

# Sicher Surfen

## Workshop-Reihe zum Datenschutz

20.6.2018



CryptoParty Graz

# Wer überwacht uns ?

- Der Staat (Sicherheitsgesetze)
- Geheimdienste (NSA, BND, ...)
- Firmen als Datensammler
- Kriminelle
- Internet der Dinge

# Corporate Surveillance in Everyday Life

Wolfie Christl

Report and web publication on how companies collect, analyze, and use personal data.

June 2017

<http://crackedlabs.org/en/corporate-surveillance>

[http://crackedlabs.org/dl/CrackedLabs\\_Christl\\_CorporateSurveillance.pdf](http://crackedlabs.org/dl/CrackedLabs_Christl_CorporateSurveillance.pdf)

Grafiken: CrackedLabs



# Digitale Profile über Milliarden: Überwachungskapitalismus 2018 und wie weiter

- Vortrag von Wolfie Christl, 21.9.2018, Konferenz netzpolitik14, Berlin
- Audio- und Videomitschnitt

[https://media.ccc.de/v/np14-10-digitale\\_profile\\_ueber\\_milliarden\\_ueberwachungskapitalismus\\_2018\\_und\\_wie\\_weiter](https://media.ccc.de/v/np14-10-digitale_profile_ueber_milliarden_ueberwachungskapitalismus_2018_und_wie_weiter)

## II.

# Analyse von Menschen durch Finanz-, Versicherungs- und Gesundheits-Unternehmen

# Data Mining und Data Analytics

„Data Mining“ und „Data Analytics“ beruhen auf statistischen Korrelationen mit Wahrscheinlichkeitsstufen

Liefere Merkmals- und Charaktereigenschaften über dem Zufallswert, sind aber jedenfalls nicht sehr exakt

Diese Methoden werden benutzt, um Menschen zu

- Sortieren
- Kategorisieren
- Kennzeichnen
- Beurteilen
- Bewerten
- Reihen

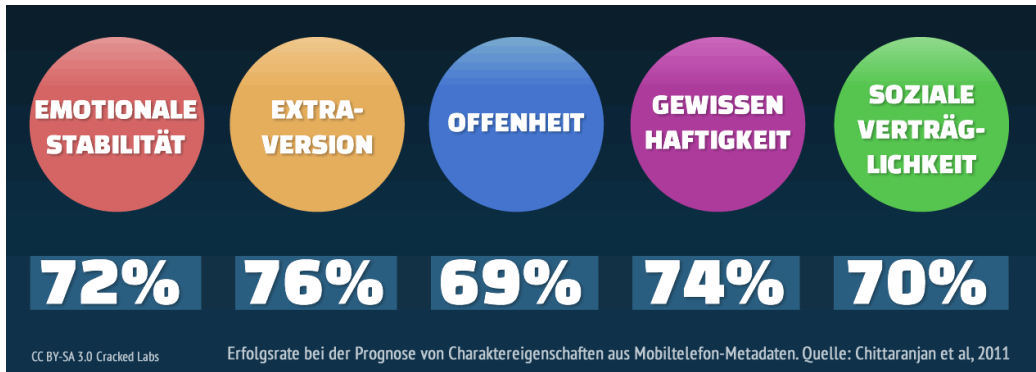
# Persönlichkeitseigenschaften aus Facebook-Likes

Eigenschaft	Zuverlässigkeit der Prognose	Was wurde genau untersucht?
Ethnischer Hintergrund	95%	Kaukasisch oder Afro-Amerikanisch?
Geschlecht	93%	Männlich oder weiblich?
Sexuelle Orientierung I	88%	Schwul?
Politische Einstellung	85%	Liberal oder konservativ?
Religion	82%	Christlich oder muslimisch?
Sexuelle Orientierung II	75%	Lesbisch?
Nikotinkonsum	73%	Raucher/Raucherin?
Alkoholkonsum	70%	Trinkt Alkohol?
Beziehung	67%	Single oder in einer Beziehung?
Drogenkonsum	65%	Konsumiert Drogen?
Trennungskind	60%	Eltern im Alter von 21 getrennt?

