ **B** A ist in dieser Arbeit als B definiert.

Verwendung

Verwendung ist jeder Umgang mit personenbezogenen Daten mit Ausnahme der Erhebung.

Weitergabe

Weitergabe ist die Übertragung personenbezogener Daten von einer Stelle auf eine andere Person oder Stelle. Die interne Weitergabe ist die Weitergabe zwischen Personen oder Stellen, die einer gemeinsamen verantwortlichen Stelle zuzuordnen sind. Gemeint ist jede Form von Übertragung: Mündliche oder fernmündliche Mitteilung, schriftliche Übersendung, drahtgebundene oder drahtlose Datenübertragung, Übergabe oder Übersendung eines Datenträgers.[2]

[2]Dammann in: Simitis, BDSG 2014, § 3 Rn. 146.

Eigene Veröffentlichungen

Bier, Christoph. „Data Protection and Security Awareness in Complex Information Systems“. In: *Future Security 2011. 6th Security Research Conference. Proceedings.* Hrsg. von Joachim Ender. Stuttgart: Fraunhofer Verlag, 2011, S. 1–4.

Vagts, Hauke und Christoph Bier. „Anonymization in Intelligent Surveillance Systems“. In: *Proceedings of the 4th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*. IEEE, 2011, S. 1–4.

Bier, Christoph, Pascal Birnstill, Erik Krempel, Hauke Vagts und Jürgen Beyerer. „How is Positive-Sum Privacy Feasible?“ In: *Future Security. 7th Security Research Conference*. Hrsg. von Nils Aschenbruck, Peter Martini, Michael Meier und Jens Tölle. CCIS 318. Berlin, Heidelberg: Springer, 2012, S. 265–268.

Bier, Christoph und Erik Krempel. „Common Privacy Patterns in Video Surveillance and Smart Energy“. In: *Proceedings of the 7th International Conference on Computing and Convergence Technology (ICCCT)*. IEEE, 2012, S. 610–615.

Bier, Christoph und Indra Spiecker genannt Döhmann. „Intelligente Videoüberwachungstechnik: Schreckensszenario oder Gewinn für den Datenschutz?“ In: *Computer und Recht* (2012), S. 610–618.

Bier, Christoph. „Das koreanische Datenschutzrecht. Ein Überblick“. In: *Datenschutz und Datensicherheit* 37.7 (2013), S. 457–460.

Bier, Christoph. „How Usage Control and Provenance Tracking Get Together - A Data Protection Perspective“. In: *Proceedings of the IEEE Symposium on Security and Privacy Workshops (SPW)*. IEEE, 2013, S. 13–17.

Salmela, Laura, Toni Ahonen, Csaba Beleznai, Rafal Knapik, Christoph Bier, Wojciech Wojciechowicz und Sirra Toivonen. „Towards User Requirements for Harmonised Automated Border Control Gates“. In: *Future Security 2013. 8th Security Research Conference. Proceedings.* Hrsg. von Michael Lauster. Stuttgart: Fraunhofer Verlag, 2013, S. 258–266.

Bier, Christoph, Pascal Birnstill, Erik Krempel, Hauke Vagts und Jürgen Beyerer. „Enhancing Privacy by Design From a Developer's Perspective“. In: *Privacy Technologies and Policy. 1st Annual Privacy Forum (APF), 2012*. Hrsg. von Bart Preneel und Demosthenes Ikonomou. LNCS 8319. Berlin, Heidelberg: Springer, 2014, S. 73–85.

Bier, Christoph und Jonas Prior. „Detection and Labeling of Personal Identifiable Information in E-mails“. In: *ICT Systems Security and Privacy Protection. 29th IFIP TC 11 International Conference, SEC 2014*. Hrsg. von Nora Cuppens-Boulahia, Frédéric Cuppens, Sushil Jajodia, Anas Abou El Kalam und Thierry Sans. IFIP AICT 428. Berlin, Heidelberg: Springer, 2014, S. 351–358.

Bier, Christoph. „Data Provenance. Technische Lösungskonzepte für das Datenschutzrecht auf Auskunft“. In: *Datenschutz und Datensicherheit* 39.11 (2015), S. 741–746.

Bier, Christoph. „Datenschutzziele im Konflikt: Eine Metrik für Unverkettbarkeit als Hilfestellung für den Betroffenen“. In: *Informatik 2016*. Hrsg. von Heinrich C. Mayr und Martin Pinzger. LNI 259. GI, 2016, S. 455–468.

Bier, Christoph und Jürgen Beyerer. „Towards Measuring the Linkage Risk in Information Flows“. In: *Security Research Conference. 11th Future Security*. Hrsg. von Oliver Ambacher, Joachim Wagner und Rüdiger Quay. Stuttgart: Fraunhofer Verlag, 2016, S. 293–299.

Bier, Christoph, Kay Kühne und Jürgen Beyerer. „PrivacyInsight: The Next Generation Privacy Dashboard“. In: *Privacy Technologies and Policy. 4th Annual Privacy Forum, APF 2016*. Hrsg. von Stefan Schiffner, Jetzabel Serna, Demosthenes Ikonomou und Kai Rannenberg. LNCS 9857. Berlin, Heidelberg: Springer, 2016, S. 135–152.

Birnstill, Pascal, Christoph Bier, Paul Wagner und Jürgen Beyerer. „Generic Semantics Specification and Processing for Inter-System Information Flow Tracking“. In: *Proceedings of the 15th International Conference on Security and Management (SAM)*. New York, NY: ACM, 2016, S. 185–191.

Bier, Christoph, Simon Kömpf und Jürgen Beyerer. „A Study on Corporate Compliance with Transparency Requirements of Data Protection Law“. In: *Data Protection and Privacy: (In)visibilities and Infrastructures*. Hrsg. von Ronald Leenes, Serge Gutwirth, Paul De Hert und Rosamunde van Brakel. Law, Governance and Technology Series 36. Cham: Springer, 2017, S. 271–289.

Betreute Abschlussarbeiten

Prior, Jonas. „Detektion und Kennzeichnung personenbezogener Daten am Beispiel Thunderbird“. Masterarbeit. Karlsruher Institut für Technologie (KIT), 2013.

Sommerfeld, Hermann. „Visualisierung und Repräsentation von Provenance-Datenmodellen in Datenschutz-Auskunftssystemen“. Masterarbeit. Hochschule Karlsruhe, 2013.

Stritzke, David. „Verfügbarkeit von Provenance-Daten im Spannungsfeld rechtlicher Anforderungen“. Masterarbeit. Karlsruher Institut für Technologie (KIT), 2014.

Kömpf, Simon. „E-Mail-basierte Analyse der Verbreitung personenbezogener Daten“. Bachelorarbeit. Karlsruher Institut für Technologie (KIT), 2015.

Helwig, Dimitri. „Implementierung von Data-Provenance-Tracking in einem e-Commerce-System und Integration in die Datenschutzauskunftsinfrastruktur“. Masterarbeit. Hochschule Karlsruhe, 2016.

Kühne, Kay. „Ergonomie und Usability in Datenschutzauskunftssystemen“. Bachelorarbeit. Karlsruher Institut für Technologie (KIT), 2016.

