Collecting, Collating, and Selling Personal Data:

Background Information and Research

During the development of the online game „Data Dealer“ we did a great amount of background research. Most of the contents and stories in our „Data Dealer“ game are based on this research.

Vienna, Austria, May 6, 2013, Wolfie Christl, Renée Winter, Barbara Schweinzer

Table of Contents

1. Introduction: Background of „Data Dealer“

Personal Data in the Digital Age

The quantity and the value of the many different types of personal data being collected today is vast: from our profiles and demographic data from bank accounts to medical records to employment data. Our Web searches, the sites we visited, our likes and dislikes and purchase histories. Our tweets, texts, emails, phone calls and photos as well as the coordinates of our real-world locations. State-of-the-art data mining technologies help to analyze and combine these massive amounts of personal data, emerging businesses in social media and mobile applications increasingly make commercial use of the collected data. While popular self-tracking services attempt to sell their tools to health insurance companies, leading old school data brokers try to link their millions of detailed personal profiles with social content. According to former European Consumer Commissioner Meglena Kuneva “personal data is the new oil of the Internet and the new currency of the digital world.”

Today’s everyday services relying on aggregated data are becoming more and more complex - even if they look simple on the surface. When we use information technology today it's more difficult than ever to estimate what the long-term consequences might be. Most people have insufficient knowledge about what may happen to their personal data when using smartphones or the internet. This leads to fear, uncertainty and a “decline in trust”, as the World Economic Forum stated in its report *Rethinking Personal Data* (2012).

Problems, Risks, and Dangers

Loss of Control: Once digitally processed, personal data cannot be easily erased or changed by individuals.

Individualized Risk: The uncontrolled transmission of sensible personal data, such as sexual orientation, political attitudes, ethnicity, mental-health problems, or illnesses can have severe consequences even in liberal societies.

Commercial Interests: More and more companies and business models operate solely on the basis of the accumulation, processing and selling of personal data.

Profiling and Scoring: In many economical sectors it is common today to rate individuals on the basis of different items of personal data. Not only banks and insurances rely on scoring systems, in more and more areas of our lives human beings are rated and classified accordingly. This can create a big imbalance between individuals on the one hand and big companies on the other hand.

Privacy and the State: A threat to civil rights and liberties is posed whenever governmental institutions get access to personal data, for example when communication service providers are obligated (within the framework of data preservation and other laws) to give access to their collected data to State or Justice.

Wrong or old data: There is a lot of wrong information out there! These flaws are difficult to correct by the individuals themselves and are particularly dangerous when it comes to being ranked on the basis of this data.

Risk of abuse: Everywhere where massive amounts of data are stored, there exists a risk of abuse. On the one hand

personal data is abused (semi) legally by companies or the government, on the other hand there is the risk of abuse through negligence, security holes, illoyal employees, or hacks.

Identity Theft: In addition to social security number, name, date of birth, license numbers etc., more and more biometric attributes (finger prints, iris-scans, facial features) are used for identity verification. If the data connected to these unchangeable attributes gets into the wrong hands, people will have problems for life.

Strengthening Digital Competence!

Besides well-considered data protection regulations, strengthening digital competence is the only way to ensure that individuals can make a positive and self-determined use of ICT in a future society.

Therefore, in 2011 the project Data Dealer was initiated, which turns privacy abuse into an online game accessible also on Facebook and thereby discovered a highly innovative and internationally unique way of providing knowledge about the personal data ecosystem. The demo version was published in German-speaking countries in 2012 and received enthusiastic feedback from the general public. In Data Dealer players switch perspectives and take on the role of ruthless data dealers. As a result this positive impact game gives answers to questions like:

- **What kinds of personal data exist?**
- **Who is collecting this data and what are their intentions?**
- **What could this data be used for and what are the possible impacts on individuals?**
- How does matching and linking large amounts of personal data work?
- How do social networks, search engines or mobile apps work?
- How does displaying targeted ads relating to searches or content posted online work?
- What is an IP address? What are cookies?
- What are the key technological principles to identify internet users?

2. Data Loss, Hacks, Cybersecurity

Where huge amounts of data are stored, there is a danger of abuse of this data.

This proved true in the case of the so called „Sony Playstation Hack". 77 Millions of personal data profiles, including name, address, email-address, date of birth and usernames and passwords, in some cases even credit card numbers, were hacked. A hack like this would mean a great leap forward for data dealers in our game!

Sony Online Entertainment PlayStation Network Hack Timeline
Summary by PC World, May 1, 2011
http://www.pcworld.com/article/226802/playstation_network_hack_timeline.html

In December 2012 1.6 million FBI and NASA accounts were hacked and records from fields such as aerospace, nanotechnology, banking, law, education, government, military, the Department of Defense, airlines, and more were released.