Erfolgsraten bei der Prognose von Persönlichkeitseigenschaften aus Facebook-Likes. Quelle: Kosinski et al, 2013

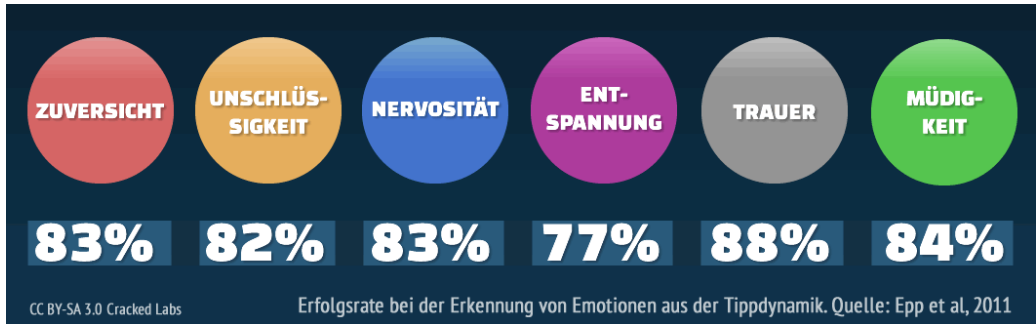
CC BY-SA 3.0 Cracked Labs

# Charaktereigenschaften aus Handy-Metadaten





# Tipp-Dynamik



Entscheidungen in Bereichen, die viele **Konsequenzen**  
haben können

Wie z.B. Finanz-, Versicherungs- und Gesundheitswesen

# Kreditvergabe, die auf digitalen Verhaltensdaten beruht

Etliche Firmen nutzen Daten von

- Sozialen Medien,
- Anfragen bei Suchmaschinen,
- Daten von Mobiltelefonen

um die **Kreditwürdigkeit** von Personen zu bestimmen.

Ohne Kenntnisse über den finanziellen Status zu haben !

# Weitere Daten-Quellen

Manchmal auch:

- Wie füllt jemand ein Online-Formular aus?
- Wie navigiert der Kunde auf Webseiten?
- Wie ist die Grammatik und Beistrichsetzung bei Textpassagen?
- Wie ist der Batterie-Status des Handys ?

Auch Daten von Freunden werden hinzugezogen.

## Firma **Cignifi**:

- Errechnet Kreditwürdigkeit aus der Zahl und Häufigkeit von **Telefonanrufen** ([http://cignifi.wpengine.com/?page\\_id=11577/](http://cignifi.wpengine.com/?page_id=11577/))

## **Facebook**:

- hat ein Patent auf Kreditbewertung beruhend auf der Kreditwürdigkeit der **facebook-Freunde** (<https://t3n.de/news/facebook-freunde-kredit-scoring-schufa-629131/>)

## Versicherung **Aviva**:

- Macht Vorhersagen über individuelles Gesundheitsrisiko (z.B. für Diabetes, Bluthochdruck, Depressionen) von 60.000 Versicherungs-Anwärtern, basierend auf Verbraucherdaten. (<https://www.soa.org/files/pdf/research-pred-mod-life-batty.pdf>)

# Spitalskosten

Consultingfirma **McKinsey**:

- Vorhersage von Spitalskosten von Krankenhaus-Patienten basierend auf Verbraucherdaten

Informationen aus den Bereichen

- Demographie,
- Familienstruktur,
- Einkaufsverhalten
- Autobesitz

wurden genutzt

# Gesundheitsrisiko

## **GNS Healthcare** (Gesundheitsanalysefirma):

Berechnet individuelles Gesundheitsrisiko von Patienten aus großem Daten-Bereich:

- Genom-Daten
- Medizinische Daten
- Labordaten
- Daten von mobilen Gesundheitsgeräten
- Konsumentenverhalten

GNS Healthcare bietet an:

- Krankheits-Fortschritte und Operationsergebnisse

Patienten werden gereiht, je nachdem, wie gross das „return on investment“ ist

<https://www.statnews.com/2015/12/15/insurance-big-data/>

# Gesundheits-Score

## **Nexis Lexis Risk Solutions:**

Großer Datenbroker und Risiko-Analyst

Gesundheits-Score, der

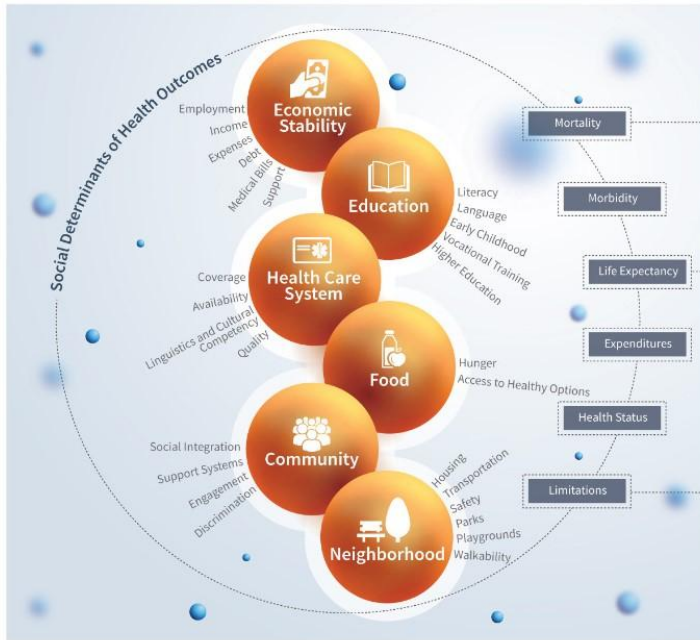
- Gesundheitsrisiko
- zu erwartende Behandlungskosten

abschätzt

Basiert auf Verbraucherdaten und Einkaufsaktivitäten

<https://www.lexisnexis.com/risk/downloads/literature/health-care/Socioeconomic-Health-Risk-Score-br.pdf>





20.6.2018

## III.

# Big Data-Sammlungen und Verwendung von Verbraucher-Daten

Beherrschende Online-Plattformen (facebook, Google):  
Haben detaillierte Informationen über das Leben von Milliarden  
Menschen

Personendaten-Industrie, Daten-Analyse-Industrie

(Ausserdem: Geheimdienste, Online-Werbefirmen, Kriminelle)

In den letzten Jahrzehnten:

- Kredit-Reporting-Firmen
- Verbraucher-Datenbanken

In den letzten Jahren:

- **Online- und offline-Daten** werden verknüpft

# Daten-Firmen und Benutzer-Profile

<b>Facebook</b>	hat Profile von	1,9 Milliarden	Facebook Nutzern
		1,2 Milliarden	WhatsApp Nutzern
		0,6 Milliarden	Instagram Nutzern
<b>Google</b>	hat Profile von	2 Milliarden	Android Nutzern
		über 1 Milliarde	Gmail Nutzern
		über 1 Milliarde	Youtube Nutzern
<b>Apple</b>	hat Profile von	1 Milliarde	iOS Nutzern

# Daten-Firmen und Benutzer-Profile

<b>Experian</b>	hat Kreditdaten von	918 Millionen	Menschen
	hat Marketingdaten von	700 Millionen	Menschen
	„Einsichten“ in	2,3 Milliarden	Menschen
<b>Equifax</b>	hat Daten über	820 Millionen	Menschen
		1 Milliarde	Geräte
<b>TransUnion</b>	hat Daten über	1 Milliarde	Menschen

