Kausalitäten, Korrelationen und Datenschutzrecht

Jörg Pohle*

3. Februar 2013

Abstract. Korrelationsbasierte Verfahren haben in den letzten Jahren einen starken Aufschwung genommen. Sie sind aber aus der Sicht des Datenschutzes besonders problematisch, da sie nicht den impliziten Annahmen über Verfahren und Verfahrensgestaltung entsprechen, die seit den 70er Jahren die spezifische Ausprägung der Phasenorientierung und die Ausgestaltung des Erforderlichkeitsbegriffes bestimmten. Neben einer Stärkung des Prinzips der Datenvermeidung und Datensparsamkeit als Ersatz für den zahnlos gewordenen Erforderlichkeitsbegriff in diesem Bereich bedarf es vor allem einer Reformulierung der rechtlichen Anforderungen unter Verwendung von Datenschutz-Schutzzielen. Außerdem sind besondere Anforderungen an die Einwilligung der Betroffenen, an die Weitergabe und die Nutzung der dabei erzeugten Daten sowie an die Datensicherheit zu stellen, wenn korrelationsbasierte Verfahren zum Einsatz kommen.

Inhaltsverzeichnis

1. Einleitung

In den letzten Jahren hat die Verwendung korrelationsgestützter Algorithmen bei der Verarbeitung personenbezogener Daten stark zugenommen. Dabei wird mit statistischen Methoden in vorhandenen Daten nach Verbindungen zwischen den Daten gesucht, die dann zur

*Humboldt-Universität zu Berlin, Institut für Informatik, Informatik in Bildung und Gesellschaft, pohle@informatik.hu-berlin.de.

Grundlage von Entscheidungen gemacht werden. Einer der Hauptgründe dafür ist die gesteigerte Benutzerfreundlichkeit von darauf basierenden Auswertungsprogrammen und ihre vermehrte Einbindung in Standardprogramme. Gemessen an ihrer steigenden praktischen Relevanz ist das Fehlen jeder eingehenden Auseinandersetzung mit ihren gesellschaftlichen Folgen im Allgemeinen und ihren datenschutzrechtlichen Folgen im Besonderen sowohl erstaunlich als auch unverzeihlich.[1]

Im Folgenden soll der Versuch unternommen werden, den Einsatz korrelationsbasierter Algorithmen datenschutzrechtlich einzuordnen. Zuerst wird ein kurzer Überblick über Informationsverarbeitung und die Gestaltung von Informationsverarbeitungprozessen und deren datenschutzrechtliche Einordnung gegeben. Anschließend wird der Übergang von kausalitätsbasierten zu korrelationsbasierten Ansätzen bei der Informationsverarbeitung mit seinen Gründen und Folgen dargestellt. Aufbauend auf diesen beiden Teilen können dann die Folgen des Einsatzes korrelationsbasierter Verfahren aus datenschutzrechtlicher Sicht dargestellt werden. Es wird erläutert, warum die derzeitigen datenschutzrechtlichen Regelungen in ihrer konkreten Ausprägung bei der rechtlichen Einhegung von korrelationsgestützten Verfahren an ihre Grenzen stoßen. Abschließend wird ein breit gefächerter Lösungsansatz skizziert.

Der Text richtet sich in erster Linie an Rechtswissenschaftlerinnen und Rechtswissenschaftler. Informatische Vorbildung wird nicht vorausgesetzt. Die damit notwendig einhergehende Verkürzung nimmt der Autor in Kauf und versucht sie zu minimieren.

2. Informationsverarbeitung und Datenschutzrecht

Das Bundesdatenschutzgesetz – und damit alle Gesetze, die seine Grundarchitektur[2] übernommen haben – basiert seit Anbeginn auf zwei „Säulen“: einer individualistisch geprägten und auf die Grundrechte zurückgehenden Menge von Betroffenenrechten und einer systemtheoretisch fundierten, strukturalistisch ausgerichteten und sich aus dem Staatsorganisationsrecht speisenden Regelung von Informationsverarbeitungsprozessen und deren Gestaltung.[3] Aus Sicht der Informatik ist diese Trennung grundsätzlich sinnvoller und hilfreicher als eine insbesondere von den meisten Zivilrechtlern geforderte Trennung zwischen öffentlichem und nicht-öffentlichem Bereich:[4] Sowohl das Vorgehen bei der Organisations- und Technikgestaltung als auch die Technik und ihr Einsatz sind im öffentlichen Bereich die gleichen wie im nicht-öffentlichen. Der im Datenschutzrecht als Verfahren bezeichnete Informationsverarbeitungsprozess findet seine Entsprechung in den Begriffen Geschäftsprozess und *use case*, die zentral für die Technikentwicklung sind. Das Recht stellt dabei – das ist die erste „Säule“ – rechtliche Anforderungen an den Ablauf der Informationsverarbeitung – und tritt damit als besondere Form eines Auftraggebers[5] auf –, die dann als Grundlage für die Technikgestaltung dienen. Die Betroffenenrechte auf der anderen Seite wirken direkt als Forderung nach Verfahren, nämlich gerade solchen, mit denen die verantwortlichen Stellen

[1] Das gilt sowohl für die Informatik als auch für die Rechtswissenschaft. Wenn korrelationsbasierte Algorithmen zum Thema gemacht werden, dann entweder sehr oberflächlich und fast in einem Nebensatz oder nur mit Verweis auf die Betroffenen und deren fehlendem Verständnis – siehe etwa Ochs und Löw 2012, S. 53 oder Birk, Reimer und Wegener 2010.

[2] Architektur wird hier in dem allgemeinen Sinn verstanden, wie er in der Informatik breit verwendet wird: ein konzeptuelles Modell, das ein System mit seinen Komponenten und deren Eigenschaften, seiner Struktur, seinem Verhalten und seinen möglichen Sichten beschreibt.

[3] Siehe umfassend dazu Steinmüller u. a. 1971, S. 60, der zwar erklärt, dass für nicht-öffentliche Stellen die Rechtmäßigkeit des Verwaltungshandelns, wie sie aus Art. 20 GG folgt, kein Maßstab sein kann, dann aber sowohl für den öffentlichen wie für den nicht-öffentlichen Bereich die gleiche Regelungsarchitektur verwendet.

[4] Zuzugeben ist aber, dass auch die Informatik – und hier insbesondere die Softwareentwickler – diesen Vorteil bisher nicht nutzen konnten.

[5] Auch bei der Softwareentwicklung werden die meisten Anforderungen als Anforderungen an einen Gebrauch der Software gestellt und nicht als Anforderungen an die Technik selbst. Der Hauptgrund dafür ist, dass zwar die meisten Auftraggeber formulieren können, was sie tun wollen, aber nur sehr wenige tatsächlich das notwendige technische Verständnis mitbringen, um Anforderungen als technische Anforderungen zu formulieren. Letzteres ist im Kern gerade auch die Aufgabe der Informatik: Analyse und (Re-)Organisation komplexer Arbeitsprozesse und ihre maschinelle Unterstützung, siehe Coy 1992, S. 18.

die Betroffenenrechte bedienen können.[6]

Zum besseren Verständnis wird im Folgenden zuerst der im Datenschutzrecht verwendete Informationsbegriff geklärt. Anschließend werden die Begriffe Zweck, Kausalität und Modell mit ihrer Verbindung zum Datenschutzrecht erläutert. Darauf aufbauend wird dargestellt, wie Informationsverarbeitungsprozesse aus Sicht von Organisationen und Datenschutz aussehen und wie sie unter Kontrolle gebracht werden sollen. Abschließend wird gezeigt, wie sich datenschutzrechtliche Anforderungen in den Prozess der Verfahrensgestaltung integrieren lassen.

