Big Data, IT-Sicherheit und Datenschutz in Hochschulen

Prof. Dr. Hannes Federrath
Sicherheit in verteilten Systemen (SVS)
http://svs.informatik.uni-hamburg.de

Campus Innovation 2014, Jahrestagung des Universitätskollegs, Universität Hamburg, 20. November 2014

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Im Vortrag geht es um…

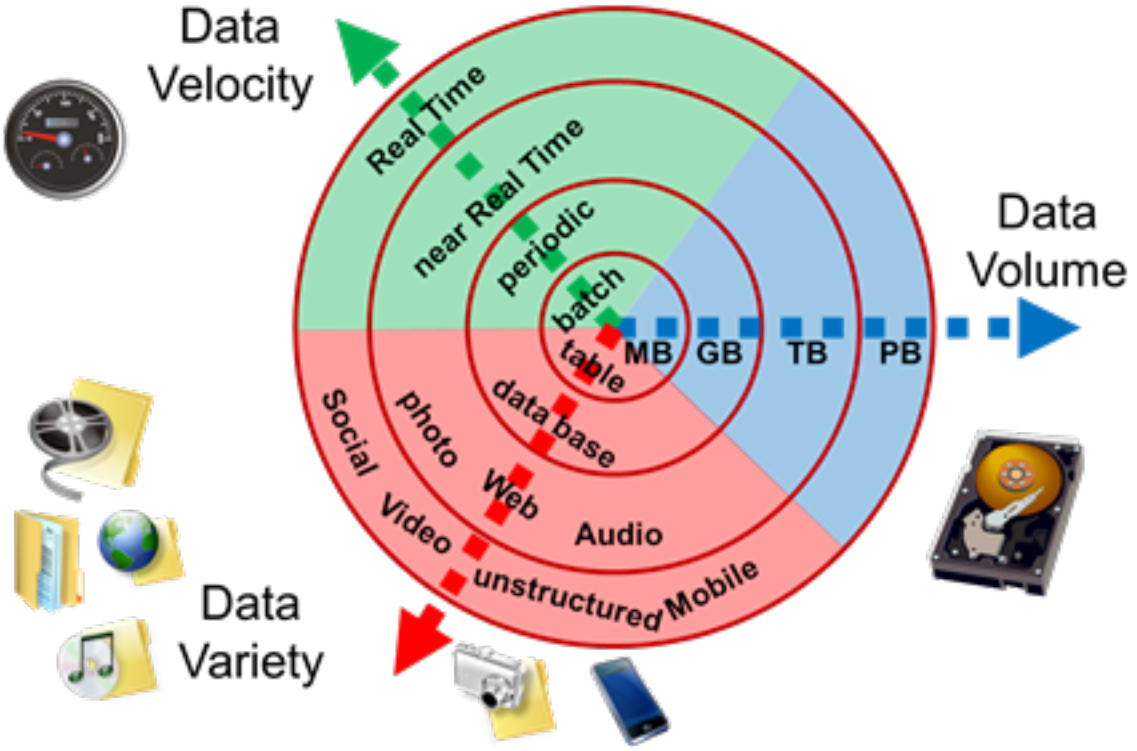
- Möglichkeiten der Datenanalyse – heute und künftig – und damit verbundenen Verletzungen der Privatsphäre im Zusammenhang mit großen Datenmengen

- Grundsätzlichen Anforderungen an den Datenschutz und Konzepte der vorbildlichen gesetzeskonformen Gestaltung von Informationsdiensten

- IT-Sicherheit als Voraussetzung für vertrauenswürdige und nachhaltige Gestaltung von Informations- und Kommunikationsprozessen

Schlagworte im Zusammenhang mit Big Data

- Always Online
- Cloud Computing
- Bring Your Own Device
- Internet of Things
- Data Mining



Quelle: http://www.gi.de/nc/service/informatiklexikon/detailansicht/article/big-data.html

Big Data – Fachbegriff oder Marketing Instrument

- Definition:

  Big Data bezeichnet Datenmengen, die

  - zu groß sind, oder/und
  - zu komplex sind, oder/und
  - sich zu schnell ändern und daher

  mit herkömmlichen Datenbanksystemen und Datenverarbeitungssystemen nicht mehr effektiv gespeichert und verarbeitet werden können oder durch Anwendung neuer Methoden neue Erkenntnisse aus diesen Datenmengen gewonnen werden können.

  in Anlehnung an http://de.wikipedia.org/wiki/Big_Data

Hochschule als Laborumgebung

These: Die Hochschule als Laborumgebung zur experimentellen Analyse neuartiger Phänomene und Erprobung unausgereifter Verfahren und Prozesse

- Big Data
  - Steuerungsprozesse gestalten: z.B. Fächeranalysen nach Beliebtheit, Studierendenströmen, Verbleibsanalysen
- Datenschutz
  - Denkbare Architektur: Datenvorverarbeitung (Anonymisierung, Aggregation/Akkumulation, Normalisierung) an der Datenquelle, Online-Partizipation und -Meinungsbildung, Ab- und Mitbestimmungsprozesse
- IT-Sicherheit
  - Single-Sign-On in sehr heterogenen, teils anarchischen IT-Umgebungen; Verschlüsselung für alle Hochschulmitglieder

Zukunftsfragen

- Wie wird Big Data und IT in Zukunft…
  - unsere Kommunikation prägen?
  - unsere Arbeit prägen?
  - unser Privatleben prägen?
- Oder eben allgemeiner:
  - unsere Gesellschaft prägen?

