Grundlagen des Standard-Datenschutzmodells (SDM), mit Anwendung auf auf BigData

Martin Rost
Kiel, 26.08.2013

ULD
Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein

„Imagination is more important than knowledge.“
(Albert Einstein)

Agenda

Generelle Linie des Vortrags:
Erläuterung SDM (65%), BigData (35%).

1. Was meint „Datenschutz“?
2. Vorstellung des Standard-Datenschutzmodells (SDM)
   - Schutzziele
   - Verfahren: Daten, Prozessen, IT-Systeme
   - Schutzbedarf, aus der Sicht der Betroffenen
3. BigData-Eigenschaften im Lichte des SDM

Was ist mit „Datenschutz“ gemeint?

Datenschutz?

1. Datenschutz ist nur das, was im Datenschutzrecht formuliert ist!
=> *Dies wäre ein juristischer Kurzschluss*

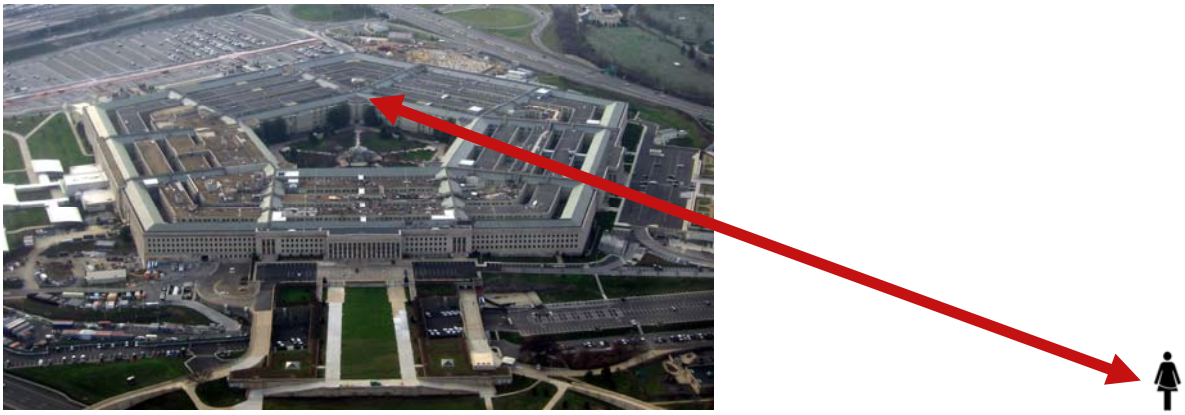
2. Praktisch werdender Datenschutz lässt sich weitgehend mit den bekannten Maßnahmen der IT-Sicherheit sicherstellen!
=> *Dies wäre ein technizistischer Kurzschluss*

3. Datenschutz entsteht unmittelbar aus dem Grundbedürfnis des Menschen nach Privatleben, nach Selbstbestimmung und Freiheit.
=> *Dies wäre ein nicht-historischer, ideologisch-psychologistischer Kurzschluss*

Was meint „Datenschutz“?

Objektbereich des Datenschutzes

Datenschutz beobachtet, bewertet und nimmt gestaltenden Einfluss auf die *asymmetrischen Machtbeziehungen* zwischen Organisationen und Personen, mit grundrechtlich-gesetzlichen Anforderungen, die Organisationen zu erfüllen haben, um diese Asymmetrie zu kompensieren.



http://commons.wikimedia.org/wiki/File:The_Pentagon_January_2008.jpg

Objektbereich des Datenschutzes

Datenschutz thematisiert ...

- ... die *externen* Machtbeziehungen zwischen...
  - öffentlicher Verwaltung und deren externen **Bürgern,**
  - privaten Unternehmen und deren **Kunden,**
  - IT-Infrastruktur-Providern und deren **Nutzern / Kunden** (bspw. Access-, Suchmaschinen-, Mail-, Socialnetwork-Betreiber);
  - Praxen / Instituten / Gemeinschaften und deren **Patienten, Mandanten, Klienten;**
  - Wissenschaftsorganisationen und deren Forschungsobjekten **Individuen, Subjekte, Menschen;**
- ... sowie die *internen* Machtbeziehungen zwischen...
  - Organisationen (Arbeitgeber, auch: Kirche, Militär, (Sport-) Verein) und deren **Mitarbeitern der Mitgliedern** (Schüler, Patienten, Gefangenen, Soldaten, ...).

Zum Verhältnis von Datenschutz und Datensicherheit

Datensicherheit:
Die Person ist der Angreifer!
Datenschutz:
Die Organisation ist der Angreifer!

Daraus folgt u.a.:

1. Personen müssen sich gegenüber Organisationen als vertrauenswürdig ausweisen.
2. Organisationen müssen sich gegenüber Personen als vertrauenswürdig ausweisen.
3. IT-Sicherheitsmanagement (ISO 27001) ist kein Datenschutzmanagement.
4. Datenschutz ist auf die Maßnahmen der Datensicherheit und des IT-Sicherheitsmanagements angewiesen, muss aber in der Lage sein, auch die IT-Sicherheit zu kontrollieren.

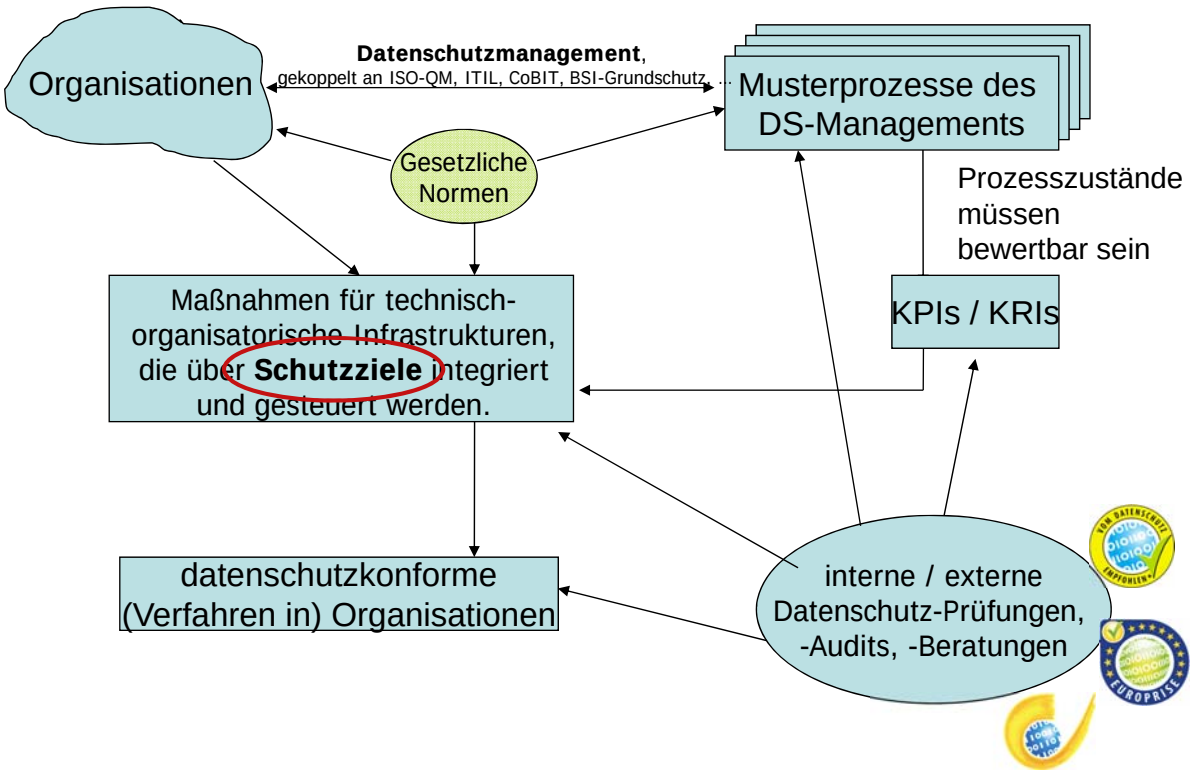
Wichtigster Grundsatz des Datenschutzrechts