Literatur

Acquisti, Alessandro, Leslie K. John und George Loewenstein. „What Is Privacy Worth?“ In: *Journal of Legal Studies* 42.2 (2013), S. 249–274.

Aldeco Pérez, Rocio und Luc Moreau. „Provenance-based Auditing of Private Data Use“. In: *Proceedings of the 2008 International Research Conference on Visions of Computer Science: BSC International Academic Conference*. Hrsg. von Erol Gelenbe, Samson Abramsky und Vladimiro Sassone. Swinton, UK: BCS, 2008, S. 141–152.

Andersson, Christer und Reine Lundin. „On the Fundamentals of Anonymity Metrics“. In: *The Future of Identity in the Information Society. Proceedings of the Third IFIP WG 9.2, 9.6/11.6, 11.7/FIDIS International Summer School on The Future of Identity in the Information Society*. Hrsg. von Simone Fischer-Hübner, Penny Duquenoy, Albin Zuccato und Leonardo Martucci. IFIP AICT 262. Boston, MA: Springer, 2008, S. 325–341.

Andersson, Henrik, Julio Angulo, Karin Bernsmed, Simone Fischer-Hübner, Christian Frøystad, Erlend Andreas Gjære, Farzaneh Karegar, Daniel Lindegren und John Sören Pettersson. *D45.4: User Interface Prototypes V2*. Project Deliverable. Accountability For Cloud und Other Future Internet Services (A4Cloud), 2015. URL: http://www.a4cloud.eu/sites/default/files/D45.4%20User%20interface%20prototypes%20V2.pdf.

Angulo, Julio, Simone Fischer-Hübner, Tobias Pulls und Erik Wästlund. „Usable Transparency with the Data Track: A Tool for Visualizing Data Disclosures“. In: *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI)*. New York, NY: ACM, 2015, S. 1803–1808.

Artikel-29-Datenschutzgruppe. *Begriff „personenbezogene Daten“*. Stellungnahme 4. Europäische Kommission, 2007.

Artikel-29-Datenschutzgruppe. *Anwendung der Begriffe der Notwendigkeit und der Verhältnismäßigkeit sowie des Datenschutzes im Bereich der Strafverfolgung*. Stellungnahme 1. Europäische Kommission, 2014.

Asghar, Muhammad Rizwan, Mihaela Ion, Giovanni Russello und Bruno Crispo. „Securing Data Provenance in the Cloud“. In: *Open Problems in Network Security. IFIP WG 11.4 International Workshop, iNetSec 2011*. Hrsg. von Jan Camenisch und Dogan Kesdogan. LNCS 7039. Berlin, Heidelberg: Springer, 2012, S. 145–160.

Bangor, Aaron, Philip Kortum und James Miller. „Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale“. In: *Journal of Usability Studies* 4.3 (2009), S. 114–123.

Basin, David, Matus Harvan, Felix Klaedtke und Eugen Zalinescu. „Monitoring Data Usage in Distributed Systems“. In: *IEEE Transactions on Software Engineering* 39.10 (2013), S. 1403–1426.

Bechini, Alessio, Mario G.C.A. Cimino, Francesco Marcelloni und Andrea Tomasi. „Patterns and Technologies for Enabling Supply Chain Traceability Through Collaborative E-business“. In: *Information and Software Technology* 50 (2008), S. 342–359.

Becker, Moritz Y., Alexander Malkis und Laurent Bussard. „A Practical Generic Privacy Language“. In: *Information Systems Security. 6th International Conference, ICISS 2010*. Hrsg. von Somesh Jha und Anish Mathuria. LNCS 6503. Berlin, Heidelberg: Springer, 2010, S. 125–139.

Bedner, Mark und Tobias Ackermann. „Schutzziele der IT-Sicherheit“. In: *Datenschutz und Datensicherheit* 34.5 (2010), S. 323–328.

Bell, Eric Temple. „Exponential Numbers“. In: *The American Mathematical Monthly* 41.7 (1934), S. 411–419.

Bellare, Mihir, Daniele Micciancio und Bogdan Warinschi. „Foundations of Group Signatures: Formal Definitions, Simplified Requirements, and a Construction Based on General Assumptions“. In: *Advances in Cryptology — EUROCRYPT 2003. International Conference on the Theory and Applications of Cryptographic Techniques*. Hrsg. von Eli Biham. LNCS 2656. Berlin, Heidelberg: Springer, 2003, S. 614–629.

Berendt, Bettina, Oliver Günther und Sarah Spiekermann. „Privacy in E-commerce: Stated Preferences vs. Actual Behavior“. In: *Communications of the ACM* 48.4 (2005), S. 101–106.

Beresford, Alastair R., Sören Preibusch und Dorothea Kübler. „Unwillingness to Pay for Privacy: A Field Experiment“. In: *IZA Discussion Paper Series* 5017 (2010).

Birnstill, Pascal und Alexander Pretschner. „Enforcing Privacy Through Usage-Controlled Video Surveillance“. In: *Proceedings of the 10th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 2013, S. 318–323.

Bock, Kirsten und Sebastian Meissner. „Datenschutz-Schutzziele im Recht“. In: *Datenschutz und Datensicherheit* 36.6 (2012), S. 425–431.

Bohli, Jens-Matthias und Andreas Pashalidis. „Relations Among Privacy Notions“. In: *ACM Transactions on Information and System Security* 14.1 (2011), S. 1–24.

Bonachela, Juan A., Haye Hinrichsen und Miguel A. Mũoz. „Entropy Estimates of Small Data Sets“. In: *Journal of Physics A: Mathematical and Theoretical* 41.20 (2008), S. 1–9.

Bose, Rajendra und James Frew. „Lineage Retrieval for Scientific Data Processing: A Survey“. In: *ACM Computing Surveys* 37.1 (2005), S. 1–28.

Brooke, John. „SUS - A Quick and Dirty Usability Scale“. In: *Usability Evaluation in Industry* 189.194 (1996), S. 4–7.

Buneman, Peter, Sanjeev Khanna und Wang-Chiew Tan. „Why and Where : A Characterization of Data Provenance“. In: *Database Theory – ICDT 2001. 8th International Conference*. Hrsg. von Jan Van den Bussche und Victor Vianu. LNCS 1973. Berlin, Heidelberg: Springer, 2001, S. 316–330.

Butin, Denis, Marcos Chicote und Daniel Le Métayer. „Log Design for Accountability“. In: *Proceedings of the IEEE Symposium on Security and Privacy Workshops (SPW)*. IEEE, 2013, S. 1–7.

Cadenhead, Tyrone, Vaibhav Khadilkar, Murat Kantarcioglu und Bhavani Thuraisingham. „A Language for Provenance Access Control“. In: *Proceedings of the 1st ACM Conference on Data and Application Security and Privacy (CODASPY)*. New York, NY: ACM, 2011, S. 133–144.

Cheney, James, Laura Chiticariu und Wang-Chiew Tan. „Provenance in Databases: Why, How, and Where“. In: *Foundations and Trends in Databases* 1.4 (2009), S. 379–474.

Chung, Lawrence und Julio Cesar Sampaio do Prado Leite. „On Non-Functional Requirements in Software Engineering“. In: *Conceptual Modeling: Foundations and Applications. Essays in Honor of John Mylopoulos*. Hrsg. von Alexander T. Borgida, Vinay K. Chaudhri, Paolo Giorgini und Eric S. Yu. LNCS 5600. Berlin, Heidelberg: Springer, 2009, S. 363–379.