2.1. Der Informationsbegriff im Datenschutzrecht

Der Begriff Datum, den das Datenschutzrecht verwendet,[7] ist ein Informationsbegriff. Entgegen anderer Annahme[8] handelt es sich weder um Gregory Batesons biokybernetischen noch um Claude Shannons technischen Informationsbegriff. Stattdessen liegt dem Datenschutzrecht der Informationsbegriff der Semiotik mit seinen vier Dimensionen Syntax, Semantik, Pragmatik und Sigmatik zugrunde.[9] Dabei wird mit Syntax die konkrete, meist zeichenmäßige Repräsentation, mit Semantik die Bedeutung und mithin der Kontext, mit Pragmatik der Zweck und mit Sigmatik der Verweis auf die betroffene Person bezeichnet und damit rechtlich regulierbar. Auch aus Informatiksicht ist das sinnvoll: Zwar können technische Systeme Zeichenfolgen ausschließlich syntaktisch verarbeiten, die Informatik hat allerdings inzwischen langjährige Erfahrung, die anderen Dimensionen datentechnisch, d. h. syntaktisch, unter Verwendung von Meta-Daten zu simulieren.[10]

2.2. Zwecke, Kausalitäten, Modelle

Im Datenschutzrecht werden Zwecke als gegeben vorausgesetzt, die von nicht-öffentlichen Stellen typischerweise selbst gesetzt werden, während sie öffentlichen Stellen meist gesetzlich vorgegeben sind.[11] Erst vor dem Hintergrund dieser gegebenen Zwecke werden die ihnen dienenden Informationsverarbeitungsprozesse datenschutzrechtlich betrachtet. Allein legitim müssen die Zwecke sein. Abstrakter Zweck der vom Datenschutzrecht betrachteten organisierten Informationsverarbeitung ist die Produktion von Entscheidungen.

Kausalitäten sind Ursache-Wirkung-Relationen zwischen Ereignissen oder Zuständen mit einer festen zeitlichen Richtung – auf die Ursache folgt die Wirkung. Es handelt sich nicht um eine zufällig auftretende Verbindung zwischen den Ereignissen oder Zuständen. Insoweit kann davon gesprochen werden, dass Kausalitäten inhaltliche Aussagen zulassen.[12]

Modelle sind Abbildungen von etwas für jemanden für einen Zweck.[13] Sie sind Mengen

[6] Dass Datenschutzrecht, Organisations- und Technikgestaltung auf so relativ einfache Weise zusammen betrachtet werden können und dürfen, liegt im Kern an der Geschichte der datenschutzrechtlichen Architekturgestaltung: Die drei Bereiche wurden damals über die Verwendung von kybernetischen und systemtheoretischen Modellen und Methoden integriert bearbeitet. Insoweit die Informatik – und dabei insbesondere die Softwareentwicklung – heute immer noch stark systemtheoretisch geprägt ist, kann sie ein systemtheoretisch beeinflusstes Datenschutzrecht mit seiner starken Prozessorientierung sinnvoll als Anknüpfungspunkt für die Technikentwicklung nutzen. Näheres dazu werde ich in meiner demnächst erscheinenden Dissertation ausführen.

[7] Das gilt nicht nur für die deutsche Sprache, siehe *data protection*, *protection des données*, *protección de datos* etc.

[8] Siehe etwa Albers 2005.

[9] Siehe grundlegend Steinmüller u. a. 1971, S. 42 f.

[10] Eine etwas ausführlichere Erklärung findet sich in Anhang A auf Seite 11.

[11] Im Datenschutzrecht wurde historisch versucht, die Begriffe „Zweck“ und „Aufgabe“ sauber voneinander zu trennen. Das verwaltungsrechtliche Verständnis dieser Begriffe wird dem nicht immer gerecht. Unsauber ließe sich sagen, dass die gesetzlichen Aufgaben einer Behörde dem datenschutzrechtlichen Zweck entsprechen, während deren Maßnahmen zur Erreichung dieses Zweck aus datenschutzrechtlicher Sicht als Aufgaben oder Verfahren beschrieben werden, siehe auch Podlech 1995, S. 20 f.

[12] Ein Beispiel dafür ist der Umgang mit mehrbändigen Werken. Menschen lesen Texte linear und kennen „die ganze Geschichte“ erst, wenn sie alle Bände gelesen haben. Wenn jemand also im Buchhandel den ersten Band bestellt, dann ist es durchaus sinnvoll, ihm auch den zweiten Band anzubieten. Wenn jemand hingegen nur den zweiten Band bestellt, dann kann nicht davon ausgegangen werden, dass er danach auch den ersten Band kaufen würde. Wahrscheinlicher ist es, dass er den zweiten Band kauft, **weil** er den ersten schon gelesen hat. Das ist dann eine inhaltliche Aussage.

[13] Umfassend dazu Podlech 1976.

von (mindestens angenommenen) Kausalitäten,[14] (möglichst) strukturäquivalent zum abgebildeten Urbild, in jedem Fall allerdings mit reduzierter Komplexität.[15] Sie werden von jemandem erstellt[16] und zwar für jemanden. Dieser ist es auch, der den Zweck vorgibt. Mit der Zweckvorgabe werden die Modelle beschränkt – tendenziell so weit, dass sie sich nicht oder nur sehr schwer für andere Zwecke einsetzen lassen. Neben der Zwecksetzung hat der Auftraggeber und Nutznießer des Modells auch weitgehend Kontrolle über den Prozess der Modellbildung selbst und kann dabei insbesondere steuern, welche Ereignisse oder Zustände entweder analysiert oder gerade von der Analyse ausgeschlossen werden, welche Prioritäten ihnen dabei jeweils zugeordnet werden etc.

2.3. Verfahren und Phasen der Informationsverarbeitung

Der Begriff Verfahren wird im Umfeld des Datenschutzrechts in zwei unterschiedlichen Bedeutungen verwendet: Einmal als Gesamtheit eines Informationsverarbeitungsprozesses vom Erheben bis zum Löschen der personenbezogenen Daten, das dem Ziel der Zweckerreichung dient,[17] und zum anderen als Art und Weise, in der einzelne Verarbeitungsschritte umgesetzt sind.[18] In seiner ersten Bedeutung entspricht ein Verfahren mit seinen Bestandteilen Informationen, Systeme und Prozesse[19] dem, was in Organisationslehre und Informatik – und dort vor allem in der Softwareentwicklung – als Geschäftsprozess oder *use case* bezeichnet wird.