- **Es dürfen keine personenbezogene Daten verarbeitet werden** PUNKT
- Eine Ausnahme von diesem Grundsatz ist zulässig, wenn eine **zweckgebundene Erhebung und Verarbeitung** durch ein **Gesetz** oder eine **Einwilligung** geregelt vorliegt.
- „Verbot mit Erlaubnisvorbehalt“, BDSG §4ff.

Personen und Rollen, Organisationen und Gesellschaft

- Personen moderner Gesellschaften, die in den Rollen als Bürger, Kunde oder Patienten auftreten müssen, haben diese Rollen, die Autonomie und Privatheit verlangen und Individualität konstituieren, nicht aus freien Stücken gewählt. **Die Rollen des Bürgers, Kunden usw. werden jeder Person gesellschaftlich zugewiesen**.
- Ob man Autonomie qua Privatheit in Anspruch nehmen will oder nicht, ist keine bloß individuelle Angelegenheit, sondern wird konstituiert durch gesellschaftliche Strukturen. **Selbstbestimmung in Privatheit ist ein Produkt aus der Gemengelage bestehend aus Märkten, Gewaltenteilung, Demokratie und freien wissenschaftlichen, ästhetischen und metaphysischen Diskursen.**
- Das was als Privatheit, Freiheit, Individualität gesellschaftlich erzeugt wird, wird zugleich von Organisationen notorisch unterlaufen. Etablierte Unternehmen haben kein Interesse am Markt, Sicherheitsbehörden kein Interesse an Gewaltenteilung, Forschungsinstitute kein Interesse an freien Diskursen. **Aus Organisationssicht sind selbstbestimmte Kunden, Bürger und Patienten Risiken.**
- Organisationen untergraben notorisch die Funktionsbedingungen moderner „funktional differenzierter“ (Luhmann) Gesellschaften. **Organisationen nutzen die IT, um die Risiken des Marktes(Kunden), der öffentlichen Ordnung (Bürger), der freien Diskurse (Subjekte) für sich selber so weit es irgend geht zu verringern.**

Big Picture Prozesse, DS-Management, Schutzziele, KPI/KRI



Der Zweck des Standard-Datenschutzmodells (SDM) besteht darin, eine integre und transparente Vermittlung bestehender Anforderungen des Datenschutzrechts mit technischen und organisatorischen Funktionen zu schaffen. Die ist nützlich und notwendig insbesondere für die Arbeit von Datenschutzbeauftragten, die sich gegenüber

- den Betroffenen,
- den zu prüfenden Organisationen,
- den anderen Kollegen im Datenschutz
- sowie nicht zuletzt gegenüber der Gesellschaft für ihre Tätigkeiten zu rechtfertigen haben.

Standardisiertes Datenschutzmodell

Wesentliche Komponenten zur Modellierung von Datenschutzanforderungen

1. **Datenschutzeigene Schutzziele**, die die Schutzmaßnahmen für die zu bearbeitenden *Risiken* ausweisen und die *Beziehung* zwischen Recht, Technik und Organisation methodisch kontrollierbar machen.
2. **Feststellung der Schutzbedürftigkeit von Personen**, denen die Verfahren in den Organisationen zu entsprechen haben.
3. **Verfahrenskomponenten** analysieren, für die entsprechend den Schutzzielen und der Schutzbedürftigkeit die angemessenen Maßnahmen auszuwählen sind.

1. Schutzziele des Datenschutzes

Schutzziele in Datenschutzgesetzen der Länder

(Bsp: NRW, RlPf, HH, B, die 6 neuen Bundesländer,

§ 10 Technische und organisatorische Maßnahmen (DSGesetz von NRW)

(1) Die Ausführung der Vorschriften dieses Gesetzes sowie anderer Vorschriften über den Datenschutz ist durch technische und organisatorische Maßnahmen sicherzustellen.

(2) Dabei sind Maßnahmen zu treffen, die geeignet sind zu gewährleisten, dass

- 1. nur Befugte personenbezogene Daten zur Kenntnis nehmen können **(Vertraulichkeit)**,
- 2. personenbezogene Daten während der Verarbeitung unversehrt, vollständig und aktuell bleiben **(Integrität)**,
- 3. personenbezogene Daten zeitgerecht zur Verfügung stehen und ordnungsgemäß verarbeitet werden können **(Verfügbarkeit)**,
- 4. jederzeit personenbezogene Daten ihrem Ursprung zugeordnet werden können **(Authentizität)**,
- 5. festgestellt werden kann, wer wann welche personenbezogenen Daten in welcher Weise verarbeitet hat **(Revisionsfähigkeit)**,
- 6. die Verfahrensweisen bei der Verarbeitung personenbezogener Daten vollständig, aktuell und in einer Weise dokumentiert sind, dass sie in zumutbarer Zeit nachvollzogen werden können **(Transparenz)**.

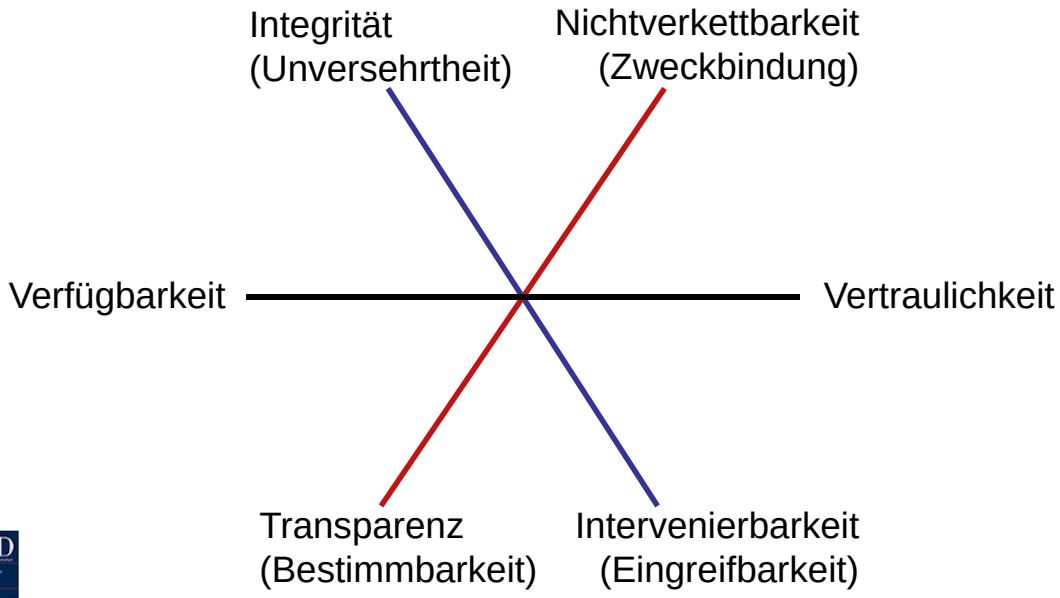
(3) (...)