Cranor, Lorrie Faith, Praveen Guduru und Manjula Arjula. „User Interfaces for Privacy Agents“. In: *ACM Transactions on Computer-Human Interaction* 13.2 (2006), S. 135–178.

Curbera, Francisco, Yurdaer Doganata, Axel Martens, Nirmal K. Mukhi und Aleksander Slominski. „Business Provenance — A Technology to Increase Traceability of End-to-End Operations“. In: *On the Move to Meaningful Internet Systems: OTM 2008. OTM 2008 Confederated International Conferences, CoopIS, DOA, GADA, IS, and ODBASE*. Hrsg. von Robert Meersman und Zahir Tari. LNCS 5331. Berlin, Heidelberg: Springer, 2008, S. 100–119.

Davidson, Susan B., Sanjeev Khanna, Sudeepa Roy, Julia Stoyanovich, Val Tannen und Yi Chen. „On Provenance and Privacy“. In: *Proceedings of the 14th International Conference on Database Theory (ICDT)*. New York, NY: ACM, 2011, S. 3–10.

Demsky, Brian. „Garm: Cross Application Data Provenance and Policy Enforcement“. In: *Proceedings of the 4th USENIX conference on Hot topics in security (HotSec)*. Montreal: USENIX, 2009, S. 10.

Demsky, Brian. „Cross Application Data Provenance and Policy Enforcement“. In: *ACM Transactions on Information and System Security (TISSEC)* 14.1 (2011), S. 1–22.

Diaz, Claudia, Stefaan Seys, Joris Claessens und Bart Preneel. „Towards Measuring Anonymity“. In: *Privacy Enhancing Technologies. Second International Workshop, PET 2002*. Hrsg. von Roger Dingledine und Paul Syverson. LNCS 2482. Berlin, Heidelberg: Springer, 2003, S. 54–68.

Dreier, Horst, Hrsg. *Grundgesetz. Kommentar*. 3. Aufl. Bd. 1. Tübingen: Mohr Siebeck, 2013.

Fabian, Benjamin, Seda Gürses, Maritta Heisel, Thomas Santen und Holger Schmidt. „A Comparison of Security Requirements Engineering Methods“. In: *Requirements Engineering* 15.1 (2010), S. 7–40.

Fernandez-Gago, Carmen, Vasilis Tountopoulos, Simone Fischer-Hübner, Rehab Alnemr, David Nuñez, Julio Angulo, Tobias Pulls und Theo Koulouris. „Tools for Cloud Accountability: A4Cloud Tutorial“. In: *Privacy and Identity Management for the Future Internet in the Age of Globalisation. 9th IFIP WG 9.2, 9.5, 9.6/11.7, 11.4, 11.6/SIG 9.2.2 International Summer School*. Hrsg. von Jan Camenisch, Simone Fischer-Hübner und Marit Hansen. IFIP AICT 457. Boston, MA: Springer, 2015, S. 219–236.

Feth, Denis und Alexander Pretschner. „Flexible Data-Driven Security for Android“. In: *Proceedings of the 6th IEEE International Conference on Software Security and Reliability*. IEEE, 2012, S. 41–50.

Fischer-Hübner, Simone, Julio Angulo und Tobias Pulls. „How can Cloud Users be Supported in Deciding on, Tracking and Controlling How their Data are Used?“ In: *Privacy and Identity Management for Emerging Services and Technologies. 8th IFIP WG 9.2, 9.5, 9.6/11.7, 11.4, 11.6 International Summer School*. Hrsg. von Marit Hansen, Jaap-Henk Hoepman, Ronald Leenes und Diane Whitehouse. IFIP AICT 421. Berlin, Heidelberg: Springer, 2014, S. 77–92.

Fischer-Hübner, Simone, Christina Köffel, John-Sören Pettersson, Peter Wolkerstorfer, Cornelia Graf, Leif Erik Holtz, Ulrich König, Hans Hedbom und Benjamin Kellermann. *D4.1.3: HCI Pattern Collection – Version 2*. Project Deliverable. Privacy und Identity Management in Europe for Life (PrimeLife), 2010. URL: http://primelife.ercim.eu/images/stories/deliverables/d4.1.3-hci_pattern_collection_v2-public.pdf.

Fischer, Lars, Stefan Katzenbeisser und Claudia Eckert. „Measuring Unlinkability Revisited“. In: *Proceedings of the 7th ACM Workshop on Privacy in the Electronic Society*. Hrsg. von Eli Biham. New York, NY: ACM, 2008, S. 105–110.

Fonseca, Rodrigo, George Porter, Randy H. Katz, Scott Shenker und Ion Stoica. „X-trace: A Pervasive Network Tracing Framework“. In: *Proceedings of the 4th USENIX Conference on Networked Systems Design & Implementation (NSDI)*. April. Cambridge, MA: USENIX, 2007, S. 20.

Franz, Matthias, Bernd Meyer und Andreas Pashalidis. „Attacking Unlinkability: The Importance of Context“. In: *Privacy Enhancing Technologies. 7th International Symposium, PET 2007*. Hrsg. von Roger Dingledine und Paul Syverson. LNCS 4776. Berlin, Heidelberg: Springer, 2007, S. 1–16.

Freire, Juliana, David Koop, Emanuele Santos, Cláudio T. Silva und English Dictionary. „Provenance for Computational Tasks: A Survey“. In: *Computing in Science & Engineering* 10.3 (2008), S. 11–21.

Fromm, Alexander, Florian Kelbert und Alexander Pretschner. „Data Protection in a Cloud-Enabled Smart Grid“. In: *Smart Grid Security. First International Workshop, SmartGridSec 2012*. Hrsg. von Jorge Cuellar. LNCS 7823. Berlin, Heidelberg: Springer, 2013, S. 96–107.

Gadelha Jr., Luiz M. R. und Marta Mattoso. „Kairos: An Architecture for Securing Authorship and Temporal Information of Provenance Data in Grid-Enabled Workflow Management Systems“. In: *Proceedings of the IEEE 4th International Conference on eScience*. IEEE, 2008, S. 597–602.

Gallwas, Hans-Ullrich. „Der allgemeine Konflikt zwischen dem Recht auf informationelle Selbstbestimmung und der Informationsfreiheit“. In: *Neue Juristische Wochenschrift* (1992), S. 2785–2790.

Gehani, Ashish und Ulf Lindqvist. „Bonsai: Balanced Lineage Authentication“. In: *Proceedings of the 23rd Annual Computer Security Applications Conference (ACSAC)*. IEEE, 2007, S. 363–373.

Gehani, Ashish und Dawood Tariq. „SPADE: Support for Provenance Auditing in Distributed Environments“. In: *Middleware 2012. ACM/IFIP/USENIX 13th International Middleware Conference*. Hrsg. von Priya Narasimhan und Peter Triantafillou. LNCS 7662. Berlin, Heidelberg: Springer, 2012, S. 101–120.

Glinz, Martin. „On Non-Functional Requirements“. In: *Proceedings of the 15th IEEE International Requirements Engineering Conference*. IEEE, 2007, S. 21–26.

