



Big Data: Von Zielgruppen zu Menschen

Joerg Blumtritt

@jbenno

Exkurs: Slow Media

- Big Data (mein Thema) und "Zukunft des Fernsehens" oder gar "Zukunft der Medien" scheint für viele Leute gleichbedeuten zu sein mit der Frage: Was ist mit Google? (Oder böse ausgedrückt: "Wir müssen etwas gegen Google unternehmen!")
- Meine beiden Ko-Autoren Sabria David, Benedikt Köhler und ich hatten 2010 begonnen, uns mit alternativen zu der scheinbar unüberwindlichen Polarität zwischen "Gute alte Medien" und "irgendwie gefährliche neue Medien" auseinanderzusetzen. Unsere Antwort war das 'Slow Media Manifest', das inzwischen in mehr als zehn Sprachen übersetzt und weltweit zum Kern von Forschungsprojekten geworden ist.
- Die Frage, ob es ein Leben im Netz "Ohne Google" gibt, bin ich damals mit einem Selbstversuch nachgegangen, indem ich mich mehrere Monate (fast ein halbes Jahr lang) komplett allen Google-artigen Produkten enthalten hatte.

- Ergebnis: für Nutzer ist Google einfach zu ersetzen. Für Publisher allerdings nicht (mehr). Die haben sich selbstverschuldet schon lange in eine einseitige Abhängigkeit zu Google manövriert.
- hier der Link:
<http://slow-media.net/ohne-google>

Blog [Slow Media Institut](#) [Das Slow Media Manifest](#) [Resonanz](#) [Manifesto/Manifeste/Манифест](#) [Kontakt](#) 



slow media

"Fast könnte man sagen, daß vom Tempo, der Geduld und Ausdauer des Verweilens beim Einzelnen, Wahrheit selber abhängt" (*twa*)

Ohne Google.

Jörg Blumtritt, 25.10.2010.

 [Read this post in English!](#)



The strongest arguments prove nothing so long as the conclusions are not verified by experience. Experimental science is the queen of sciences and the goal of all speculation.

Roger Bacon

Ich habe mich entschlossen, ein Experiment zu machen: ich werde von heute an die Suche von Google nicht mehr nutzen.

SLOW MEDIA

Das Blog für diskursiven Essayismus, medienübergreifendes Forschen und collective theorism: Medienwandel in Theorie und Praxis.

AUTOREN



Sabria David (*sabria*, [Slow Media Institut](#))



Jörg Blumtritt (*jbenno*, [Galerie Royal](#), [Memetic Turn](#))



Benedikt Köhler (*benedikt*, [viralmythen](#), [Isarrunde](#))

Von Zielgruppen zu Menschen

- "Zielgruppe" (wie viele weitere Marketing-Begriffe - z.B. "Kampagne") ist martialischer Jargon.
- Big Data ist der dritte große Paradigmenwechsel, der Publishing schrittweise von Menschenmassen und ihren Massenmedien wegbewegt:
- Mitte der 90er kam das Web mit einer wesentlichen Demokratisierung des Zugangs zu Inhalten.
- Mitte der 2000er folgten Social Media. Wikipedia, Facebook, Twitter etc. und einfache Content Management Systeme wie Wordpress machen - nicht nur potenziell - jeden Menschen zum Publisher.
- Big Data schließlich macht es möglich, selbst bei Millionen (oder gar Milliarden) von Nutzerinnen und Nutzern Distribution, Kommunikation und Inhalte komplett auf jede einzelne Person zuzuschneiden. Dabei geht es nicht mehr um die Auswahl aus vorgefertigten Produktvarianten wie sie schon heute im Bereich Automobilneuwagen bereitgestellt wird. Es geht darum das gesamte Produkt potenziell auf jede einzelne Person zuschneiden.
- Auch wenn die offensichtlich wichtigsten Anwendungsfälle aus dem Bereich Medizin und personalisierte, genom- oder mikrobiomspezifische Medikamente stammen, gilt der Paradigmenwechsel für Medien gleicher Weise.
- Auf den nächsten Charts kommen ein paar einleitende Bemerkungen, was Big Data bedeutet.



Big Data

- Paradigmenwechsel in Verarbeitung, Analyse und Interpretation von Daten
- Von festen Strukturen zu beliebigen Inputs
- Von vorgefassten Hypothesen zu Mustererkennung
- Völlig neue Player im Bereich Software und Infrastruktur
- "Demokratisierung" von Datenanalyse - "Ende des Herrschaftswissens"
- Wechsel Von Statistik und Business Intelligence zu Data Science

Das erste Chart zeigt eine Karte von Berlin, die Eric Fisher aus den Geodaten von Tweets (Blau) und Flickr-Bildern erstellt hat. Es sind die *Metadaten*, die die interessante Information tragen. Die Karte ist völlig ohne "geografisches Knowhow" entstanden.

Dieses Beispiel ist typisch für Big Data: Strukturen aus Daten erkennen, die nicht explizit dafür gesammelt worden waren.