Schutzziele im LDSG-SH, §5 *(Stand: Januar 2012)*

(1) Die Ausführung der Vorschriften dieses Gesetzes sowie anderer Vorschriften über den Datenschutz im Sinne von § 3 Abs. 3 ist durch technische und organisatorische Maßnahmen sicherzustellen, die nach dem Stand der Technik und der Schutzbedürftigkeit der Daten erforderlich und angemessen sind. Sie müssen gewährleisten, dass

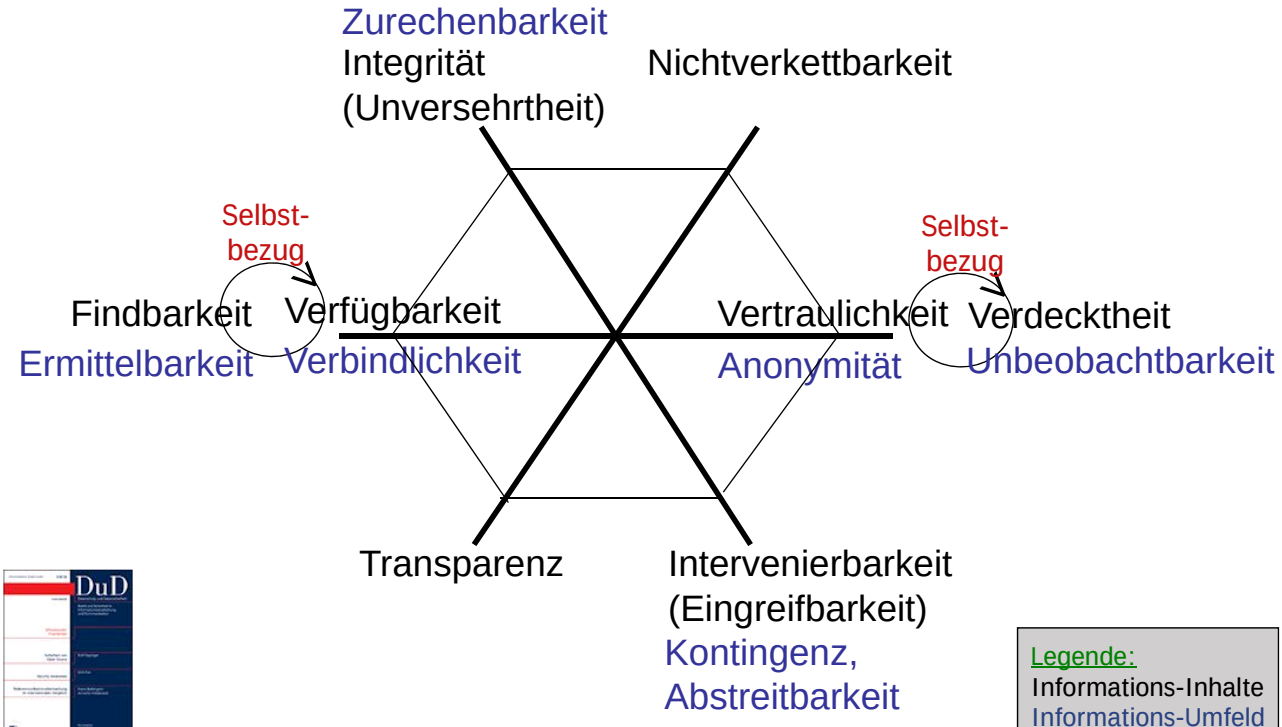
- Verfahren und Daten zeitgerecht zur Verfügung stehen und ordnungsgemäß angewendet werden können **(Verfügbarkeit),**
- Daten unversehrt, vollständig, zurechenbar und aktuell bleiben **(Integrität),**
- nur befugt auf Verfahren und Daten zugegriffen werden kann **(Vertraulichkeit)**,
- die Verarbeitung von personenbezogenen Daten mit zumutbarem Aufwand nachvollzogen, überprüft und bewertet werden kann **(Transparenz)**,
- personenbezogene Daten nicht oder nur mit unverhältnismäßig hohem Aufwand für einen anderen als den ausgewiesenen Zweck erhoben, verarbeitet und genutzt werden können **(Nicht-Verkettbarkeit)** und
- Verfahren so gestaltet werden, dass sie den Betroffenen die Ausübung der ihnen zustehenden Rechte nach den §§ 26 bis 30 wirksam ermöglichen **(Intervenierbarkeit)**.