Organisationen haben ein Interesse daran, ihre Verfahren kontrollierbar zu gestalten. Nur Kontrollierbarkeit ermöglicht den Nachweis von Verfahrenseigenschaften. Während öffentliche Stellen an die Rechtmäßigkeit des Verwaltungshandelns gebunden sind und diese dementsprechend nachweisen müssen, bezieht sich die Nachweispflicht in nicht-öffentlichen Stellen vor allem auf die Erreichung von Organisationszielen (etwa Produktionsmenge, Umsatz, Gewinn, Marge o. ä.) und die dafür aufgewendeten Mittel (dabei vor allem ihre Effizienz). An diese schon aus Organisationssicht notwendige Kontrollierbarkeit von Verfahren knüpfte in den 70er Jahren der Datenschutz an und formulierte eine datenschutzspezifische Nachweispflicht für Verfahren: Verantwortliche Stellen müssen die Datenschutzkonformität ihrer Informationsverarbeitungsverfahren gegenüber Betroffenen und Datenschutzaufsicht nachweisen können.[20]

Die Architekten des Datenschutzrechts gingen zu Beginn der 70er Jahre davon aus, dass alle Verfahren in einzelne Verfahrensschritte zerlegt werden könnten.[21] Ein jeder Verfahrensschritt entspricht dabei einer der Verarbeitungsphasen des Datenschutzrechts: Erheben, Verarbeiten (Speichern, Verändern, Übermitteln, Sperren, Löschen), Nutzen, Anonymisieren, Pseudonymisieren. An jede Phase werden dann datenschutzrechtliche Anforderungen gestellt.[22] Die Kontrollierbarkeit des Verfahrens insgesamt ist dabei gerade dann gegeben,

[14] Andere als Kausalitätsmodelle wie etwa solche zur Veranschaulichung (z. B. Architekturmodelle) oder zur physikalischen Simulation (z. B. Flugzeugmodelle im Windkanal) werden hier nicht betrachtet.

[15] Beispiele für solche Modelle gibt es in allen Bereichen: Schiffsmodelle sind gegenüber dem Original verkleinert und die meisten davon dienen vor allem der Veranschaulichung. Wahlumfragen basieren auf der Befragung ausgewählter Wahlbürger, die sich statistisch ähnlich wie das Wahlvolk insgesamt verhalten. Die Menge der ausgewählten Wahlbürger ist dann das Modell. Ein Verkehrsmodell bildet Verkehrsströme ab, ohne dass dabei einzelne Fahrzeuge betrachtet werden müssten.

[16] Die damit möglicherweise einhergehende strukturelle Beschränktheit jedes Modells hat Podlech damals übersehen, wahrscheinlich weil Auftraggeber und Auftragnehmer im Modellbildungsprozess in den 1970er Jahren noch weitgehend gleich waren. Sie wird sogar heute noch viel zu wenig reflektiert, vor allem in der Informatik, obwohl sie häufig in Modellbildungsprozesse involviert ist und dabei vor allem als Auftragnehmer wirkt.

[17] Siehe etwa §§ 4d Abs. 1, 4e BDSG.

[18] Siehe etwa §§ 3 Abs. 4, 28b sowie die Anlage zu § 9 Satz 1 BDSG.

[19] Siehe dazu Bock und Meissner 2012.

[20] Insoweit sind moderne Organisationen und Datenschutz natürliche Verbündete: Die Nachweispflicht zwingt die Organisation zur Selbstbeobachtung bezüglich ihrer eigenen Prozesse, die sich als Grundlage für deren Rationalisierung eignen und mithin Basis einer möglichen Maschinisierung sind, aus der die Organisation wiederum Nutzen durch Effizienzsteigerung ziehen kann.

[21] Grundlegend dazu Steinmüller u. a. 1971, S. 57 ff.

[22] Im Gegensatz dazu steht der Verarbeitungsbegriff in Art. 2b) EU-DSRL für „jeden mit oder ohne Hilfe automatisierter Verfahren ausgeführten Vorgang oder jede Vorgangsreihe im Zusammenhang mit personenbezogenen Daten“. Im Ergebnis können mit der Auftrennung in verschiedene Phasen, wie sie das BDSG vornimmt, inhaltlich angemessenere Regelungen formuliert werden als mit der EU-DSRL. Gleiches gilt für den derzeit diskutierten Entwurf der Datenschutz-Grundverordnung.

wenn jeder einzelne Schritt kontrollierbar ist. Das Ganze ist also – jedenfalls in den Vorstellungen des Datenschutzrechts – genau die Summe seiner Teile.

2.4. Verfahrensgestaltung und Datenschutzrecht

In Bezug auf das Vorgehen bei der Verfahrensgestaltung, an das sich das Datenschutzrecht mit seinen Anforderungen anbindet, müssen die Vorstellungen der 70er Jahre zugrunde gelegt werden.[23]

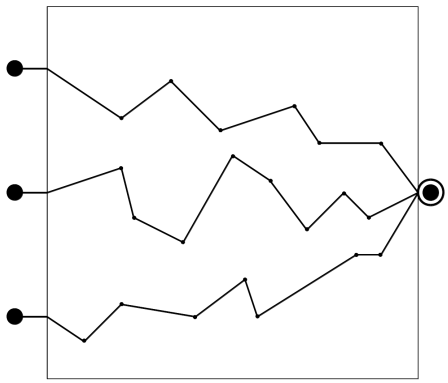

Abbildung 1: Modell und Zweck. In dem hier dargestellten Modell gibt es drei mögliche Verfahren, die mit unterschiedlich vielen Informationen (links als „Input“ des Modells) den gleichen Zweck („Output“ des Modells) erreichen können.

Gegeben sei ein legitimer Zweck, zu dessen Erreichen die Organisation Informationen verarbeiten muss. Ob sie dazu personenbezogene Informationen braucht, entscheidet sich auf der Basis der von der Organisation entwickelten Modelle. Die Modellbildung unterliegt damit datenschutzrechtlichen Anforderungen. In neuerer Zeit werden diese Anforderungen mit den Begriffen Datenvermeidung und Datensparsamkeit bezeichnet.[24] Historisch sollte die Modellbildung durch die Forderung nach Erforderlichkeit gesteuert werden.[25]

In den Modellen sind Mengen von Kausalitätsbeziehungen zwischen Ereignissen eingespeichert, mit deren Hilfe sich feststellen lässt, welche personenbezogenen Informationen zur Zweckerreichung dienen können. Die Kausalitäten müssen dabei nicht wahr sein. Es reicht, wenn diejenigen, die diese Modelle einsetzen wollen, hinreichend davon überzeugt sind, dass die Kausalitäten korrekt ermittelt wurden. Selbst für eine rationale Organisation ist es hinreichend, dass es sich um sinnvolle Annahmen handelt.[26] Nicht nur lassen sich für einen gegebenen Zweck häufig unterschiedliche Modelle erstellen, in einem gegebenen Modell gibt es meistens auch mehrere Wege, d. h. Ketten von Kausalitätsbeziehungen, die zur Zweckerreichung tauglich sind. Als Eingabe erwartet jeder dieser Wege eine Menge von Informationen, entweder mit oder ohne Personenbezug,[27] und produziert als Ausgabe gerade

[23] Dass solche Vorstellungen zwar dem Datenschutzrecht zugrunde gelegt wurden, gleichzeitig aber nicht immer, und nie vollständig, expliziert wurden, kritisiert schon Leib 1985. Daran hat sich bis heute leider auch nichts geändert.

[24] Leider werden diese Anforderungen in der Praxis nur an die Verfahren angelegt, nicht aber auch an die den Verfahren zugrunde liegenden Modelle.

[25] Siehe vor allem Podlechs Forderung nach einer Beschreibbarkeit der Erforderlichkeitsrelation, Podlech 1982, S. 455 f., wobei es sich gleichzeitig um eine frühe Form einer Schutzzieldefinition handelt. Im Gegensatz dazu steht die heute selbst von „radikalen“ Datenschützern vertretene Ansicht, das Erforderlichkeitsprinzip beziehe sich nicht nur auf einen gegebenen Zweck, sondern auch ein gegebenes technisches System und einen gegebenen Datenverarbeitungsprozess, siehe Scholz in Simitis 2011, § 3a Rn. 33.

[26] Für nicht-öffentliche Stellen ist offenkundig, dass deren Modelle nur *relativ* gut sein müssen: Sie müssen erstens besser sein als die aus dem letzten Jahr, dann verbessert sich das betriebswirtschaftliche Ergebnis. Und sie müssen zweitens besser sein als die der Konkurrenz, und die kocht auch nur mit Wasser.