Gola, Peter, Rudolf Schomerus und Barbara Körffer. *Bundesdatenschutzgesetz. Kommentar*. 12. Aufl. München: C.H. Beck, 2015.

Grossklags, Jens und Alessandro Acquisti. „When 25 Cents is Too Much: An Experiment on Willingness-To-Sell and Willingness-To-Protect Personal Information“. In: *Proceedings of the 6th Annual Workshop on the Economics of Information Security (WEIS)*. 2007.

Grunwell, Daniel, Randike Gajanayake und Tony Sahama. „The Security and Privacy of Usage Policies and Provenance Logs in an Information Accountability Framework“. In: *Proceedings of the 8th Australasian Workshop on Health Informatics and Knowledge Management (HIKM)*. Hrsg. von Anthony Maeder und Jim Warren. CRPIT 164. 2015, S. 33–40.

Gurlit, Elke. „Verfassungsrechtliche Rahmenbedingungen des Datenschutzes“. In: *Neue Juristische Wochenschrift* (2010), S. 1035–1041.

Hammer, Volker und Reinhard Fraenkel. „Datenschutzfreundliches Smart Metering“. In: *Datenschutz und Datensicherheit* 35.12 (2011), S. 890–895.

Hammer, Volker, Ulrich Pordesch und Alexander Roßnagel. *Betriebliche Telefon- und ISDN-Anlagen rechtsgemäß gestaltet*. Berlin: Springer, 1993.

Hansen, Marit. „Putting Privacy Pictograms into Practice – a European Perspective“. In: *Informatik 2009. Im Focus das Leben*. Hrsg. von Stefan Fischer, Erik Maehle und Rüdiger Reischuk. LNI 154. GI, 2009, S. 1703–1716.

Hansen, Marit. „Top 10 Mistakes in System Design from a Privacy Perspective and Privacy Protection Goals“. In: *Privacy and Identity Management for Life. 7th IFIP WG 9.2, 9.6/11.7, 11.4, 11.6/PrimeLife International Summer School*. Hrsg. von Jan Camenisch, Bruno Crispo, Simone Fischer-Hübner, Ronald Leenes und Giovanni Russello. IFIP AICT 375. Berlin, Heidelberg: Springer, 2011, S. 14–31.

Hansen, Marit, Meiko Jensen und Martin Rost. „Protection Goals for Privacy Engineering“. In: *Proceedings of the IEEE Symposium on Security and Privacy Workshops (SPW)*. IEEE, 2015, S. 159–166.

Harvan, Matús und Alexander Pretschner. „State-based Usage Control Enforcement with Data Flow Tracking Using System Call Interposition“. In: *Proceedings of the 3rd International Conference on Network and System Security (NSS)*. IEEE, 2009, S. 373–380.

Hedbom, Hans. „A Survey on Transparency Tools for Enhancing Privacy“. In: *The Future of Identity in the Information Society. 4th IFIP WG 9.2, 9.6/11.6, 11.7/FIDIS International Summer School*. Hrsg. von Vashek Matyáš, Simone Fischer-Hübner, Daniel Cvrcek und Petr Švenda. IFIP AICT 298. Berlin, Heidelberg, New York: Springer, 2009, S. 67–82.

Heinemann, Oliver und Florian Wäßle. „Datenschutzrechtlicher Auskunftsanspruch bei Kreditscoring - Inhalt und Grenzen des Auskunftsanspruchs nach § 34 BDSG“. In: *Multimedia und Recht* (2010), S. 600–604.

Herdegen, Matthias. *Europarecht*. 18. Aufl. Grundrisse des Rechts. München: C.H. Beck, 2016.

Herkenhöner, Ralph, Hermann de Meer, Meiko Jensen und Henrich C. Pöhls. „Towards Automated Processing of the Right of Access in Inter-organizational Web Service Compositions“. In: *Proceedings of the 6th World Congress on Services (SERVICES)*. IEEE, 2010, S. 645–652.

Herrmann, Dominik und Jens Lindemann. „Obtaining Personal Data and Asking for Erasure: Do App Vendors and Website Owners Honour Your Privacy Rights?“ In: *Sicherheit 2016. Sicherheit, Schutz und Zuverlässigkeit*. Hrsg. von Michael Meier, Delphine Reinhardt und Steffen Wendzel. LNI 256. GI, 2016, S. 149–160.

Hevia, Alejandro und Daniele Micciancio. „An Indistinguishability-Based Characterization of Anonymous Channels“. In: *Privacy Enhancing Technologies. 8th International Symposium, PETS 2008*. Hrsg. von Nikita Borisov und Ian Goldberg. LNCS 5134. Berlin, Heidelberg: Springer, 2008, S. 187–201.

Hilty, Manuel, Alexander Pretschner, David Basin, Christian Schaefer und Thomas Walter. „A Policy Language for Distributed Usage Control“. In: *Computer Security – ESORICS 2007. Proceedings*. Hrsg. von Joachim Biskup und Javier López. LNCS 4734. Berlin, Heidelberg: Springer, 2007, S. 531–546.

Holland, David A., Margo Seltzer, Uri Braun und Kiran-Kumar Muniswamy-Reddy. „PASSing the Provenance Challenge“. In: *Concurrency and Computation: Practice and Experience* 20.5 (2008), S. 531–540.

Holtz, Leif-Erik, Katharina Nocun und Marit Hansen. „Towards Displaying Privacy Information with Icons“. In: *Privacy and Identity Management for Life. 6th IFIP WG 9.2, 9.6/11.7, 11.4, 11.6/PrimeLife International Summer School*. Hrsg. von Simone Fischer-Hübner, Penny Duquenoy, Marit Hansen, Ronald Leenes und Ge Zhang. IFIP AICT 352. Heidelberg, Dordrecht, London, New York: Springer, 2011, S. 338–348.

Hughes, Dominic und Vitaly Shmatikov. „Information Hiding, Anonymity and Privacy: A Modular Approach“. In: *Journal of Computer Security* 12.1 (2004), S. 3–36.

Iannella, Renato. „The Open Digital Rights Language : XML for Digital Rights Management“. In: *Information Security Technical Report* 9.3 (2004), S. 47–55.

Janic, Milena, Jan Pieter Wijbenga und Thijs Veugen. „Transparency Enhancing Tools (TETs): An Overview“. In: IEEE, 2013, S. 18–25.

Jureta, Ivan J., John Mylopoulos und Stéphane Faulkner. „Revisiting the Core Ontology and Problem in Requirements Engineering“. In: IEEE, 2008, S. 71–80.

Kahlert, Anna. „Rechtsgestaltung mit der Methode KORA“. In: *Datenschutz und Datensicherheit* 38.2 (2014), S. 86–92.

Kani-Zabihi, Elahe und Martin Helmhout. „Increasing Service Users' Privacy Awareness by Introducing On-Line Interactive Privacy Features“. In: *Information Security Technology for Applications. 16th Nordic Conference on Secure IT Systems, NordSec 2011*. Hrsg. von Peeter Laud. LNCS 7161. Berlin, Heidelberg: Springer, 2012, S. 131–148.

Kelbert, Florian und Alexander Pretschner. „Data Usage Control Enforcement in Distributed Systems“. In: *Proceedings of the 3rd ACM Conference on Data and Application Security and Privacy (CODASPY)*. New York, NY: ACM, 2013, S. 71–82.