Schutzziele Systematik



2009/06

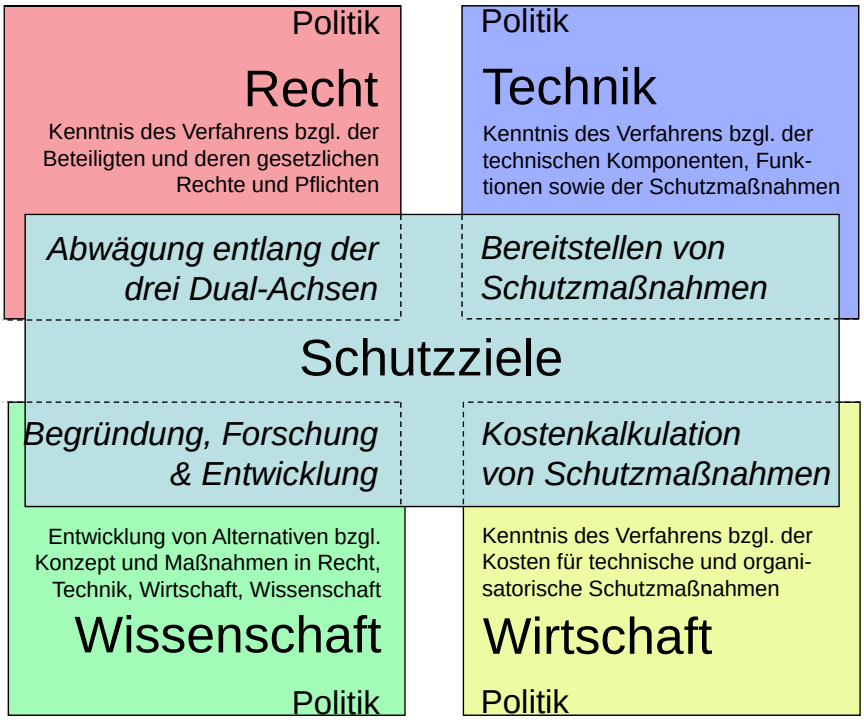
Schutzziele Systematik, vollständig



Legende:
Informations-Inhalte
Informations-Umfeld

2009/06

Schutzziele als generalisierte Vermittler



Schutzziele und Maßnahmen zu deren Umsetzung

- Sicherstellung von *Verfügbarkeit*
  Daten/Prozesse: **Redundanz, Schutz, Reparaturstrategien**
- Sicherstellung von *Integrität*
  Daten: **Hash-Wert-Vergleiche**
  Prozesse: **Festlegen von Min./Max.-Referenzen**, Steuerung der Regulat.
- Sicherstellung von *Vertraulichkeit*
  Daten: **Verschlüsselung**
  Prozesse: **Rollentrennungen**, Abschottungen, Containern
- Sicherstellen von *Transparenz* durch Prüffähigkeit
  Daten: **Protokollierung**
  Prozesse: **Dokumentation** von Verfahren
- Sicherstellen von *Nichtverkettbarkeit* durch Zweckbestimmung/-bindung
  Daten: **Pseudonymität, Anonymität** (anonyme Credential)
  Prozesse: **Identitymanagement, Anonymitätsinfrastruktur**
- Sicherstellen von *Intervenierbarkeit* durch Ankerpunkte
  Daten: Zugriff auf Betroffenen-Daten durch den Betroffenen
  Prozesse: **SPOC für Änderungen, Korrekturen, Löschen**, Aus-Schalter, **Changemanagement** auf Seiten der Organisationen

Getrieben wird die Sicherstellung der Maßnahmen durch organisations-**externe Auditierungen** und organisations-**internes Datenschutzmanagement.**

2. Schutzbedarfe

Schutzbedarfe für „Person“

Die Ermittlung des Schutzbedarfs für Personen orientiert sich an der BSI-Grundschutzdefinition(*), wechselt die Perspektive jedoch von Geschäftsprozessen einer Organisation zur betroffenen Person:

- **Schutzbedarf „normal“:**
  Die Schadensauswirkungen sind begrenzt und überschaubar und etwaig eingetretene Schäden für Betroffene relativ leicht durch eigene Aktivitäten zu heilen.
- **Schutzbedarf „hoch“:**
  Die Schadensauswirkungen werden für Betroffene als beträchtlich eingeschätzt, z.B. weil der Wegfall einer von einer Organisation zugesagten Leistung die Gestaltung des Alltags nachhaltig veränderte und der Betroffene nicht aus eigener Kraft handeln kann sondern auf Hilfe angewiesen wäre.
- **Schutzbedarf „sehr hoch“:**
  Die Schadensauswirkungen nehmen ein unmittelbar existentiell bedrohliches, katastrophales Ausmaß für Betroffene an.

((*) https://www.bsi.bund.de/cae/servlet/contentblob/471452/publicationFile/30748/standard_1002_pdf.pdf, S. 49.)

3. Komponenten „personenbezogener Verfahren“

Verfahrens-komponenten

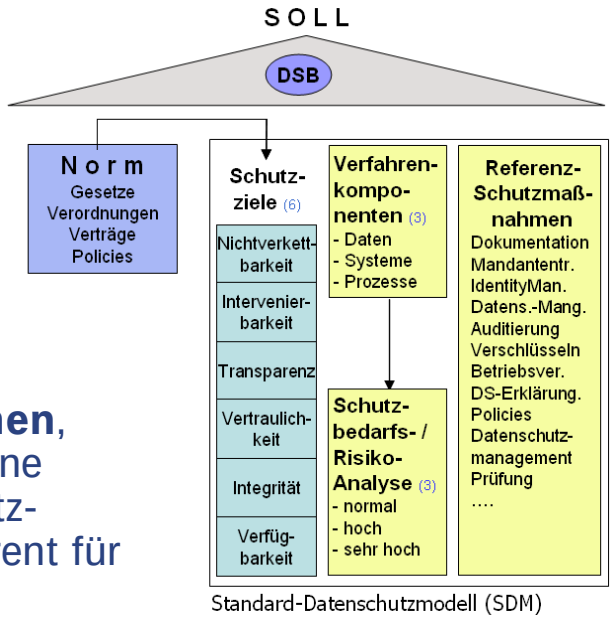
Zu jedem Verfahren mit Personenbezug lassen sich **drei Komponenten** unterscheiden, die einen spezifischen funktionalen Beitrag zur Steigerung / Verringerung des Datenschutz-Risikos leisten:

- Daten (und Datenformate)
- IT-Systemen (und Schnittstellen)
- Prozesse (und adressierbare Rollen)

Nun das Modell in der Gesamtsicht...

Das Standard-Datenschutzmodell (SDM)

- 6 Schutzziele - hinterlegt mit einem Maßnahmenkatalog!
- 3 Schutzbedarfsabstufungen aus der Personenperspektive!
- 3 Verfahrenskomponenten



Standard-Datenschutzmodell (SDM)

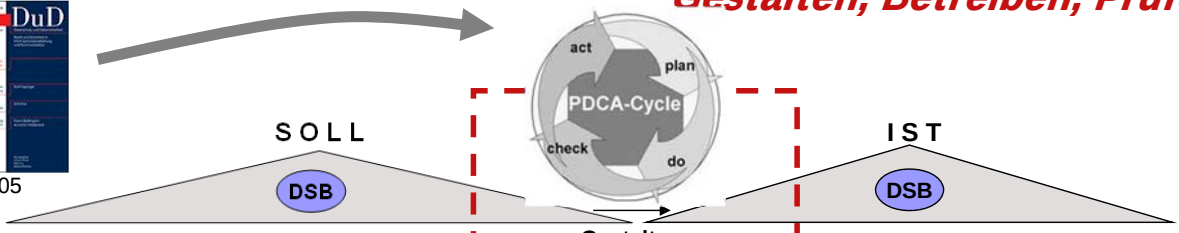
Das ergibt ein Referenz-Datenschmodell mit 54 spezifischen Schutzmaßnahmen, gegen das jedes personenbezogene Verfahren und dessen Datenschutz-maßnahmen integer und transparent für alle Beteiligten prüfbar ist!