[27] Das Gebot der Datenvermeidung verlangt Verfahren ohne Verwendung personenbezogener Informationen. Dafür winkt das Privileg der Nichtanwendbarkeit des Datenschutzrechts. Die mit dieser explizit eingeräumten Möglichkeit zur „Flucht aus dem Datenschutz“ verbundene Erwartung an den Durchbruch datenvermeidender Verfahren gilt inzwischen als gescheitert, siehe Ohm 2010.

die Zweckerreichung. Die Mengen der jeweils notwendigen Informationen sind unterschiedlich groß. Das Gebot der Datensparsamkeit fordert dann die Wahl desjenigen Weges, der die kleinste Datenmenge zur Entscheidungsproduktion erfordert.

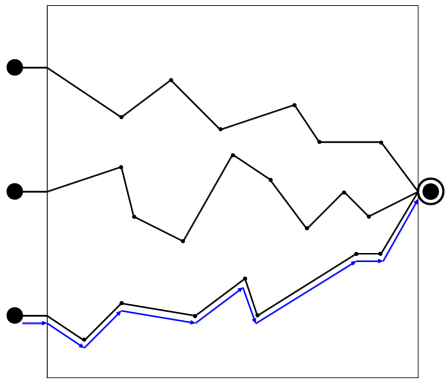

Abbildung 2: Modell, Zweck und Verfahren.

Der Weg wird in Schritten – oder Operationen oder Arbeitsaufgaben – abgelaufen bzw. abgearbeitet.[28] Für die einzelnen Schritte kann daraufhin entschieden werden, ob und wie sie automatisiert oder generell technisch unterstützt werden können. Für die Technikgestaltung werden dazu inzwischen fast durchgängig graphische Werkzeuge und Modellierungssprachen wie UML eingesetzt. Deren Vorteil besteht vor allem darin, dass die damit erstellten Abbilder der Informationsverarbeitungsprozesse zumindest für Domänenexpertinnen und -experten sowie Datenschutzbeauftragte vergleichsweise leicht verständlich sind, selbst wenn sie sie nicht selber erstellen können. Die einzelnen Verarbeitungsschritte werden dann so zusammengefasst, dass die dabei entstehenden Gruppen jeweils einer Phase im Sinne des Datenschutzrechts entsprechen, d. h. alle Schritte, die sich mit dem Erheben von personenbezogenen Informationen beschäftigen, werden der Phase „Erheben“ zugeordnet usw. Am Ende muss das gesamte Verfahren, zumindest wo es auf die Verarbeitung personenbezogener Informationen setzt, vom datenschutzrechtlichen Phasenmodell überdeckt werden.

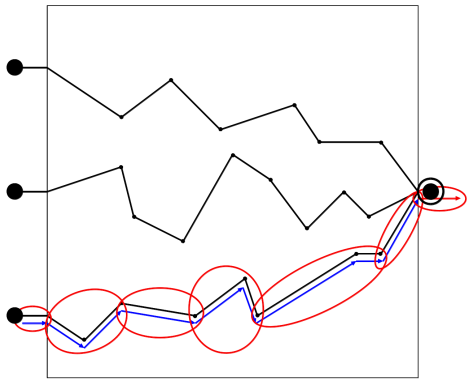

Abbildung 3: Modell, Zweck, Verfahren und Phasen. Jede Phase kann dabei durchaus mehr als einmal im Laufe eines Verfahrens vorhanden sein.

Unter Verwendung der datenschutzrechtlichen Anforderungen an die einzelnen Verarbei-

[28] Der jeweils gewählte Bezeichner stellt einen oder mehrere Aspekte besonders in den Vordergrund, während er andere eher verdeckt: „Schritt“ und „Operation“ verweisen eher auf den prozessartigen Charakter des Verfahrens, während „Arbeitsaufgabe“ vor allem darauf verweist, dass solche Verfahren immer nach Vorgaben ablaufen, also normativ definiert sind. Solche Bezeichner sind also immer *cum grano salis* zu verstehen, bieten aber gleichzeitig die Möglichkeit zur transdisziplinären Kommunikation.

tungsphasen muss das Verfahren dann überarbeitet werden. Neben den Geboten der Datenvermeidung und Datensparsamkeit, die in dieser Phase der Verfahrensgestaltung vor allem durch die Verwendung von Anonymisierungs- und Pseudonymisierungstechniken umgesetzt werden können, sind hier insbesondere die Folgen für die Organisationsgestaltung zu beachten, die sich dann auch wieder technisch abbilden lassen, etwa als rollenbasiertes Rechtemodell. Der praktisch aufwändigste Abschnitt dieser Phase der Verfahrensgestaltung ist die Ableitung konkreter rechtlicher Anforderungen aus den weit verstreuten Rechtsquellen des Datenschutzes. Im Normalfall muss jedes Verfahren genau zwei Mal hinsichtlich der Erfüllung datenschutzrechtlicher Anforderungen überarbeitet werden: Die bei der ersten Überarbeitung aus Datenschutzgründen eingefügten Verfahrensschritte – dazu gehören etwa eine ordentliche Benutzerverwaltung und ein funktionierender Protokolldienst – müssen im zweiten Schritt selbst datenschutzkonform ausgestaltet werden.

Die entstandenen Verfahren können dann manuell oder (teil-)automatisiert umgesetzt werden. Die Verfahren sind datenschutzkonform gestaltet, weil und soweit sie aus datenschutzkonformen Verfahrensschritten zusammengesetzt sind. Sie sind kontrollierbar, weil jeder Verfahrensschritt kontrollierbar und die Schrittfolge deterministisch – und damit auch: kontrollierbar – ist. Damit – und mit Hilfe passender Kontrolltechniken wie Protokollen – kann die verantwortliche Stelle dann die Datenschutzkonformität der Verfahrenspraxis nachweisen.

Soweit die Theorie!

3. Von Kausalitäten zu Korrelationen

Auf Kausalitäten und Kausalitätsannahmen basierende Ansätze stoßen seit langem an Grenzen, die nachfolgend kursorisch dargestellt werden sollen. Anschließend werden korrelationsbasierte Ansätze als Alternative vorgestellt, gefolgt von den sich daraus ergebenden Änderungen für die Kontrollierbarkeit von Verfahren.

3.1. Grenzen kausalitätsbasierter Ansätze

Kausalitätsbasierte Ansätze setzen die Bildung geeigneter Modelle voraus. Dazu müssen zumindest sinnvolle Annahmen über die Kausalitäten zwischen Ereignissen getroffen werden.[29] Zwei Aspekte sind dabei aus Sicht der Organisation besonders kritisch: Erstens hat die technische Entwicklung inzwischen technische Systeme hervorgebracht, die auch große Modelle effizient verarbeiten können. Dem folgend sind die typischen Modellgrößen stark gestiegen: Heute benutzte Modelle stellen oft nicht nur einen größeren Ausschnitt der Welt dar, sie sind auch gleichzeitig feiner granuliert. Mit der Vergrößerung der Modelle steigt der Aufwand zu ihrer Erstellung mindestens im gleichen Maße, sehr oft allerdings überproportional.[30] Zweitens sinkt der Preis für Rechenleistung, während Personalkosten steigen. Das Interesse von Organisationen an großen, validen Modellen wird damit sowohl vom erforderlichen Aufwand wie auch von den zu erwartenden Kosten in Grenzen gehalten.