Big Data als Instrument für

- Operations Research
- Business Intelligence
- Behavioral Targeting
- Human Resources Management
- Signal Intelligence

Typische Zukunftszenarien: Mitarbeiter und Bedienstete

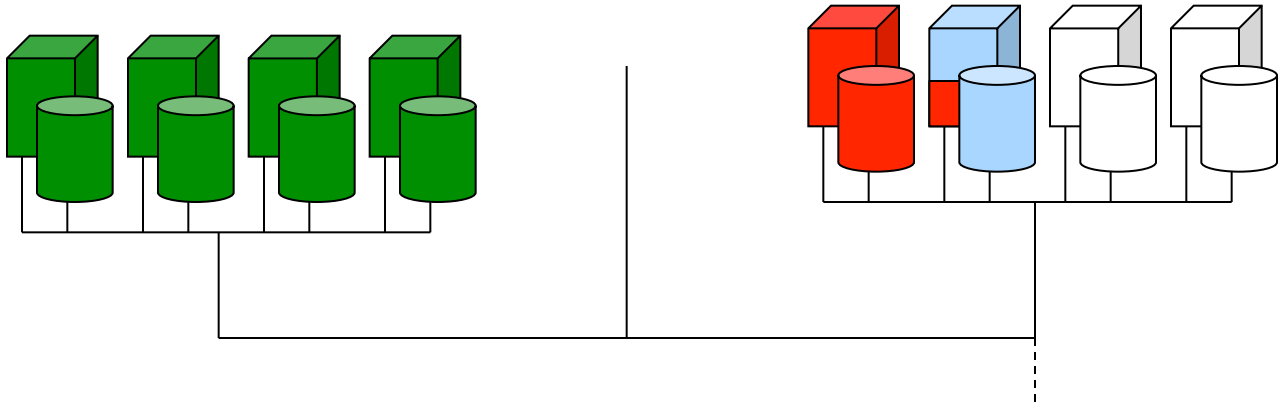
- Die Mitarbeiterin/der Mitarbeiter ist ständig erreichbar. Das «Goodie» ist ein Gerät, das nicht nur dienstlich, sondern auch privat benutzt werden darf/soll/muss.
- Alle betrieblichen und privaten Informationen sind für alle Mitarbeiter stets verfügbar.
- Abschalten wird erlaubter Luxus. Auch das wird aufgezeichnet.
- Präsenz im Büro nimmt ab; zeitliche und örtliche Flexibilität gehört zu den erlaubten und erwarteten Eigenschaften eines «guten Mitarbeiters».
- Vorgesetzte können nur dann adäquat auf ihre Mitarbeiter eingehen, wenn sie genügend Informationen über deren Lebensumstände haben.

Typische Zukunftszenarien: Studierende

- Studierende speichern und greifen alle Informationen über ein mobiles Gerät ab, das nicht nur fürs Studium, sondern auch privat benutzt wird.
- Alle Hochschulinformationen und privaten Daten sind für alle Studierenden stets verfügbar.
- Abschalten findet nicht mehr statt. Man könnte ja etwas verpassen.
- Präsenz in der Hochschule nimmt ab. Kurse werden Online besucht, Prüfungsleistungen werden Online erbracht und abgenommen.
- Lehrende können nur dann adäquat auf ihre Studierenden eingehen, wenn sie genügend Informationen über deren Lebensumstände haben.

Always Online

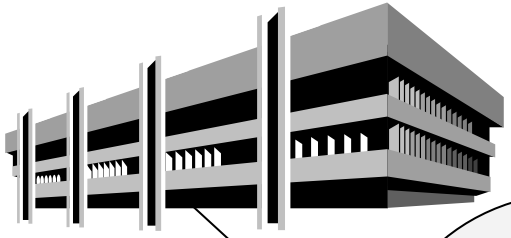
- Verschiebung des Speicherorts von Daten vom lokalen Datenträger ins Netz
- Datenträger als Massenspeicher verlieren ihre Bedeutung
- Daten werden nur noch einmal beim Cloud-Anbieter langfristig gespeichert und bei Bedarf abgerufen (Streaming)
- Duplikate gleicher Inhalte unabhängiger Nutzer werden erkannt und nur einmal gespeichert
- Datenfragmente werden an vielen Orten im Netz kurzzeitig (von Millisekunden bis zu Tagen) zwischengespeichert (Caching)



Cloud Computing

- Speicherort kann sich dynamisch ändern
- Es kann mehrere Speicherorte geben
  - Effizienz, Datensicherheit

z.B. Dropbox,
Google Docs
...



vm1
vm2
...
z.B.

Europa

Amerika

Asien

Bring Your Own Device

– + ? Legalisiert die Benutzung eigener Geräte für dienstliche Aufgaben

+ – Zwingt Nutzer, die saubere Trennung von privater und beruflicher IT-Nutzung aufzugeben

  - Gerät kann ferngewartet, -gelöscht, -beobachtet werden

  - Private Nutzung des Internets am Arbeitsplatz:
    - meist verboten wegen technischer Schwierigkeiten in der Umsetzung des Fernmeldegeheimnisses
    - Proxies, Filter, Log-Files machen private Kommunikation beobachtbar

- Technisch wird es jetzt noch viel schwieriger werden, Privatheit umzusetzen, es sei denn, eine Organisation ist auch bereit, Kontrolle aufzugeben

Sensorik in Smartphones ermöglicht neue Apps

- GPS
- WiFi
- GSM/3G/LTE
- Bluetooth
- Mikrofone
- Kameras
- Zeit
- Temperatur, Luftdruck
- Bewegungs- und Lagesensoren
- Beschleunigungssensoren
- Anschlussmöglichkeiten für weitere Sensoren
  - Persönliche: Herzschlag, Atemfrequenz, Muskelkontraktion, Blutzucker, …
  - Umgebung
    - Autos: CAN-Bus-Adapter
    - Haus: Smart Meter, Heizung, Alarmanlage

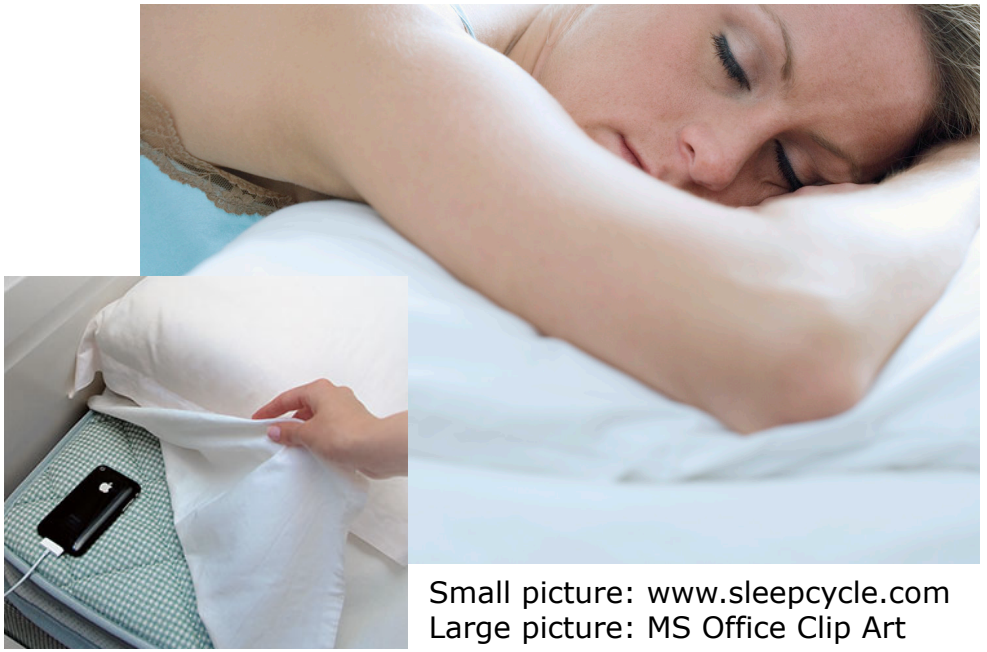
http://blog.digifit.com/wp-content/uploads/2011/02/