	Daten	Systeme	Prozesse
Verfügbarkeit	D1.1 Einschränkung von Lösch-/Veränderungsrechten D 1.2 Schutz vor Schadsoftware D 1.3 Backup der Daten	S 1.1: Schutz vor Schadsoftware S 1.2: Backup von Konfigurationen und Software S 1.3: Hardwareredundanz S 1.4: Ausweichräume, und -Netze	P 1.1: Vertretungspersonal P 1.2: Fähigkeit zur Aufgabenerledigung durch Drit (Vorbereitung Outsourcing) P 1.3: Ausweichprozesse, Amtshilfe
Vertraulichkeit	D 2.1: Einschränkung von Leserechten (für Datenverarbeiter, ggf. durch den Nutzer selbst) D 2.2: Protokollierung lesender Zugriffe D 2.3: Verschlüsselung der Daten D 2.4: Ende-zu-Ende-Verschlüsselung	S 2.1: Einschränkung von lesenden Zugriffsrechten auf IT-Systeme (z. B. Netztrennung durch Sicherheitsgateways) S 2.2: Verschlüsselung auf Systemebene (Festplatten, Datenbank)	P 2.1: Verpflichtung auf das Datengeheimnis (BDS P 2.2: Verschwiegensheitsvereinbarungen P 2.3: Geeignete Organisation bei der Vergabe von Zugriffsrechten („need-to-know“)
Integrität	D 3.1: Einschränkung von Schreib- und Änderungsrechten D 3.2: Protokollierung von schreibenden/ändernden Zugriffen D 3.3: Protokollierung geänderter Daten D 3.4: Nachberichtigung D 3.5: technische Integritätskontrollen (Signaturen/Hashes)	S 3.1: Einschränkung von schreibenden Zugriffen/Konfigurationmöglichkeiten auf IT-Systeme (z. B. Netztrennung durch Sicherheitsgateways) S 3.2: Regelmäßige Integritätsprüfungen/Audits	P 3.1: Detaillierte Planung von Verfahren und Verfahrensschritten P 3.2: Geordnete Zuweisung von Rechten und Rolle P 3.3: Geordnete Änderung von Verfahren und Verfahrensschritten P 3.4: Regelmäßige Überprüfung
Nicht-Verkettbarkeit	D 4.1: Einschränkung von Verarbeitung-s-/Nutzungs- /Übermittlungsrechten für einzelne Daten D 4.2: Kennzeichnung der Zwecke auf Ebene der Daten D 4.3: Einschränkung von identifizierenden Daten; Pseudonymisierung D 4.4: Anonymisierung von Daten	S 4.1: Kennzeichnung der Zwecke auf Ebene des Systeme S 4.2: Trennung von Datenbeständen S 4.3: Einschränkungen von Verarbeitungs-, Nutzungs- und Übermittlungsmöglichkeiten (Funktionalitätseinschränkung) S 4.4: Trennung auf Systemebene (Software, Hardware; Mandantenfähigkeit)	P 4.1: Trennung auf Verfahrensebene P 4.2: Rechte + Rollenvergabe, ggf. an eine andere rechtliche Entität (z. B. Personalvertretung) P 4.3: Gewaltenteilung
Transparenz	D 5.1: Dokumentation der Datenfelder einschließlich Erforderlichkeit D 5.2: Protokollierung von Datenverarbeitungen mit Schutzbedarf zunehmender Detaillierungsgrad und Speicherdauer D 5.3: Integritätsschutz der Protokolle (separater Protokollierungsserver)	S 5.1: Dokumentation der Systeme (Hardware, Software, Algorithmen) S 5.2: Protokollierung von Konfigurationsänderungen S 5.3: zunehmende Kontrolldichte bei höherem Schutzbedarfen; automatisiertes Monitoring	P 5.1: Dokumentation des Verfahren und einzelner Prozesse (einschließlich beteiligter Organisationsein Rollen und Übermittlungen an Dritte) P 5.2: Dokumentation der Änderungsprozesse
Intervenierbarkeit	D 6.1: Schaffung notwendiger Datenfelder (z. B. für Gegendarstellungen)	S 6.1 Funktionalitäten in den Systemen für die Bearbeitung von Sperrungen, Widersprüchen, Beauskunftungen S 6.2 Funktionalitäten in den Systemen für die Umsetzung von weiteren Rechten Betroffener (z. B. Rufnummerunterdrückung, Pseudonyme Nutzungsmöglichkeit, etc.) S 6.3 Funktionalitäten für Betroffene, einzelne Betroffenenrechte direkt wahrzunehmen (z.B. Auskunftsportal, „Datenbrief“, Zusendung von Protokollen, eigene Änderungsmöglichkeiten) S 6.4 Steuerungsmöglichkeiten für einzelne Funktionen („Overrride“) bei automatisierten Einzelentscheidungen S 6.5 Deaktivierungsmöglichkeit einzelner Funktionalitäten ohne Mitleidenschaft für das Gesamtsystem	P 6.1: Organisation der Umsetzung der Betroffenen (Rechte + Rollen für Auskunft, Sperrungen) P 6.2: Organisation der Umsetzung der Betroffenen (Rechte und Rollen bei der Bearbeitung von Gegendarstellungen und Einwänden; Übersteuer automatisierter Einzelfallentscheidungen) P 6.3: Single Point of Contact für Datenschutzfrage

2012/06

SDM
Gestalten, Betreiben, Prüfen

2013/05



Norm
Gesetze
Verordnungen
Verträge
Policies

Schutzziele (6)
- Nichtverkettbarkeit
- Intervenierbarkeit
- Transparenz
- Vertraulichkeit
- Integrität
- Verfügbarkeit

Verfahrenskomponenten (3)
- Daten
- Systeme
- Prozesse

Schutzbedarfs- / Risiko-Analyse (3)
- normal
- hoch
- sehr hoch

Referenz-Schutzmaßnahmen
Dokumentation
Mandantentr.
IdentityMan.
Datens.-Mang.
Auditierung
Verschlüsseln
Betriebsver.
DS-Erklärung.
Policies
Datenschutzmanagement
Prüfung
....

Reale Schutzmaßnahmen, Verantwortlichkeiten
Dokumentation
Mandantentr.
IdentityMan.
Datens.-Mang.
Auditierung
Verschlüsseln
Betriebsver.
DS-Erklärung.
Policies
Datenschutzmanagement
Prüfung, …

Gesicherte Verfahrens-Verantwortlichkeit

Schutz-Mechanismen
Logging, Monitoring
Rollendef. im LDAP
Pseudonymisierung
Anonymisierung
PGP / SSL, TSL
MD5-Hashwertvergl.
Aus-Schalter
Spiegelsystem und Vertretungsregel
Virensc-Signatur…

Rechte/ Pflichten und Rollen
- Rechtsgrundl.
- Betroffener
- Organisation
  - Betreiber
  - Auftr-DV

Betriebssystem
Applikation
Firewall
Virenscanner
Betriebsvereinb.
Verträge mit Dienstleistern
…

Technik und Organisation

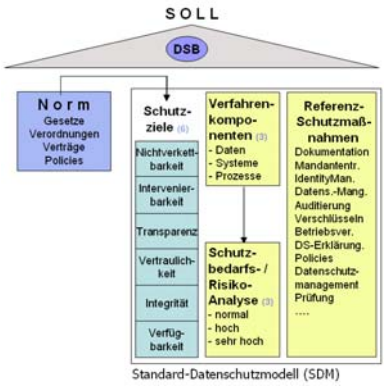
Standard-Datenschutzmodell (SDM)

Tatbestand bzw. Verfahren mit Personenbezug

2012/06

Stand: 2013-0812mr

SDM
Zum aktuellen Status



Standard-Datenschutzmodell (SDM)

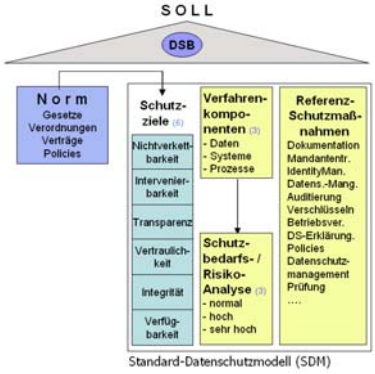
Das SDM wurde bislang u.a. in folgenden Projekten genutzt:

- Datenschutz bei Ambient Assistet Living
- Standardisierungsroadmap AAL im DIN
- Orientierungshilfe Smart Metering
- Cyber Physial Systems
- Datenschutz bei Cloud-Anwendungen
- Privacy Impact Assessment (RFID, BSI)
- Spezifikation des Transportadapters XTA (XÖV-Standard)

Die Konferenz der Datenschutzbeauftragten empfielt die Nutzung und Weiterentwicklung des Modells (DSK, 13.3.2013)

Der „Arbeitskreis SDM“ hat sich konstitutiert. Erste deutschlandweit abgestimmte Ergebnisse werden ab November 2013 vorliegen.

