Informatische Anmerkungen zum Stand der Technik und zu ihrem Gebrauch

Jörg Pohle*

13. September 2012

1 Einleitung

Der Begriff Informationsverarbeitung bezieht sich im Folgenden auf die Organisationsebene und entspricht dem datenschutzrechtlichen Begriff der Datenverarbeitung. Der verwendete Informationsbegriff, der auch dem Datenschutzrecht zugrunde liegt, entstammt der Semiotik und besitzt vier Dimensionen: Syntax, Semantik, Pragmatik und Sigmatik. Wenn im Folgenden von Daten gesprochen wird, sind damit Daten im technischen Sinne gemeint.

2 Kausalitäten, Korrelationen und die Kontrolle über Verfahren

Eine rationale Verwaltung (*Bürokratie*) zeichnet sich dadurch aus, dass sie strukturell besser als Individuen oder Gruppen in der Lage ist, Informationen zu verarbeiten und Entscheidungen zu treffen. Als *rationale* Verwaltung kann sie nur über Informationen verfügen, wenn diese datenmäßig abgebildet sind, andernfalls wäre sie auf bestimmte Menschen als Informationsträger angewiesen. Die Informationen verlieren dabei mindestens ihre semantische (*Kontext*) und ihre pragmatische (*Zweck*) Dimension. Die Einführung des Computers als Datenverarbeitungsmaschine (Werkzeug) zur Unterstützung der Informationsverarbeitung der Organisation wirkt dabei entgrenzend auf die Informationsverarbeitung: Alle Beschränkungen, die vorher notwendige Folge der manuellen Datenverarbeitung waren, werden damit strukturell aufgehoben, sowohl in zeitlicher und örtlicher Hinsicht als auch in Bezug auf die Menge der verarbeitbaren Informationen.

Organisationen, die Entscheidungen über Betroffene treffen, müssen die Menschen als Objekte intern abbilden. Wächst der Bedarf nach Entscheidungen, müssen notwendig immer vollständigere Abbilder der Betroffenen entstehen. Dabei treffen Organisationen Entscheidungen nicht auf der Basis von Geeignetheit, Vollständigkeit und Korrektheit von Daten, sondern in der Erwartung, dass diese gegeben sind.

In den letzten Jahren werden *ex ante* prüfbare Verfahren auf der Basis von (mindestens angenommenen) Kausalitäten mehr und mehr durch Datenkorrelationsmechanismen ersetzt, deren Qualität Organisationen nur noch *ex post* anhand der Folgen darauf basierender Entscheidungen bewerten können. Mit dem Verlust an Kontrolle über Verfahren und deren Ersatz durch Kontrolle über Ergebnisse endet tendenziell die insoweit bestehende Koalition zwischen Organisation und Datenschutz, weil letzterer die rechtliche Regulierung an Verfahren und deren Kontrollierbarkeit gebunden hat.

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

3 Kausalitäten, Korrelationen und datenschutzrechtliche Erforderlichkeit

Verfahren sind über (angenommene) Kausalitäten (*Modelle*) auf Zwecke bezogen und dienen unter Verwendung von Informationen dem Erreichen dieser Zwecke. Modelle sind Abbildungen von etwas, für jemanden, für einen Zweck, die von jemandem erstellt sind. Auf der Basis von Modellen kann festgestellt werden, welche Informationen für die Erreichung des Zwecks erforderlich sind. Auf diese Erforderlichkeit bezieht sich das Datenschutzrecht. Das heißt aber auch, dass eine vermehrte Verwendung von Datenkorrelationsmechanismen das Prinzip der Erforderlichkeit grundlegend aushebelt, weil sich Korrelationen nicht auf Modellen, sondern nur auf (großen Mengen von) bereits erhobenen Informationen bestimmen lassen.

4 Korrelationen und Persönlichkeitsprofil

Ursprünglich ging man von der Annahme aus, dass Persönlichkeitsprofile gleichbedeutend mit extrem ausführlichen „Dossiers“ über die Betroffenen sein würden. Die davon ausgehenden Gefahren waren allerdings nie allein auf die Existenz solcher Dossiers ausgerichtet – der Schwerpunkt lag vielmehr immer auf ihrer Nutzbarkeit. Informatisch gesprochen: Dossiers sind Informations- oder Datensammlungen, die zu beliebigen Anfragen zu konkreten Betroffenen (hinreichend) korrekte und vollständige Antworten liefern. Daraus folgt aber auch, dass beliebige Informations- oder Datenmengen – auch wenn sie nicht in Form eines Dossiers organisiert sind –, die die gleichen Antworten auf die Anfragen und gleiche Menge der Antworten liefern, rechtlich als Persönlichkeitsprofile zu werten sind.

