Karlsruher Institut für Technologie

Datenschutz und Privatheit in vernetzten Informationssystemen

Kapitel 3: Digitale Identitäten

Erik Buchmann (buchmann@kit.edu)

IPD, Systeme der Informationsverwaltung, Nachwuchsgruppe „Privacy Awareness in Information Systems“

Inhalte und Lernziele dieses Kapitels

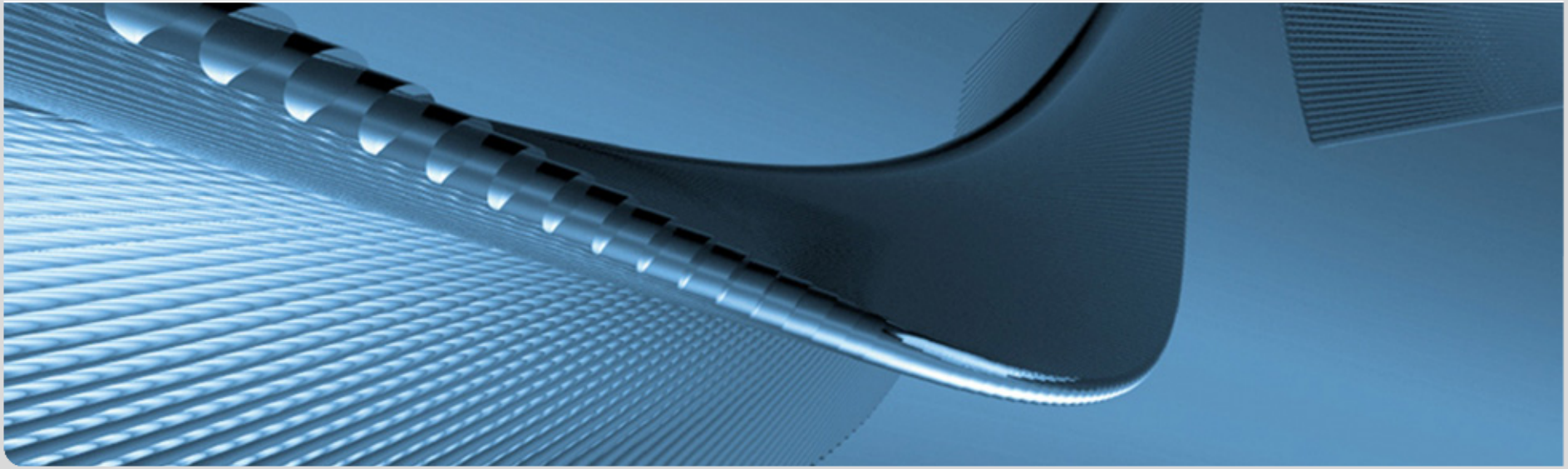
- Digitale Identitäten
- Bedrohungspotential
- Identitätsdiebstahl
- Fallbeispiel: Internet-Suchmaschinen

- Lernziele
  - Erkennen Sie, auf welche unterschiedlichen Arten die eigene Identität im Netz repräsentiert sein kann.
  - Sie können die Probleme erklären, die durch eine unkontrollierte Verbreitung dieser (Teil-)Identitäten entstehen können.
  - Sie können Aussagen dazu machen, welche Gefahren für die persönliche Handlungsfreiheit insbesondere durch ein Zusammenführen dieser Teilidentitäten entstehen.

Karlsruher Institut für Technologie

Die Digitale Identität

IPD, Systeme der Informationsverwaltung, Nachwuchsgruppe „Privacy Awareness in Information Systems“

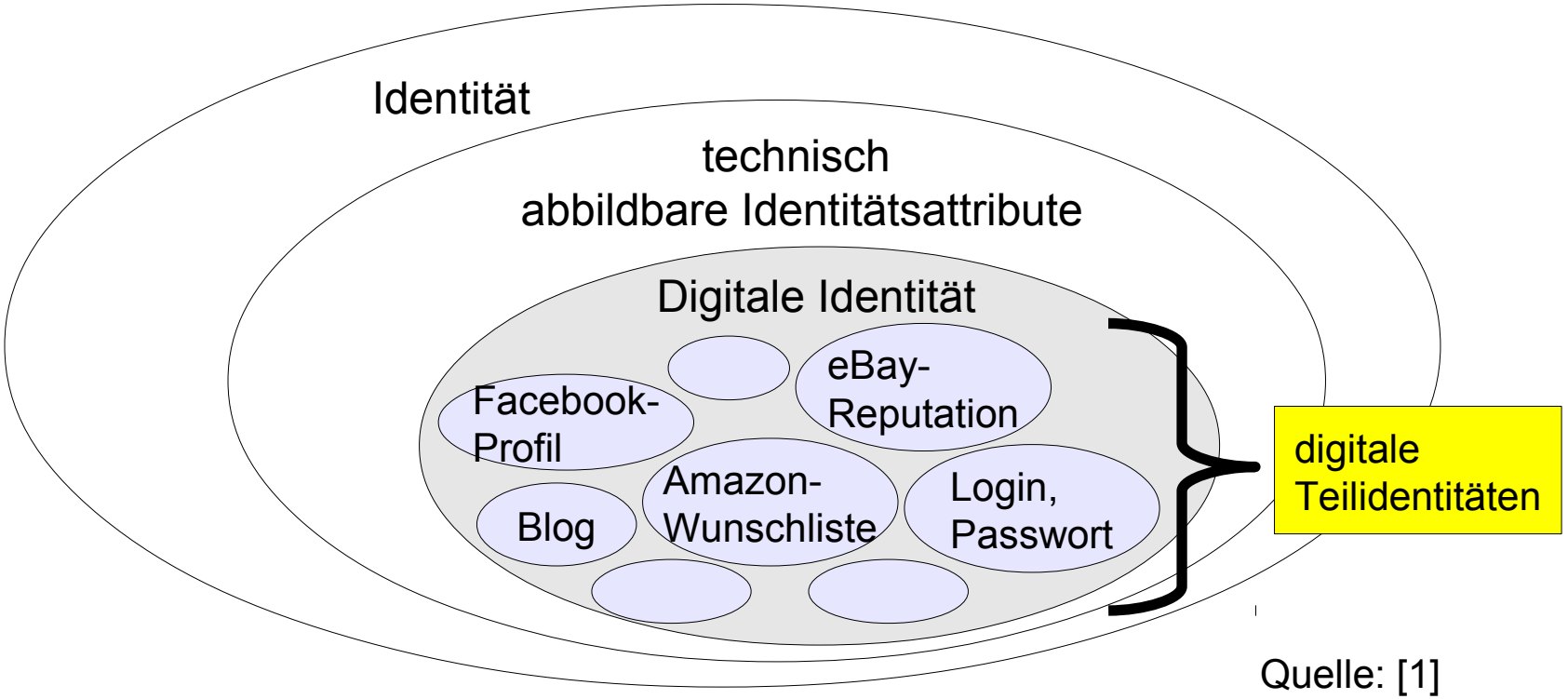


Die Digitale Identität

- Definition: “Jede mögliche Form von technisch abgebildeten Daten, die zu einer Person gehören” [1]
  - Daten zur eindeutigen Authentifizierung,
    z.B. Adresse, Name, biometrische Daten
  - Daten zur pseudonymen Identifizierung,
    z.B. Login, Passwörter, Nicknames, Foren-Namen
  - Persönliche Merkmale,
    z.B. Vorlieben, Hobbies, Religion, Lebensumfeld
  - nicht unbedingt von jedem einer Person zuordnbar
    - Beispiel: IP-Adresse ist Teil der digitalen Identität, aber nur vom Internet-Provider zuordnbar

Übersicht: Identität im Netz

- Digitale Teilidentität: Untermenge der digitalen Identität, die eine Untermenge der abbildbaren Attribute sind
  - viele separate **digitale Teilidentitäten** möglich



Quelle: [1]