Fragen zu BigData aus Sicht des SDM



Standard-Datenschutzmodell (SDM)

BigData und SDM

Das Standard-Datenschutzmodell zur Analyse von BigData-Analyse heranzuziehen bedeutet:

1. BigData dem Licht der Schutzziele auszusetzen;
2. die Einnahme der Sicht des Betroffenen;
3. BigData-Analyse-Verfahren nach Daten, Prozessen und IT-Systemen zu differenzieren.

Facebook Social Graph

Was Facebooks Graph Search für den Webauftritt bedeutet

Die Vorlieben der Freunde

Frank Puscher

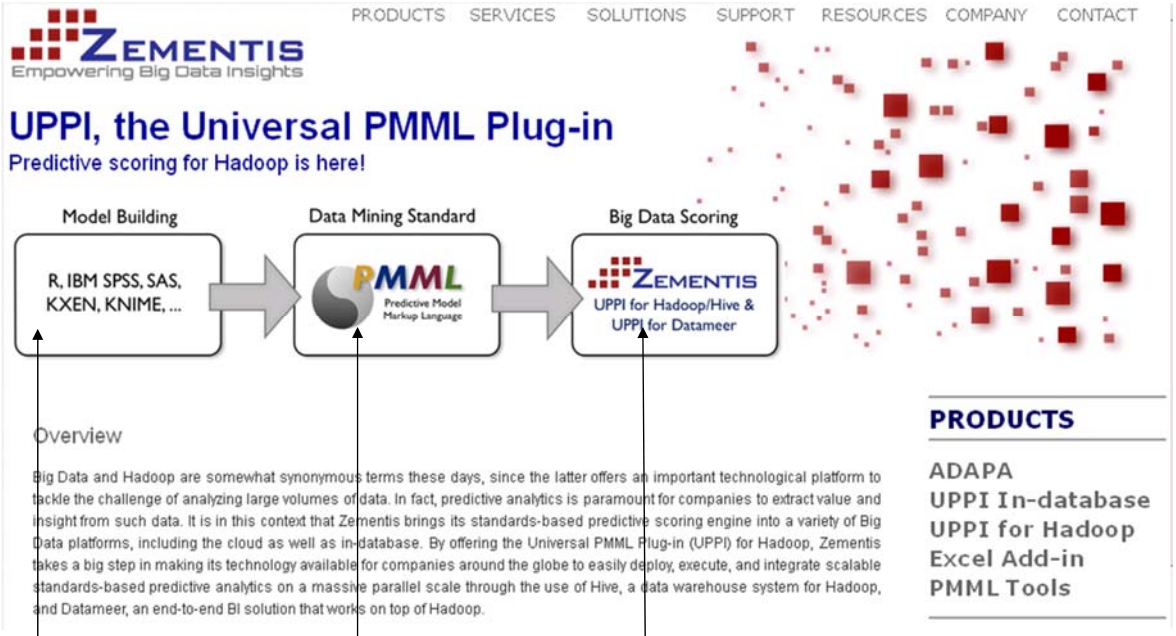
Allmählich tauchen in Deutschland erste Konten auf, in denen Facebooks Graph Search aktiviert wurde. Die neue Suchfunktion ist unglaublich mächtig, mitunter datenschutzrechtlich bedenklich. In ihrem Fahrwasser könnte Microsofts Suchmaschine Bing mehr Bedeutung bekommen.

- Welcher meiner verheirateten Freunde mag die Herbertstraße?
- Welche Seiten mögen die Fans meiner Marke noch?
- Wer hat mein urheberrechtlich geschütztes Foto geteilt?

Solche und ähnliche Fragen dürften die Nutzer in naher Zukunft an Facebook stellen. Nicht an den Kundendienst, sondern an Facebooks neue Suchfunktion Graph Search, die derzeit in den USA im Betastadium läuft. Experten erwarten, dass sie in zwei bis drei Monaten allen Amerikanern zur Verfügung stehen wird. Wann und in welcher Konfiguration sie nach Deutschland kommt, ist derzeit noch nicht klar. Absehbar ist hingegen, dass die Datenschützer dagegen Sturm laufen werden.

Quelle: http://www.heise.de/ix/artikel/Die-Vorlieben-der-Freunde-1936983.html

Was meint Big Data?



Mächtige Modellierungs- und Statistiktools für „predictive analytics“

PMML
standardisiertes Austauschformat

Hadoop / Hive
Massiv parallele Speicherung/ Verwaltung von Datenmassen mit SQL Standard-Methoden auf billiger Standard-Hardware

NOSQL
Die Performanz der neuen Datenbank-Generation

Menge der Personen

Partner and Vukotic's experiment seeks to find friends-of-friends in a social network, to a maximum depth of five. Given any two persons chosen at random, is there a path that connects them that is at most five relationships long? For a social network containing 1,000,000 people, each with approximately 50 friends, the results strongly suggest that graph databases are the best choice for connected data, as we see in Table 2-1.

Table 2-1. Finding extended friends in a relational database versus efficient finding in Neo4j

Depth	RDBMS execution time (s)	Neo4j execution time (s)	Records returned
2	0.016	0.01	~2500
3	30.267	0.168	~110,000
4	1543.505	1.359	~600,000
5	Unfinished	2.132	~800,000

Friends-of-Friends Pfadtiefe: 5 Schritte

Oracle / MySQL: „Unfinished“

Neo4j: 2.2 Sekunden!

Das Ergebnis

At depth two (friends-of-friends), both the relational database and the graph database perform well enough for us to consider using them in an online system. Although the Neo4j query runs in two-thirds the time of the relational one, an end user would barely notice the difference in milliseconds between the two. By the time we reach depth three (friend-of-friend-of-friend), however, it's clear that the relational database can no longer deal with the query in a reasonable time frame: the 30 seconds it takes to complete would be completely unacceptable for an online system. In contrast, Neo4j's response time

Quelle: Robinson et al, 2013: Graph Databases, S. 20

BigData

BigData bezeichnet...

- ... allgemein *besonders große Datenmengen*, die per IT zugänglich sind.
- ... im Kontext des Datenschutzes besonders große Datenmengen mit Personenbezug, die aus beliebigen Quellen stammen können und die Organisationen (Unternehmen, Behörden, Interessensverbände, Wissenschaftsinstitute) automatisiert verarbeiten.

BigData-Analyse („Data Mining im Data Warehouse“) bezeichnet...