Hacker group GhostShell claims attack on FBI, Interpol, NASA, and Pentagon, theft of 1.6M accounts
Emil Protalinski, The next web, December 10, 2012
http://thenextweb.com/insider/2012/12/10/hacker-group-ghostshell-claims-attack-on-fbi-interpol-nasa-and-pentagon-theft-of-1-6m-accounts

DataLoss Data Base
DataLossDB is a research project aimed at documenting known and reported data loss incidents world-wide.
http://datalossdb.org

3. Concepts and terminology: Big data, Data mining, Privacy

The term „Big data" refers to the huge data sets that can not be as easily stored, processed and accessed as former collections of data. According to Wikipedia „the world's technological per-capita capacity to store information has roughly doubled every 40 months since the 1980s; as of 2012, every day 2.5 quintillion bytes of data were created." To be able to handle these large data sets, „data mining", the computational process of discovering patterns in large data sets is required. Through data mining it is possible to identify structures and patterns within these massive amounts of data. Being commissioned for specific aims and uses, these data mining processes search for specific patterns, such as buying habits, political preferences or credit history. This is where the notion of privacy comes into play. There are serious concerns among people, consumer organizations, NGOs, journalists and activists all over the world about a future where governments, health insurances, employers, landlord associations or other organizations are able to deliberately access personal data of their citizens, costumers or employees. The accessibility of these data has increased particularly because of digital media and communication technologies.

See:

http://en.wikipedia.org/wiki/Personally_identifiable_information

http://en.wikipedia.org/wiki/Big_data

http://en.wikipedia.org/wiki/Data_mining

http://en.wikipedia.org/wiki/Internet_privacy

http://en.wikipedia.org/wiki/Information_privacy

We're losing control of our digital privacy

(Rebecca MacKinnon, CNN, January 29, 2012)

http://edition.cnn.com/2012/01/26/opinion/mackinnon-sopa-government-surveillance/index.html

4. Consumer Data

Collecting information about consumers is one of the most lucrative fields of data brokers. In an article about Acxiom, one of the leading companies in the field of collecting consumer data, Natasha Singer describes methods of data-gathering and the potential dangers to privacy and worries about a ranking system that classifies people as "waste":

„Few consumers have ever heard of Acxiom. But analysts say it has amassed the world's largest commercial database on consumers — and that it wants to know much, much more. Its servers process more than 50 trillion data "transactions" a year. Company executives have said its database contains information about 500 million active consumers worldwide, with about 1,500 data points per person. That includes a majority of adults in the United States."

Mapping, and Sharing, the Consumer Genome

Natasha Singer, New York Times, June 16, 2012)

http://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-of-consumer-database-marketing.html

4.1 The Value of Consumer Data

Data about consumers is extremely valuable. Consumer targeting is among the most frequent reasons for collecting and collating personal data. The fantasy of being able to predict consumer behaviour leads to the greatest efforts.

See for example: Wall Street Journal's investigation that finds that one of the fastest-growing businesses on the Internet is the business of spying on consumers:

The Web's New Gold Mine: Your Secrets

(Julia Angwin, Wall Street Journal, July 30, 2010)

http://online.wsj.com/article/SB10001424052748703940904575395073512989404.html

EPIC, **the Electronic Privacy Information Center,** carried out an interesting survey of what kinds of information are collected, how they are classified, who uses this information for what reasons and gives hints about what you can do to avoid having your data sold and being profiled.

See: http://epic.org/privacy/profiling/

4.2 The Sources of Consumer Data

The sources of consumer data are manifold and range from loyalty cards to online search histories to face and motion recognition.

The initiative RFID1984 has collected a vast amount of background information and articles on loyalty cards, which are one of the most important sources for companies to get information about consumer behaviour, buying habits, household income, and much more. And what's most valuable: every data comes in combination with names and addresses:
http://www.rfid1984.com/supermarket.html

Another page connected to the initiative lists various information about uses of RFID-chips:
http://www.spychips.com

The Privacy Rights Clearinghouse issued a Fact Sheet about what information you should and should not give to companies:

What Personal Information Should You Give to Merchants?
https://www.privacyrights.org/fs/fs15-mt.htm

In 2012 an online marketing company illegally exploited a browser flaw to get the browsing history of its website users:

Online marketer tapped browser flaw to see if visitors were pregnant
(Dan Goodin, Ars Technica, Dec 5, 2012)
http://arstechnica.com/security/2012/12/online-marketer-tapped-browser-flaw-to-see-if-visitors-were-pregnant/