Appification

- Für jeden Zweck eine eigene App
  - Taxi
  - Weather
  - Wikipedia
  - Shopping list
  - Writing app
  - Notebook
  - Doc scanning
  - Sleep rhythm
  - Running app

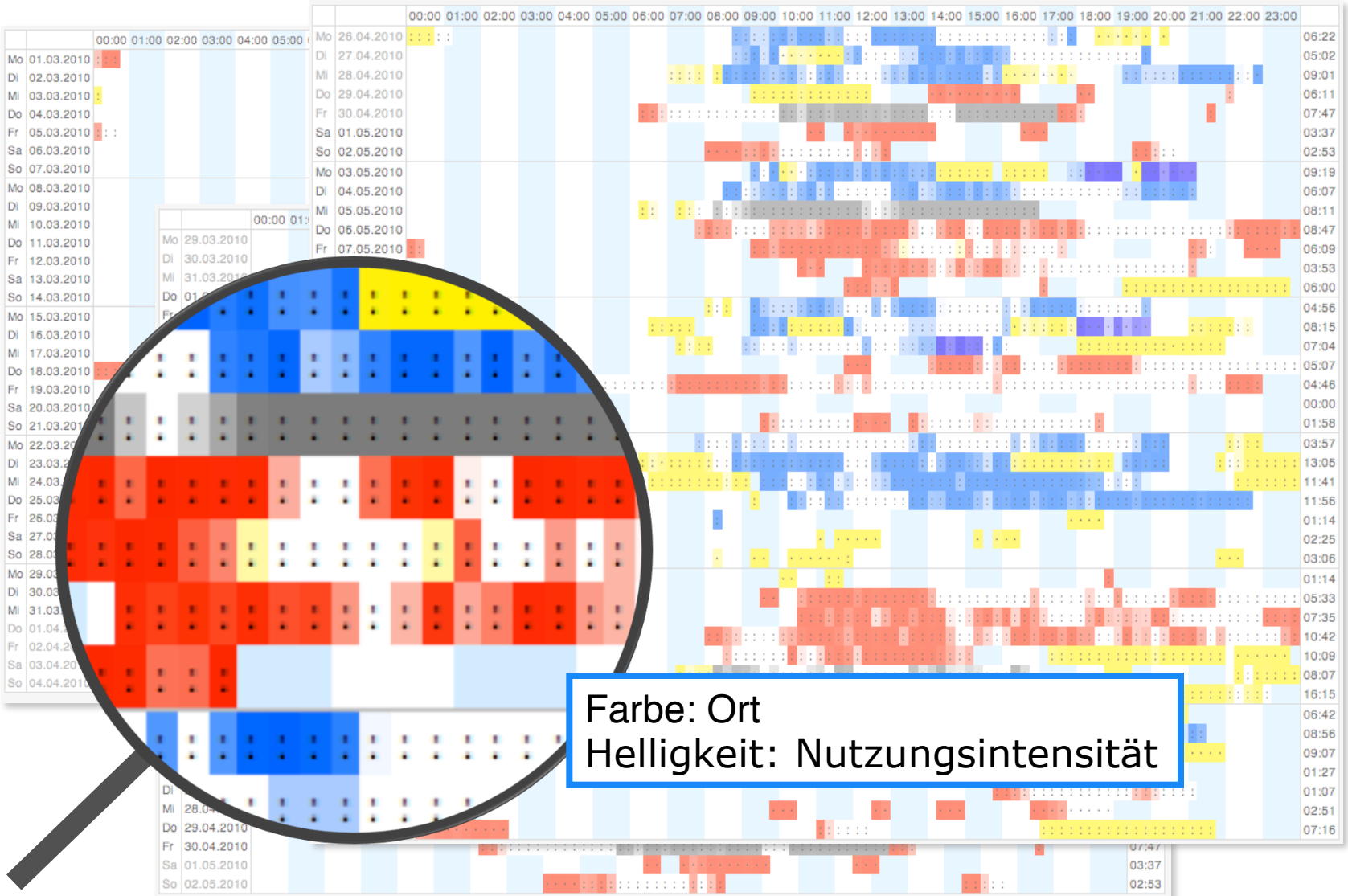
    …

  - Video apps (product advertisement)
  - Torch apps



Small picture: www.sleepcycle.com
Large picture: MS Office Clip Art

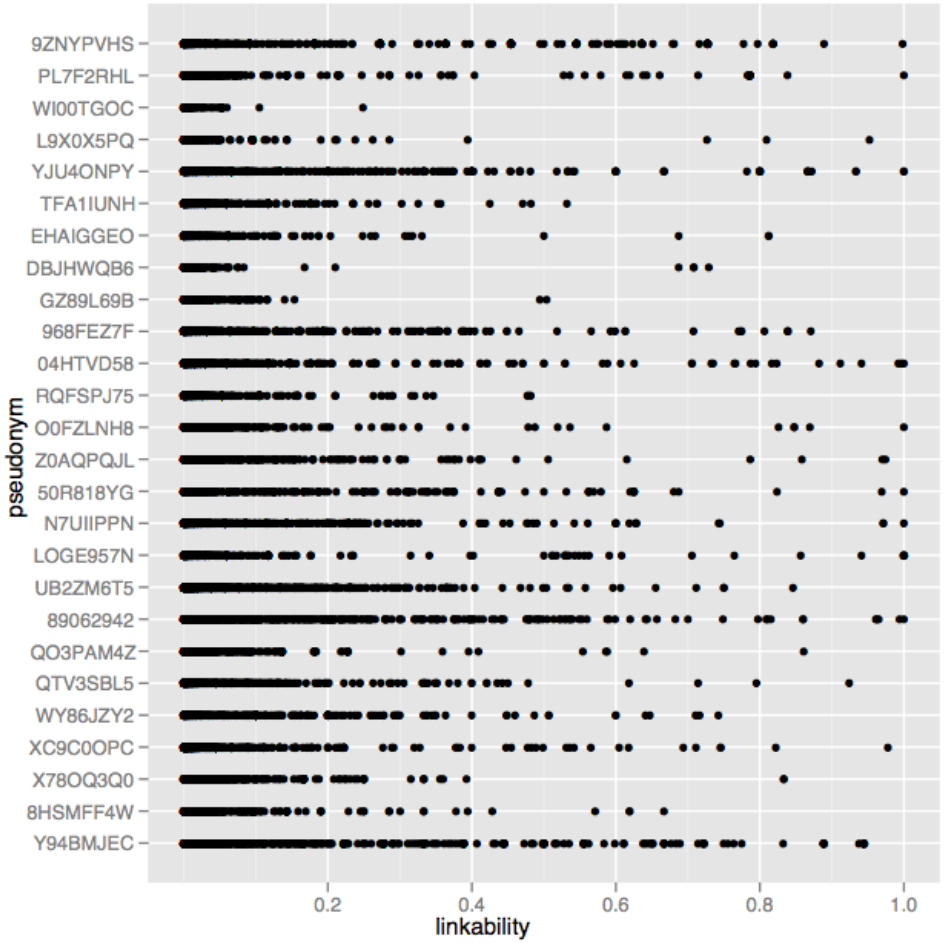
Usage Timelines inkl. ip-geo-tagging



Data Mining und Big Data

- Analyse der Daten auf eine bestimmte Fragestellung hin
- Big Data: Vor Auswertung keine Vorverarbeitung (Aufbereitung) der Daten. Gründe:
  - nicht möglich
  - zu ineffizent
  - statistisch nicht signifikant
- Analyse von Log-Files,
- Verknüpfung von Datenbanken
- Maschinelles Lernen
- Komplexitätstheorie kann auch bei Big Data nicht ausgeschaltet werden
  - Kryptographie: Knacken von Schlüsseln bleibt exponentiell

Bild: Gerber 2009



§ 3a BDSG · Datenvermeidung und Datensparsamkeit

«Gestaltung und Auswahl von Datenverarbeitungssystemen haben sich an dem Ziel auszurichten, keine oder so wenig personenbezogene Daten wie möglich zu erheben, zu verarbeiten oder zu nutzen. Insbesondere ist von den Möglichkeiten der Anonymisierung und Pseudonymisierung Gebrauch zu machen, soweit dies möglich ist und der Aufwand in einem angemessenen Verhältnis zu dem angestrebten Schutzzweck steht.»

- Konsequenzen:
  - Daten, die nicht erhoben werden, können nicht gespeichert und ausgewertet werden.
  - Sensorik: Es sind inzwischen mehr Daten verfügbar, als uns bekannt, uns recht, gesetzlich erlaubt ist.

Recht auf informationelle Selbstbestimmung