Unterschiedliche digitale Teilidentitäten

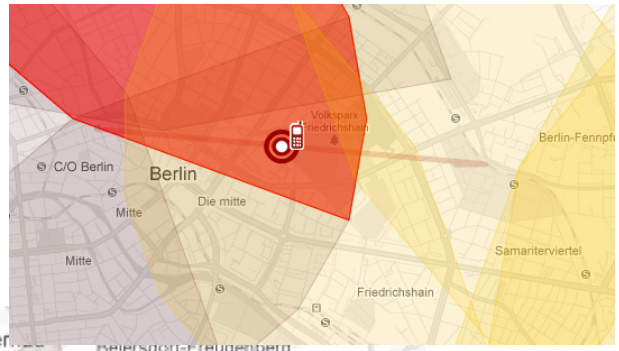
- Datenspuren im täglichen Leben
  - Blogs, Soziale-Netzwerk-Portale, Internet-Communities
  - Finanzamt, Wählerlisten, Anträge bei Behörden
  - Verkehr mit Wirtschaftsunternehmen, z.B. Einkäufe im Supermarkt oder Web-Shops

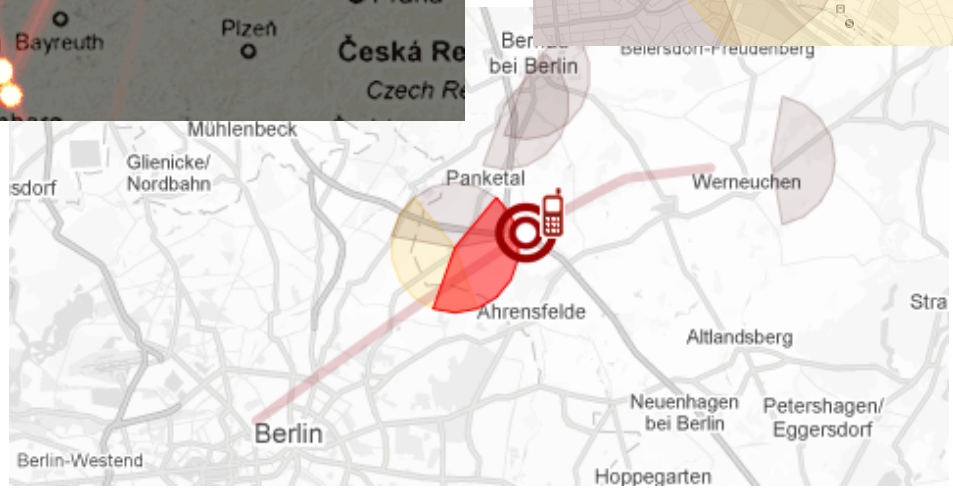
- Nicht alle Datenspuren werden wissentlich hinterlassen
  - Was weiß der
    - Supermarkt (Rabatt-System)
    - Mobilfunkprovider (Positionsdaten vom Handy)
    - Gerätehersteller (iPhone 4 speicherte 2010 und 2011 Bewegungsdaten)

→ durch **Verkettung von Teilidentitäten** Aufbau umfassender Persönlichkeitsprofile möglich!

Konkretes Beispiel: Malte Spitz, B90/Grüne

- Visualisierung eines Auskunftsersuchens bei der Telekom





Weniger relevant für Privatheit	Potentiell gefährlich für die Privatheit
anonym	eindeutig identifizierend
nicht wiedererkennbar	wiedererkennbar
ändert sich im Verborgenen über die Zeit	unveränderlich
leicht änderbar	nicht änderbar
weitergebbar / übertragbar	nicht weitergebbar/übertragbar
flüchtig	langfristig gespeichert
nur einmal verwendet	häufig wiederverwendet
Authentizität unklar	authentisch / bestätigt durch Dritte
Zugriff anderer nicht möglich bzw. kontrollierbar	Zugriff anderer möglich oder intransparent
ermöglicht keinen direkten Kontakt	ermöglicht unmittelbaren Kontakt
unauffällig/geht in der Masse unter	unnormal oder herausragend
für wenige Teile des eigenen Lebens relevant / als trivial empfunden	betrifft zentrale Bereiche des täglichen Lebens
keine zusätzlichen Informationen enthaltend	enthält für Weitergabe nicht abtrennbare Zusatzinformationen

Erläuterung am Beispiel

- Biometrische Merkmale (Fingerabdruck, Gesicht)
  - stabil über die Zeit, kaum änderbar, nicht übertragbar, langfristig speicherbar (Reisepass), häufig wiederverwendet (Passkontrolle), authentisch, andere können darauf zugreifen (sofern nicht verschleiert)
  - **potentiell gefährlich für die Privatheit**

- selbstbestimmte Identitätsmerkmale (Login, Passwort)
  - anonym oder pseudonym, leicht änderbar, übertragbar an Dritte, Zugriff anderer nicht möglich, oft nur für unwichtige Teile des pers. Lebens relevant
  - **weniger gefährlich für die Privatheit**

Karlsruher Institut für Technologie

Bedrohungen durch die Digitale Identität

IPD, Systeme der Informationsverwaltung, Nachwuchsgruppe „Privacy Awareness in Information Systems“

Probleme mit digitalen (Teil-)Identitäten

- Fehlentscheidungen durch Falschinformationen
- Datenmißbrauch
- Zweitverwertung
- Langfristige Aufbewahrung
- Verknüpfbarkeit
- Öffentliche Zugänglichmachung

(Anm.: Liste ohne Anspruch auf Vollständigkeit)

Falschinformationen

- Offensichtliches Problem: falsche Daten können
  - zu einem falschen Bild in der Öffentlichkeit führen
  - wirtschaftlich schädlich sein, z.B. falsche Schufa-Einträge, falsche Steuer-Berechnung

- Problem
  - Korrektur schwierig umzusetzen, wenn Daten vielfach redundant vorhanden sind
    - Daten in Backups
    - Usenet-Nachrichten, P2P-Dateien vielfach gespiegelt
  - Korrektur schwierig einzufordern, wenn Daten nicht gewerbsmäßig verarbeitet werden, sondern von privaten Anwendern (Wikipedia, Facebook)
  - Datennutzung ist häufig Geschäftsgeheimnis; Fehlentscheidungen für Betroffenen intransparent

Datenmißbrauch

- Personenbezogene Daten werden ohne Wissen oder Zustimmung des Betroffenen an Dritte übermittelt
  - einfachste Form: Handel mit Adresslisten
- Ausnahme: einige Arten der Datenübermittlung ohne Wissen und Zustimmung sind gesetzlich legitimiert, z.B.
  - Steuerfahndung
  - polizeiliche Ermittlungen

- <u>Problem</u>
  - unklar, wer Kenntnis von persönlichen Daten hat
  - Verbleib persönlicher Informationen für den Betroffenen nicht nachvollziehbar

Zweitverwertung

- Daten vom Geschäftsbetrieb für andere Zwecke nutzen
  - Data Mining auf Kundendaten
  - Autobahn-Maut (Toll Collect) → Strafverfolgung
  - Handy-Verbindungsdaten → Positionsbestimmung
  - Werbung

- Problem
  - Daten fallen oft unvermeidlich an, z.B. muss Lieferant die Lieferadresse kennen
  - Grenze zwischen legitimer Nutzung für den Geschäftszweck und illegitimer Nutzung oft unklar
  - für den Betroffenen aufgrund fehlender Transparenz oft nicht nachvollziehbar [2]

Verknüpfbarkeit

- Zusammenführen von “harmlosen” digitalen Teilidentitäten aus unterschiedlichen Quellen
  - öffentliche Quellen: Telefonbuch, Schufa-Daten, Handelsregister, Liegenschaftsbuch, etc.
  - nichtöffentlich: Daten aus dem Geschäftsbetrieb, Informationen von Behörden, Banken
- Anreichern von eigenen Daten mit Zusatzinformationen
- Suche nach diskriminierenden Merkmalskombinationen in verlinkten Daten → Rasterfahndung

- Problem
  - Aufbau von komplexen Persönlichkeitsprofilen
  - Fehler in den Daten können zu falschen Schlüssen führen

Langfristige Aufbewahrung

- Daten können beliebig lange gespeichert werden
- Rekonstruierbarkeit und Nachvollziehbarkeit von Aussagen oder Handlungen wird möglich
- Anmerkung: schon aus Praktikabilitätsgründen umfasst die Löschpflicht des BDSG keine Backups

- <u>Problem</u>
  - Selbstdarstellung in der Vergangenheit kann drastisch von aktuell gewünschter Selbstdarstellung abweichen
    - während der Schulzeit Partylöwe, jetzt Lehre zum Bankkaufmann
  - Kontrollverlust über einmal preisgegebene Daten;
    was einmal im Internet steht ist 'verbrannt'