Credit Reporting Agencies

# Daten-Firmen und Benutzer-Profile

<b>Acxiom</b>	hat Daten über	700 Millionen	Menschen
		1 Milliarde	Cookies und mobile Geräte
	managt	3,7 Milliarden	Benutzerprofile für Klienten
<b>Oracle</b>	hat Daten über	1 Milliarde	Mobile Nutzer
		1,9 Milliarden	Webseiten-Besucher
	liefert Zugang zu	5 Milliarden	„eindeutigen“ Konsumenten-IDs

Konsumenten-Daten Broker

20.6.2018

# Beispiel facebook

- Nutzt mindestens 50.000 Persönlichkeits-Merkmale
- 1,9 Milliarden Nutzer
- Analysiert werden:  
Postings, likes, shares, Freunde, Fotos, Bewegungsdaten,  
...
- Kategorisierung nach:  
politischen Ansichten, Ethnizität, Einkommen, ...
- Facebook kauft zusätzlich Daten von anderen Daten-Brokern:  
Acxiom, Epsilon, Datalogix und BlueKai
- Partnerschaft mit facebook
- Damit können noch bessere Profile gebildet werden



# Online-Werbe-Industrie

Ökosystem von tausenden Firmen

Aufgaben:

- Menschen tracken,
- Profile bilden

2011: 150 Firmen

2014: 1000 Firmen

2018: 7000 Firmen

<https://chiefmartec.com/2018/04/marketing-technology-landscape-supergraphic-2018/>



# IV.

## Daten-Broker und das Geschäft mit Personen-Daten

## Verbraucherdaten-Broker:

- Aggregieren
- kombinieren
- Handeln mit riesigen Mengen an angesammelten Informationen von online- und offline-Quellen
- Kaufen oder lizenzieren Verbraucherdaten von anderen Firmen

Diese Datensammlung geschieht meist **ohne Kenntnis** der Betroffenen

Nicht nur Informationen über

- Ausbildung
- Beziehungsstatus
- Kinder
- Religionszugehörigkeit
- Ethnizität
- Politische Ansichten
- Aktivitäten
- Interessen
- Mediennutzung

Sondern auch das **online-Verhalten:**

- Webseiten-Aufrufe
- Online-Einkäufe
- Kreditkartenkäufe
- Einkommen
- Bank- und  
Versicherungspolizzen
- Eigentum
- Fahrzeugbesitz
- ...

20.6.2018

Auch:

**Scores** (Wertungen), die ein zukünftiges Verhalten vorhersagen

z.B.

- Ökonomische Stabilität
- Job-Wechsel
- Baby-Planung

# Datensammler Acxiom



1969 gegründet

- Bis zu 3000 Datenelemente von 700 Millionen Menschen
- Daten von 1000en Quellen aus den USA, UK, Deutschland

## **AbliTec Link system:**

<https://developer.myacxiom.com/code/api/data-bundles/bundle/ablitec>

- Register von Personen, Haushalten, Gebäuden
- Wird mit Geburten, Sterbedaten, Heiraten, Scheidungen, Namen- und Adressänderungen aktuell gehalten
- 13 Religionszugehörigkeiten
- 200 ethnische Codes



# Datensammler Acxiom



Acxiom verkauft diese Datenprofile

Verwaltet 150.000 Verbraucher-Datenbanken mit Milliarden Profilen

Kunden sind Grossbanken, Versicherungen, Gesundheitsvorsorge-Firmen, Behörden

Acxiom liefert

- Identitätsfeststellung
- Risikomanagement
- Betrugsaufdeckungs-Service

Acxiom liefert Informationen an Facebook, Google, Twitter, ...