Link: <https://www.flickr.com/photos/walkingsf/>

(unter Creative Commons-Lizenz cc-by-sa veröffentlicht)



Ernesto Ramirez

@eramirez

 Follow

Data is made of people. Read this great post by [@wilbanks](#) on Larry Page and Pooling Medical Data: del-fi.org/post/801749671... [#quantifiedself](#)

 from Los Angeles, CA

 Reply  Retweet  Favorite  More

- Der "Rohstoff" für Big Data sind Menschen. Die Daten widerspiegeln unser Leben, entwerfen ein Bild von uns, machen uns analysierbar.
- Daten sind nicht "virtuell", sondern Teil der Wirklichkeit (wörtlich: sie haben Wirkung auf unser Leben). Man kann sich das versinnbildlichen, dass Daten uns wie eine, für unsere Augen unsichtbare, zusätzliche Dimension umgeben.
- Vergleichbar zum Mikroskop, das uns die unsichtbare Welt der Mikroorganismen um uns herum sichtbar macht, werden die Daten für uns durch Datenanalyse und Visualisierung greifbar.
- Daten zeichnen ein extrem präzises Bild unseres Verhaltens. Sie ermöglichen uns, persönliche Vorhersagen auf einzelne Menschen zu machen, statt wie bisher auf grobe Aggregate wie "Geschlecht" oder "Alter" zurückfallen zu müssen.
- Marketing, mit Search, Online-Retail, Online-Targeting und Social Media, ist der wichtigste Treiber für den technologischen Fortschritt bei Big Data.

Data

- Stichproben
- begrenzte Zahl an Variablen
- Konsistenz
- Stammdaten
- Metadaten

- Big Data

- Totalerhebung
- flexible Zahl an Variablen
- „Immunability“
- Daten
- Daten

Nicht-Repräsentativität

Zielgruppe

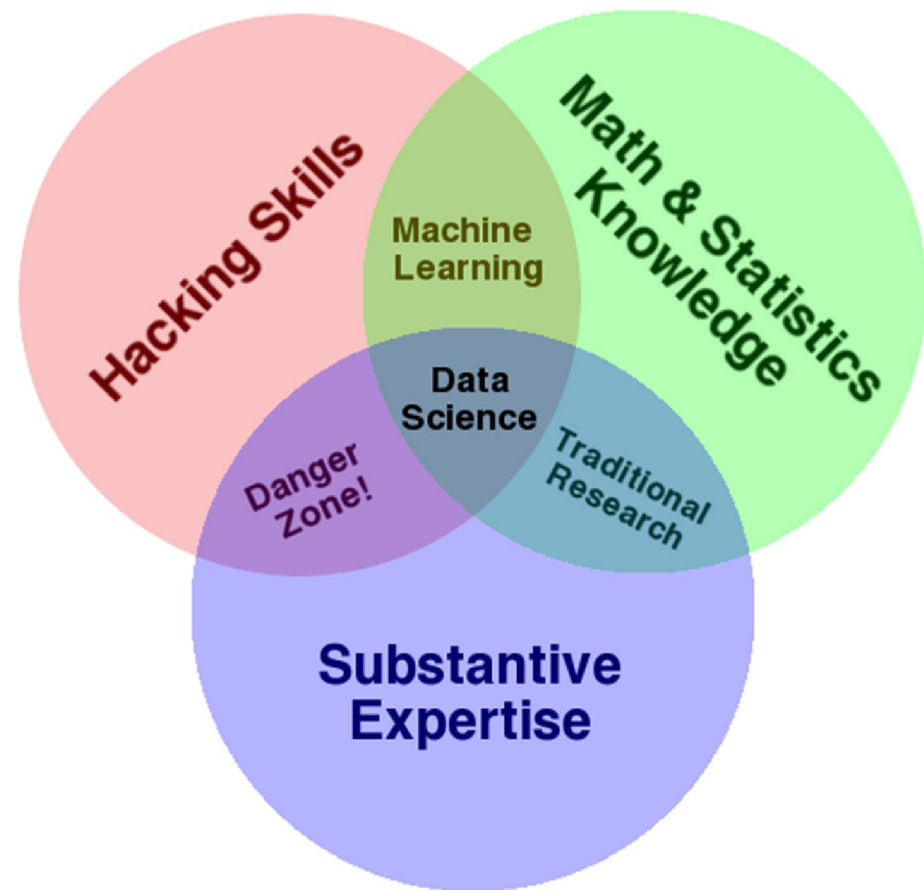
- Menge, die durch ein typisches Element repräsentiert wird
- Anonym
- Losgelöst vom konkreten Handeln der einzelnen Personen
- Aggregierte Modelle

Big Data Ansatz

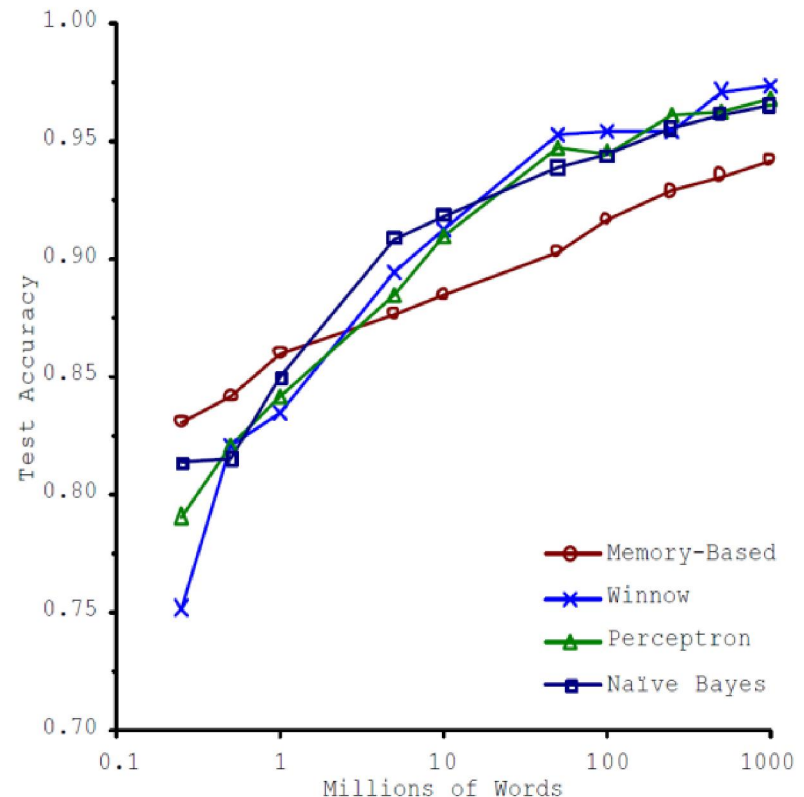
- Einzelne Personen werden nicht zu Mengen zusammengefasst
- Personalisiert
- In direkter Beziehung zum persönlichen Handeln
- Modelle auf Basis von Einzelfällen