Kelbert, Florian und Alexander Pretschner. „Decentralized Distributed Data Usage Control“. In: *Cryptology and Network Security. 13th International Conference, CANS 2014*. Hrsg. von Dimitris Gritzalis, Aggelos Kiayias und Ioannis Askoxylakis. Bd. 8813. LNCS. Berlin, Heidelberg: Springer, 2014, S. 353–369.

Kelbert, Florian und Alexander Pretschner. „A fully Decentralized Data Usage Control Enforcement Infrastructure“. In: *Applied Cryptography and Network Security. 13th International Conference, ACNS 2015, Revised Selected Papers*. Hrsg. von Tal Malkin, Vladimir Kolesnikov, Allison Bishop Lewko und Michalis Polychronakis. LNCS 9092. Berlin, Heidelberg: Springer, 2015, S. 409–430.

Kim, Hyung Chan, Angelos D. Keromytis, Michael Covington und Ravi Sahita. „Capturing Information Flow with Concatenated Dynamic Taint Analysis“. In: IEEE, 2009, S. 355–362.

Kiyavitskaya, Nadzeya, Alzbeta Krausová und Nicola Zannone. „Why Eliciting and Managing Legal Requirements Is Hard“. In: *Proceedings of the 1st International Workshop on Requirements Engineering and Law (RELAW)*. IEEE, 2008, S. 26–30.

Klas, Benedikt und Christine Möhrike-Sobolewski. „Digitaler Nachlass – Erbenschutz trotz Datenschutz“. In: *Neue Juristische Wochenschrift* (2015), S. 3473–3478.

Kloepfer, Michael und Holger Greve. „Das Informationsfreiheitsgesetz und der Schutz von Betriebs- und Geschäftsgeheimnissen“. In: *NVwZ* (2011), S. 577–584.

Kolter, Jan, Michael Netter und Günther Pernul. „Visualizing Past Personal Data Disclosures“. In: *Proceedings of the International Conference on Availability, Reliability and Security (ARES)*. IEEE, 2010, S. 131–139.

Kompetenzzentrum für angewandte Sicherheitstechnologie (KASTEL). *Begriffsdefinitionen in KASTEL*. Techn. Ber. Karlsruher Institut für Technologie (KIT), 2013. URL: http://www.kastel.kit.edu/downloads/Begriffsdefinitionen_in_KASTEL.pdf.

Kühling, Jürgen, Anastasios Sivridis, Mathis Schwuchow und Thorben Burghardt. „Das datenschutzrechtliche Vollzugsdefizit im Bereich der Telemedien – ein Schreckensbericht“. In: *Datenschutz und Datensicherheit* 33.6 (2014), S. 335–342.

Kumaraguru, Ponnurangam und Lorrie F. Cranor. *Privacy Indexes: A Survey of Westin's Studies*. Techn. Ber. CMU-ISRI-5-138. Institute for Software Research International, Carnegie Mellon University, 2005. URL: http://repository.cmu.edu/cgi/viewcontent.cgi?article=1857&context=isr.

Kumari, Prachi und Alexander Pretschner. „Model-Based Usage Control Policy Derivation“. In: *Engineering Secure Software and Systems. 5th International Symposium, ESSoS 2013*. Hrsg. von Jan Jürjens, Benjamin Livshits und Riccardo Scandariato. LNCS 7781. Berlin, Heidelberg: Springer, 2013, S. 58–74.

Kumari, Prachi, Alexander Pretschner, Jonas Peschla und Jens-Michael Kuhn. „Distributed Data Usage Control for Web Applications: A Social Network Implementation“. In: *Proceedings of the 1st ACM Conference on Data and Application Security and Privacy (CODASPY)*. New York, NY: ACM, 2011, S. 85–96.

Lakshmanan, Geetika T., Francisco Curbera, Juliana Freire und Amit Sheth. „Guest Editors' Introduction: Provenance in Web Applications“. In: *IEEE Internet Computing* 15.1 (2011), S. 17–21.

Lamsweerde, Axel van. „Requirements Engineering in the Year 00: A Research Perspective“. In: *Proceedings of the 22nd International Conference on Software Engineering*. New York, NY: ACM, 2000, S. 5–19.

Langheinrich, Marc. „Privacy by Design - Principles of Privacy-Aware Ubiquitous Systems“. In: *Ubicomp 2001: Ubiquitous Computing. International Conference, Proceedings*. Hrsg. von Gregory D. Abowd, Barry Brumitt und Steven Shafer. LNCS 2201. Berlin, Heidelberg: Springer, 2001, S. 273–291.

Laue, Philip, Judith Nink und Sascha Kremer, Hrsg. *Das neue Datenschutzrecht in der betrieblichen Praxis*. 1. Aufl. Baden-Baden: Nomos, 2016.

Laugwitz, Bettina, Theo Held und Martin Schrepp. „Construction and Evaluation of a User Experience Questionnaire“. In: *HCI and Usability for Education and Work. 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008*. Hrsg. von Andreas Holzinger. LNCS 5298. Berlin, Heidelberg: Springer, 2008, S. 63–76.

Lazouski, Aliaksandr, Gaetano Mancini, Fabio Martinelli und Paolo Mori. „Architecture, Workflows, and Prototype for Stateful Data Usage Control in Cloud“. In: *Proceedings of the IEEE Symposium on Security and Privacy Workshops (SPW)*. IEEE, 2014, S. 23–30.

Lazouski, Aliaksandr, Fabio Martinelli und Paolo Mori. „Usage Control in Computer Security: A Survey“. In: *Computer Science Review* 4.2 (2010), S. 81–99.

Lazouski, Aliaksandr, Fabio Martinelli und Paolo Mori. „A Prototype for Enforcing Usage Control Policies Based on XACML“. In: *Trust, Privacy and Security in Digital Business. 9th International Conference, TrustBus 2012*. Hrsg. von Simone Fischer-Hübner, Sokratis Katsikas und Gerald Quirchmayr. LNCS 7449. Berlin, Heidelberg: Springer, 2012, S. 79–92.

Leucker, Franziska. „Die zehn Märchen der Datenschutzreform“. In: *Privacy in Germany* 3.5 (2015), S. 195–202.

Lewis, James R. und Jeff Sauro. „The Factor Structure of the System Usability Scale“. In: *Human Centered Design. First International Conference, HCD 2009*. Hrsg. von Masaaki Kurosu. LNCS 5619. Berlin, Heidelberg: Springer, 2009, S. 94–103.

Lörscher, Michael. „Usage Control for the Thunderbird Mail Client“. Masterarbeit. Technische Universität Kaiserslautern, 2012.

Lovat, Enrico. „Cross-layer Data-centric Usage Control“. Dissertation. Technische Universität München (TUM), 2015.

Lovat, Enrico und Florian Kelbert. „Structure Matters - A New Approach for Data Flow Tracking“. In: *Proceedings of the IEEE Symposium on Security and Privacy Workshops (SPW)*. IEEE, 2014, S. 39–43.

Lovat, Enrico, Johan Oudinet und Alexander Pretschner. „On Quantitative Dynamic Data Flow Tracking“. In: *Proceedings of the 4th ACM Conference on Data and Application Security and Privacy (CODASPY)*. New York, NY: ACM, 2014, S. 211–222.