»Freie Entfaltung der Persönlichkeit setzt unter den modernen Bedingungen der Datenverarbeitung den *Schutz des Einzelnen gegen unbegrenzte Erhebung, Speicherung, Verwendung und Weitergabe seiner persönlichen Daten* voraus. …

Wer nicht mit hinreichender Sicherheit überschauen kann, welche ihn betreffenden Informationen in bestimmten Bereichen seiner sozialen Umwelt bekannt sind, und wer das Wissen möglicher Kommunikationspartner nicht einigermaßen abzuschätzen vermag, kann in seiner Freiheit wesentlich gehemmt werden, aus eigener Selbstbestimmung zu planen oder zu entscheiden. *Mit dem Recht auf informationelle Selbstbestimmung wäre eine Gesellschaftsordnung nicht vereinbar, in der Bürger nicht mehr wissen können, wer was wann und bei welcher Gelegenheit über sie weiß.«*

aus dem Volkszählungsurteil des Bundesverfassungsgerichts vom 15. Dezember 1983 1. BvR 209/83 Abschnitt C II.1, S. 43

Recht auf informationelle Selbstbestimmung

Mit dem Begriff Datenschutz wird das Recht des Einzelnen auf informationelle Selbstbestimmung umschrieben.

- Recht auf informationelle Selbstbestimmung = Grundrecht

- Herleitung des Rechts auf informationelle Selbstbestimmung
  - aus dem Allgemeinen Persönlichkeitsrecht gemäß Art. 2 Abs. 1 i.V.m. Art. 1 Abs. 1 Grundgesetz (GG) durch das

    Bundesverfassungsgericht im Volkszählungsurteil

- «Volkszählungsurteil» des Bundesverfassungsgerichts vom 15.12.1983:
  - «Das Grundrecht gewährleistet [...] die Befugnis des Einzelnen, grundsätzlich selbst über die Preisgabe und Verwendung seiner persönlichen Daten zu bestimmen.»