Data Science

- Data Science unterscheidet sich von Statistik und empirischer Sozialforschung:
- Data Science bedeutet nicht nur, Verfahren aus vorgegebenen Packages wie SPSS oder SAS einsetzen zu können, sondern den gesamten Prozess, von der Datenerhebung bis zur Programmierung des Produktivsystems zu überblicken.
"Statisticians use software, data scientists code."
- Im Focus der Data Science steht Recheneffizienz und Anwendbarkeit. Modelle müssen agil angepasst werden können. Ausgabe der Ergebnisse erfolgt in der Regel als interaktives Web-Dashboard. Neben Standard-Tools wie Tableau Software werden diese Dashboards in JavaScript D3, HTML5 etc. umgesetzt.



- Egal welche statistische Methode zur Analyse gewählt wird - ab einer ausreichend großen Datenmenge ist der Unterschied der Verfahren nicht mehr relevant. Entscheidend ist vielmehr die Rechengeschwindigkeit. Relativ grobe Methoden wie der Text-Vektor-Vergleich (s. "Textmining" unten) sind intelligenten aber aufwändigen Verfahren schnell im Vorteil.
- Bei Millionen oder gar Milliarden von Datenpunkten sind bereits kleinste Unterschiede zwischen Untermengen einer Stichprobe signifikant. Klassische Hypothesen-Tests führen daher bei Big Data nicht weiter.
- Statt eine Hypothese aufzustellen, zu testen und dann für wahr zu halten, geht man bei Big Data iterativ vor: die Brauchbarkeit der Modelle wird ständig gemessen und mehrere Modelle mit Variation der Parameter gegeneinander getestet ("a/b-Testing, evolutionäre Methode); das zum jeweiligen Zeitpunkt bessere Modell wird weitergeführt, bis es seinerseits abgelöst wird.



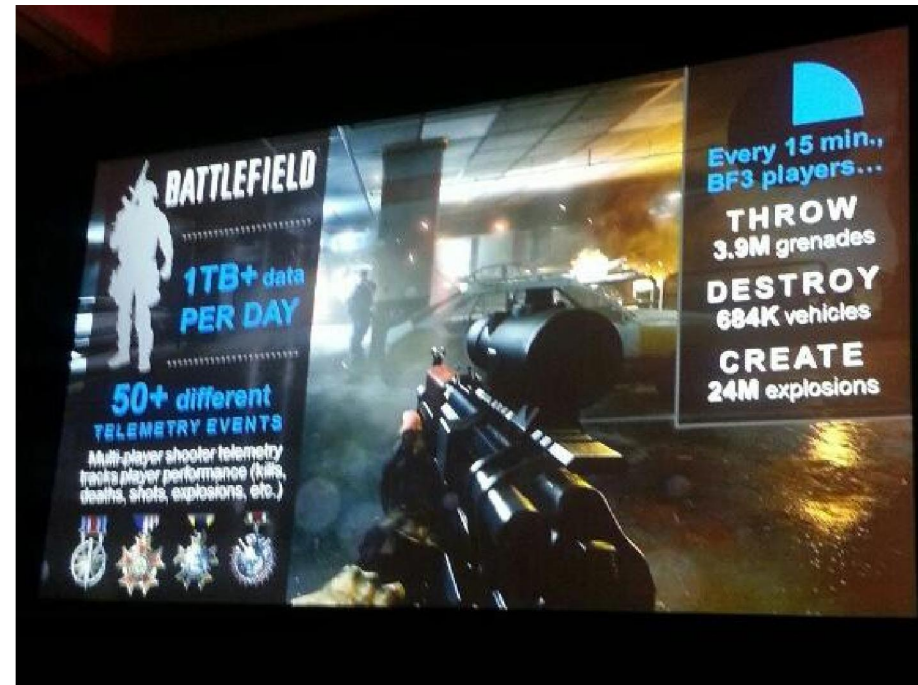
Michele Banko und Eric Brill 2001:
<http://acl.ldc.upenn.edu/P/P01/P01-1005.pdf>

- Alle Daten können relevante Aussagen für unsere Zwecke liefern. Was früher Meta-Daten waren, sind heute häufig viel aussagekräftigere Inputs für unsere Modelle.
- Es ist auch nicht sinnvoll, zwischen Stammdaten und Bewegungsdaten zu unterscheiden. Während sich Stammdaten über die Zeit sehr wohl ändern können (z.B. die Adresse durch Umzug, die Firmierung, etc.), sind Bewegungsdaten wertvolle Beschreibungen von Einstellung und Verhalten von Menschen: Was jemand kauft, sagt oft mehr über ihn aus, als sein Alter.
- Beim Modelling sind natürlich Kollinearitäten und Redundanzen zu beachten. Moderne Modelling-Packages unterstützen allerdings die sinnvolle Reduzierung der Variablen.