5 Anonymität, Anonymisierbarkeit und Datenschutzrecht

Zu Beginn der modernen Datenschutzdiskussion gab es zwei Annahmen zu anonymen Informationen: Sie sind erstens objektiv – d. h. für beliebige Verarbeiter – anonym, und zweitens lassen sie keinen Zugriff der Organisation auf das Individuum zu und damit auch keine Ausübung von Macht über das Individuum. Aus diesem Grund wurden anonyme Daten nicht den Datenschutzgesetzen unterworfen. Mehr noch: Wer gespeicherte Daten anonymisiert, kann sich dadurch den Datenschutzgesetzen auch nachträglich noch entziehen.

Inzwischen hat sich einhellig die Auffassung durchgesetzt, dass Informationen allenfalls relativ anonymisiert werden können: Für verschiedene Organisationen können sie jeweils anonym oder personenbezogen sein, abhängig ist das allein vom möglichen Zusatzwissen und der jeweiligen Fähigkeit zur technischen Datenverarbeitung. Seit etwa 15 Jahren gibt es verstärkte Forschungstätigkeiten auf dem Gebiet der Anonymisierung von Informationen, wobei für alle vorgestellten Verfahren bisher immer auch Möglichkeiten zur nachträglichen De-Anonymisierung gefunden wurden. Das Datenschutzrecht hat darauf schon vor langer Zeit mit einer Absenkung der Anforderungen an eine Anonymisierung geantwortet: Gefordert wird jetzt nur noch die *faktische* Anonymisierung. Die Rechtsfolgen sind allerdings die gleichen geblieben.

Bezeichnender Weise wurde bei der Diskussion zum Institutionaldatenschutz keine derartige Privilegierung für die Organisation bei der Verwendung anonymer (*statistischer*) Informationen vorgesehen.

Auch die zweite Annahme ist nicht zu halten: Sie unterstellt unter anderem implizit, dass Organisationen jeweils nur ein Zugriffsinteresse auf ein bestimmtes Individuum haben. Anonymität ist jedoch immer nur relativ gegeben, nämlich zu einer sogenannten Anonymitätsmenge. Informatisch ist Anonymität daher auch immer nur in Bezug auf diese Menge definiert: Anonym ist, wer in dieser Menge nicht identifiziert werden kann. Praktisch kann nun zweierlei passieren. Erstens kann die Menge durchaus klein sein. Zweitens kann für die Organisation der Nachweis ausreichend sein, dass ein Individuum Teil dieser Anonymitätsmenge ist. In beiden Fällen ist das Individuum gegenüber der Organisation ungeschützt.

6 Strukturierte und unstrukturierte Daten

Auch die strikte Unterscheidung zwischen strukturierten und unstrukturierten Daten, wie sie von Prof. Öman am zweiten Tag in die Debatte eingebracht wurde, lässt sich aus informatischer Sicht nicht aufrecht erhalten, wo sie zur Grundlage für unterschiedliche rechtliche Bewertungen gemacht wird.

Zwar unterscheidet auch die Informatik Daten nach Strukturgesichtspunkten (strukturiert, semistrukturiert, unstrukturiert), für die Möglichkeiten zur automatisierten Verarbeitung ist das jedoch (quasi) bedeutungslos. So durchsuchen etwa moderne Desktop-Suchsysteme auch unstrukturierte Daten, soweit das jeweilige Programm (oder Betriebssystem) das betreffende Dateiformat lesen kann. Beispiele dafür sind *Windows Search*, *Spotlight* oder *Beagle*. Systeme, die sich an professionelle Anwender wie größere Organisationen richten, sind deutlich leistungsfähiger. Am Beispiel von *Google Images* wird deutlich, dass solche Systeme nicht nur für die Verarbeitung von Text in unstrukturierten Datenmengen geeignet sind, sondern auch Bilder verarbeiten können: Bei *Google Images* können Bilddateien als „Suchbegriffe“ verwendet werden, für die Google dann ähnliche Bilder als Suchergebnisse zurück liefert.