Öffentliche Zugänglichmachung

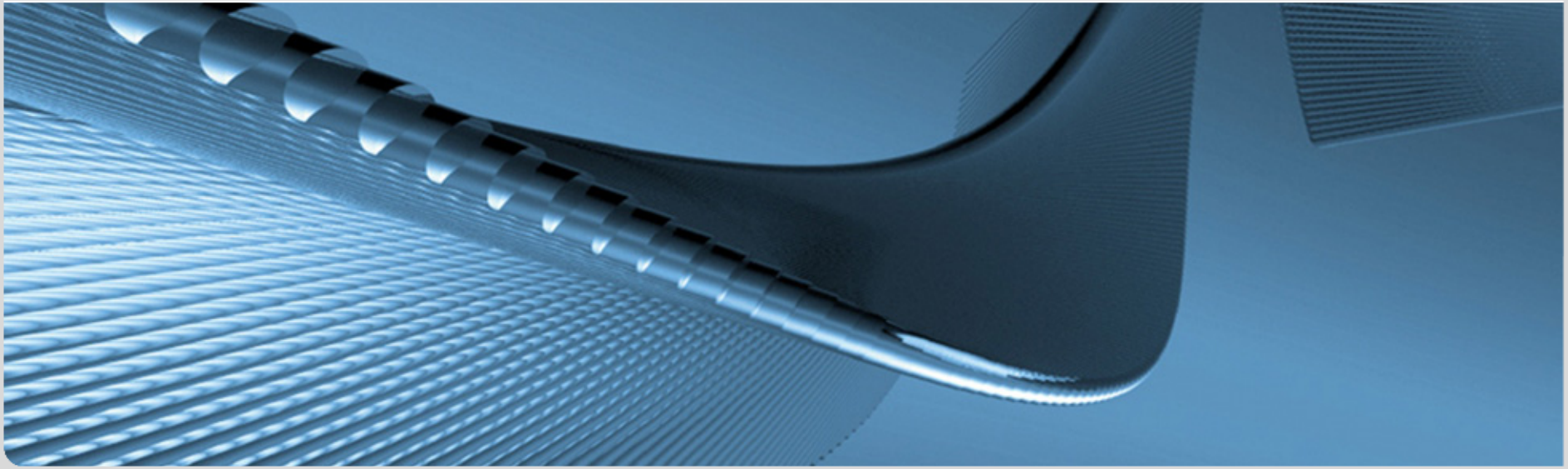
- Auch objektive Fakten können zu einer falschen Selbstdarstellung führen
  - Auswahl an Informationen entscheidend, z.B. nur wenig schmeichelhafte oder veraltete Daten
- Selbst-Inszenierung des persönlichen Auftritts und Selbst-Bild sind wichtig, z.B.
  - soziale Kontakte, Selbstwertgefühl
  - Vorstellungsgespräche, Verkaufsgespräche

- Problem
  - auch persönliche Daten, die der Betroffene selbst veröffentlicht hat, sind schützenswert
  - Preisgabe von persönlichen Daten durch Dritte, z.B. Freundeslisten in Web-Communities

Karlsruher Institut für Technologie

Identitätsdiebstahl

IPD, Systeme der Informationsverwaltung, Nachwuchsgruppe „Privacy Awareness in Information Systems“



Mißbrauch personenbezogener Daten

- Personenbezogene Daten werden ohne Wissen oder Zustimmung des Betroffenen zu Zwecken gesammelt, gespeichert, verarbeitet oder übermittelt, die
  - den Interessen des Betroffenen zuwiderlaufen, z.B.
    - **Identitätsdiebstahl**
    - Stalking, Mobbing
    - Kreditbetrug
  - aber nicht gesetzlich legitimiert sind, z.B.
    - Steuerfahndung
    - polizeiliche Ermittlungen
- Für den Betroffenen nicht zu verhindern, da ohne Wissen und Zustimmung erfolgt

Identitätsdiebstahl

- *“A fraud committed or attempted using the identifying information of another person without authority.”*
  (Definition der Federal Trade Commission, USA)

  - unlegitimierte Nutzung einer fremden Identität
  - betrügerischer Vermögensvorteil unter Inkaufnahme von Nachteilen für den “Inhaber” der Identität, Kreditgebern, Händlern etc.

Hoofnagle, Chris Jay, *Identity Theft: Making the Known Unknowns Known.* Harvard Journal of Law and Technology, Vol. 21, 2007.

Identitätsdiebstahl USA 2008

- Zahlen 2008: Data Breach Stats (“Einbruchstatistik”)

	# of Breaches	*# of Consumer Records*
Banking/Credit/Financial	78	18,731,947
Business	240	5,886,960
Educational	131	806,142
Government/Military	110	2,954,373
Medical/Healthcare	97	7,311,833
total:	**656**	**35,691,255**

- Quelle: http://www.idtheftcenter.org, Breach Database
  (Anm.: Zahlen beschreiben nur Verlust von pers. Daten, nicht die Mißbrauchsfälle)

Identitätsdiebstahl USA 2009

- Zahlen 2009: Data Breach Stats (*"Einbruchstatistik"*)

	# of Breaches	*# of Consumer Records*
Banking/Credit/Financial	57	8,364
Business	208	132,402,177
Educational	78	803,667
Government/Military	90	79,470,963
Medical/Healthcare	65	10,461,818
total:	**498**	**223,146,989**

- Quelle: http://www.idtheftcenter.org, Breach Database
  (Anm.: Zahlen beschreiben nur Verlust von pers. Daten, nicht die Mißbrauchsfälle)

Identitätsdiebstahl USA 2010

- Zahlen 2010: Data Breach Stats (*"Einbruchstatistik"*)

	# of Breaches	*# of Consumer Records*
Banking/Credit/Financial	54	4,853,708
Business	279	6,626,435
Educational	65	1,598,266
Government/Military	104	1,214,773
Medical/Healthcare	160	1,874,360
total:	**662**	**16,167,542**

- Quelle: http://www.idtheftcenter.org, Breach Database
  (Anm.: Zahlen beschreiben nur Verlust von pers. Daten, nicht die Mißbrauchsfälle)

Identitätsdiebstahl USA 2011

- Zahlen 2011: Data Breach Stats (*"Einbruchstatistik"*)

	# of Breaches	*# of Consumer Records*
Banking/Credit/Financial	28	413,348
Business	198	7,917,907
Educational	59	818,458
Government/Military	48	10,036,657
Medical/Healthcare	86	3,732,071
total:	**419**	**22,918,441**

- Quelle: http://www.idtheftcenter.org, Breach Database
  (Anm.: Zahlen beschreiben nur Verlust von pers. Daten, nicht die Mißbrauchsfälle)