Insoweit Organisationen ihre eigenen Interessen vor die Interessen aller anderen stellen, mit denen sie in Kontakt treten, reicht es, wenn Verfahren aus ihrer Sicht folgende Eigenschaften erfüllen: Der Anteil falscher Entscheidungen soll relativ klein sein. Die einzelne Fehlentscheidung soll für die Organisation selbst relativ unbedeutend bleiben. Die Gesamtkosten sollen minimiert werden. Aus der Sicht von Organisationen sind statistische Verfahren, die auf der Basis automatisiert berechneter Korrelationen zwischen Ereignissen Ergebnisse produzieren, eine annehmbare Lösung.[31]

[29] Siehe oben das Beispiel der mehrbändigen Werke und der Kaufreihenfolge.

[30] Beispiel: Wer sollte etwa sinnvolle inhaltliche Verbindungen zwischen allen Büchern bzw. anderen Verkaufsartikeln, die über Amazon angeboten werden, erstellen, um Kundinnen und Kunden passende Angebote unterbreiten zu können, wenn andererseits unter Verwendung korrelationsbasierter Verfahren auf inhaltliche Verbindungen verzichtet werden kann?

[31] Dass Organisationen nach solchen Verfahren streben, hat schon Luhmann Ende der 60er Jahre festgestellt, der sie als Zweckprogramme bezeichnete, siehe Luhmann 1977. Ihre breite Einsatzmöglichkeit verdanken sie allerdings dem oben angeführten technischen Fortschritt.

3.2. Korrelationen

Korrelationen sind statistische Verbindungen zwischen Ereignissen. Im Gegensatz zu Kausalzusammenhängen lassen sie sich immer berechnen und setzen insbesondere keine inhaltliche Verbindung zwischen den Ereignissen voraus. So lässt sich unproblematisch eine Korrelation zwischen der Anzahl der jährlichen Meteoriteneinschläge auf der Erde und der Geburtenrate in süddeutschen Mittelstädten berechnen, ohne dass irgendwer auch nur im entferntesten auf die Idee käme, die beiden Ereignisse hätten irgendetwas inhaltlich miteinander zu tun.

Der große Vorteil von korrelationsbasierten gegenüber kausalitätsbasierten Ansätzen ist, dass sich Korrelationen zwischen Ereignissen vollständig automatisiert berechnen lassen. Es ist dabei insbesondere nicht notwendig, dass vorher Annahmen formuliert werden müssen. An die Stelle einer vorherigen Modellbildung tritt eine nachträgliche Bewertung der berechneten Ergebnisse.[32] Selbst diese nachträgliche Bewertung muss jedoch nicht mehr explizit vorgenommen werden – auch sie lässt sich automatisieren: als Rückkopplung. Im Falle eingeblendeter Werbung für ein Produkt bei Amazon kann etwa die Anzahl der Klicks, die dem Computer ja bekannt ist, selbst als Maßstab für die Bewertung[33] derjenigen Algorithmen benutzt werden, die auf der Basis von Korrelationsmechanismen darüber entscheiden, welche Werbung einem Benutzer eingeblendet wird.

Mathematisch sind Korrelationsberechnungen weder neu noch besonders schwer, allenfalls zeitaufwendig. Auch gehören sie schon lange zu Ausbildungskanon in der Informatik. Lange Zeit war ihre praktische Benutzung jedoch relativ selten, weil die zu korrelierenden Daten oft umständlich vorbereitet und die Algorithmen nicht selten selbst implementiert werden mussten. Inzwischen sind solche statistischen Methoden zur Korrelationsberechnung in großer Zahl in Standard-Programmen implementiert, mit einer relativ einfach zu bedienenden Oberfläche und ausgefeilten Hilfsfunktionen.

3.3. Folgen für die Kontrollierbarkeit

Im Gegensatz zu kausalitätsbasierten Verfahren sind korrelationsbasierte tendenziell nur noch in ihrer Gesamtheit kontrollierbar, nämlich gerade über die Rückkopplung der produzierten Ergebnisse und deren Verwendung als zusätzliche Eingabe für die Korrelationsalgorithmen. Aus Sicht der informationsverarbeitenden Organisation ist das unproblematisch, insbesondere wenn es sich um eine privatwirtschaftliche Organisation handelt, weil sie auch viele andere für sie relevante Informationen wie Geschäftszahlen oder Rückläuferquoten erst nachträglich erhält und daher schon lange gelernt hat, damit umzugehen. Aus der Sicht des Datenschutzrechts als Gefahrenabwehrrecht ist das zu spät.[34]

Da das Datenschutzrecht direkt an den Einzelschritten und deren Kontrollierbarkeit anknüpft und für diese spezifische rechtliche Anforderungen formuliert, greift es gegenüber korrelationsbasierten Ansätzen tendenziell ins Leere. In jedem Fall verliert der Datenschutz dabei einen – vormals – natürlichen Verbündeten: Es ist für die Durchsetzung der Interessen der verantwortlichen Stelle nicht mehr notwendig, ihre eigenen Verfahren über die Sicherstellung der Kontrollierbarkeit der einzelnen Verfahrensschritte unter Kontrolle zu bringen. Das Datenschutzrecht ist demnach in Bezug auf solche Verfahren der einzige Interessent für diese spezifische Art der Kontrollierbarkeit.

Mit dem Fehlen einer *ex ante* stattfindenden Modellbildung geht auch einher, dass Modelle nicht mehr der Begrenzung der Informationsmenge dienen können, weil der Erforderlichkeitsbegriff leer läuft. Bei der Verwendung von Korrelationsverfahren kann frühestens nach der Berechnung – und damit nach der Verarbeitung im Sinne des Datenschutzrechts

[32] Damit sind korrelationsbasierte Ansätze allerdings besonders empfänglich für Fehlinterpretationen. Eine oft zu beobachtende Fehlinterpretation ist die Gleichsetzung von berechneten Korrelationen mit Kausalitäten, also die Annahme, dass das Vorliegen einer Korrelation zwischen zwei Ereignissen bedeutet, dass die beiden Ereignisse in einem inhaltlichen Zusammenhang zueinander stehen. Dazu und zu den Folgen für die Betroffenen sowie mit einer grundsätzlicher Kritik an einem verstärkten Einsatz korrelationsbasierter Ansätze aus Datenschutzgründen, siehe Clarke 1988, S. 507. Umfassend dazu auch Gandy 1993.

[33] In der Informatik wird das als „Fitness“ bezeichnet.

[34] Siehe § 1 BDSG, wonach schon Beeinträchtigungen des Persönlichkeitsrechts der Betroffenen verhindert werden sollen. Der Schutz wird damit auch zeitlich sehr weit nach vorne verlagert. Diese Anforderung ist unmöglich zu erfüllen, wenn die Beeinträchtigung erst nachträglich festgestellt werden kann.

– bestimmt werden, welche Daten erforderlich sind.[35] Im engeren Sinne handelt es sich bei dieser nachträglich vorgenommenen Einschätzung nicht um eine Bestimmung der Erforderlichkeit von Informationen, sondern um eine Bestimmung ihrer Signifikanz. Die bisher bestehende Möglichkeit, mit dem Erforderlichkeitsbegriff schon bei der Erhebung einzugreifen, indem Informationen, die nicht verwendet werden müssen, um einen Zweck zu erreichen, gar nicht erst erhoben werden, geht damit verloren. Schon die ersten *privacy*-Anhörungen vor dem US-Repräsentantenhaus Mitte der sechziger Jahre haben deutlich gemacht, dass nur Datenvermeidung einen erfolgversprechenden Regulierungsansatz darstelle, nicht aber Verwendungsbeschränkung.[36]

4. Korrelationen und Datenschutzrecht

Die Alternative zu einem allgemeinen Verbot korrelationsbasierter Verfahren aus Datenschutzgründen kann ausschließlich in einem breit gefächerten Ansatz zur Eindämmung der aus ihnen erwachsenden Risiken für die Betroffenen und die Gesellschaft liegen.