Maiden, Neil. „User Requirements and System Requirements“. In: *IEEE Software* 25.2 (2008), S. 90–91.

Malinka, Björn, Erin Donaldson, Karsten Neumann, Amelie Sophia Schleip, Sophie Tchanyou, Dominik Wied und Dominik Zier. *Datenschutzpraxis in Unternehmen 2015*. Studie. 2B Advice GmbH, 2015. URL: https://www.2b-advice.com/GmbH-en/DS-Praxis-2015.pdf.

Masing, Johannes. „Herausforderungen des Datenschutzes“. In: *Neue Juristische Wochenschrift* (2012), S. 2305–2312.

Maunz, Theodor und Günter Düring, Begr. *Grundgesetz. Kommentar*. Hrsg. von Roman Herzog, Matthias Herdegen, Rupert Scholz und Hans H. Klein. Bd. 1. München: C. H. Beck, 2016.

Merten, Detlev und Hans-Jürgen Papier, Hrsg. *Handbuch der Grundrechte in Deutschland und Europa. Band IV Grundrechte in Deutschland: Einzelgrundrechte I*. 1. Aufl. Heidelberg: C.F. Müller, 2011.

Miles, Simon, Paul Groth, Miguel Branco und Luc Moreau. *The Requirements of Recording and Using Provenance in E-Science Experiments*. Techn. Ber. School of Electronics und Computer Science, University of Southampton, 2005. URL: http://eprints.ecs.soton.ac.uk/11189.

Moreau, Luc, Ben Clifford, Juliana Freire, Joe Futrelle, Yolanda Gil, Paul Groth, Natalia Kwasnikowska, Simon Miles, Paolo Missier, Jim Myers, Beth Plale, Yogesh Simmhan, Eric Stephan und Jan Van den Bussche. „The Open Provenance Model Core Specification (v1.1)“. In: *Future Generation Computer Systems* 27.6 (2011), S. 743–756.

Moreau, Luc, Paul Groth, Simon Miles, Javier Vazquez-Salceda, John Ibbotson, Sheng Jiang, Steve Munroe, Omer Rana, Andreas Schreiber, Victor Tan und Laszlo Varga. „The Provenance of Electronic Data“. In: *Communications of the ACM* 51.4 (2008), S. 52–58.

Muniswamy-Reddy, Kiran-Kumar, Uri Braun, David A. Holland, Peter Macko, Diana Maclean, Daniel Margo, Margo Seltzer und Robin Smogor. „Layering in Provenance Systems“. In: *Proceedings of the 2009 USENIX Annual Technical Conference (USENIX)*. Berkeley, CA: USENIX, 2009.

Ni, Qun, Elisa Bertino und Ravi Sandhu. „A Characterization of the Problem of Secure Provenance Management“. In: IEEE, 2009, S. 310–314.

Otto, Paul N. und Annie I. Anton. „Addressing Legal Requirements in Requirements Engineering“. In: *Proceedings of the 15th IEEE International Requirements Engineering Conference*. IEEE, 2007, S. 5–14.

Paal, Boris P. und Daniel A. Pauly, Hrsg. *Datenschutz-Grundverordnung*. 1. Aufl. München: C.H. Beck, 2017.

Pahlen-Brandt, Ingrid. „Datenschutz braucht scharfe Instrumente Beitrag zur Diskussion um „personenbezogene Daten““. In: *Datenschutz und Datensicherheit* 32.1 (2008), S. 34–40.

Park, Jaehong und Ravi Sandhu. „Towards Usage Control Models: Beyond Traditional Access Control“. In: *Proceedings of the 7th ACM Symposium on Access Control Models and Technologies (SACMAT)*. ACM, 2002, S. 57–64.

Park, Jaehong und Ravi Sandhu. „The UCON ABC Usage Control Model“. In: *ACM Transactions on Information and System Security* 7.1 (2004), S. 128–174.

Parnas, David L. und Paul C. Clements. „A Rational Design Process: How and Why to Fake It“. In: *IEEE Transactions on Software Engineering* 12.2 (1986), S. 251–257.

Pashalidis, Andreas. „Measuring the Effectiveness and the Fairness of Relation Hiding Systems“. In: *Proceedings of the Asia-Pacific Services Computing Conference (APSCC)*. IEEE, 2008, S. 1387–1394.

Pfitzmann, Andreas und Marit Hansen. *A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management*. v0.34. 2010. URL: http://dud.inf.tu-dresden.de/literatur/Anon%5C_Terminology%5C_v0.34.pdf.

Pfitzmann, Andreas und Marit Köhntopp. „Anonymity, Unobservability, and Pseudonymity – A Proposal for Terminology“. In: *Designing Privacy Enhancing Technologies. International Workshop on Design Issues in Anonymity and Unobservability*. Hrsg. von Hannes Federrath. LNCS 2009. Berlin, Heidelberg: Springer, 2001, S. 1–9.

Plath, Kai-Uwe, Hrsg. *BDSG/DSGVO. Kommentar zum BDSG und zur DSGVO sowie den Datenschutzbestimmungen des TMG und TKG*. 2. Aufl. Köln: Dr. Otto Schmidt, 2016.

Preneel, Bart. „Post-Snowden Threat Models“. In: *Proceedings of the 20th ACM Symposium on Access Control Models and Technologies (SACMAT)*. New York, NY: ACM, 2015, S. 1.

Pretschner, Alexander, Matthias Büchler, Matús Harvan, Christian Schaefer und Thomas Walter. „Usage Control Enforcement with Data Flow Tracking for X11“. In: *Proceedings of the 5th International Workshop on Security and Trust Management (STM)*. 2009, S. 124–137.

Pretschner, Alexander, Manuel Hilty und David Basin. „Distributed Usage Control“. In: *Communications of the ACM* 49.9 (2006), S. 39–44.

Pretschner, Alexander, Manuel Hilty, David Basin, Christian Schaefer und Thomas Walter. „Mechanisms for Usage Control“. In: *Proceedings of the 2008 ACM Symposium on Information, Computer and Communications Security (ASIACCS)*. New York, NY: ACM, 2008, S. 240–244.

Pretschner, Alexander, Manuel Hilty, Florian Schutz, Christian Schaefer und Thomas Walter. „Usage Control Enforcement. Present and Future“. In: *IEEE Security & Privacy* 6.4 (2008), S. 44–53.

Pretschner, Alexander, Enrico Lovat und Matthias Büchler. „Representation-Independent Data Usage Control“. In: *Data Privacy Management and Autonomous Spontaneus Security. 6th International Workshop, DPM 2011, and 4th International Workshop, SETOP 2011*. Hrsg. von Nora Cuppens-Boulahia Joaquin Garcia-Alfaro Guillermo Navarro-Arribas und Sabrina de Capitani di Vimercati. LNCS 7122. Berlin, Heidelberg: Springer, 2011, S. 122–140.

Probst, Thomas. „Generische Schutzmaßnahmen für Datenschutz-Schutzziele“. In: *Datenschutz und Datensicherheit* 36.6 (2012), S. 439–444.