A very informative article about Consumer Data in general and info graphics about which websites share what kind of personal data appeared in the Wall Street Journal:

They Know What You're Shopping For
'You're looking at the premium package, right?' Companies today are increasingly tying people's real-life identities to their online browsing habits.
(Jennifer Valentino-Devries and Jeremy Singer-Vine, Wall Street Journal, December 7, 2012)
http://online.wsj.com/article/SB10001424127887324784404578143144132736214.html

EyeSee- mannequins, produced by the Italian company Almax, have been used in stores since 2012. They have an implemented camera that uses facial recognition to identify the age, gender and ethnicity of people doing (window) shopping:

No dummy: This mannequin is spying on you
(Tim Hornyak, C'NET, November 21, 2012)
http://news.cnet.com/8301-17938_105-57553272-1/no-dummy-this-mannequin-is-spying-on-you/

Department Store Mannequins Are Watching You. No, Really.
(Joanna Stern, ABC News, November 26, 2012)
http://abcnews.go.com/Technology/department-store-mannequins-watch-eyesee-analyzes-shoppers-webcams/story?id=17813441#.UYJp5MpNGE4

EyeSee mannequins used to spy on shoppers, confirm paranoid fears
(Nicole Lee, engadget.com, November 20, 2012)
http://www.engadget.com/2012/11/20/eyesee-mannequins/

5. Data Brokers and the uses of personal data

Data Brokers are companies living off collecting, classifying, analyzing and selling personal data. Emerging from offline address-collecting services, they nowadays find sheer endless possibilities through the emerging digital media.

A good introduction to the topic and compact summary can be found on the ProPublica Webpage:

Everything We Know About What Data Brokers Know About You
(Lois Beckett, ProPublica, March 7, 2013)
http://www.propublica.org/article/everything-we-know-about-what-data-brokers-know-about-you

Natasha Singer from the New York Times described how she tried to get the information that the database marketing company Acxiom had collected about her, a task that proved really difficult: "Data brokers like Acxiom have developed advanced techniques to collect and collate information about consumers' offline, online and mobile behavior. But they have been slow to develop innovative ways for consumers to gain access to the information that companies obtain, share and sell about them for marketing purposes."

Consumer Data, but Not for Consumers
(Natasha Singer, New York Times, July 21, 2012)
http://www.nytimes.com/2012/07/22/business/acxiom-consumer-data-often-unavailable-to-consumers.html

The Privacy Rights Clearinghouse published a huge list of (online) data brokers and possibilities to opt out:
Online Data Vendors: How Consumers Can Opt Out of Directory Services and Other Information Brokers
https://www.privacyrights.org/online-information-brokers-list

5.1 Commercial Use: Advertising, Targeting, Marketing

As we already showed in the chapter „Consumer Data", one big sphere of activity for data collectors is marketing, advertising and consumer targeting.

The company Euclid for example came up with a software that tracks consumers motions while shopping via their smart phones:

Meet The Real-Life Tracking Database That Could Include You
(Andrea Peterson; Think Progress, March 14, 2013)
http://thinkprogress.org/economy/2013/03/14/1717711/euclid-retail-analytics-tracking-franken-privacy/

5.2 Entrepreneurial Use: Employee Screening, Resident Screening

Entrepreneurs, companies, landlord associations – many profit-oriented organizations are very interested in detailed information about their employees, customers or residents.

In some industries it is common that before a person is hired, accurate background checks are carried out. Of course only with prior agreement of the job candidate (but anyone who doesn't want to have a background check is immediately suspected to have someting to hide).

One of the many enterprises that conduct background checks is HireRight. HireRight has already had several charges filed against it because it didn't deliver accurate information and failed to hand over information it collected to the concerned persons:

Employment Background Screening Company to Pay $2.6 Million Penalty for Multiple Violations of the Fair Credit Reporting Act
(Federal Trade Commission, August 8, 2012)
http://www.ftc.gov/opa/2012/08/hireright.shtm

HireRight Sued Again
(Steve Brownstein, The Background Investigator, December 27, 2012)
http://www.thebackgroundinvestigator.com/Articles/HireRight-Sued-Again/1023/

Background Checks can destroy existences, as showed an Associated Press Report by Jordan Robertson:
When your criminal past isn't yours
(Jordan Robertson, Associated Press, December 16, 2011)
http://news.yahoo.com/ap-impact-criminal-past-isnt-yours-182335856.html

5.3 Governmental Use: Ruling and Regulating

For various reasons personal data can be very interesting for governments or governmental institutions: From surveilling personal communications under the cover of fighting against crime and terror to regulating migration on the basis of income or socioeconomic background rather than on the grounds of human rights and freedom to move.