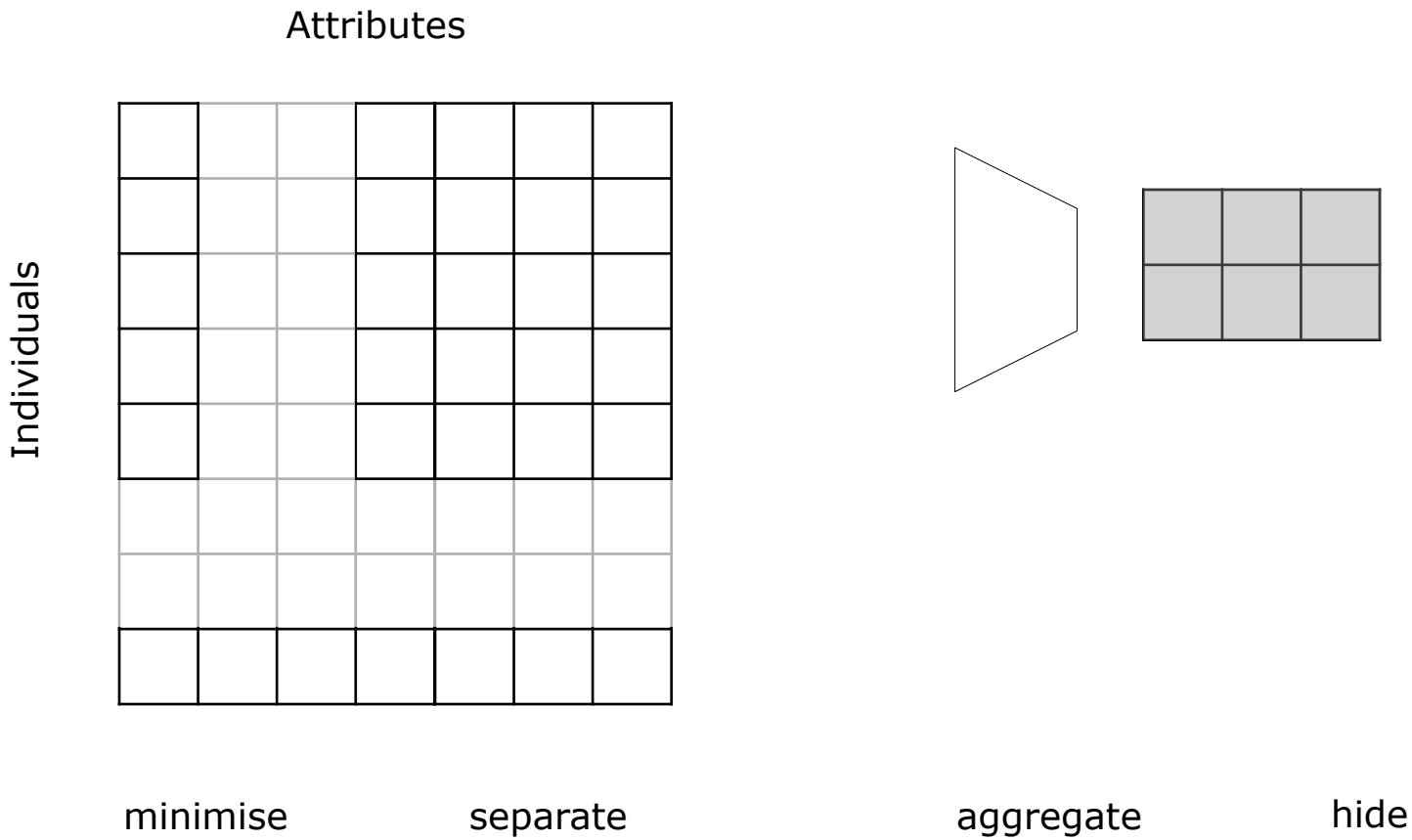
Goldene Regeln zur Umsetzung von Datenschutz

- Aus Sicht der IT-Sicherheit:
  - Informieren (Transparenz)
  - Auskunftsverfahren etablieren
  - Einwilligung, wo nötig
  - Weniger (speichern) ist mehr (Datenschutz)
  - Regelmäßige Sensibilisierung (wie Umwelt- und Arbeitsschutz)
  - Sanktionen bei Verstößen klarmachen
  - Aber: Kontrollieren und beraten, nicht gleich bestrafen!

- Immer fragen: Was ist die Grundlage der Erhebung/Verarbeitung/Speicherung?
  - Einwilligung?
  - Gesetzliche Vorgabe?
  - Aufrechterhaltung des laufenden Betriebs? (IT-Sicherheit)