Prütting, Hans, Gerhard Wegen und Gerd Weinreich, Hrsg. *BGB. Kommentar*. 10. Aufl. Köln: Wolters Kluwer Luchterhand, 2015.

Pulls, Tobias und Roel Peeters. „Balloon: A Forward-Secure Append-Only Persistent Authenticated Data Structure“. In: *Computer Security – ESORICS 2015. Proceedings, Part II*. Hrsg. von Pernul, Günther and Ryan, Peter Y. A. and Weippl, Edgar. LNCS 9327. Berlin, Heidelberg: Springer, 2015, S. 622–641.

Pulls, Tobias, Roel Peeters und Karel Wouters. „Distributed Privacy-Preserving Transparency Logging“. In: *Proceedings of the 12th ACM Workshop on Privacy in the Electronic Society (WPES)*. New York, NY: ACM, 2013, S. 83–94.

Ralph, Paul. „The Illusion of Requirements in Software Development“. In: *Requirements Engineering* 18.3 (2013), S. 293–296.

Rasthofer, Siegfried, Steven Arzt, Enrico Lovat und Eric Bodden. „DroidForce: Enforcing Complex, Data-centric, System-wide Policies in Android“. In: *Proceedings of the 9th International Conference on Availability, Reliability and Security (ARES)*. IEEE, 2014, S. 40–49.

Rauschenberger, Maria, Jörg Thomaschewski und Martin Schrepp. „User Experience mit Fragebögen messen. Durchführung und Auswertung am Beispiel des UEQ“. In: *Usability Professionals 2013*. Hrsg. von Henning Brau, Andreas Lehmann, Kostanija Petrovic und Matthias C. Schroeder. German UPA, 2013, S. 72–76.

Robrecht, Bettina, Hrsg. *EU-Datenschutzgrundverordnung: Transparenzgewinn oder Information-Onverkill*. Bd. 7. Beiträge zum Informationsrecht. Edewecht: Oldenburger Verlag für Wirtschaft, Informatik und Recht, 2015.

Roßnagel, Alexander. *Handbuch Datenschutzrecht*. München: C.H. Beck, 2003.

Roßnagel, Alexander, Hrsg. *Europäische Datenschutz-Grundverordnung. Vorrang des Unionsrechts – Anwendbarkeit des nationalen Rechts*. 1. Aufl. Baden-Baden: Nomos, 2017.

Roßnagel, Alexander, Andreas Pfitzmann und Hansjürgen Garstka. *Modernisierung des Datenschutzrechts*. Gutachten. Bundesministerium des Inneren, 2001.

Rost, Martin und Kirsten Bock. „Privacy By Design und die Neuen Schutzziele“. In: *Datenschutz und Datensicherheit* 35.1 (2011), S. 30–35.

Rost, Martin und Andreas Pfitzmann. „Datenschutz-Schutzziele – revisited“. In: *Datenschutz und Datensicherheit* 33.6 (2009), S. 353–358.

Russell, Stuart J. und Peter Norvig. *Artificial Intelligence. A Modern Approach*. 3. Aufl. Upper Saddle River, NJ: Prentice Hall, 2010.

Schätzle, Daniel. „Ein Recht auf die Fahrzeugdaten. Das Recht auf Datenportabilität aus der DS-GVO“. In: *Privacy in Germany* 4.2 (2016), S. 71–75.

Scheiwe, Kirsten. „Informationsrechte von Patienten hinsichtlich der medizinischen und psychiatrischen Dokumentation. Eine Diskussion der Grenzen des vertraglichen Einsichtsrecht nach der BGH-Rechtsprechung im Verhältnis zu datenschutzrechtlichen Auskunftsansprüchen“. In: *Kritische Vierteljahresschrift für Gesetzgebung und Rechtswissenschaft* 81.3 (1998), S. 313–335.

Schulz, Thomas, Hendrik Skistims, Julia Zirfas, Diana Comes und Christoph Evers. „Vorschläge zur rechtskonformen Gestaltung selbst-adaptiver Anwendungen“. In: *Informatik 2011. Informatik schafft Communities*. Hrsg. von Hans-Ulrich Heiß, Peter Pepper, Holger Schlingloff und Jörg Schneider. LNI 192. GI, 2011, S. 182–182. URL: http://www.user.tu-berlin.de/komm/CD/paper/040515.pdf.

Schulze, Reiner, Manfred Zuleeg und Stefan Kadelbach, Hrsg. *Europarecht. Handbuch für die deutsche Rechtspraxis*. 3. Aufl. Baden-Baden: Nomos, 2015.

Schulzki-Haddouti, Christiane. „Des Kaisers neue Kleider – Wie sieht eine angemessene Datenschutzkontrolle aus?“ In: *Zukunft der informationellen Selbstbestimmung*. Hrsg. von Stiftung Datenschutz. Bd. 1. DatenDebatten, Schriftenreihe der Stiftung Datenschutz. Berlin: Erich Schmidt Verlag, 2016, S. 111–126.

Schwenke, Matthias. *Individualisierung und Datenschutz. Rechtskonformer Umgang mit personenbezogenen Daten im Kontext der Individualisierung*. Wiesbaden: Deutscher Universitäts-Verlag, 2006.

Seemann, Michael. „Informationelle und andere Selbstbestimmungen – Wie das Internet unsere Freiheiten umsortiert“. In: *Zukunft der informationellen Selbstbestimmung*. Hrsg. von Stiftung Datenschutz. Bd. 1. DatenDebatten, Schriftenreihe der Stiftung Datenschutz. Berlin: Erich Schmidt Verlag, 2016, S. 127–135.

Serjantov, Andrei und George Danezis. „Towards an Information Theoretic Metric for Anonymity“. In: *Privacy Enhancing Technologies. Second International Workshop, PET 2002*. Hrsg. von Roger Dingledine und Paul Syverson. LNCS 2482. Berlin, Heidelberg: Springer, 2003, S. 41–53.

Simitis, Spiros, Hrsg. *Bundesdatenschutzgesetz*. 7. Aufl. Baden-Baden: Nomos, 2011.

Simitis, Spiros, Hrsg. *Bundesdatenschutzgesetz*. 8. Aufl. Baden-Baden: Nomos, 2014.

Simmhan, Yogesh L., Beth Plale und Dennis Gannon. „A Survey of Data Provenance in e-Science“. In: *ACM SIGMOD Record* 34.3 (2005), S. 31–36.

Solmecke, Christian, Thomas Köbrich und Robin Schmitt. „Der digitale Nachlass - haben Erben einen Auskunftsanspruch? Überblick über den rechtssicheren Umgang mit den Daten von Verstorbenen“. In: *Multimedia und Recht* (2015), S. 291–295.

Spiecker genannt Döhmann, Indra. „Protection of Confidential Business Data in the Age of Convergence“. In: *Communications Regulation in the Age of Digital Convergence. Legal and Economic Perspectives*. Hrsg. von Jan Krämer und Stefan Seifert. Karlsruhe: Universitätsverlag Karlsruhe, 2009, S. 29–42.

Spiecker genannt Döhmann, Indra. „Datenschutzrecht im Internet in der Kollision“. In: *Zukunft der informationellen Selbstbestimmung*. Hrsg. von Stiftung Datenschutz. Bd. 1. DatenDebatten, Schriftenreihe der Stiftung Datenschutz. Berlin: Erich Schmidt Verlag, 2016, S. 137–149.