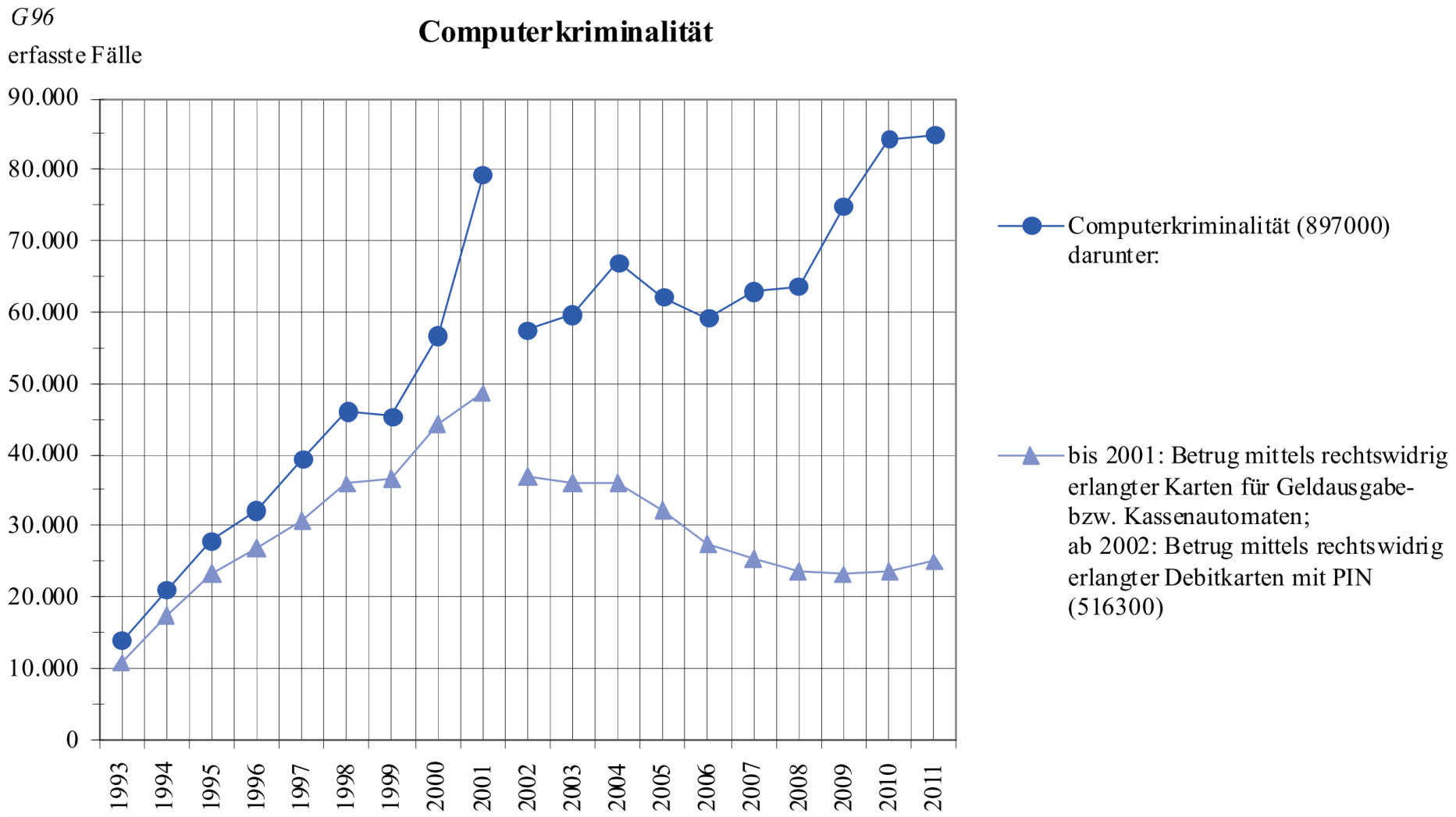
Identitätsdiebstahl USA 2012

- Zahlen 2012: Data Breach Stats (*"Einbruchstatistik"*)

	# of Breaches	*# of Consumer Records*
Banking/Credit/Financial	17	470,048
Business	165	4,615,893
Educational	61	2,304,663
Government/Military	50	7,688,707
Medical/Healthcare	154	2,237,873
total:	**447**	**17,317,184**

- Quelle: http://www.idtheftcenter.org, Breach Database
  (Anm.: Zahlen beschreiben nur Verlust von pers. Daten, nicht die Mißbrauchsfälle)

Polizeiliche Kriminalstatistik 2011



Polizeiliche Kriminalstatistik 2011

Schlüs-sel	Straftaten(gruppen)	erfasste Fälle		Veränderung		Aufklärungsquote	
		2011	2010	absolut	in %	2011	2010
897000	Computerkriminalität	84.981	84.377	604	0,7	32,6	35,8
	davon:						
516300	Betrug mittels rechtswidrig erlangter Debitkarten mit PIN	24.923	23.612	1.311	5,6	37,4	40,7
517500	Computerbetrug -§263a StGB-	26.723	27.292	-569	-2,1	27,0	30,2
517900	Betrug mit Zugangsberechtigungen zu Kommunikationsdiensten	4.730	7.993	-3.263	-40,8	37,8	44,0
543000	Fälschung beweiserheblicher Daten, Täuschung im Rechtsverkehr bei Datenverarbeitung -§§ 269, 270 StGB-	7.671	6.840	831	12,1	47,0	52,0
674200	Datenveränderung, Computersabotage -§§ 303a, 303b StGB-	4.644	2.524	2.120	84,0	41,2	32,1
678000	Ausspähen, Abfangen von Daten einschl. Vorbereitungshandlungen	15.726	15.190	536	3,5	21,3	24,0
715100	Softwarepiraterie (private Anwendung z.B. Computerspiele)	412	794	-382	-48,1	92,5	94,1
715200	Softwarepiraterie in Form gewerbsmäßigen Handelns	152	132	20	15,2	92,8	97,0

- Identitätsdiebstahl in Deutschland kein eigener Straftatbestand, aber Schlüssel 517900, 543000, 674200 sind einschlägig

Methoden zum Identitätsdiebstahl (1/2)

- Physische Ansätze
  - Diebstahl von Brieftaschen, Post aus dem Briefkasten, durchwühlen des Mülls nach Kontoauszügen
- Social Engineering
  - Täuschung und Überredung am Telefon, Erraten von schwachen Passwörtern (Name der Katze)
- Phishing
  - “Wegen <*irgendwas*> braucht <*eine URL, die fast so aussieht wie die die URL Ihrer Bank*> dringend Passwörter, Kontonummer, Pin, Tan und Name”
- Spyware/Viren/Trojaner
  - Mitschneiden der Tastenanschläge, Ausspähen von Passwörtern

Methoden zum Identitätsdiebstahl (2/2)

- Hacking
  - Einbruch in Firmensysteme, auf denen Kundendaten liegen
- Pharming
  - Angriff eines DNS-Servers, Umleitung von Bank-Webseiten
- Durchforsten des Webs
  - Impressum auf Webseiten, veröffentlichte Lebensläufe auf Bewerbungsportalen, Social-Network-Portalen etc.
- Sorglosigkeit
  - Ausnutzen von versehendlich ins Netz gestellten Daten
- Underground Economy
  - Kauf von Identitätsdaten von Hacker-Portalen
- *(Liste erhebt keinen Anspruch auf Vollständigkeit)*

Varianten des Identitätsdiebstahls

- Generieren einer (teilweise) künstlichen Identität
  - New Account Fraud
- Übernahme einer bestehenden Identität
  - Account Takeover

New Account Fraud

- Generieren einer künstlichen Identität
  - echte Daten, um Validierungsverfahren zu täuschen
    - Kontoeröffnung USA: Social Security Number
    - Kontoeröffnung Deutschland: Name, Postanschrift für Schufa-Auskunft
  - künstliche Daten zum Vervollständigen
    - Geschlecht, Alter, Beruf, Einkommen, Familienstand
  - künstliche Daten, um Plausibilität zu erhöhen
    - Vermögenslage (Rechnungen, Hypotheken, Kreditkartenkonten etc.)
    - Lebenslauf

Erkennung ist schwierig

- Künstliche Identitäten bestehen aus Mischung von echten und falschen Daten
- Inhaber der echten Daten erfahren oftmals nur indirekt vom Mißbrauch
  - Korrespondenz, Unterlagen, Mahnungen gehen an den Betrüger
  - Schaden entsteht meist indirekt,
    z.B. wenn Schufa-Auskunft belastet
- Aus Händler- oder Bankensicht kein Unterschied zwischen flüchtigem Schuldner und gefälschter Identität
  - unklar, ob Betrugsfall oder gewöhnlicher Kreditausfall

Account Takeover

- Übernahme einer bestehenden digitalen Identität
  - *Pishing*, gefälschte EMails erfragen Kontodaten, eBay-Konten, Kreditkartennummern,
    www.meinebank.de.pisher.org
  - *Pharming*, Webbrowser wird durch DNS-Spoofing o.ä. auf manipulierte Webseiten umgeleitet, die eBay- oder Banken-Webseiten gleichen
  - *Malware* auf dem Rechner protokolliert Anmeldeinformationen
  - *Social Engineering*,
    beliebtes Passwort: Name der Freundin,
    beliebte PIN: Geburtsdatum des Kindes