Privacy design strategies

nach: Hoepmann, 2003



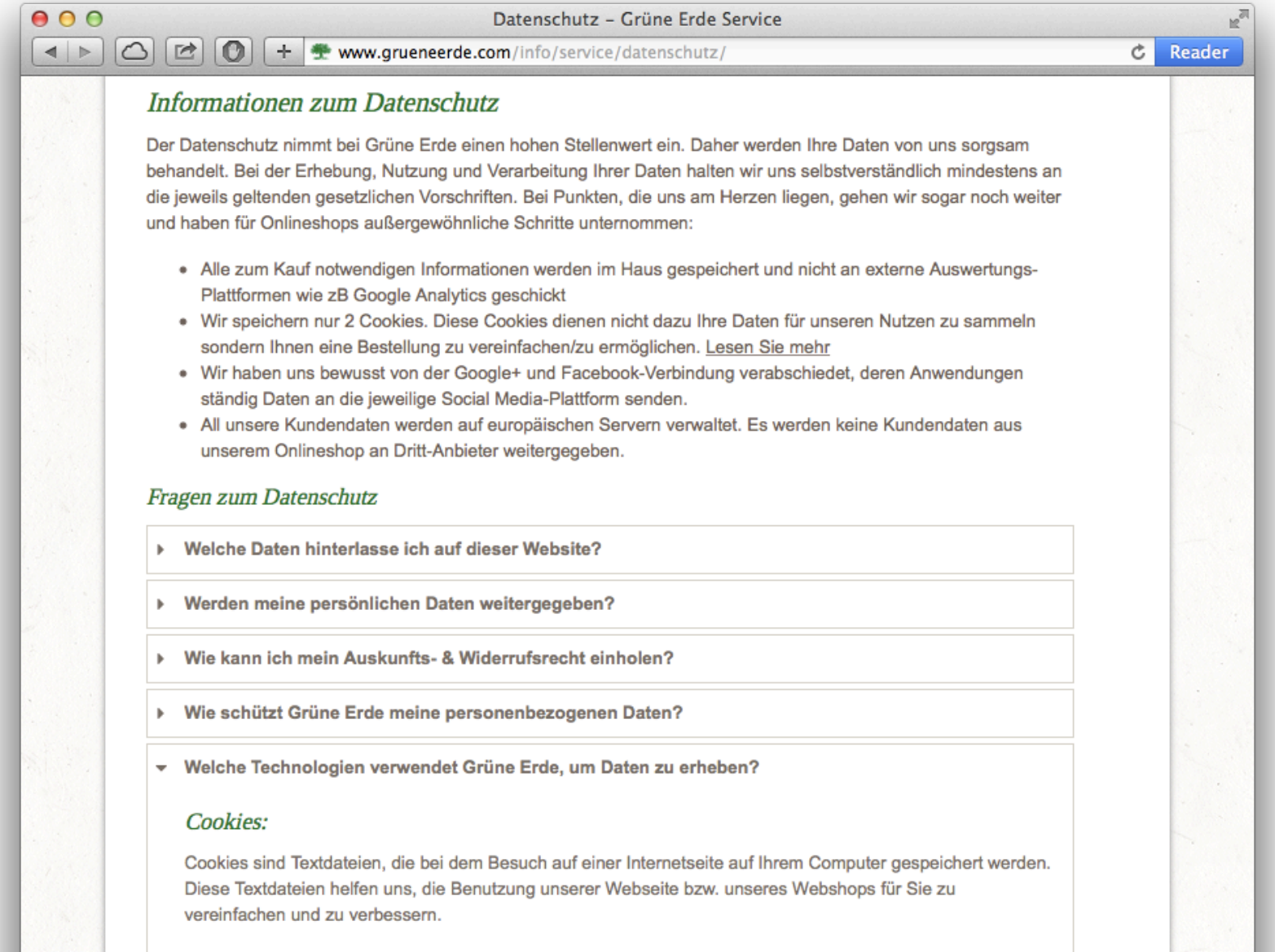
Privacy design strategies

nach: Hoepmann, 2003

- Technisch
  - Minimise: Nur notwendige Daten speichern und verarbeiten
  - Separate: Daten verteilt verarbeiten und speichern
  - Aggregate: Daten auf das notwendige Maß zusammenfassen
  - Hide: Daten nicht in offener Form speichern

- Organisatorisch
  - Enforce: Existenz und Durchsetzung einer Datenschutz-Policy (access control)
  - Inform: Betroffene informieren, wann und wie ihre Daten verwendet werden (P3P)
  - Control: Eingriffsmöglichkeit der Betroffenen in die Datenverarbeitung (informed consent)
  - Demonstrate: Überprüfbarkeit der Datenverarbeitung (privacy management, logging)

Beispiel einer datenschutzfreundlichen Policy



Grüne Erde verwendet 2 Cookies:

- Session-Cookie: Dieses Cookie dient zum Beispiel dazu, Ihren Warenkorb während Ihres Besuches in unserem Webshop zu speichern und ist notwendig um Online-Bestellungen durchzuführen. Das Session-Cookie wird innerhalb von 7-10 Tagen wieder gelöscht.
- Footprint-Cookie: Dieses Cookie erinnert Sie während Ihres Stöberns durch den Grüne Erde-Onlineshop welche Artikel Sie bereits angesehen haben. Es wird nach Ihrem Besuch wieder gelöscht.

Cookies können keine E-Mail-Adressen in Erfahrung bringen oder auf Ihre Festplatte zugreifen. Cookies übertragen auch keine Viren und versenden keine E-Mails unbemerkt. Es werden keine personenbezogenen Daten durch die Speicherung von Cookies erfasst.

Sie haben die Möglichkeit Cookies zu blockieren in dem Sie die nötigen Einstellungen bei Ihrem Browser vornehmen.

Bsp.: Internet Explorer:

1. Wählen Sie im Menü den Punkt „Internetoptionen".
2. Klicken Sie auf die Registerkarte „Datenschutz".
3. Nun können Sie einstellen, ob Cookies angenommen, selektiert oder abgelehnt werden sollen.
4. Mit „OK" bestätigen Sie Ihre Einstellungen

Bsp.: Firefox:

1. Wählen Sie im Menü den Punkt „Einstellungen".
2. Klicken Sie auf die Registerkarte „Datenschutz".
3. Wählen Sie im Drop-Down-Menü den Eintrag „nach benutzerdefinierten Einstellungen anlegen" aus.
4. Nun können Sie einstellen, ob Cookies akzeptiert werden sollen, wie lange Sie diese Cookies behalten wollen und Ausnahmen hinzufügen, welchen Websites Sie immer bzw. niemals erlauben möchten, Cookies zu benutzen.
5. Mit „OK" bestätigen Sie Ihre Einstellungen.

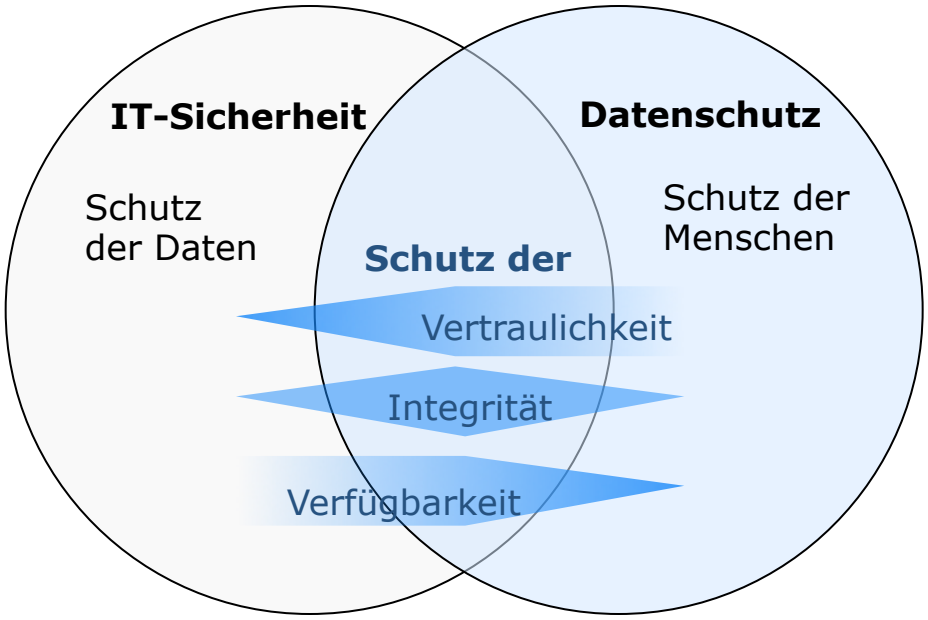
Weitere Infos zum Blockieren, Zulassen oder Löschen von Cookies finden Sie bei der Hilfe-Funktion in der Menüleiste Ihres Browsers.

Wir empfehlen Ihnen die Speicherung von Cookies zuzulassen, da sonst die Funktionen unserer Webseite bzw. unseres Webshops etwas eingeschränkt sind.