Spiekermann, Sarah. „The Desire for Privacy: Insights into the Views and Nature of the Early Adopters of Privacy Services“. In: *International Journal of Technology and Human Interaction* 1.1 (2005), S. 74–83.

Spiekermann, Sarah und Lorrie Faith Cranor. „Engineering Privacy“. In: *IEEE Transactions on Software Engineering* 35.1 (Aug. 2009), S. 67–82.

Statistisches Bundesamt. *Statistisches Jahrbuch 2015. Deutschland und Internationales*. Studie. Wiesbaden: Statistisches Bundesamt, 2015. URL: https://www.destatis.de/DE/Publikationen/StatistischesJahrbuch/StatistischesJahrbuch2015.pdf?__blob=publicationFile.

Steinbrecher, Sandra und Stefan Köpsell. „Modelling Unlinkability“. In: *Privacy Enhancing Technologies. Third International Workshop, PET 2003*. Hrsg. von Roger Dingledine. LNCS 2760. Berlin, Heidelberg: Springer, 2003, S. 32–47.

Streinz, Rudolf, Hrsg. *EUV/AEUV. Vertrag über die Europäische Union und Vertrag über die Arbeitsweise der Europäischen Union*. 2. Aufl. Bd. 57. Beck'sche Kurz-Kommentare. München: C.H. Beck, 2012.

Sultana, Salmin und Elisa Bertino. „A File Provenance System“. In: *Proceedings of the 3rd ACM Conference on Data and Application Security and Privacy (CODASPY)*. New York, NY: ACM, 2013, S. 153–156.

Sydow, Gernot. „Vorwirkungen von Ansprüchen auf datenschutzrechtliche Auskunft und Informationszugang“. In: *Neue Zeitschrift für Verwaltungsrecht* (2013), S. 467–471.

Taeger, Jürgen und Detlev Gabel, Hrsg. *Kommentar zum BDSG. und zu den Datenschutzvorschriften des TKG und TMG*. 2. Aufl. Frankfurt am Main: Fachmedien Recht und Wirtschaft in Deutscher Fachverlag, 2013.

Tangens, Rena. „Tausche Bürgerrechte gegen Linsengericht. Die Wir-Wollen-Alles-Über-Sie-Wissensgesellschaft". In: *Wissen als Begleiter!?* Hrsg. von Rita Herwig, Jens Uhlig und Johannes Küstner. Bd. 4. diagonal denken. Berlin: LIT Verlag, 2008, S. 139–152.

Tariq, Dawood, Maisem Ali und Ashish Gehani. „Towards Automated Collection of Application-Level Data Provenance". In: *Proceedings of the 4th USENIX Conference on Theory and Practice of Provenance (TaPP)*. USENIX, 2012, S. 16–16.

Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein (ULD) und Technische Universität Dresden, Professur Datenschutz und Datensicherheit. *Verkettung digitaler Identitäten*. Report. Bundesministerium für Bildung und Forschung, 2007.

Vicknair, Chad, Michael Macias, Zhendong Zhao, Xiaofei Nan, Yixin Chen und Dawn Wilkins. „A Comparison of a Graph Database and a Relational Database: A Data Provenance Perspective". In: *Proceedings of the 48th Annual Southeast Regional Conference (ACM SE)*. New York, NY: ACM, 2010, 42:1–42:6.

Wächter, Michael. „Datenschutz als „Software-Routine" - Ein datenschutzrechtlicher Implementierungsvorschlag". In: *Datenschutz und Datensicherheit* 20.5 (1996), S. 272–278.

Wästlund, Erik, Simone Fischer-Hübner, Staffan Gustafsson, Peter Wolkerstorfer, Cornelia Graf, Tobias Pulls, Hans Hedbom und Marit Hansen. *D4.2.2: End User Transparency Tools: UI Prototypes*. Project Deliverable. Privacy und Identity Management in Europe for Life (PrimeLife), 2010. URL: http://primelife.ercim.eu/images/stories/deliverables/d4.2.2-transparency_tools_ui_prototypes-public.pdf.

Wästlund, Erik, Peter Wolkerstorfer und Christina Köffel. „PET-USES: Privacy-Enhancing Technology – Users' Self-Estimation Scale". In: *Privacy and Identity Management for Life. 5th IFIP WG 9.2, 9.6/11.4, 11.6, 11.7/PrimeLife International Summer School*. Hrsg. von Michele Bezzi, Penny Duquenoy, Simone Fischer-Hübner, Marit Hansen und Ge Zhang. IFIP AICT 320. Berlin, Heidelberg: Springer, 2010, S. 266–274.

Weichert, Thilo. „Auskunftsanspruch in verteilten Systemen". In: *Datenschutz und Datensicherheit* 30.11 (2006), S. 694–699.

Weichert, Thilo. „Der Datenschutzanspruch auf Negativauskunft". In: *Neue Zeitschrift für Verwaltungsrecht* (2007), S. 1004–1007.

Weichert, Thilo. „Der Personenbezug von Geodaten". In: *Datenschutz und Datensicherheit* 31.2 (2007), S. 113–119.

Weichert, Thilo. „Datenschutz bei Internetveröffentlichungen“. In: *Verbraucher und Recht* 24.9 (2009), S. 323–330.

Woodruff, Allison, Vasyl Pihur, Sunny Consolvo, Lauren Schmidt, Laura Brandimarte und Alessandro Acquisti. „Would a Privacy Fundamentalist Sell Their DNA for $1000... If Nothing Bad Happened as a Result? The Westin Categories, Behavioral Intentions, and Consequences“. In: *Proceedings of the 10th Symposium On Usable Privacy and Security*. USENIX, 2014, S. 1–19.

Wüchner, Tobias und Alexander Pretschner. „Data Loss Prevention Based on Data-Driven Usage Control“. In: *Proceedings of the IEEE 23rd International Symposium on Software Reliability Engineering (ISSRE)*. IEEE, 2012, S. 151–160.

Zhang, Qing, John Mccullough, Justin Ma, Nabil Schear, Michael Vrable, Amin Vahdat, Alex C. Snoeren, Geoffrey M. Voelker und Stefan Savage. „Neon: System Support for Derived Data Management“. In: *Proceedings of the 6th ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments (VEE)*. New York, NY: ACM, 2010, S. 63–74.

Zhang, Xinwen, Jaehong Park, Francesco Parisi-Presicce und Ravi Sandhu. „A logical Specification for Usage Control“. In: *Proceedings of the 9th ACM Symposium on Access Control Models and Technologies (SACMAT)*. New York, NY: ACM, 2004.

Zimmermann, Christian, Rafael Accorsi und Günter Müller. „Privacy Dashboards: Reconciling Data-Driven Business Models and Privacy“. In: *Proceedings of the 9th International Conference on Availability, Reliability and Security (ARES)*. IEEE, 2014, S. 152–157.

Zurawski, Nils. *D5: Exercising Democratic Rights Under Surveillance Regimes. Germany Country Reports*. Project Deliverable. Increasing Resilience in Surveillance Societies (IRISS), 2014. URL: http://irissproject.eu/wp-content/uploads/2014/06/Germany-Composite-Reports-Final1.pdf.