# Datensammler Acxiom



## LiveRamp:

Verbindet die digitalen Profile von hunderten Daten- und Werbe-Firmen

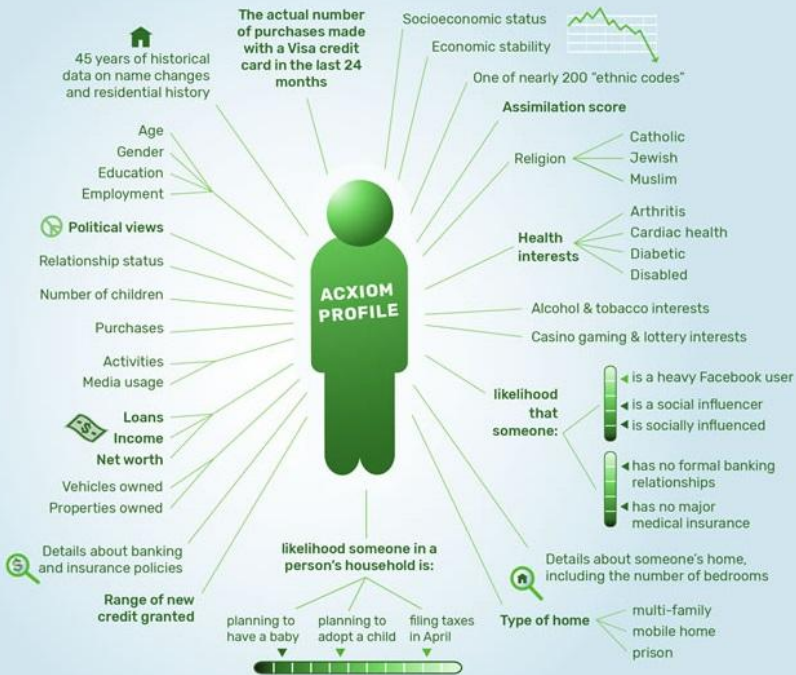
→ Personen können **über Datenbanken hinweg** erkannt werden





© Cracked Labs CC BY-SA 4.0, April/May 2017. Disclaimer: the mentioned companies typically keep information about their activities secret. This illustration is based on publicly available information, mainly the companies' own statements. Every effort has been made to accurately interpret and represent the companies' activities, but we cannot accept any liability in the case of eventual errors. Sources: Axiom website, press releases, brochures, annual reports, and response to US congress inquiry. LiveRamp website, brochures, press releases, presentations. ibotta website. Crossix website, press releases. FreckleIoT press releases. For details about the sources see the report "Corporate Surveillance in Everyday Life".





**Acxiom provides of up 3,000 attributes and scores**  
on 700 million people in the US, Europe, and other regions.

# Zusammenführen der Identitäten



20.6.2018

Data in diagram is fictional - for illustrative purposes only.

# HOW COMPANIES IDENTIFY PEOPLE

to link profile information from various sources and monitor individuals throughout the day

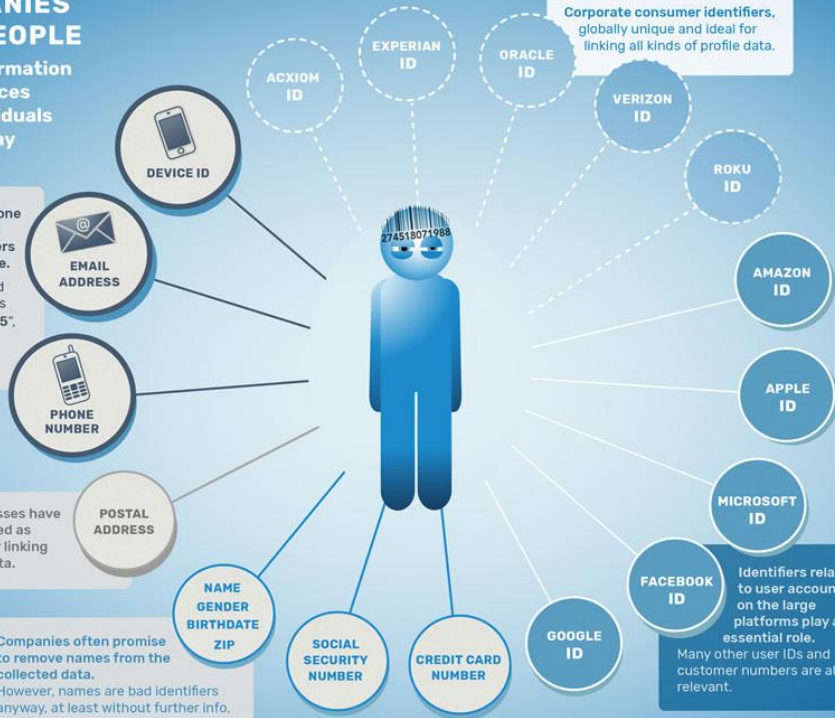
Email addresses and phone numbers are among the most important identifiers used to recognize people.