Erstens muss das Gebot zur Datenvermeidung und Datensparsamkeit als „echte“ Pflicht ausgestaltet werden,[37] um den zahnlos gewordenen Erforderlichkeitsbegriff beim Umgang mit korrelationsbasierten Verfahren zu ersetzen.

Zweitens muss gesetzlich geregelt werden, dass sich – in Anlehnung an § 4a Abs. 3 BDSG – die Einwilligung ausdrücklich auch auf diese spezifische Verarbeitungsform beziehen muss.[38]

Drittens müssen Ergebnisse korrelationsbasierter Verfahren, die gespeichert oder weiterverarbeitet werden sollen, besonders gekennzeichnet werden, und diese Kennzeichnung muss auch bei einer Weitergabe aufrechterhalten werden.[39]

Viertens bedarf es eines Verbotes zur Rückspeicherung der Ergebnisse in den ursprünglichen Datenpool; Korrelationen dürfen mithin nur auf der Basis richtiger Daten berechnet werden.[40]

Fünftens sind solche Ergebnisse einer besonderen Zweckbindung zu unterwerfen, wobei die Zwecke eng zu definieren sind.

Sechstens sind wegen der großen Menge der gespeicherten Daten und den damit einhergehenden Sicherheitsrisiken besonders hohe Anforderungen an die technischen und organisatorischen Maßnahmen zur Datensicherheit zu stellen. Die einzelnen Maßnahmen müssen dabei dem Stand der Technik entsprechen, während ihre Auswahl auf dem Stand von Wissenschaft und Technik erfolgen muss.[41]

Insoweit diese großen Datenmengen auch Informationen aus unterschiedlichen Lebensbereichen der Betroffenen enthalten und mithin die Rollentrennung aufheben und die Betroffenen damit vollständig transparent machen können,[42] gibt es ein großes Interesse von privater, aber vor allem auch von öffentlicher Seite, auf diese Daten zugreifen zu können, sei es aus ökonomischen oder aus Gründen der Staatssicherheit. Siebtens müssen daher sowohl die Daten, auf deren Basis Korrelationen berechnet werden, als auch die dabei erzeugten Ergebnisse einer strengen Weitergabebeschränkung unterliegen, wie sie auch für personenbezogene Daten in Sozialen Netzwerken zu fordern ist.[43]

[35] Vgl. Podlechs Definition der Erforderlichkeit: „Eine Information ist zur Erfüllung einer Aufgabe erforderlich, wenn die Aufgabe ohne Kenntnis der Information nicht [...] erfüllt werden kann“, Podlech 1995, S. 21. Weil diese Bestimmung der Erforderlichkeit erst *nach* der Verarbeitung möglich ist, sind die Informationen *vor* der Verarbeitung *quasi immer* erforderlich.

[36] Siehe etwa Miller 1969, S. 1220 f.

[37] Siehe Scholz in Simitis 2011, § 3a Rn. 57 ff. zum derzeitigen Charakter des Gebots. Im eigentlichen Sinne handelt es sich derzeit gar nicht um ein „Gebot“, sondern um einen „Wunsch“.

[38] Unbestreitbar bleibt dabei, dass die Einwilligung selbst in den meisten Fällen immer eine Fiktion bleiben muss, siehe etwa Kamp und Rost 2013. Allerdings handelt es sich um eine Fiktion, die unbedingt erhalten bleiben muss, will man nicht das Menschenbild des Grundgesetzes selbst – der Mensch als informiertes, aufgeklärtes, selbstbestimmtes Wesen – aufgeben.

[39] Vorbild könnte hier die Regelung des § 25 Abs. 5a ASOG sein, die Gleiches für verdeckt erhobene oder dem Kernbereichsschutz oder dem Amts- und Berufsgeheimnis unterliegende personenbezogene Daten fordert.

[40] Siehe Mallmann in Simitis 2011, § 20 Rn. 9 ff. zur Frage, wann Daten unrichtig sind. Den Ergebnissen korrelationsbasierter Verfahren ist daher auch jeder Beweiswert abzusprechen.

[41] Zu den einzelnen Technikklauseln siehe BVerfGE 49, 89, 134 ff. – Kalkar I.

[42] Siehe dazu grundlegend Karhausen und Müller 1972, aktualisiert Rost 2013.

[43] Siehe ausführlich Pohle 2012.

Achtens muss die Phasenorientierung gestärkt werden, gerade auch auf europäischer Ebene, wo sie etwa im Vorschlag zur Datenschutz-Grundverordnung unangemessen stark beschränkt ist.[44] Aus dezidiert rechtsinformatischer Sicht bietet die Phasenorientierung zwei große Vorteile: Erstens können für die einzelnen Phasen sachlich angemessene und risikoadäquate Regelungen formuliert werden. Zweitens kann damit sichergestellt werden, dass der Umgang mit personenbezogenen Informationen jederzeit datenschutzkonform ist, weil jeder Verarbeitungsschritt von einem datenschutzkonformen Zustand wieder in einen datenschutzkonformen Zustand führt.[45]

Neuntens sind die Anforderungen, die an die einzelnen Phasen der Informationsverarbeitung zu stellen sind, unter Verwendung von Datenschutz-Schutzzielen zu formulieren.[46] Dieser im Bereich der IT-Sicherheit seit mehr als zwanzig Jahren verbreitete zielorientierte Ansatz formuliert in seiner auf den Datenschutzbereich ausgedehnten Fassung abstrakte Gestaltungsziele (neben den drei traditionellen Zielen der IT-Sicherheit – Verfügbarkeit, Integrität und Vertraulichkeit – auch drei explizite Datenschutz-Schutzziele – Transparenz, Intervenierbarkeit und Nicht-Verkettbarkeit), die für drei Verfahrenskomponenten (Informationen, technische Systeme und Prozesse) und drei Schutzbedarfe (normal, hoch und sehr hoch) gegeneinander abgewogen und dann organisatorisch und technisch umgesetzt werden müssen. Während dieser Ansatz aus informatischer Sicht erfolgversprechend erscheint, wenn die erforderlichen Kenntnisse und Fähigkeiten bei den informationsverarbeitenden Systemen vorhanden sind, und aus juristischer Sicht zumindest als umsetzbar,[47] fehlt den Datenschutzbeauftragten jede Durchsetzungsmacht, solange die Schutzziele nicht gesetzlich geregelt sind.

In jedem Fall ist offenkundig, dass korrelationsbasierte Verfahren dem Datenschutzrecht unterworfen werden müssen, um den Schutz der Betroffenen beim Umgang mit ihren personenbezogenen Informationen sicherstellen zu können.