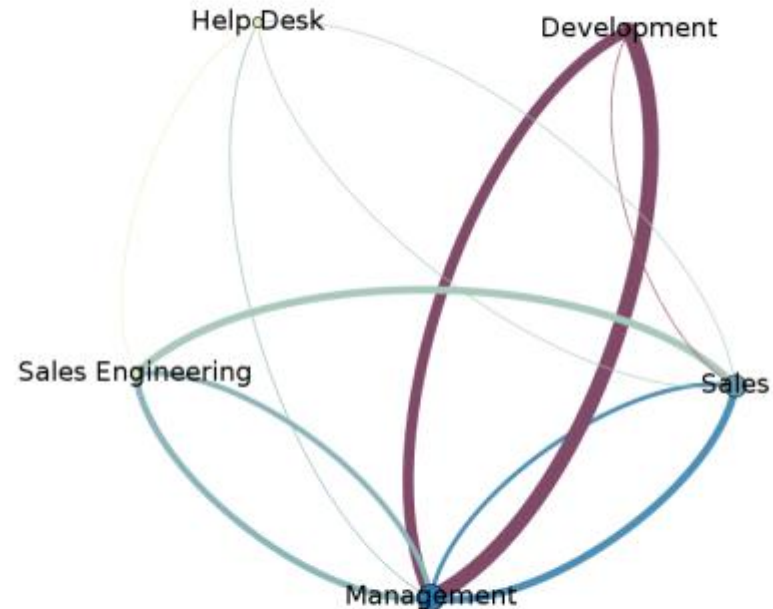
Inhalte in Echtzeit personalisieren

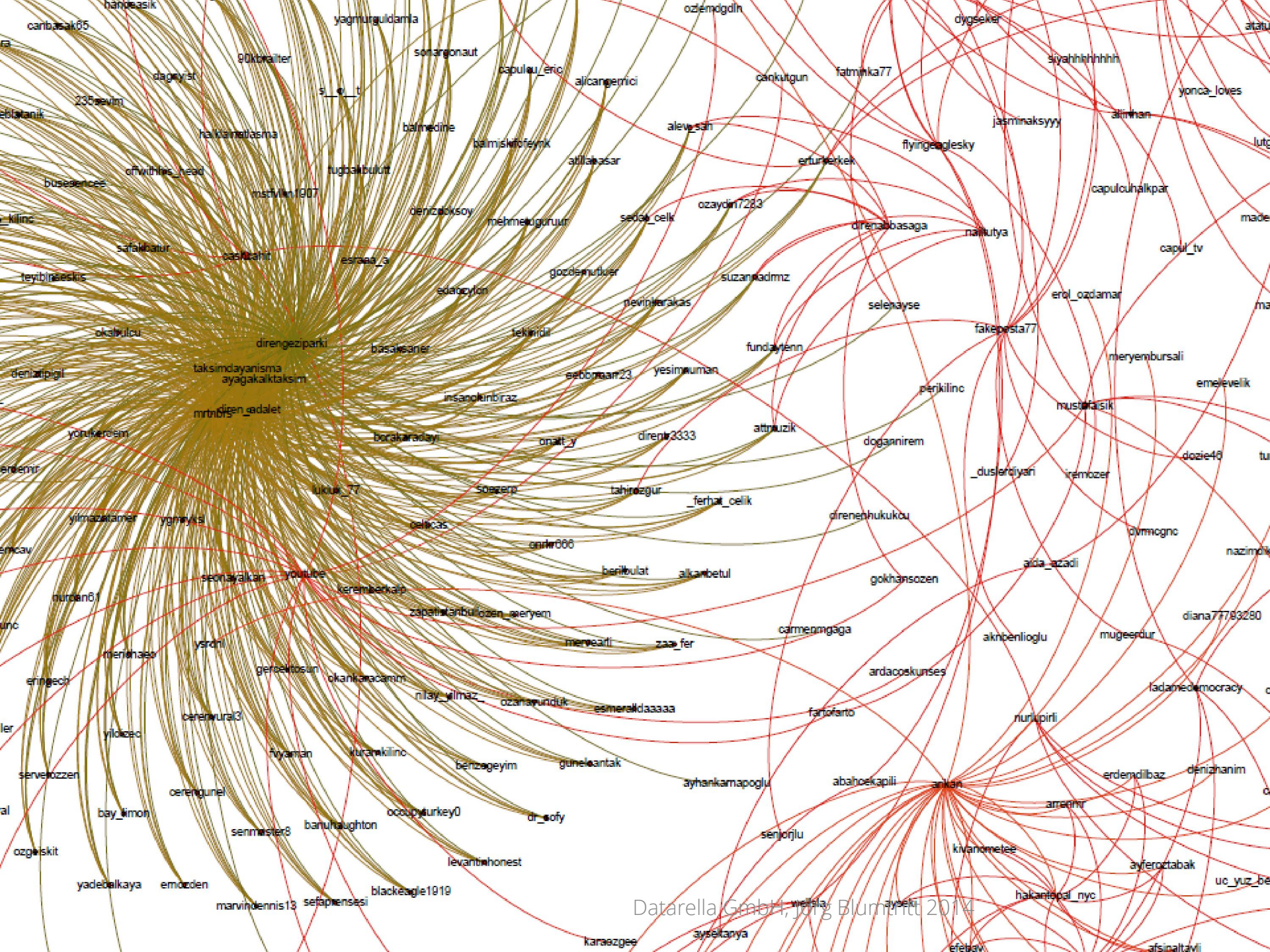
- Personalisiert e Spiele in Echtzeit.
- Während laufenden Spiels werden unterschiedliche Einstellungen und Rahmenbedingungen der Spiele auf das individuelle Spielerlebnis angepasst.
- Jede Spielerin und jeder Spieler erhält potenziell das Spiel, dass optimal passt.
- Das System skaliert für millionen Spieler parallel.
- (das Bild habe ich während des Vortrags von EA bei der Strata Conference 2011 gemacht).