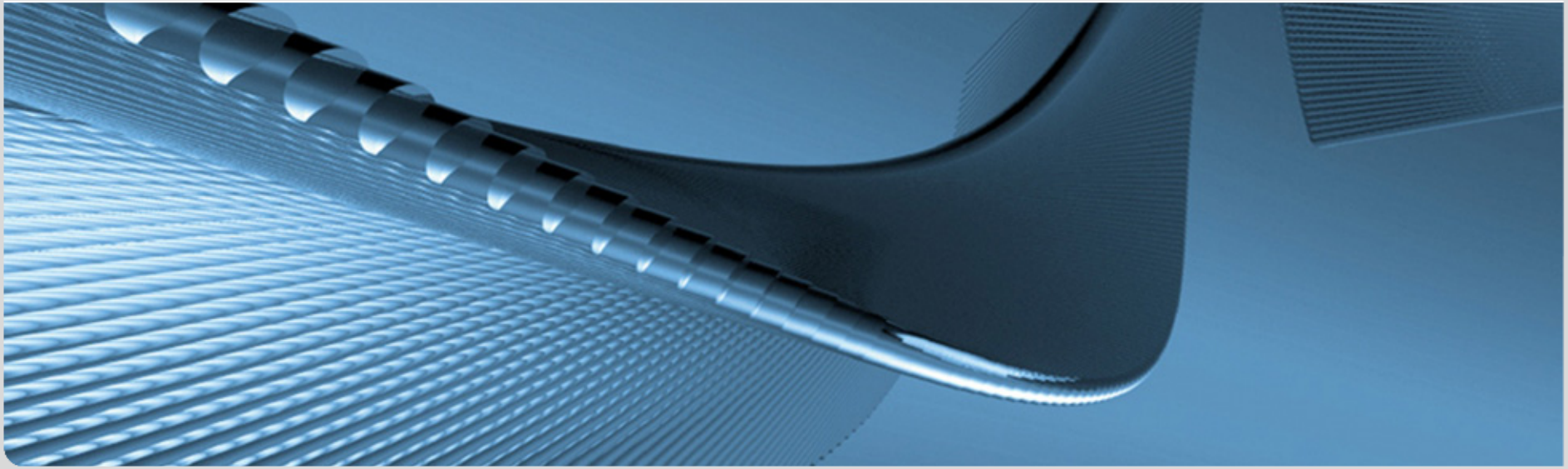
Erkennbarkeit

- Account Takeover oft leichter zu erkennen als New Account Fraud
  - die Betroffenen erhalten zumeist Mahnungen, Rechnungen etc.
- gesetzlicher Schutz des Betroffenen
  - Rückbuchung von per Lastschrift eingezogenen Beträgen
  - Stornierung von Kreditkartenrechnungen
  - Strafanzeige gegen Unbekannt

Karlsruher Institut für Technologie

Fallbeispiel: Identitätsattribute in Suchmaschinen

IPD, Systeme der Informationsverwaltung, Nachwuchsgruppe „Privacy Awareness in Information Systems“



Suchmaschinen: Google, AOL, MSN und Co.

- Suche übermittelt an Suchmaschinenbetreiber:

Kategorie	Attribute
Suchterme (technisch unvermeidlich)	Aneinanderreihung von Schlüsselworten
Browser-Kommunikation (technisch unvermeidlich)	Zeitstempel, IP, Browser, Betriebssystem, Spracheinstellungen, zuletzt besuchte Seite
Browser-Kommunikation (technisch vermeidlich)	ausgewähltes Suchergebnis, (Implementiert als Redirect)
Zusatzinformationen (vom Nutzer vermeidbar)	Cookie-Informationen, Session-ID
	über JavaScript ermittelte Daten, Verweildauer auf der Seite, Bildschirmauflösung
	Nutzer-ID (z.B. Google-Login, Yahoo-ID)

Mögliche Verkettung dieser Attribute zur Profilbildung

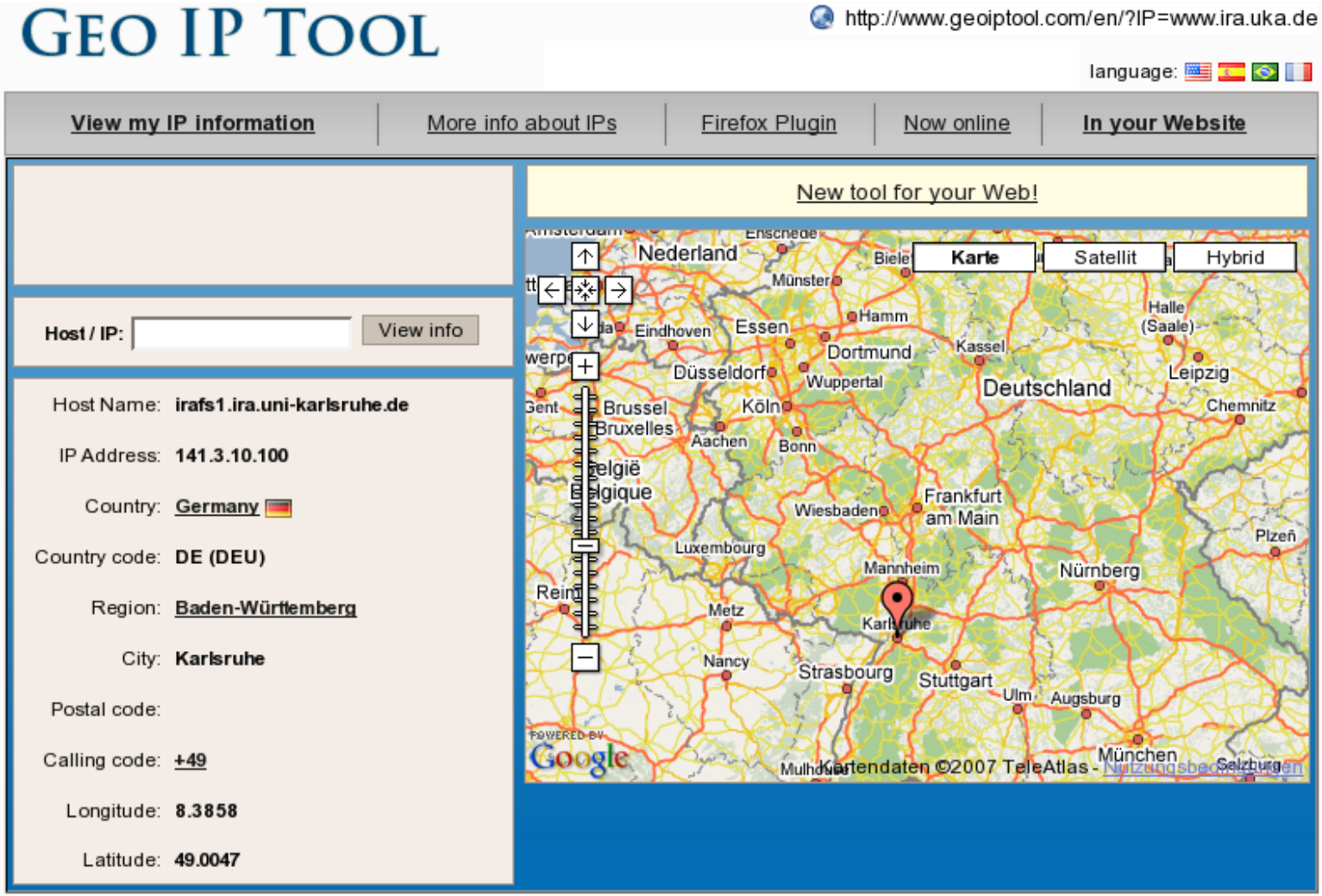
Was verrät mein Browser?

- https://panopticlick.eff.org

Test: Opera, SuSE Linux

browser characteristic	bits of identifying information	one in x browsers have this value	value
User Agent	13.41	10913.75	Opera/9.80 (X11; Linux x86_64) Presto/2.12.388 Version/12.14
HTTP_ACCEPT	8.58	382.27	text/html, */* gzip, deflate en-US,en;q=0.9
Browser Plugins	21.45+	2859402	Plugin 0: DivX Web Player; DivX Web Player version 1.4.0.233; libtotem-mully-plugin.so; (; video/divx; divx). Plugin 1: Gnome Shell Integration; […]
Time Zone	2.85	7.22	-120
Screen Size and Color Depth	7.1	136.97	1600x1200x24
System Fonts	21.45+	2859402	LM Mono Slanted 10, GFS Baskerville, FreeMono, LM Mono 10, LM Mono 12, Droid Sans, […]
Are Cookies Enabled?	0.42	1.34	Yes
Limited supercookie test	0.99	1.99	DOM localStorage: Yes, DOM sessionStorage: Yes, IE userData: No

Was verrät die IP-Adresse?