7 Der Mensch und sein Computer

Gerade der zweite Tag des Workshops hat deutlich gemacht, dass viele bereit sind, Aussagen über Technik zu treffen, sie rechtlicher Regulierung zu unterwerfen oder sie schlicht selbst zu benutzen, ohne dass sie Funktionsweise und (individuelle und gesellschaftliche) Implikationen der Technik verstehen. Psychologisch handelt es sich dabei um unberechtigte Selbstsicherheit. Im Datenschutzbereich stehen Datenverarbeiter und Betroffene häufig vor dem gleichen Problem.

In den 1970er Jahren gingen die Datenschützer davon aus, dass die aufkommende Informationsgesellschaft immer vor dem Problem komplexer und kompliziert zu bedienender Technik stehen würde. Daher war die Annahme verbreitet, dass Menschen nur dann in großem Umfang mit Computern interagieren würden können – und genau das ist in einer Informationsgesellschaft offenkundig sowohl notwendig als auch erwünscht –, wenn sie mit dem Computer in „seiner“ Sprache kommunizieren, d. h. programmieren, können. Diese Fähigkeit setzt zumindest ein gewisses Verständnis für die Funktionsweise des betreffenden technischen Artefakts voraus. Die Geschichte hat sich jedoch völlig anders entwickelt als erwartet: Nicht die Menschen haben sich dem Computer, sondern der Computer hat sich den Menschen angepasst. Insbesondere die graphischen Benutzeroberflächen haben die Bedienung trivialisiert und damit die weiterhin vorhandene Komplexität der Maschine mehr und mehr vor den Benutzern versteckt. Das, was sowohl Datenverarbeiter als auch Betroffene nun vor allem sehen – und damit meinen, es unter Kontrolle zu haben –, ist nur noch das farbenprächtige Spiel von Oberflächen. Die Informatik hat im Bereich der Softwareentwicklung Jahrzehnte gebraucht, sich darauf einzustellen, dass die Technik nicht von Informatikern bedient wird, sondern vor allem von technischen Laien. Viel schwerer fällt es ihr, die Konsequenzen dieser Entwicklung zu verstehen und damit umzugehen: Zwar lässt sich so gut wie jede Software weitgehend konfigurieren, aber mehr als 40% der Nutzerinnen und Nutzer ändern Voreinstellungen nie. Daraus ergibt sich auch die in den letzten Jahren vermehrt geäußerte Forderung nach *Privacy by Default*.

Auch für die Datenverarbeiter sieht die Situation nicht wirklich besser aus. Während an vielen Stellen fälschlich behauptet wird, das Datenschutzrecht sei stark auf Großrechner zugeschnitten, wurde eher implizit angenommen, dass nur Organisationen automationsgestützt Informationsverarbeitung betreiben. Solche Organisationen differenzieren sich dann intern aber notwendig aus – es entsteht eine eigene DV, EDV- oder IT-Abteilung. Und diese kann und wird sich dann sowohl spezialisieren als auch professionalisieren. Mit anderen Worten: Dem Datenschutzrecht liegt die Annahme zugrunde, dass die Informationsverarbeitung professionell betrieben wird. Nur daraus lassen sich bestimmte Eigenheiten des Datenschutzrechts erklären. Diese Annahme lässt sich vor dem Hintergrund der geschichtlichen Entwicklung aber nicht halten: Die Mehrzahl der privaten Datenverarbeiter betreibt die Informationsverarbeitung nicht professionell, sondern laienhaft. Das zeigt sich etwa daran, dass

viele sowohl Betriebssystem als Anwendersoftware *Basic*-, *Home*- oder *Standard*-Versionen benutzen und nicht *Professional*-, *Server*- oder gar *Enterprise*-Versionen.

Im Gegensatz zur ersten Generation der Datenschützer – und insbesondere im Gegensatz zu denen, die die Architektur des Datenschutzrechts entworfen haben –, äußern sich heute viele zur weiteren Entwicklung des Datenschutzrechts, die weder die Prinzipien oder die Methoden der automationsgestützten Informationsverarbeitung noch die technischen Datenverarbeitungssysteme hinreichend verstehen.