Kommunikation im Team

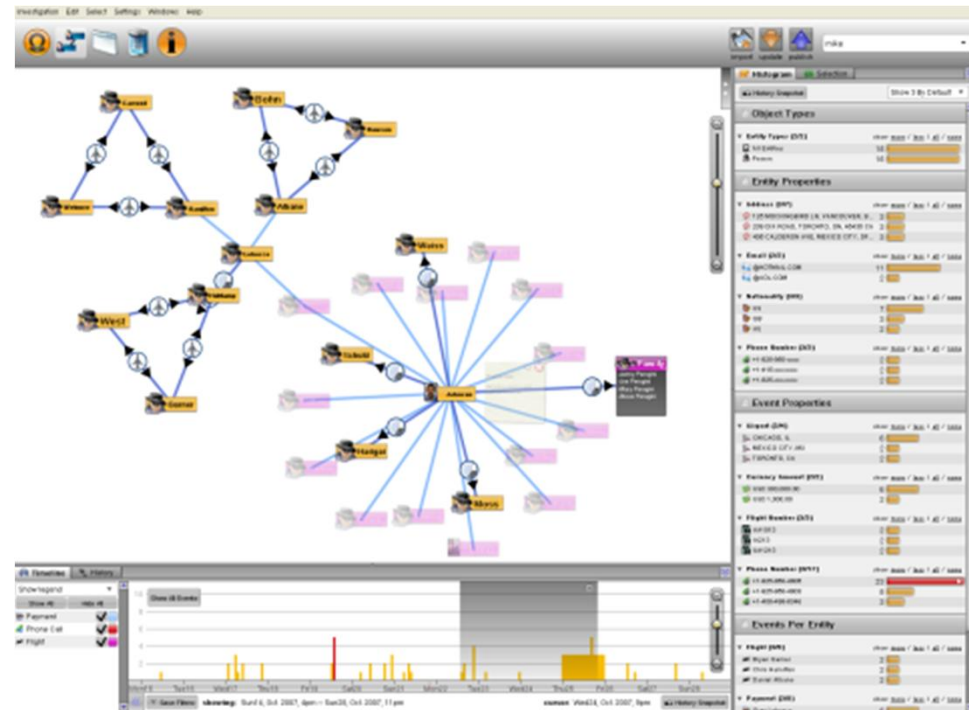
- Analyse "Wer schickt wem Mails"
- Ergebnis: sofort erkennbar ist die ungleichmäßige Kommunikation zwischen den Abteilungen (die sich genau entsprechend in den auftretenden Qualitätsproblemen zeigt).
- Die Analyse kann automatisiert für alle Teams eines Unternehmens und alle Filialen durchgeführt werden und fördert zu Tage, was durch einfache Zählstatistik der Kommunikation nicht leicht zu sehen ist.
- Auf der nächsten Seite : Twitter-Nutzer, verbunden durch den Hashtag #direngezi (Visualisierung mit Gephi)





Beispiel: Palantir

- Palantir liefert die Dateninfrastruktur hinter der NSA.
- Palantir liefert Netzwerkanalysen auf Basis von Verbindungen zwischen Personen, die aus allen möglichen Datenquellen gezogen werden.
- Das Gründer-Team ist ein Teil der Gründer von Paypal. Paypal wurde als Online-Zahlungssystem erfolgreich, weil es viel weniger Zahlungen als potenziell betrügerisch ablehnen musste, als die Systeme der klassischen Abwickler. Das Prinzip ist dasselbe: Daten einfach darstellen und von Experten manuell beurteilen lassen.





Import Update Publish

Search: make



Program Selection

History Snapshot Show 3 By Default

Object Types

Entity Types (NT)

All Entities	14
Persons	14

Entity Properties

Address (NY)

125 WASHINGTON ST, MANHATTAN, N.Y.	3
226 CHAMBERS ST, MANHATTAN, N.Y.	3
400 CALDERON AVE, NEW YORK, NY	3

Email (NY)

@GMAIL.COM	11
@OL.COM	2

Natality (NY)

NY	7
NY	3
NY	2

Phone Number (NY)

+1-820-950-XXXX	2
+1-820-950-XXXX	2
+1-820-950-XXXX	2

Event Properties

Event (NY)

NY	4
NY	2
NY	2

Primary Account (NY)

NY	6
NY	2

Flight Number (NY)

NY	2
NY	2
NY	2

Phone Number (NY)

+1-820-950-XXXX	23
+1-820-950-XXXX	6
+1-820-950-XXXX	2

Events Per Entity

Event (NY)

All Flight Number	2
All Phone Number	2
All Email Address	2

Payment (NY)