They are often converted into pseudonyms such as "e907c95ef289bxw2345", which can still serve as personal ID numbers.

Postal addresses have long been used as key nodes for linking consumer data.

Companies often promise to remove names from the collected data. However, names are bad identifiers anyway, at least without further info.

Corporate consumer identifiers, globally unique and ideal for linking all kinds of profile data.



Identifiers related to user accounts on the large platforms play an essential role. Many other user IDs and customer numbers are also relevant.



People can also be (re)identified through calculating digital fingerprints from behavioral data:



# Datensammler ORACLE

Oracle ist weltweit einer der größten Anbieter an Business Software und Datenbanken

Seit kurzem auch: Datenbroker

The Oracle logo is displayed in a bold, red, sans-serif font. The word "ORACLE" is written in all caps, with a registered trademark symbol (®) at the end.

- 3 Milliarden User Profile von 15 Millionen Webseiten
- Daten von 1 Milliarde Usern
- Milliarden Einkäufe von Lebensmittelhändlern und 1500 Einzelhändlern
- 700 Millionen Nachrichten von Sozialen Netzwerken , Blogs und Consumer Review Seiten **täglich**

# Datensammler Oracle

100 Datenlieferanten (Acxiom, Experia, TransUnion, i-Behaviour, Visa, MasterCard, ...)

- Web-Tracker
- Mobile Tracker
- Bewegungsdaten
- Online Quizzes

30.000 Nutzer-Datenkategorien

Teilt Daten mit Facebook, Twitter, ...

Oracle ID graph: ID-Kennzeichnung



## ORACLE

One of the world's largest IT companies

ORACLE DATA CLOUD

# ORACLE DATA CLOUD

aggregates data on billions of consumers

**CROSSWISE**  
identifies individuals across billions of PCs, phones, tablets, and TVs

Companies acquired by Oracle

**ADDTHIS**  
tracks 900 million users across 15 million websites as well as 1 billion mobile users

**DATALOGIX**  
collects purchase data from 1,500 large retailers

### BLUEKAI

Data management platform and data marketplace

### Social Data

Oracle analyzes 700 million messages from social media and blogs daily



Clients can combine their own customer data with profile data from Oracle and its partners, and then share and utilize it in many ways

Profile and sort customers

Find and target people

Personalize content & offers

Measure how people behave

Sell data

### Visa

provides "audience data" for profiling, based on 16 billion credit card payments in the US

### MasterCard

provides data based on payments from cardholders that can be "associated with cookie populations"

### i-Behavior

Has purchase and psychographic data on 339 million people from 5,820 data providers in the US, Canada, UK, France, and Brazil

### Experian

Has data on 700 million people in America, Europe and Asia, including financial information and purchase data

### Axiom

Has up to 3,000 data elements about 700 million people, including from the US, UK, and Germany

### TransUnion

Has data on 1 billion people in America, Africa, and Asia from 90,000 sources, including financial institutions

Listed as data providers in Oracle's "data directory"

### Lotame

Collects profile data about website visitors and mobile app users linked to 3 billion cookies and 1 billion device IDs

### PlacIQ

Collects location data, movement patterns, and activity profiles from 100 million mobile devices