- ... Analyseverfahren, die eine Grundlage für Entscheidungen zur Gestaltung von Verfahren bilden, mit denen Organisationen ihre Geschäftsprozesse in Bezug zu ihrem Klientel (Bürger, Kunden, Patienten, Klienten, Menschen, Individuen, Subjekte) effektivieren.

Big Data
Stakeholder, Objektbereich

- Abstrakt: Erstellung von Prognosen und Trends, auch Visualisierung von Strukturen, Simulation und Optimierung von Prozessen (SocialWeb, Markt- und Sozialforschung, Sicherheitsbehörden). Der Zweck: *Erforschung nicht-offensichtlicher, komplexer Zusammenhänge menschlichen Verhaltens.*
- Konkret: Bonität-Scorings, Auswertung von Geo-, Audio-(Sprach), Video-und Sozial-Daten, Lagebeurteilungen bei Armee, Polizei, Nachrichten- bzw. Geheimdiensten. Der Zweck: *Sicherung des flexiblen Zugriffs auf Massendaten, zum Zweck des individualisierten Zugriffs.*

Big Data Vertrauenswürdigkeit?

1. **Standardisierte Personenmodelle** ermöglichen Unternehmen, Behörden, Instituten Verhaltensprognosen und die Industrialisierung der Interaktions-formen des Umgang mit Personen (= „automatisierte Einzelentscheide“).
2. **Jederzeit möglicher Zugriff auf Daten einer Einzelperson** bedeutet Vollüberwachung von Kunden, Bürgern, Patienten, Mitarbeitern.

Beherrschen die Organisationen die BigData-Prozesse nachgewiesenermaßen funktional korrekt?

Ist **Fairness** bei der faktischen Vollüber-wachung gegeben? Wer bestimmt die Regeln?

These: Der **Nachweis(!)** der *Beherrschbarkeit* von BigData und insbesondere BigData-Analyse sowie die *faire* (= zumindest rechtskonforme) Verarbeitung der Daten und insbesondere die Nutzung der Ergebnisse sind die maßgeblichen Voraussetzungen für deren **Vertrauenswürdigkeit.**

BigData Verfügbarkeit?

Verfügbarkeit adressiert das Problem, dass die Entscheidungen nach BigData-Analysen durch Behörden, Unternehmen und Wissenschaftsinstitute das Risiko der Fremdbestimmung von Bürgern, Kunden, Individuen und Patienten steigern.

- Gibt es einen verallgemeinerbaren, vernünftigen Nutzen aus BigData-Analysen?
- Wem **„gehören“ die Daten**, Ergebnisse, Folgerungen, Entscheidungen und Anwendungen aus BigData-Analysen?
- Wer bestmmt die Regeln, mit denen bspw. **Risikoschwellen** (für wen?) bestimmt werden?
- Oligopolunternehmen (wie Microsoft, IBM, HP, Cisco, Intel...) verfügen wahrscheinlich durchgängig über **Backdoors, um sich Zugriff** auf die von ihnen technisch kontrollierten IT-Systeme zu verschaffen („Dennis-Ritchie-Backdoor“).
- Der Staat kann sich letztlich Zugriff auf alle Daten und deren Auswertungen verschaffen (Gewaltmonopol).

Mögliche Entschärfung der Risiken: Politischer Diskurs mit breiter Beteiligung, Verrechtlichung und gesellschaftliche Kontrolle der Nutzung von BigData.

BigData Integrität?

Integrität adressiert das Problem der Sicherung der Korrektheit der Daten und deren Berechnungen, der Programme sowie vor allem der Folgerungen aus den Analysen bzw. den Entscheidungen, die Organisationen aus den Ergebnissen ableiten und in Aktivitäten gegenüber Personen transformieren.

BigData-Analysen erzeugen statistische **Korrelationen, keine Kausalitäten**.

Die wissenschaftlich meist mangelhafte Qualität bzgl. Validität und Reliabilität der Daten sowie die Methoden der Prognoseerstellungen bleibt mit allen negativen Folgen für die Betroffenen bestehen, solange auch das Scheitern der Modellierungen nicht wissenschaftlich gesichert festzustellen ist und die Konkurrenz auf einem ähnlich niedrigen Niveau agiert.

Mögliche Entschärfung der Risiken:

- Festlegung und fortgesetztes Monitoring wissenschaftlicher Standards zu Methodiken des Datenerhebens, der Auswertung, der Transformation von Ergebnissen in Entscheidungen, von Entscheidungen in Aktivitäten der Organisationen.
- Die Zweckbindung zumindest im Auswertungsvorgang erfordert Kausalität. Die Verbindung zwischen Kausalität und Korrelation muss durch explizite theoretische Annahmen oder Modellierungen erfolgen.

BigData Vertraulichkeit und Nichtverkettbarkeit?

Vertraulichkeit und Nichtverkettbarkeit adressieren beide das Problem, dass BigData-Analyse im Wesentlichen gleichbedeutend ist mit dem beliebigen Zugriff beliebiger Interessenten auf beliebige personenenbezogene Daten und dem Durchbrechen einer verallgemeinerungsfähig definierten Zweckbindung.

Weder eine gesicherte vertrauliche Verarbeitung durch Hochtreiben der IT-Sicherheit noch eine **Anonymisierung** personenbezogener Daten durch hochgetriebene Anonymisierungstechniken verhindern die Auswirkungen von BigData-Analysen auf die Betroffenen.

Eine ggfs. wirksame Anonymisierung im Erhebungszusammenhang ist durch Kontextierungen durch Verschneiden mit anderen (anonymiserten) Daten im Auswertungszusammenhang im Ergebnis für den Betroffenen wirkungslos.

Mögliche Entschärfung der Risiken:
Die Nutzung von BigData-Analysen wäre so einzurichten, dass sich auf dem großen Datenpool nur mit enger Zweckbindung Analysen durchführen lassen. (Prinzip „Zerberus“). Das Motiv, die Zweckbindung, die Methodik und Durchführung der Analyse sind dabei fortwährend zu kontrollieren. Dies setzt wiederum ein reifes Datenschutzmanagementsystem voraus.

BigData Transparenz?

Transparenz adressiert das Problem, dass BigData-Analyse verspricht, zumindest theoretisch auf eine beherrschbare Weise Transparenz in intransparent-komplexe Zusammenhänge zu bringen. Die Komplexität der Daten, der Analysemethoden und Applikationen bleiben jedoch praktisch ebenso intransparent wie letztlich die Nutzung der Ergebnisse, die Profizeure und die unabsehbaren Folgen für die Betroffenen.

- So weisen bspw. Faktorenanalysen nur Clusterbildungen auf. Die Behauptung von Ursachen für diese Cluster erfordert Modellierungen auf Theoriebasis. Methoden und Modellierung können Organisation berechtigt als Geschäftsgeheimnis bzw. nötig zur Sicherung der öffentlichen Ordnung deklarieren.
- Genetische Algorithmen und Neuronale Netze erzeugen „selbstgewichtete“ Berechnungen, so dass trotz rechtlich korrekter Dokumentation trivialer Algorithmen der Zusammenhang von **Ursache und Wirkung eines Systems im Unklaren** bleibt.