NY	4
----	---

Entities History

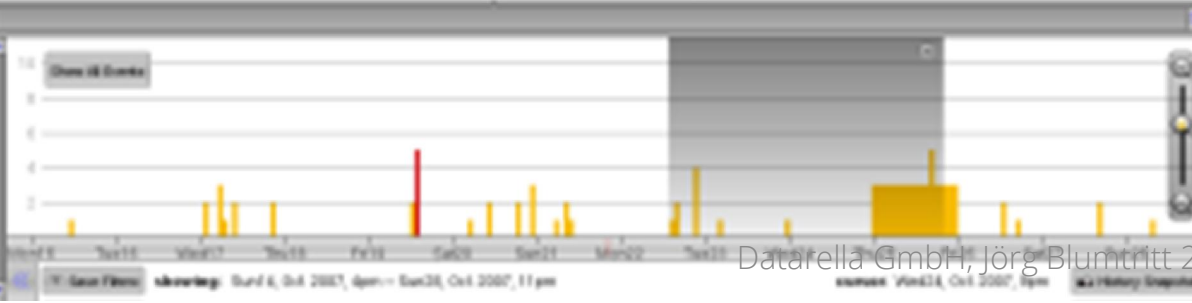
Show Legend

Show All Hide All

Payment

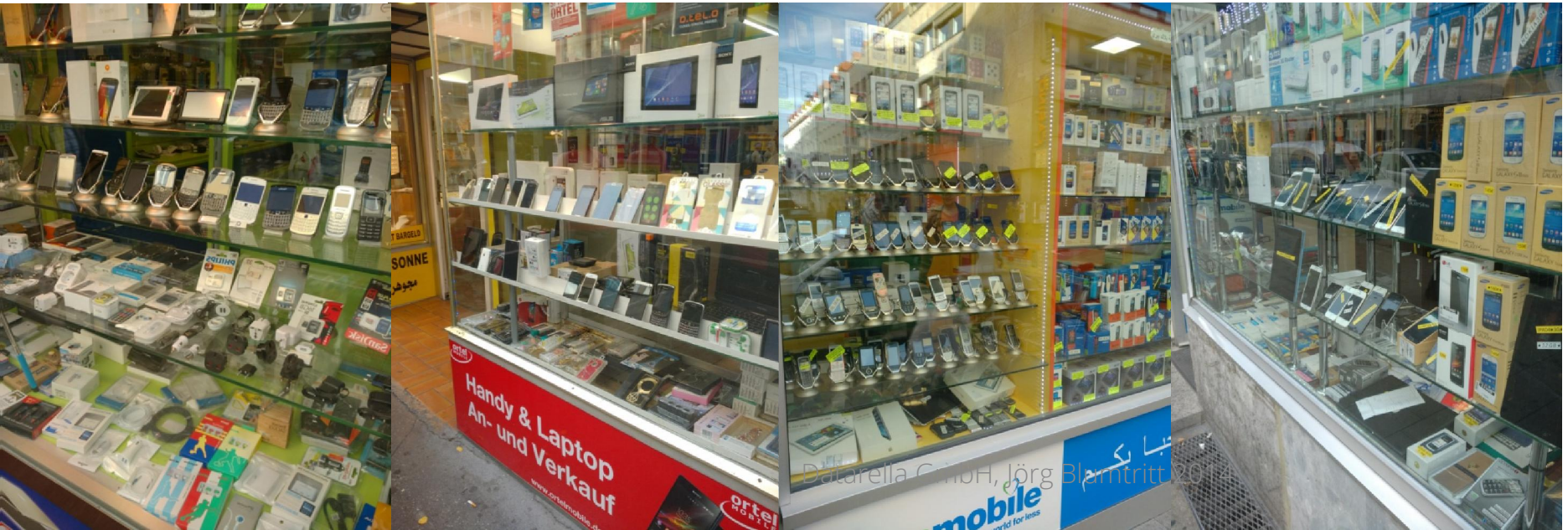
Phone Call

Flight



Mobile

- Zwei Milliarden Menschen nutzen Smartphones (das sind dreimal mehr, als einen PC nutzen).
- Smartphones sind aber weit mehr als "mobile Computer": Sie tragen mehr als zwanzig Sensoren, die kontinuierlich Daten über unser Verhalten, und unsere Umgebung messen, aufzeichnen und sammeln.
- Mobile wird die wichtigste Datenquelle für Big Data und läuft damit den Social Networks den Rang ab.
- Über die Nutzung durch Menschen, hinaus gibt es bereits mehr als fünf Milliarden Mobilfunkgeräte, die mit Objekten verbunden sind. Beispiele sind Autos und andere Maschinen, die zum "Internet of Things" verbunden werden.



London Evening Standard

[HOME](#) | [NEWS](#) | [SPORT](#) | [FOOTBALL](#) | [SHOWBIZ](#) | [GOING OUT](#) | [LIFESTYLE](#) | [FASHION & BEAUTY](#) | [VIDEOS](#) | [BUSINESS](#) | [COMME](#)

[London](#) | [Crime](#) | [Politics](#) | [Mayor](#) | [Transport](#) | [Education](#) | [Health](#) | [TECHNOLOGY](#) | [UK](#) | [World](#) | [Campaigns](#) | [Diary](#)

BOMB DISPOSAL OFFICERS BLOW UP BAG OUTSIDE IMPERIAL WAR MUSEUM AMID REMEMBRANCE DAY SECURITY

Average smartphone user checks device 221 times a day, according to research



MOST READ



system



afterparty





1

**Computing as an extension
of ourselves**

After several decades of 'convergence', where more and more functions were squeezed into single devices such as smartphones, a new trend began in the 2010s. Inexpensive electronics meant that small, single-purpose devices could be produced, enabling a wide range of connected, wearable computing products to become a practical reality.

- 1 Sony SmartBand wearable fitness band, 2014 A range of wearable health monitors like this were introduced in the mid-2010s. The band connects to an app to track activity, location and communication.

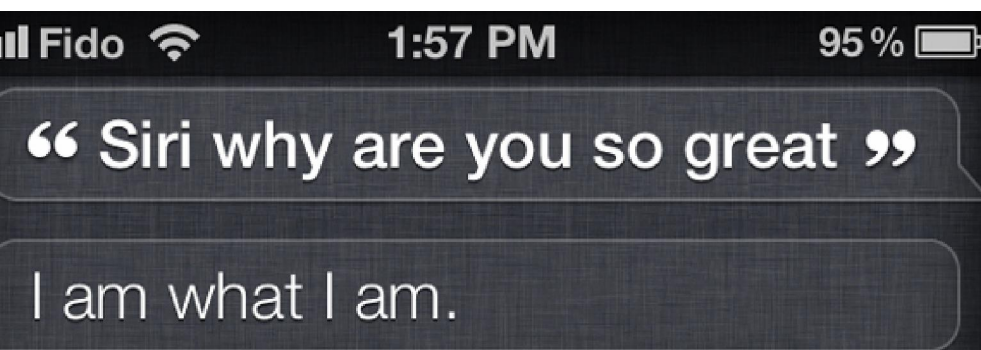
Donated by Sony
Object No: 2014-145

“ Siri why are you so great

I am what I am.