5. Schluss

Während der Erforderlichkeitsbegriff in den letzten Jahrzehnten viel von seiner ursprünglich vorhandenen Fähigkeit zur strukturellen Begrenzung der Informationsmenge[48] verloren hat, weil sich die herrschende Meinung zur angemessenen Auslegung des Begriffs geändert hat, haben die technische Entwicklung und der vermehrte Einsatz korrelationsbasierter Verfahren dazu geführt, dass Begriff und Prinzip der Erforderlichkeit an jeder juristischen Auseinandersetzung vorbei zahnlos geworden sind. Die Folge davon sind gravierend: Nicht mehr der demokratisch legitimierte Gesetzgeber bestimmt die wesentlichen Fragen auf der Basis rechtlicher Abwägungen, sondern informationsverarbeitende Organisationen – und mit ihnen Technikerinnen und Techniker – entscheiden einseitig und nur unter Berücksichtigung spezifischer Interessen. Sowohl die Betroffenen, aber auch Juristinnen und Juristen als technische Laien und in ihrem eigenen methodischen Zugang werden damit übergangen.[49]

Auch die konkrete Ausprägung der Phasenorientierung im derzeitigen Recht stößt mit der Verbreitung korrelationsbasierter Ansätze an ihre Grenzen. Die Phasenorientierung selbst ist aber als analytisches Mittel zur Komplexitätsreduktion – und damit als Grundlage Nachweis datenschutzkomformer Informationsverarbeitung sowohl gegenüber den Betroffenen als auch gegenüber den Aufsichtsbehörden – sowohl für die Formulierung angemessener gesetzlicher

[44] Mit der Definition des Verarbeitungsbegriffs in Art. 4 Abs. 3 der Datenschutz-Grundverordnung durch eine ausdifferenzierte Liste unterschiedlicher Verarbeitungsarten sind die Grundlagen für eine fundierte Phasenorientierung gelegt. Das Problem ist, dass diese Möglichkeiten in der weiteren Verordnung nicht aufgegriffen wird, und stattdessen jede Verarbeitungsart grundsätzlich gleich behandelt wird.

[45] Eine wichtige Menge von Eigenschaften von Verarbeitungsschritten wird in der Informatik mit dem Akronym *ACID* bezeichnet: *Atomicity* (der Verarbeitungsschritt findet ganz oder gar nicht statt), *Consistency* (ein Verarbeitungsschritt führt wieder zu einem konsistenten Zustand, falls der vorherige Zustand auch konsistent war), *Isolation* (nebenläufige Verarbeitungsschritte beeinflussen sich nicht) und *Durability* (das Ergebnis des Verarbeitungsschrittes ist dauerhaft). Eine für den Datenschutzbereich sinnvolle Adaptation dieser Prinzipien befindet sich derzeit in der Entwicklung.

[46] Grundlegend dazu Rost und Pfitzmann 2009 und Rost 2012.

[47] Siehe Bock und Meissner 2012.

[48] Siehe dazu Simitis 2000.

[49] Siehe zur Diskussion über das Verhältnis von Recht und Technik, deren Ergebnisse die aufkommende Datenschutzdiskussion stark beeinflusst haben, Bull 1964, S. 44 ff., der ein klares Primat des Rechts fordert.

Regelungen als auch für deren praktische Umsetzung unverzichtbar. Die Anforderungen an die einzelnen Phasen müssen dazu aber unter Verwendung von Datenschutz-Schutzzielen neu formuliert werden. Anders lassen sich die spezifischen Risiken, die sich bei der Nutzung korrelationsbasierter Verfahren ergeben, und die bisher einseitig von den Betroffenen zu tragen sind, nicht in gesellschaftlich akzeptabler Weise rechtlich regeln.

A. Informationen, Daten und Datenverarbeitung

Anhand eines Beispiels soll nachfolgend der Unterschied zwischen Informationen und Daten deutlich gemacht werden. Darauf aufbauend wird erläutert, wie sich die verschiedenen Dimensionen von Informationen datentechnisch simulieren lassen und welche Probleme dabei auftreten.

In der Semiotik, deren Informationsbegriff dem Datenschutzrecht zugrunde liegt, werden Informationen durch vier Dimensionen beschrieben: Syntax, Semantik, Pragmatik und Sigmatik. Mit Syntax wird dabei die konkrete, meist zeichenmäßige Repräsentation bezeichnet. Unterschiedliche Zeichenfolgen (Worte) wie „cooky“ und „spooky“ bezeichnen deshalb grundsätzlich unterschiedliche Dinge. Einem Computer fällt es sehr leicht, diesen Unterschied zu erkennen – er muss die beiden Zeichenfolgen nur buchstabenweise vergleichen. Andererseits können unterschiedliche Zeichenfolgen sich auf das gleiche Ding beziehen – das gilt etwa für „cooky“ und „cookie“. Dieser Bezug auf das gleiche Ding kann ein Computer grundsätzlich nicht feststellen. Eine Zeichenfolge wie „cookie“ kann auch mehrere Bedeutungen haben – die Dinge unterscheiden sich dann nicht ihrer Syntax nach, sondern nur nach ihrer Semantik. Auch das lässt sich grundsätzlich mit einem Computer nicht feststellen, da sich die Bedeutung von Wörtern erst aus dem Kontext ergibt, in dem die Wörter gebraucht werden. Mit der pragmatischen Dimension wird der Zweck bezeichnet. Ein „cookie“ mit der Bedeutung „Keks“ kann für verschiedene Zwecke eingesetzt werden und bezeichnet dann unterschiedliche Dinge: Ein Keks, der zum Essen hergestellt wird, ist ein grundsätzlich anderer Keks als derjenige, der für eine Theateraufführung gebraucht wird und damit auch noch aus der letzten Reihe sichtbar, aber eben nicht essbar sein muss. Trotzdem ist ihre Bedeutung als Lebensmittel genauso gleich wie das Wort, das sie bezeichnet. Mit der Sigmatik wird der Verweis auf ein konkretes Ding bezeichnet. Die Information, ob „der Keks“ schmeckt, hängt eben auch davon ab, auf welchen Keks sie sich bezieht.

Computer können – stark vereinfacht ausgedrückt – nur Binärzahlen speichern, d. h. Bitfolgen wie `0110 0001`. Diese Bitfolge (Syntax) kann unterschiedliche Bedeutungen (Semantik) haben und etwa eine Zahl (97) oder ein Zeichen (`a`) darstellen. Auf der nächsthöheren Abstraktionsebene ist dann das Zeichen `a` wieder nur Syntax und es kann daher verschiedene Bedeutungen haben, etwa als Kleinbuchstabe *a* des lateinischen oder des kyrillischen Alphabets. Die Zeichenfolge „`mama`“ kann deshalb sowohl *mama* als auch *tata* darstellen. Welche Darstellung korrekt ist, kann der Computer nicht wissen. Sie wird von Außen (bei der Programmierung oder bei der Eingabe) vorgegeben.

Als Beispiel werde die Zeichenfolge `s1 = "cookie"` betrachtet. Der Computer kann sie von der Zeichenfolge `s2 = "cooky"` unterscheiden. Mehr nicht. Wenn der Computer entscheiden können soll, dass beide Zeichenfolgen `s1` und `s2` die gleiche Bedeutung haben sollen, nämlich „Keks“ bzw. allgemein „Lebensmittel“ – einmal in der allgemeinen englischen Form, einmal in der nur im *American English* gebräuchlichen –, dann muss ihm das explizit mitgeteilt und gespeichert werden. Und das muss wieder in Datenform geschehen, z. B. `s1-sem = "lebensmittel"` und `s2-sem = "lebensmittel"`. Die Zeichenfolge `s1 = "cookie"` kann aber auch eine andere Bedeutung haben und bezeichnet im Zusammenhang mit dem Internet einen Datensatz, der Angaben zum Besuch einer Webseite enthält, so etwa Zugangsdaten oder Zeitangaben. In beiden Fällen kann der Computer die Bedeutung von `s1` immer noch nicht verstehen, aber das Verstehen – und damit eine Form der Informationsverarbeitung – kann jetzt in Form von Datenverarbeitung simuliert werden. Das gleiche gilt für die pragmatische und die sigmatische Dimension, z. B. `s1-pragm = "rezept_en"` oder `s2-pragm = "rezept_us"`. Die Daten `s1-sem` und `s1-pragm` beschreiben dann Eigenschaften von `s1` und werden Meta-Daten genannt.