Quelle: http://www.grueneerde.com/info/service/datenschutz/

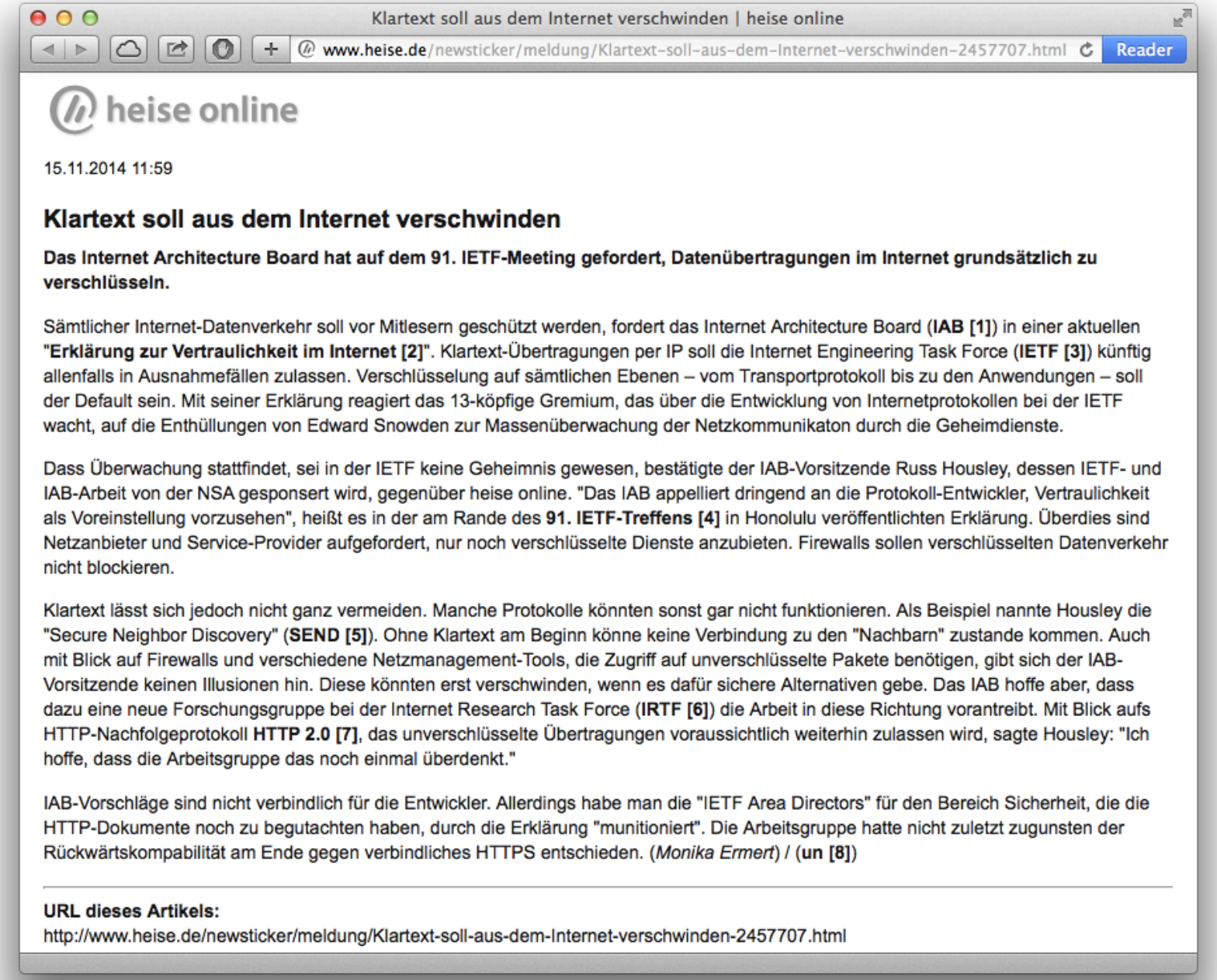
BDSG § 9 Technische und organisatorische Maßnahmen

- Öffentliche und nicht-öffentliche Stellen, die selbst oder im Auftrag personenbezogene Daten erheben, verarbeiten oder nutzen, haben die Maßnahmen zu treffen, die erforderlich sind, um die Ausführung der Vorschriften dieses Gesetzes, insbesondere die in der Anlage zu diesem Gesetz genannten Anforderungen, zu gewährleisten. Erforderlich sind Maßnahmen nur, wenn ihr Aufwand in einem angemessenen Verhältnis zu dem angestrebten Schutzzweck steht.



Anlage zu § 9 Abs. 1 BDSG

1. Zutrittskontrolle (räumlicher Zutritt, Gebäude)
2. Zugangskontrolle (Benutzung, Passwort)
3. Zugriffkontrolle (Berechtigung, Administratoren)
4. Weitergabekontrolle (Transport, Netze)
5. Eingabekontrolle (Nutzer-Protokoll)
6. Auftragskontrolle (Outsourcing, Wartung)
7. Verfügbarkeitskontrolle (Zerstörung)
8. Trennungsgebot (Zwecktrennung)



heise online

15.11.2014 11:59

Klartext soll aus dem Internet verschwinden

Das Internet Architecture Board hat auf dem 91. IETF-Meeting gefordert, Datenübertragungen im Internet grundsätzlich zu verschlüsseln.

Sämtlicher Internet-Datenverkehr soll vor Mitlesern geschützt werden, fordert das Internet Architecture Board (**IAB [1]**) in einer aktuellen "**Erklärung zur Vertraulichkeit im Internet [2]**". Klartext-Übertragungen per IP soll die Internet Engineering Task Force (**IETF [3]**) künftig allenfalls in Ausnahmefällen zulassen. Verschlüsselung auf sämtlichen Ebenen – vom Transportprotokoll bis zu den Anwendungen – soll der Default sein. Mit seiner Erklärung reagiert das 13-köpfige Gremium, das über die Entwicklung von Internetprotokollen bei der IETF wacht, auf die Enthüllungen von Edward Snowden zur Massenüberwachung der Netzkommunikaton durch die Geheimdienste.

Dass Überwachung stattfindet, sei in der IETF keine Geheimnis gewesen, bestätigte der IAB-Vorsitzende Russ Housley, dessen IETF- und IAB-Arbeit von der NSA gesponsert wird, gegenüber heise online. "Das IAB appelliert dringend an die Protokoll-Entwickler, Vertraulichkeit als Voreinstellung vorzusehen", heißt es in der am Rande des **91. IETF-Treffens [4]** in Honolulu veröffentlichten Erklärung. Überdies sind Netzanbieter und Service-Provider aufgefordert, nur noch verschlüsselte Dienste anzubieten. Firewalls sollen verschlüsselten Datenverkehr nicht blockieren.