Werbung

- Programmatic Buying - der voll automatisierte Einkauf von Werbeplätzen und die voll automatisierte Platzierung der Werbung ist die Zukunft der Mediaplanung.
- Solides Big-Data-Know-How ist die Voraussetzung, hier erfolgreich zu sein und sich nicht "über den Tisch ziehen zu lassen".
- Unternehmen, die Big Data im Marketing für sich einsetzen, werden wesentlich erfolgreicher für sich werben.
- Siri und Google Now sind "Intelligente Agenten" mit millionenfacher Verbreitung. Zunehmend werden Nutzerinnen und Nutzer lästige Entscheidungen (wie z.B. Filtern von Werbung) an Agenten abgeben. Die Distribution von Inhalten (und Werbung) wird ein Wettrennen mit den Algorithmen dieser Agenten.

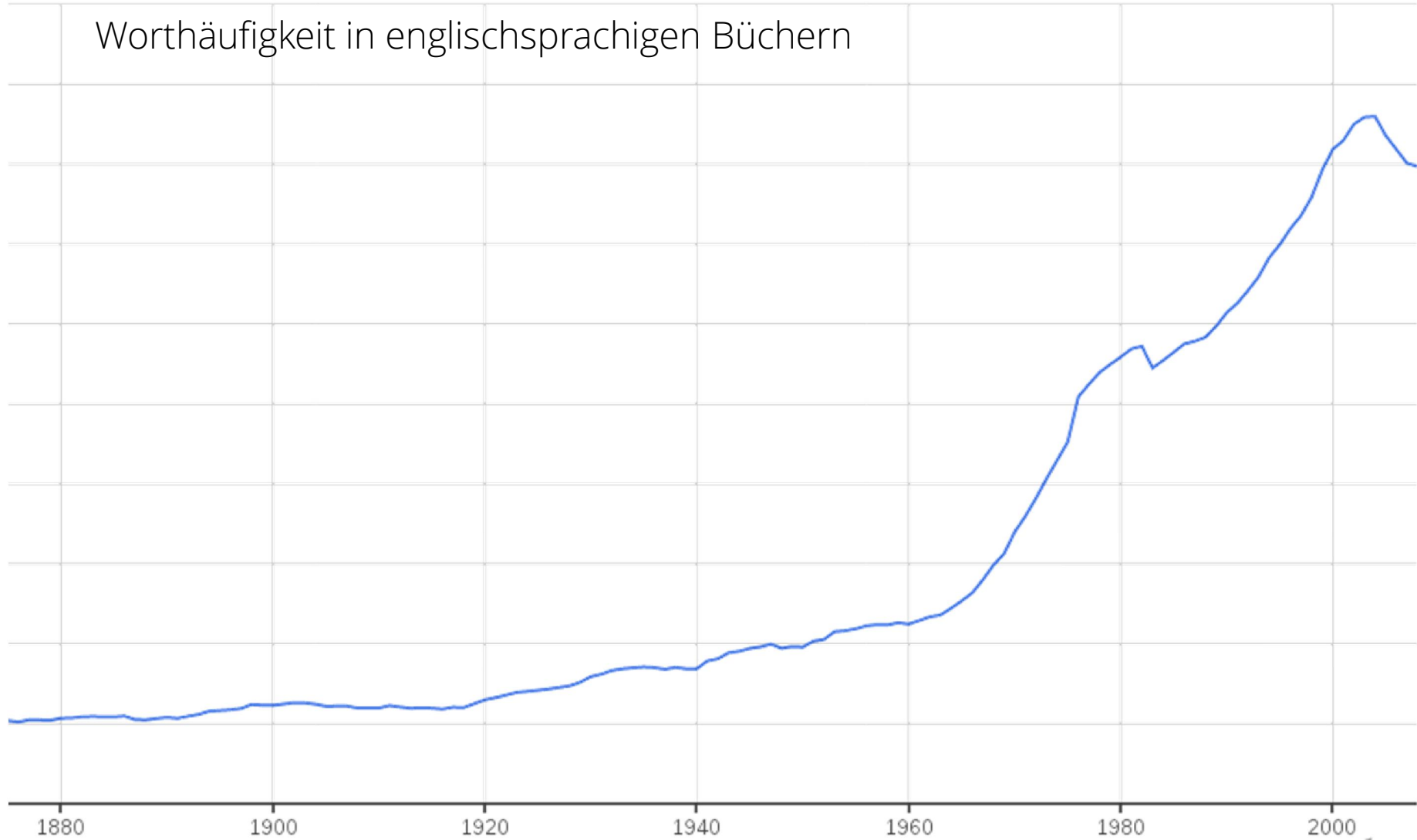


Datenschutz und informationelle Selbstbestimmung

- Die Snowden-Enthüllungen haben deutlich gemacht, dass Datenschutz nichts hilft, gegen staatliche Überwachung. Schon heute haben die Finanzämter vollen Zugriff auf unsere Konten. Die Voratsdatenspeicherung wurde zwar vom Europäischen Gerichtshof vor kurzem für rechtswidrig und nichtig erklärt. Dennoch werden insbesondere Mobilfunkdaten in unglaublichem Umfang von den Sicherheitsbehörden ausgewertet. Zwei Millionen Funkzellenabfragen gab es 2012 alleine in Schleswig-Holstein!
- Unternehmen wie Google und selbst das viel gescholtene Facebook scheinen dagegen bis heute relativ pfleglich mit unseren Daten umzugehen - ihre Furcht vor Imageverlust und daraus folgendem wirtschaftlichem Schaden ist zu groß, als dass sie sich Fehler erlauben könnten.
- Durch die "Heartbleed"-Sicherheitslücke, die im April bekannt wurde, sind Passwörter, vertrauliche Nachrichten und alle möglichen anderen persönlichen Daten im gesamten Internet über Monate, wenn nicht Jahre mehr oder weniger ungeschützt zugänglich gewesen, - zumindest für die, die von der Sicherheitslücke wussten.
- Datenschutz wiegt uns in falscher Sicherheit, macht uns unaufmerksam. Statt Daten zu schützen, sollten wir stärker auf offene Daten drängen. Wenn jeder Zugriff hat, gibt es kein hierarchisches Gefälle mehr,. Wenn wir uns bewusst sind, dass die Daten potenziell von jedem Menschen eingesehen werden können, werden wir uns so verhalten, dass wir nicht mehr erpressbar sind.
- Die Vorstellung von "persönlichen Daten", von Privatheit, ist zu tiefst bürgerlich (im wörtlichen Sinne); sie ist dem Wunsch der Bürger entsprungen, sich vom Adel abzugrenzen. Wie öffentlich die Aristokratie im Feudalismus lebte, kann man bis heute in den offenen Zimmerfluchten der Barockschlösser bewundern. Für Menschen auf den Dörfern gab es keine Privatsphäre. Alle lebten in einer Kammer zusammen. Erst mit der bürgerlichen "Familienwohnung" des 19 Jhd. konnte sich Privatsphäre als kultureller Wert entwickeln.
- Die Digitalisierung ist nicht mehr rückgängig zu machen. Die Vorteile, Daten zu teilen sind so überwältigend, dass die Gesellschaften einen "Datenschutz" oder "Informationelle Selbstbestimmung" wie im 20. Jahrhundert nicht mehr zulassen werden. Lernen wir, damit zu leben!

Privacy

Worthäufigkeit in englischsprachigen Büchern



Data Protection

Häufigkeit des Suchwortes in der Google Suche



THE PRINCIPLES OF DATA LOVE

Love data

Data is essential
 Data must flow
 Data must be used
 Data is neither good nor bad
 There is no illegal data
 Data is free
 Data can not be owned
 No man, machine or system shall interrupt the flow of data