## FACEBOOK

- > to help the platform better sort and categorize its users
- > to let Facebook track what its users are purchasing in stores
- > Oracle also sends its clients' customer data to Facebook to find and target them on the platform

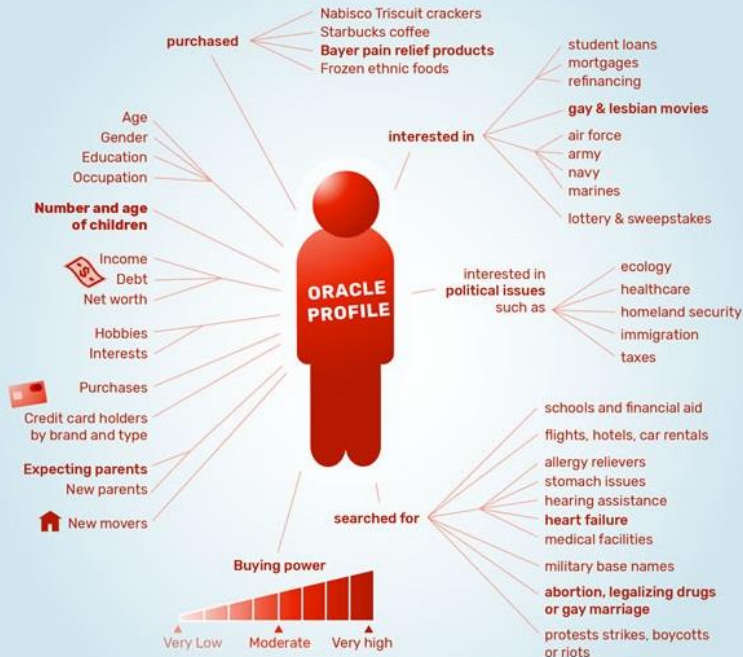
### Proxama

Detailed location data collected using Bluetooth beacons placed across buses, trains, stores, malls, stadiums, cinemas

### VisualDNA

Psychographic profiles about 500 million users in the US, UK, Germany, Russia, and other regions, in part based on online quizzes

© Cracked Labs CC BY-SA 4.0, April/May 2017. Disclaimer: the mentioned companies typically keep information about their activities secret. This illustration is based on publicly available information, mainly the companies' own statements. Every effort has been made to accurately interpret and represent the companies' activities, but we cannot accept any liability in the case of eventual errors. Sources: Oracle website, press releases, data directory, brochures, presentations, MasterCard website, Axiom annual report, TransUnion annual report, Lotame website, VisualDNA brochure, Facebook website, ProPublica article. For details about the sources see the report "Corporate Surveillance in Everyday Life".



**Oracle sorts people into thousands of categories and provides > 30,000 attributes on 2 billion consumer profiles**

V.

# Real-time Verhaltens-Monitoring während des gesamten Lebens

- Online-Plattformen
  - Werbetechologie-Anbieter
  - Daten-Broker
- monitoren  
erkennen  
analysieren
- Menschen  
in vielen  
Situationen

Daraus kann man schliessen,

- was Menschen interessiert
- was sie heute getan haben
- was sie voraussichtlich morgen tun werden
- wie viel sie als Kunde wert sind

# Daten über das Online- und Offline-Leben

Seit vielen Jahrzehnten werden Daten über Menschen gesammelt

- Vor Internet: Kreditbüros, Direktmarketing-Firmen
- Seit 1990er: systematische Konsumenten-Beobachtung durch Datenbank-Marketing, Kunden-Treue-Programme
- Anfang 2000er: Online-Werbung nimmt zu
- Ende 2000er: Soziale Netze, Smartphones, Online-Werbung
- 2010er: Online- und Offline-Daten werden kombiniert

# DATA COLLECTION ON CONSUMERS

SINCE DECADES

RECENT YEARS



Public available data, court records, vehicle records, voter records



Name and address changes, marriages, divorces, births, deaths



Lists from surveys, warranty registrations, and sweepstakes



Newspaper and magazine subscribers, mailorder buyers, travel bookers



Purchases from retailers, pharmacies, loyalty programs, clubs



Credit histories, data from banks, credit card networks, insurers, healthcare, employers