Mögliche Entschärfung der Risiken: Im Rahmen von Datenschutz-Audits bzw. Datenschutz-Impact-Assessments, durch Hinzuziehen von Spezialisten, sisyphos-artig Transparenz der Daten, der Methoden, der Prozesse, der Ergebnisse sowie der organisatorischen Zusammenhänge der Stakeholder zumindest verbessern.

BigData Intervenierbarkeit?

Intervenierbarkeit adressiert das Problem der Steuerung von BigData-Prozessen sowohl bei der Erhebung der Daten, als auch bei der Analyse und Verarbeitung sowie der Umsetzung der Ergebnisse in Aktivitäten.

Eine betroffene Person hat **keinerlei Chancen auf eine aussichtsreiche Intervention** in den verschiedenen Stadien von BigData-Verfahren. Das Vorenthalten bspw. einer Einwilligung verhindert nicht die Erzeugung von Ergebnissen und deren Anwendung auf die Person.

Mögliche Entschärfung der Risiken:

Organisationen sind gehalten, die Beherrschbarkeit und faire Umsetzung der Ergebnisse sowohl in den Personenmodellen als auch bei der Gestaltung der Informationsverarbeitung und Kommunikationen nachzuweisen. Dies ist eine wichtige Voraussetzung, um auf (Veränderungen) rechtliche Anforderungen reagieren zu können.

Fazit

- Der Ansatz von BigData und BigData-Analysen **verstößt gegen Grundrechte** bzw. Datenschutzrecht, weil die Anforderungen nach Vertraulichkeit, Integrität, Zweckgebundenheit, Transparenz, Intervenierbarkeit durchgängig nicht erfüllt werden (können).
- **Weder Anonymisierung noch wissenschaftlich korrekte Methoden bieten Schutz** vor organisierter Fremdbestimmung.
- Organisationen verringern mit BigData die Risiken, die durch Freiheit des Bürgers und Kunden entstehen, einseitig zu ihren Gunsten. **Organisationen gefährden insofern die Konzepte der Gewaltenteilung und Demokratie, des Marktes und der freien Diskurse** der modernen Weltgesellschaft, die wiederum die Quellen der Freiheit bilden.

22.08.2013 08:48 « Vorige | Nächste »

Juristen zu PRISM und Tempora: Aushöhlung der Grundrechte

vorlesen / MP3-Download

Auf dem Rechtsweg können Bundesbürger nur schwer vor dem Ausspähen durch ausländische Geheimdienste geschützt werden. Das ist das Ergebnis eines juristischen Fachgesprächs der Fraktion der Grünen im Bundestag zum US-Überwachungsprogramm PRISM und seinem britischen Ableger Tempora. Verfassungsrechtler etwa sehen zwar einen klaren Auftrag der Bundesregierung, Bürger vor einer anlasslosen und flächendeckenden Bespitzelung zu bewahren. Dessen Umsetzung sei aber schwierig, da sich aus der Vorgabe keinen konkreten Folgen ableiten ließen.

Durch das verdachtsunabhängige Sammeln von Datenströmen werde der Grundrechtsschutz völlig ausgehöhlt, erklärte die frühere Richterin am Bundesverfassungsgericht Lerke Osterloh. Soweit "ausländische Figuren" entsprechend im Inland tätig würden, müsse die Politik direkt dagegen einschreiten. Da hälfe auch kein Verweis auf Möglichkeiten zum Selbstschutz etwa durch Verschlüsselung, führte die Juristin aus. Ein solches Verfahren sei oft übermäßig aufwändig und der Wettlauf zwischen Kryptographen und Codeknackern verlaufe immer zu Lasten der Mehrheit der Anwender.

Quelle: http://www.heise.de/newsticker/meldung/Juristen-zu-PRISM-und-Tempora-Aushoehlung-der-Grundrechte-1940114.html

Forderungen aus Datenschutzsicht

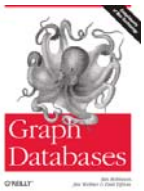
- Die **Aufgabe des Staates** besteht darin, vertrauenswürdige und sichere IT-Systeme und Infrastrukturen als Grundlage von Grundrechtsausübung zu gewährleisten.
- **Es sind deshalb Regelungen zu treffen** in Bezug auf
  - *Beschränkung der Zwecke*, zu denen BigData-Analysen eingesetzt werden dürfen;
  - *Beschränkung der Daten*, die mit BigData-Analysen erhoben und verarbeitet werden dürfen;
  - *Beschränkung der Methoden* zur Analyse von BigData-Datenbeständen;
- **Kontrolle der Umsetzung in der Praxis** durch
  - *Datenschutz-Impact-Assessments;*
  - *Genehmigungsvorbehalte*;
  - *Prüfungen / Auditierungen* im Rahmen reifer interner Datenschutzmanagementsysteme sowie einer wirkungsvollen externen Datenschutzaufsicht.

Literatur

Big Data

Sven Fessler, 2013: Die Datenflut steigt – wie können wir sie nutzen?
http://www.dataport.de/ueber-uns/publikationen/Seiten/Datareport%202013-1/2013-1-Datenflut-steigt.aspx

Robinson, I.; Webber, J.; Eifrem, E., 2013: Graph Databases, O'Reilly Media
http://www.goodreads.com/book/show/17465372-graph-databases



Datenschutzrechtliche Analyse von Big Data

Thilo Weichert 2013: Big Data und Datenschutz
https://www.datenschutzzentrum.de/bigdata/20130318-bigdata-und-datenschutz.pdf

Juristen zu PRISM und Tempora: Aushöhlung der Grundrechte
http://www.heise.de/newsticker/meldung/Juristen-zu-PRISM-und-Tempora-Aushoehlung-der-Grundrechte-1940114.html

Standard-Datenschutzmodell

Martin Rost, 2012: Standardisierte Datenschutzmodellierung
in: DuD 2012/06: 433-438, http://www.maroki.de/pub/privacy/2012-06-DuD-SDM.html

Weitere, verwendete Literatur

Atteslander, Peter, 2000: Methoden empirischer Sozialforschung, 11. Aufl., de Gruyter

Bock, Kirsten; Meissner, Sebastian: Datenschutz-Schutzziele im Recht;
in: DuD 2012/06: 425-431.

Probst, Thomas: Generische Schutzmaßnahmen für Datenschutz-Schutzziele;
in: DuD 2012/06.: 439-444

Rost, Martin / Pfitzmann, Andreas: Datenschutz-Schutzziele - revisited;
in: DuD 2009/06: 353-358

Rost, Martin: Impact Assessment im Lichte des Standard-Datenschutzmodells;
in: DuD 2012/10: 472-477.

Rost, Martin: Zur Konditionierung von Technik und Recht;
Tagungsband der GI 2013 (im Erscheinen)

Rost, Martin: Zur Soziologie des Datenschutzes;
in: DuD 2013/02: 85-91

Rost, Martin: Datenschutzmanagementsystem;
in: DuD 2013/05: 295-300

Vielen Dank für Ihre Aufmerksamkeit!

ULD
Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein

Unabhängiges Landeszentrum für
Datenschutz Schleswig-Holstein
Martin Rost
Telefon: 0431 988 – 1200
uld32@datenschutzzentrum.de
http://www.datenschutzzentrum.de/