 Locking data is a crime against datanity

Love data

DATA LOVE

Resources
 Flickr Group
 #DataLove IRC-Chatroom
 Twitter
 Telecomix Status.net

No Copyright, no license. Sorry, no kitties.
 Part of the Telecomix Network.



info [at] datalove [dot] me
 info [at] werebuild [dot] eu
 info [at] telecomix [dot] org

"Data begs to be used" (Bruce Schneier)

- "We're living in a world with more transparency, we need to learn to do intelligence with more transparency too." (Bruce Schneier)
- In "Snow Crash", Neal Stephenson imagines a Central Intelligence Corporation, the CIA and NSA becoming a commercial service where everybody just purchases the information they'd need. This was, what came immediately to my mind when I read through the transcript of the talk on "Intelligence Gathering and the Unowned Internet" that was held by the Berkman Center for Internet and Society at Harvard, starring Bruce Schneier, who for every Mathematician in my generation is just the godfather of cryptography. Bruce has been arguing for living "beyond fear" for more than a decade, advocating openness instead of digging trenches and winding up barbed wire. I am convinced, that information does not want to be free (as many of my comrades in arms tend to phrase). However I strongly belief Bruce is right: We can hear data's call.

Link to the event: <http://cyber.law.harvard.edu/events/2014/04/unownedinternet>

Link to the transcript:

<https://docs.google.com/document/d/1djjE3lBfl2aiO7DGQvEGKx4p0qVcCMD5OH6tWuRPh9g/edit>

Cyborg Rights

- My friend Enno Park, shown in this picture, is deaf. But cochlea implants made him hear again. Enno is a hacker, so he wants to access the technology that has become part of his body. But he can't. It is proprietary, DRM protected, a black box.
- So he founded the German Cyborg Association to fight for the rights of people who entangle their lives with technology.
- We should keep reminding people that our self-tracking is just a very visible and striking aspect of something that has long become ubiquitous. Technology as always been an "extension of man" (sorry for the gendered term, but it is a quote by Teilhard from the 1930s). As it would have been useless for someone in the 19th century to separate "the railroad world" from "the real world", it is useless to maintain the fiction of data being somehow virtual. We have long become cyborgs.

Cyborgs e.V.: <http://cyborgs.cc/>

079: Hörspiele, Schweine und Maschinenmenschen



Elektrischer Reporter



079: Hörspiele, Schweine und Maschinenmenschen



Links zu Big Data / Data Science

- <http://beautifuldata.net> Blog zu Big Data mit einigem Code
- <http://datarella.com/blog> Blog zu Quantified Self und Behavioral Data
- <http://radar.oreilly.com/> Beste Quelle für Big-Data-News
- <http://twitter.com/jbenno/bigdata> Twitter-Liste zu Big Data
- <https://github.com/ptwobrussell/Mining-the-Social-Web-2nd-Edition/>
... sehr gutes Buch zum Do-it-Yourself-Lernen mit iPython
- <http://shop.oreilly.com/product/0636920028529.do>
"Doing Data Science", Cathy O'Neil, Rachel Schutt sehr gutes Buch aus dem Google-Universum mit vielen Übungen
- <http://www.amazon.de/Automate-This-Algorithms-Markets-World-ebook/dp/B0064W5UAS/>
"Automate This", Christoph Steiner



Jörg Blumtritt

@jbenno

Datarella GmbH

Oskar-von-Miller-Ring 36

80333 München

089/44 23 69 99

info@datarella.com