Es gibt nun zwei grundsätzliche Möglichkeiten, mit Computern Informationen zu verarbeiten. Einerseits können die Programme so gestaltet werden, dass sie zu allen Daten auch

jeweils alle relevanten Meta-Daten speichern und bei jeder Verarbeitung prüfen, ob die zu speichernden oder gespeicherten Meta-Daten zu den erwarteten oder geforderten passen. Andererseits können nur die Daten gespeichert und verarbeitet werden, während die eigentliche Informationsverarbeitung an die vor dem Computer sitzenden Menschen übertragen werden. Im zweiten Fall kann etwa in den Lebenslauf einer Informatikerin einfach geschrieben werden, sie sei Expertin für die Verarbeitung von „cookies“, ohne dass der Zeichenfolge das Meta-Datum `internettechnik"` mitgegeben wird. Hier werden auch schon die zentralen Probleme deutlich, die sich in beiden Fällen ergeben. Im ersten Fall sind die Programme entweder besonders aufwendig zu erstellen, weil Informationen intern in allen Dimensionen verarbeitet werden müssen, oder sie sind extrem aufwendig in der Bedienung, weil zu allen einzugebenden Daten auch immer alle relevanten Meta-Daten mit eingegeben werden müssen. Die oft als Alternative dargestellte semantische Analyse durch den Computer selbst ist in vielen Fällen noch nicht praxistauglich: Weil Menschen nicht nur „cookies“ löschen, sondern auch „files“, und „cookies“ automatisch als Backwaren klassifiziert wurden, mussten dann wohl auch „files“ Backwaren sein – das ist der Stand der Wissenschaft![50] Im zweiten Fall muss organisatorisch sichergestellt werden, dass sich die Informationen nicht ändern, nur weil sie ein anderer Mensch aus dem Computer ausliest, oder weil sie an eine andere Abteilung oder eine andere Organisation weitergegeben wurden, oder weil sich die Welt verändert hat. Auch das ist ein bislang grundsätzlich ungelöstes Problem.

Literatur

Albers, Marion (2005). *Informationelle Selbstbestimmung*. Baden-Baden: Nomos Verlagsgesellschaft.

Birk, Dominik, Helmut Reimer und Christoph Wegener (2010). „Soziale Netze – neue Impulse zum Datenschutz“. In: *Datenschutz und Datensicherheit* 34.7, S. 492.

Bock, Kirsten und Sebastian Meissner (2012). „Datenschutz-Schutzziele im Recht“. In: *Datenschutz und Datensicherheit* 36.6, S. 425–431.

Bull, Hans Peter (1964). *Verwaltung durch Maschinen – Rechtsprobleme der Technisierung der Verwaltung*. 2. Aufl. Köln, Berlin: G. Grote'sche Verlagsbuchhandlung KG.

Clarke, Roger A. (Mai 1988). „Information technology and dataveillance“. In: *Communications of the ACM* 31.5, S. 498–512. ISSN: 0001-0782. DOI: 10.1145/42411.42413. URL: http://doi.acm.org/10.1145/42411.42413.

Coy, Wolfgang (1992). „Für eine Theorie der Informatik!“ In: *Sichtweisen der Informatik*. Hrsg. von Wolfgang Coy u. a. Braunschweig/Wiesbaden: Vieweg, S. 17–32.

Gandy Jr., Oscar H. (1993). *The Panoptic Sort*. Boulder, San Francisco, Oxford: Westview Press.

Kamp, Meike und Martin Rost (2013). „Kritik an der Einwilligung“. In: *Datenschutz und Datensicherheit* 37.2, S. 80–83.

Karhausen, Mark O. und Paul J. Müller (Aug. 1972). „Datenbank, Datentransparenz und Datenschutz“. In: *Nachrichten für Dokumentation* 23.4, S. 148–153.

Leib, Hans-Jürgen (1985). „Technische Entwicklung und Datenschutzrecht“. In: *Datenschutz und Datensicherung im Wandel der Informationstechnologien*. Hrsg. von Peter Paul Spies. Informatik-Fachberichte 113. Berlin: Springer-Verlag, S. 218–228.

Luhmann, Niklas (1977). *Zweckbegriff und Systemrationalität*. 2. Aufl. Suhrkamp Verlag.

Miller, Arthur Raphael (Apr. 1969). „Personal Privacy in the Computer Age: The Challenge of a New Technology in an Information-Oriented Society“. In: *Michigan Law Review* 67.6, S. 1089–1246. URL: http://www.jstor.org/stable/1287516.

Ochs, Carsten und Martina Löw (Sep. 2012). „Un/Faire Informationspraktiken: Internet Privacy aus sozialwissenschaftlicher Perspektive“. In: *Internet Privacy. Eine multidisziplinäre Bestandsaufnahme*. Hrsg. von Johannes Buchmann. acatech – Deutsche Akademie der Technikwissenschaften. Darmstadt, S. 15–62.

Ohm, Paul (2010). „Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization“. In: *UCLA Law Review* 57, S. 1701–1777.

[50] Siehe „Aiming to Learn as We Do, a Machine Teaches Itself“, New York Times, 04.10.2010. URL: http://www.nytimes.com/2010/10/05/science/05compute.html; Stand: 31.01.2013.

Podlech, Adalbert (1976). „Information – Modell – Abbildung – Eine Skizze“. In: *Informationsrecht und Rechtspolitik*. Hrsg. von Wilhelm Steinmüller. Rechtstheorie und Informationsrecht 1. München, Wien: Oldenbourg Verlag, S. 21–24.

— (1982). „Individualdatenschutz – Systemdatenschutz“. In: *Beiträge zum Sozialrecht – Festgabe für Grüner*. Hrsg. von Klaus Brückner und Gerhard Dalichau. Percha: Verlag R. S. Schulz, S. 451–462.

— (1995). *Der Informationshaushalt der Krankenkassen: Datenschutzrechtliche Aspekte*. Baden-Baden: Nomos Verlagsgesellschaft.

Pohle, Jörg (2012). „Social Networks, Functional Differentiation of Society, and Data Protection“. In: *Arxiv preprint arXiv:1206.3027*.

Rost, Martin (2012). „Standardisierte Datenschutzmodellierung“. In: *Datenschutz und Datensicherheit* 36.6, S. 433–438.

— (2013). „Zur Soziologie des Datenschutzes“. In: *Datenschutz und Datensicherheit* 37.2, S. 85–91.

Rost, Martin und Andreas Pfitzmann (2009). „Datenschutz-Schutzziele – revisited“. In: *Datenschutz und Datensicherheit* 6, S. 353–358.

Simitis, Spiros (2000). „Das Volkszählungsurteil oder der lange Weg zur Informationsaskese – (BVerfGE 65, 1)“. In: *Kritische Vierteljahresschrift für Gesetzgebung und Rechtswissenschaft* 83, S. 359–375.

— Hrsg. (2011). *Bundesdatenschutzgesetz*. 7. Aufl. Nomos Verlagsgesellschaft.

Steinmüller, Wilhelm u. a. (1971). *Grundfragen des Datenschutzes*. Techn. Ber. Gutachten im Auftrag des Bundesministeriums des Innern, BT-Drs. VI/3826, Anlage 1. Bundesministerium des Innern.