Computer security specialist and writer Bruce Schneier warns of increasing possibilities for governments to buy personal data:

Do You Want the Government Buying Your Data From Corporations?
A new bill moving through Congress would give the authorities unprecedented access to citizens' information
(Bruce Schneier, The Atlantic, April 30, 2013)
http://www.theatlantic.com/technology/archive/2013/04/governments-wont-need-to-issue-ids-data-brokers-will-identify-you-for-them/275431/

6. Methods of data collection

6.1 Old School address-collecting and loyalty cards

Having already talked about the loyalty card systems and their potential to collect data, there still exists good old address-collecting and selling via sweepstakes, all sorts of forms where you have to put in your address etc. But much more possibilities are offered by digital media and communication technologies.

6.2 Mobile Phone Data

German politician Malte Spitz obtained the information that his cell phone operator collected about his activities. He combined this data with other information about him, available on the net, and created an infographic that shows a detailed account of his life in the six months covered. He talked about his experience at TEDGlobal 2012:

Malte Spitz: TEDxGlobal June 2012: Your phone company is watching
http://www.ted.com/talks/malte_spitz_your_phone_company_is_watching.html

In a recent article in the New York Times, Peter Maass and Megha Rajagopalan argue: „The device in your purse or jeans that you think is a cellphone — guess again. It is a tracking device that happens to make calls. Let's stop calling them phones. They are trackers."

That's No Phone. That's My Tracker
(Peter Maass and Megha Rajagopalan, NY Times, July 13, 2012)
http://www.nytimes.com/2012/07/15/sunday-review/thats-not-my-phone-its-my-tracker.html

A study of mobile phone data shows that the mobility patterns produced are not only very detailed, but also highly unique: „We study fifteen months of human mobility data for one and a half million individuals and find that human mobility traces are highly unique. In fact, in a dataset where the location of an individual is specified hourly, and with a spatial resolution equal to that given by the carrier's antennas, four spatio-temporal points are enough to uniquely identify 95% of the individuals. We coarsen the data spatially and temporally to find a formula for the uniqueness of human mobility traces given their resolution and the available outside information. This formula shows that the uniqueness of mobility traces decays approximately as the 1/10 power of their resolution. Hence, even coarse datasets provide little anonymity. These findings represent fundamental constraints to an individual's privacy and have important implications for the design of frameworks and institutions dedicated to protect the privacy of individuals."

Unique in the Crowd: The privacy bounds of human mobility
Yves-Alexandre de Montjoye, César A. Hidalgo, Michel Verleysen & Vincent D. Blondel, Scientific Reports 3, 25 March 2013
www.nature.com/srep/2013/130325/srep01376/full/srep01376.html

6.3 Apps

It's not only the phones or cell phone operators themselves, but in the age of smart phones, apps also have access to manifold types of personal data. A Wall Street Journal report discovered that Android Apps are sharing personal data without the consent of the users:

Your Apps Are Watching You.
A WSJ Investigation finds that iPhone and Android apps are breaching the privacy of smartphone users
(Scott Thurm and Yukari Iwatani Kane, Wall Street Journal, 17.12.2010)
http://online.wsj.com/article/SB10001424052748704694004576020083703574602.html

6.4 Face recognition and Biometrics

Once mainly collected for security reasons by the police and governmental organizations, after expanding it's use on the field of (im-)migration and asylum, biometric data is increasingly used to identify people in other areas as well.

A few years ago several articles expressed their concern about Disney's introduction of Finger Scanners in their Theme Parks. In a blog entry on Boing Boing it was argued that, „what these readers are effective at is conditioning kids to accept surveillance and routine searches and identity checks without particularized suspicion."

Biometrics Comes To Disney World
(David Utter, Boing Boing, September 1, 2006)
http://boingboing.net/2008/03/15/fingertip-biometrics.html

6.5 Self Tracking

By far the cheapest alternative to collecting data is to get all those people to collect and collate their data themselves.