Name  
Birthdate  
**LINKING**  
Postal Address  
Phone ZIP

CENSUS DATA, MARKET RESEARCH

DATA ABOUT NEIGHBORHOODS & BUILDINGS

HOUSEHOLD-LEVEL DATA

INDIVIDUAL-LEVEL DATA

UNIQUE IDENTIFIERS



Profile data is stored by many companies from retailers to data brokers to online platforms.

AGGREGATED, ANONYMIZED "BIG DATA"

METADATA ABOUT WEBSITES, APPS, CONTENT, PLACES

BROWSER OR DEVICE SPECIFIC DATA

INDIVIDUAL-LEVEL DATA

UNIQUE IDENTIFIERS

Location  
Device IDs  
**LINKING**  
Accounts Email  
Cookie IDs



Publicly available data on the web



Website visits, searches and other behaviors



Social media, platforms and user accounts



Smartphone and app usage



Online Purchases

Smart TVs, activity trackers, game consoles, e-readers, cars and other devices



© Cracked Labs CC BY-SA 4.0



# Konsumentendaten-Broker

Konsumentendaten-Broker haben Informationen von

- Zeitungs- und Zeitschriften-Abos
- Buch- und Film-Clubs
- Versandhändlern
- Reisebüro-Buchungen
- Seminar- und Konferenzteilnahme
- ausgefüllten Produkt-Garantie-Karten
- Kunden-Treue-Programme

Zusätzlich: Informationen über Nachbarschaft, Wohngebäude

Heute:

Metadaten über

- Aufruf von Videos
- Apps, die benutzt wurden
- Orte, die besucht wurden
- ...

Wesentlich **mehr** und **genauere** Daten !



# Allgegenwärtiges Digitales Tracking und Profiling

Webseiten haben Tracking-Services eingebaut

→ liefern Daten an Dritte weiter (3<sup>rd</sup> Party)

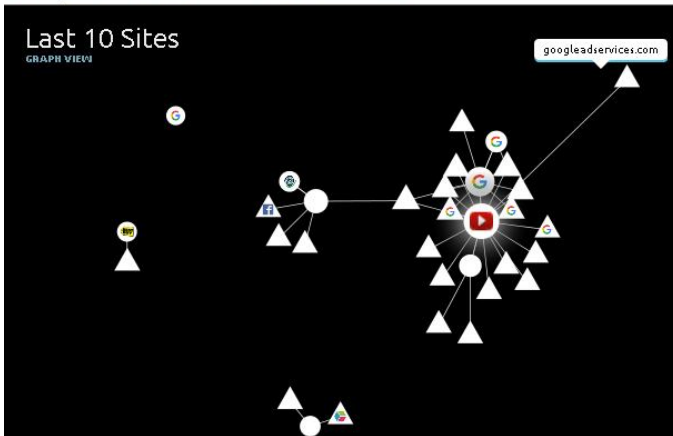
z.B. facebook-Like-Button, eingebettetes Youtube-Video

→ liefern Daten an facebook bzw. Google

Viele andere Tracking-Services sind „unsichtbar“, sammeln aber trotzdem Daten

**Visualisierung** mit Lightbeam oder redmorph-Spyderweb

DATA GATHERED SINCE SEPT 12, 2016    YOU HAVE VISITED 9 SITES    YOU HAVE CONNECTED WITH 27 THIRD PARTY SITES



[https://commons.wikimedia.org/wiki/File:Lightbeam\\_firefox-addon\\_screenshot\\_120916.png](https://commons.wikimedia.org/wiki/File:Lightbeam_firefox-addon_screenshot_120916.png)



20.6.2018

<https://redmorph.com/assets/img/complete-awareness-mobile.jpg>