Verkettung von Suchanfragen

- Über die **Browser-Kommunikation**, insbes. IP-Adresse
  - dynamische IP-Adressen: überdauern mindestens eine Such-Session
    - Kombination aus Standort des IP-Adressbereichs, Betriebssystem, Sprache, Browser etc. können als Quasi-Identifier ausreichen
  - statische IP-Adressen, z.B. Uni-Netz: länger gültig
- Über **Cookies, Session-IDs, Nutzer-Login**
  - Identifiziert einen Browser (und damit oft dessen Benutzer) über lange Zeiträume eindeutig
  - auch bei wechselnder IP-Adresse
- Über die **Suchterme**
  - z.B. Suche nach eigenem Namen, seltene Hobbies
- **Kombinationen** aus allem

Verkettung von Suchanfragen

- Über die **Browser-Kommunikation**, insbes. IP-Adresse
  - dynamische IP-Adressen: überdauern mindestens eine Such-Session
    - Kombination aus Standort des IP-Adressbereichs, Betriebssystem, Sprache, Browser etc. können als Quasi-Identifier ausreichen
  - statische IP-Adressen, z.B. Uni-Netz: länger gültig
- Über **Cookies, Session-IDs, Nutzer-Login**
  - Identifiziert einen Browser (und damit oft dessen Benutzer) über lange Zeiträume eindeutig
  - auch bei wechselnder IP-Adresse
- Über die **Suchterme**
  - z.B. Suche nach eigenem Namen, seltene Hobbies
- **Kombinationen** aus allem

Verkettung mit Zusatzdiensten

- Suchmaschinenanbieter oft Anbieter weiterer Dienste
  - Google: Youtube, Maps, Email, Verzeichnisdienste, Google Docs, Google Earth, News, Usenet
- *separate* digitale Teilidentitäten werden verkettbar
  - über IP-Adresse Nutzerbewegungen über mehrere Dienste hinweg nachvollziehbar
  - oft übergreifendes Login für viele Dienste,
    - Microsoft Passport, Windows Live ID
    - Google Authentication for Web Applications (OAuth)
- Informationen über viele Lebensbereiche
  - Arbeit, Privatleben, Hobbies, Kommunikationspartner etc.

Haben Suchanfragen Personenbezug?

- kommt ganz drauf an...
  - zur Erinnerung: *“...Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person.”*
  - Heute übliche Interpretation: kein Personenbezug, wenn “*für Einzelangaben zu einer Person die Wahrscheinlichkeit, dass diese der Person zugeordnet werden können, so gering ist, dass sie nach der Lebenserfahrung oder dem Stand der Wissenschaft praktisch ausscheidet.*”

Quelle: Roßnagel, A.; Scholz, P.: Datenschutz durch Anonymität und Pseudonymität, MMR 2000

Personenbezug von Suchbegriffen

- auch hier: kommt drauf an...
  - “Katzenfutter billig” → nein;
  - “Erik Buchmann Urlaub Italien” → ja

- Personenbezug ist *abhängig von den Benutzereingaben*; nicht ohne weiteres automatisch vom Betreiber entscheidbar!
  - je umfangreicher die Suchhistorie eines Nutzers, desto wahrscheinlicher kommt eine identifizierende Kombination von Identitätsattributen zusammen
    (Beispiel folgt)

Personenbezug der Kommunikation

- Dynamische IP-Adresse
  - eher nicht personenbezogen
    - *Personenbezug erfordert Mithilfe des Kenners der Zuordnungsregel*, d.h., Internet Service Provider
- Statische IP-Adresse für festen Rechner
  - Suchmaschinenbetreiber kann Suchanfragen über lange Zeiträume einer Person zuordnen
  - mit zunehmender Zahl der Suchvorgänge steigt Wahrscheinlichkeit, dass sich der Suchende offenbart
    (vgl. vorangegangene Folie)
  - daher: oftmals personenbezogen

Anmerkung: derzeit unterschiedliche Rechtsauffassungen;
hier vorgestellt wird akutell gängige Praxis – kann sich aber ändern

Personenbezug von Zusatzinformationen

- Bildschirmauflösung, Sprache, Verweildauer auf der Seite
  - nicht als identifizierendes Merkmal geeignet
- Nutzer-Login, Session-ID, Cookie-Informationen
  - Betreiber hat Login, Session-ID bzw. Cookie selbst dem Nutzer zugeordnet
  - daher: oftmals personenbezogen, da *Suchmaschinenbetreiber gleichzeitig Kenner der Zuordnungsregel*

Aug. 2006: AOL-Datenleck (1/2)

- AOL Research veröffentlicht für Forschung 20 Mio. Suchanfragen von 650.000 Usern, gesammelt über 3 Monate
  - IDs pseudonymisiert, künstl. Schlüssel

User ID	Search Keywords	Date	Website
4417749	numb fingers	2006-03-06 18:35:02	http://irgendwas.de
...	...	...	...

- zwar ist keiner der Datensätze unmittelbar personenbezogen, aber schnell werden einzelne Identitäten und komplette Persönlichkeitsprofile offenbar
- 3 Tage später: Datenbank ist vom Netz, aber schon vielfach in Tauschbörsen kopiert, bis heute verfügbar

Aug. 2006: AOL-Datenleck (2/2)

- Nutzer 4417749: hunderte Suchanfragen in 3 Monaten
  - “dog urinates on everything”: *Hundebesitzer*
  - “60 single man”: *einsame ältere Frau*
  - “numb fingers”: *körperliche Gebrechen*
  - “homes sold in gwinnett county”: *Wohnung*
  - “xxx Arnold, yyy Arnold”: *suche nach Verwandten*
  - “school supplies for Iraq children”: *karitativ*
  - “best season to visit Italy”: *nächster Urlaub*

- identifiziert als Thelma Arnold:
  “My goodness, its my whole personal life!”
  - Privat- und Alltagsleben, Ängste, Gebrechen
  - falsches Selbstbild durch Suchanfragen für Freunde

Quelle: http://www.nytimes.com/2006/08/09/technology/09aol.html

Wie handhaben es die Betreiber?

- Google
  - nutzt Cookies, verknüpft Daten der Suchmaschine und aller hauseigenen Dienste (Google Docs, Gmail, Google+, Google Analytics)
  - Angabe 2009: nach 9 Monaten werden Suchlogs anonymisiert (was immer das heißt)
- Microsoft, AOL, Yahoo
  - vergleichbare Praxis der Datensammlung
  - Speicherdauer Yahoo, AOL: 6 Monate
  - Speicherdauer Microsoft: *“After 18 months, we will completely anonymize all Search queries [...] by irreversibly removing all cross-session identifiers [...] including the full IP address and all cookie IDs.”*
    → schützt nicht vor persönlichen Suchtermen

Alternativen existieren

- Es gibt Suchmaschinen, die keine IP-Adressen oder Suchangaben protokollieren

Zukunft: Collaborative Search Engines (CSE)