As early as 2001 an article in the Wall Street Journal showed for example how BodyMedia, a company that creates and develops self tracking devices tried selling their products to employers, health insurers and others:

BodyMedia Tries to Sell Health Care With High-Tech Style to Insurers
(Wall Street Journal, April 17, 2001)
http://online.wsj.com/article/SB987459566451890634.html

A more recent article based on a story by Aarti Shahani on the National Public Radio answers the question „Who Could Be Watching Your Figure?" with „Your Boss"

Who Could Be Watching You Watching Your Figure? Your Boss
(npr.org, Aarti Shahani, December 26, 2012)
http://www.npr.org/blogs/alltechconsidered/2012/12/26/167970303/who-could-be-watching-you-watching-your-figure-your-boss

For some companies and insurers, tracking the health of their employers or customers has already become reality in 2013, as shows another article in the Wall Street Journal: „Your company already knows whether you have been taking your meds, getting your teeth cleaned and going for regular medical checkups. Now some employers or their insurance companies are tracking what staffers eat, where they shop and how much weight they are putting on—and taking action to keep them in line." See:

How the Insurer Knows You Just Stocked Up on Ice Cream and Beer
(Jen Wieczner, Wall Street Journal: February 25, 2013)
http://online.wsj.com/article/SB10001424127887323384604578326151014237898.html

7. Personal Data and the Social Web

Having started with online marketing strategies and targeting ads, the possibilities to identify users and to make use of (or exploit) the many different types of personal data they voluntarily or unvonluntarily leave on the internet are keeping many people, enterprises and software developers busy.

7.1 Social Networks

A huge amount of reports and stories has already been published about what kind of data social networks (and especially Facebook) collect about their users.

To mention only a few, we want to cite the following:

Facebook Is Using You
(Lori Andrews, The Learning Network@NYTimes, February 4, 2012)
http://www.nytimes.com/2012/02/05/opinion/sunday/facebook-is-using-you.html

The One Click Personality test: You are what you like
http://www.youarewhatyoulike.com/page.php?p=how_does_it_work

A study about how Facebook gets it's users to refine his Face recognition:

Faces of Facebook: Privacy in the Age of Augmented Reality
By: Alessandro Acquisti (Heinz College, Carnegie Mellon University), Ralph Gross (Heinz College, Carnegie Mellon University), Fred Stutzman (Heinz College, Carnegie Mellon University)
http://www.heinz.cmu.edu/~acquisti/face-recognition-study-FAQ/acquisti-faces-BLACKHAT-draft.pdf

To be able to sell more ads by promising better targeting, Facebook partnered with old school data brokers like Acxiom:

Facebook improves ad targeting with Partner Categories, using third-party data about what you buy
(Ken Yeung, The Next Web, April 10, 2013)
http://thenextweb.com/facebook/2013/04/10/facebook-improves-ad-targeting-with-partner-categories-using-third-party-data-about-what-you-buy/

The Electronic Frontier Foundation published an informative overview about Facebook's cooperation with data brokers, possibilities of opting out and how data brokers get information about people:

The Disconcerting Details: How Facebook Teams Up With Data Brokers to Show You Targeted Ads
(Kurt Opsahl and Rainey Reitman, Electronic Frontier Foundation,April 22, 2013)
https://www.eff.org/deeplinks/2013/04/disconcerting-details-how-facebook-teams-data-brokers-show-you-targeted-ads

RIOT is the name of a software that links together social network data on behalf of the US defense contractor Raytheon:

Software that tracks people on social media created by defence firm
(Ryan Gallagher, The Guardian, February 10, 2013)
http://www.guardian.co.uk/world/2013/feb/10/software-tracks-social-media-defence

Right after the release of Facebook Home, several articles were published about it's possibilities to collect even more personal data:

Facebook Home Means You'll Never Check Facebook Again (It'll Check You)
(Bianca Bosker, Huffington Post, April 5, 2013)
http://www.huffingtonpost.com/2013/04/05/facebook-home_n_3020808.html

Facebook Home doesn't impress, but its potential as a data collection powerhouse does
(Ken Yeung, The Next Web, April 5, 2013)
http://thenextweb.com/facebook/2013/04/05/facebook-home-doesnt-impress-but-its-potential-as-a-data-collection-powerhouse-does/

7.2 Search Engines

Apart from the fact that search engines – like every other website – know your „digital fingerprint", they get a lot more information about who is using a particular computer or device.

Panopticlick - How Unique and Trackable is your Browser?
http://panopticlick.eff.org/

Google knows too much about you
(Frida Ghitis, CNN, February 9, 2012)
http://edition.cnn.com/2012/02/09/opinion/ghitis-google-privacy

In 2013 Google street view cars tracked and saved data like passwords and emails from private networks they came by without permission:

Google Concedes That Drive-By Prying Violated Privacy
(David Streitfeld, NY Times, March 12, 2013)
http://www.nytimes.com/2013/03/13/technology/google-pays-fine-over-street-view-privacy-breach.html

PS: Life is fast and in digital media it is even faster. New devices, apps and technologies are developed incessantly. This demands constant research on the handling of personal data in new communication technologies. We are eager to stay up-to-date and to integrate new developments and incidents directly into the game.