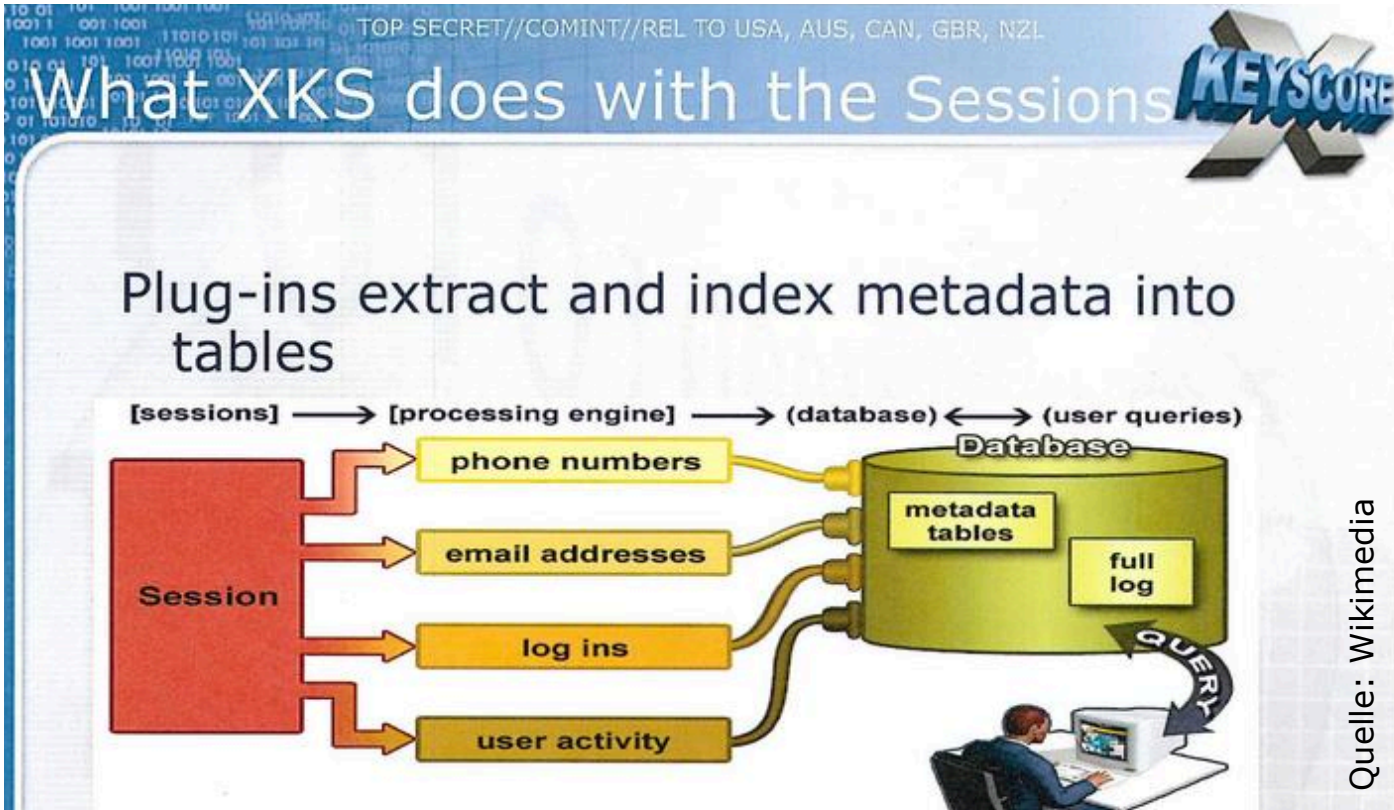
Klartext lässt sich jedoch nicht ganz vermeiden. Manche Protokolle könnten sonst gar nicht funktionieren. Als Beispiel nannte Housley die "Secure Neighbor Discovery" (**SEND [5]**). Ohne Klartext am Beginn könne keine Verbindung zu den "Nachbarn" zustande kommen. Auch mit Blick auf Firewalls und verschiedene Netzmanagement-Tools, die Zugriff auf unverschlüsselte Pakete benötigen, gibt sich der IAB-Vorsitzende keinen Illusionen hin. Diese könnten erst verschwinden, wenn es dafür sichere Alternativen gebe. Das IAB hoffe aber, dass dazu eine neue Forschungsgruppe bei der Internet Research Task Force (**IRTF [6]**) die Arbeit in diese Richtung vorantreibt. Mit Blick aufs HTTP-Nachfolgeprotokoll **HTTP 2.0 [7]**, das unverschlüsselte Übertragungen voraussichtlich weiterhin zulassen wird, sagte Housley: "Ich hoffe, dass die Arbeitsgruppe das noch einmal überdenkt."

IAB-Vorschläge sind nicht verbindlich für die Entwickler. Allerdings habe man die "IETF Area Directors" für den Bereich Sicherheit, die die HTTP-Dokumente noch zu begutachten haben, durch die Erklärung "munitioniert". Die Arbeitsgruppe hatte nicht zuletzt zugunsten der Rückwärtskompabilität am Ende gegen verbindliches HTTPS entschieden. (*Monika Ermert*) / (**un [8]**)

URL dieses Artikels:
http://www.heise.de/newsticker/meldung/Klartext-soll-aus-dem-Internet-verschwinden-2457707.html

Big Data – Die Technik im Mittelpunkt!

- unbegrenzter Speicher
- viel Bildschirmfläche
- large-scale computing
- hoher Stromverbrauch
- starke Verkettbarkeit



Quelle: Wikimedia

Big Data – Der Mensch im Mittelpunkt?

- große Transparenz (wünschenswert)
- starke Offenheit (erwartet)
- starke Kontrolle (notwendig)

- Orwells Metapher in 1984
  - Lückenlose Überwachung und Bevormundung von Menschen durch Menschen (Big Brother)

- Neue Metapher in 20xx
  - Lückenlose Überwachung und Profilbildung von Menschen durch Computer (Big Data) … durch Menschen

Hochschule als Laborumgebung

These: Die Hochschule als Laborumgebung zur experimentellen Analyse neuartiger Phänomene und Erprobung unausgereifter Verfahren und Prozesse

- Big Data
  - Steuer
    Beliebt
- Datenschu
  - Denkb
    Aggreg
    Online
    Mitbestimmungsprozesse
- IT-Sicherheit
  - Single-Sign-On in sehr heterogenen, teils anarchischen IT-Umgebungen; Verschlüsselung für alle Hochschulmitglieder

IT (und Big Data) als Enabler,

Datenschutz als Chance und

IT-Sicherheit als Herausforderung

verstehen.