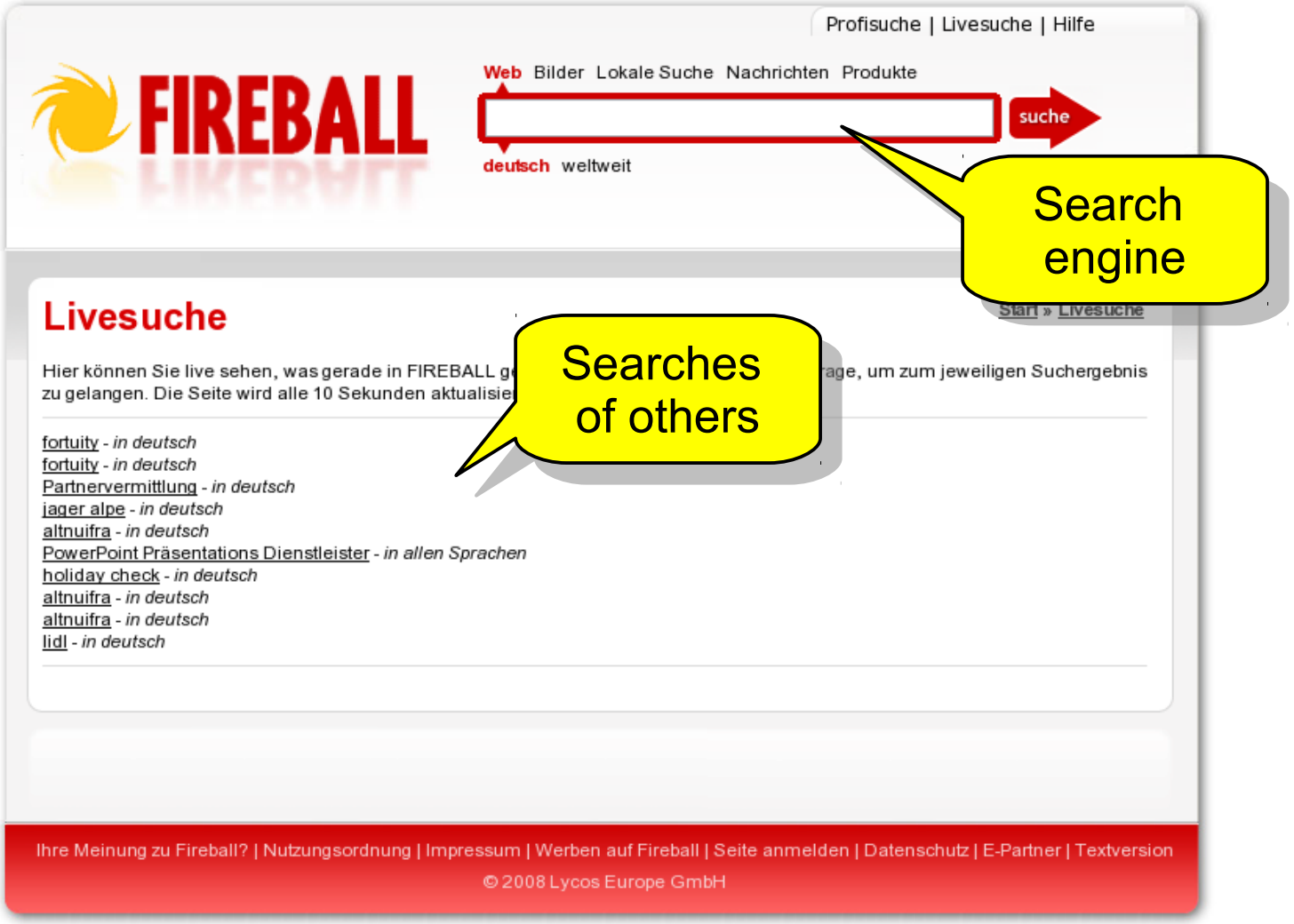
- Neuer Trend für Internet-Suche
  - effizient gemeinsam mit Freunden, Kollegen suchen
  - Unterstützen von unerfahrenen Nutzern
  - spannende Suchen und Suchergebnisse weitergeben
  - andere am Alltagsleben teilhaben lassen

Effiziente Suche

Intensive Nutzerinteraktion

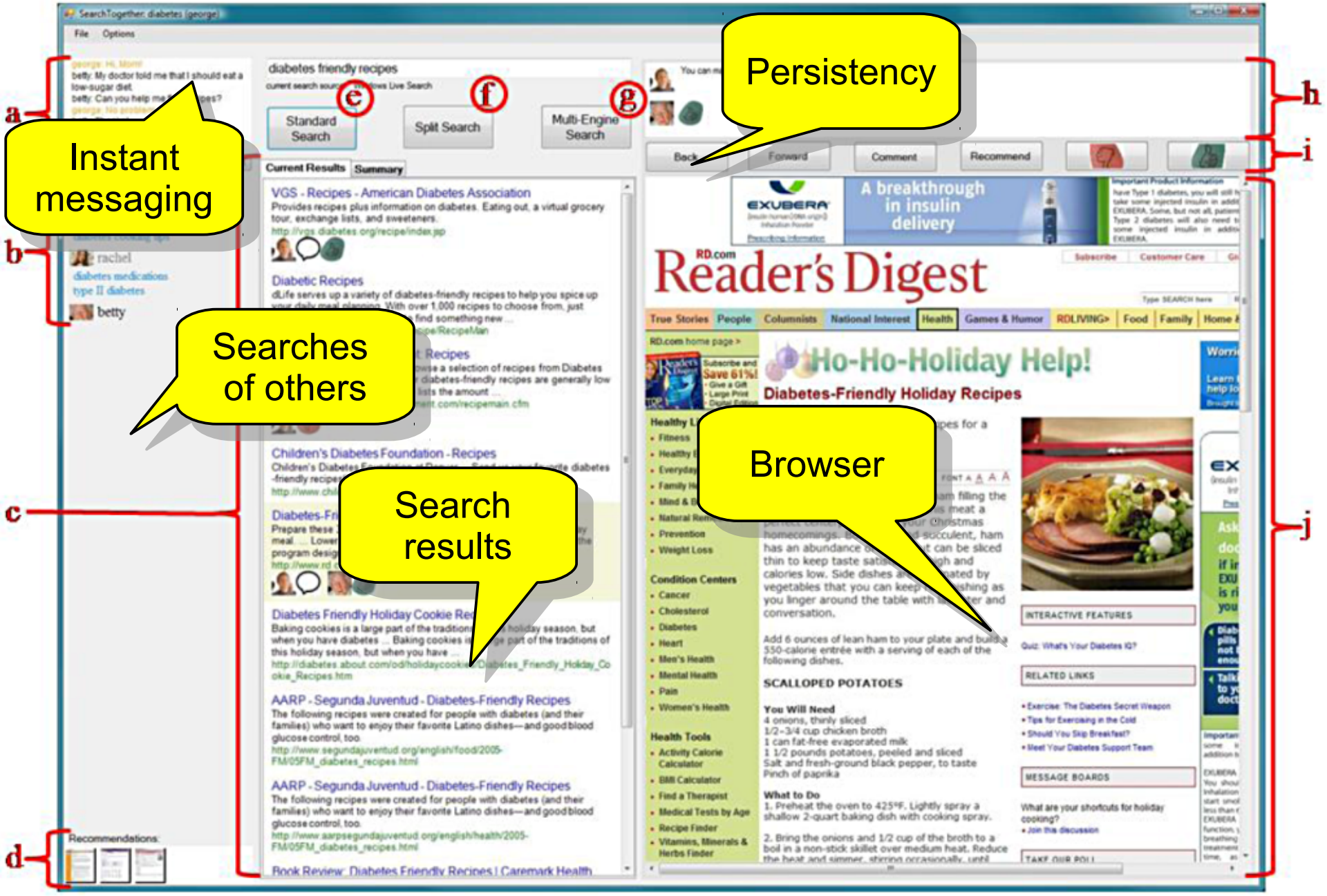
Datenschutz?

Einfachstes Beispiel: Fireball Livesearch*

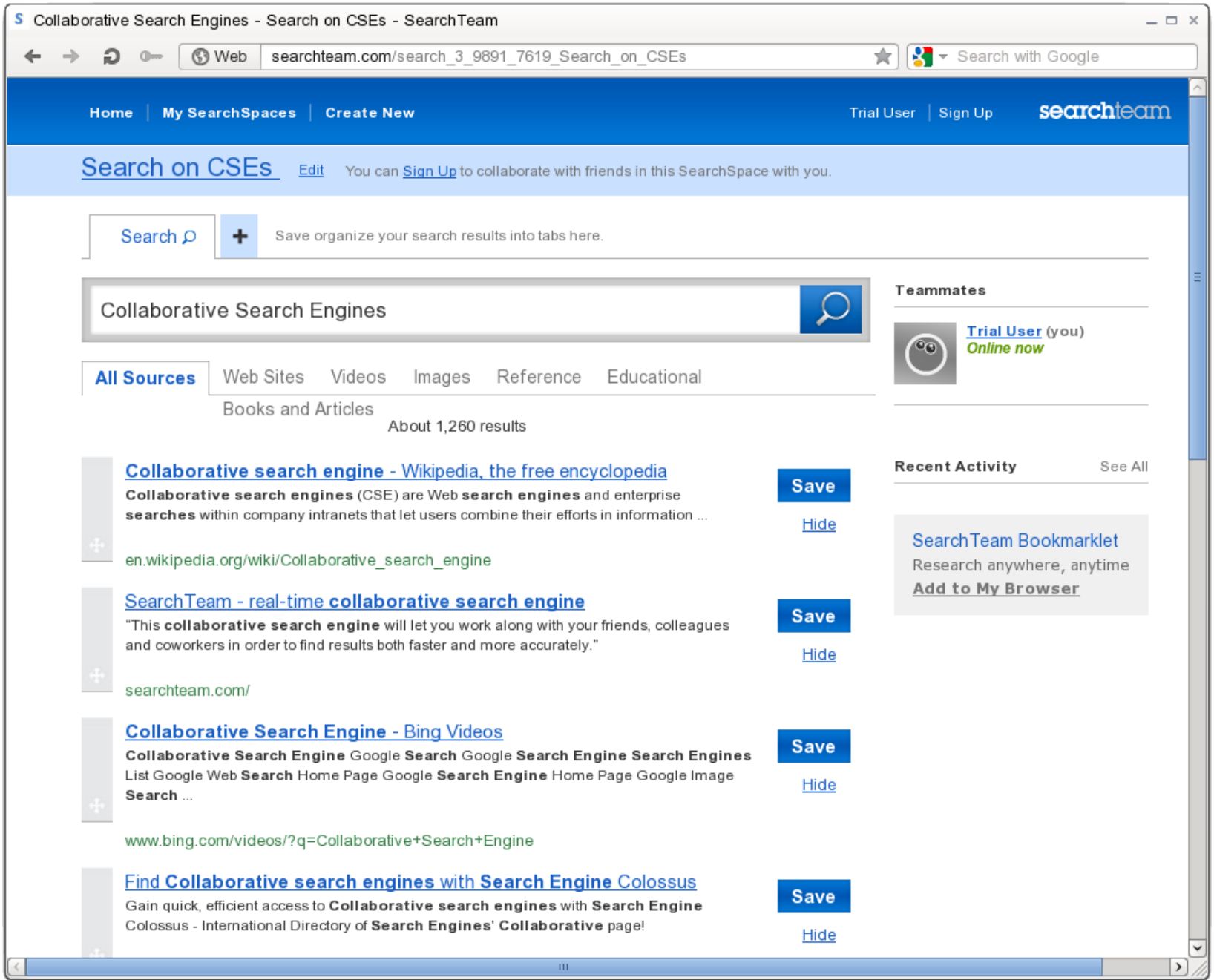


*) inzwischen deaktiviert

Komplexes Beispiel 1: SearchTogether



Komplexes Beispiel 2: SearchTeam



Datenschutz in aktuellen CSE-Prototypen (1/2)

- SearchTogether

Microsoft SearchTogether (Beta)

April 24, 2008: SearchTogether is now available for download!
See below for details on how to install and use SearchTogether.

Privacy Policy

SearchTogether allows you to share your web search results and activity with others. When you join a SearchTogether session, your profile, search queries, results, and web site navigation, along with how you use the SearchTogether tool will be sent to Microsoft for research purposes. In addition, this information will be shared with all the people in that session, regardless of when they join. For more information, read our privacy statement.

Datenschutz in aktuellen CSE-Prototypen (2/2)

- Search Team

Privacy Policy

Your privacy is very important to us. Accordingly, we have developed this Policy in order for you to understand how we collect, use, communicate and disclose and make use of personal information. The following outlines our privacy policy.

- Before or at the time of collecting personal information, we will identify the purposes for which information is being collected.
- We will collect and use of personal information solely with the objective of fulfilling those purposes specified by us and for other compatible purposes, unless we obtain the consent of the individual concerned or as required by law.
- We will only retain personal information as long as necessary for the fulfillment of those purposes.
- We will collect personal information by lawful and fair means and, where appropriate, with the knowledge or consent of the individual concerned.
- Personal data should be relevant to the purposes for which it is to be used, and, to the extent necessary for those purposes, should be accurate, complete, and up-to-date.
- We will protect personal information by reasonable security safeguards against loss or theft, as well as unauthorized access, disclosure, copying, use or modification. [...]

Bedrohungen für den Datenschutz

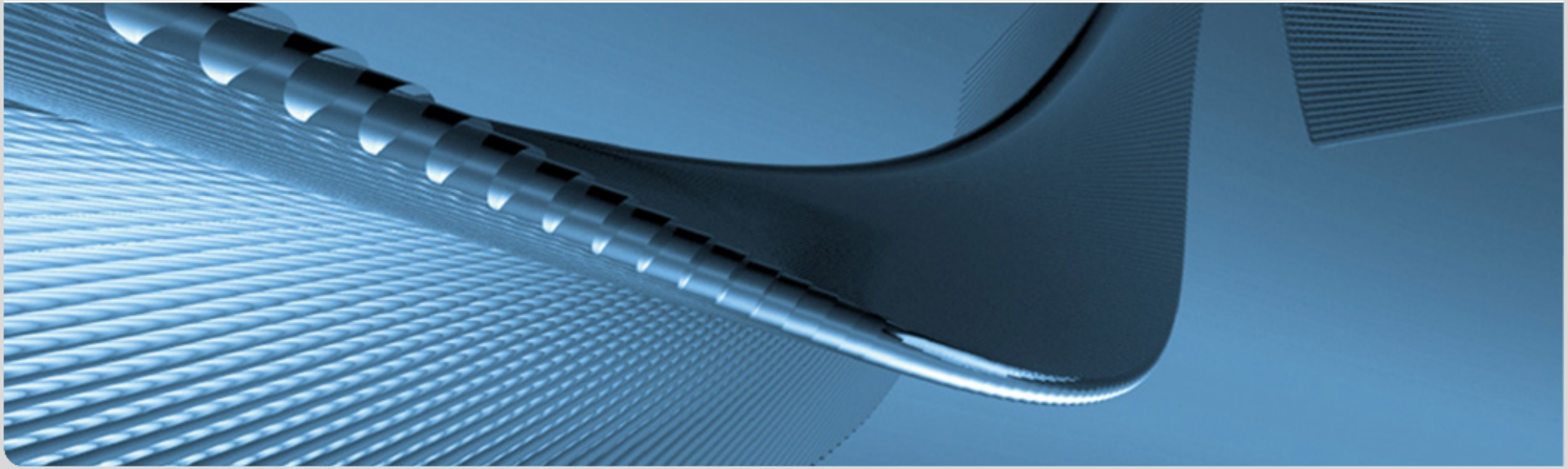
- Jede CSE-Komponente selbst ist problematisch
  - **Suchmaschine**: Nutzerinteressen, Absichten
  - **Weitergabe von Anfragen, Links**: unkontrollierte Verbreitung persönlicher Infos
  - **Kommunikation**: private Gespräche, Kontakte
  - **Speicherung**: Anfragen, angeklickte Links, Kommunikation bleiben für lange Zeit verfügbar

- Verkettung von Informationen potenziert Gefahren
  - “Speichere dauerhaft, wer sich wann für was interessiert hat, und mit wem er das diskutiert hat.”

Karlsruher Institut für Technologie

Zusammenfassung

IPD, Systeme der Informationsverwaltung, Nachwuchsgruppe „Privacy Awareness in Information Systems“



Zusammenfassung

- digitale Identitäten als Untermenge aller technisch abbildbaren Attribute mit Personenbezug
- digitale (Teil-)Identitäten sind grundsätzlich bedrohlich, so harmlos sie auch zunächst scheinen mögen
- Identitätsdiebstahl als derzeit größte Bedrohung durch den unbedachten Umgang mit pers. Informationen
- Verkettung digitaler Identitäten führt zu Informationsanreicherung und Profilbildung
- Aktuelles Fallbeispiel: Verkettung in Suchmaschinen

Literatur

[1] Verkettung Digitaler Identitäten, *Unabhängiges Landeszentrum für Datenschutz Schleswig Holstein, 2007*

[2] Thorben Burghardt, Erik Buchmann, and Klemens Böhm. *Why Do Privacy-Enhancement Mechanisms Fail, After All?,* W2Trust'08