Chapter 0.1: Lecture Orga

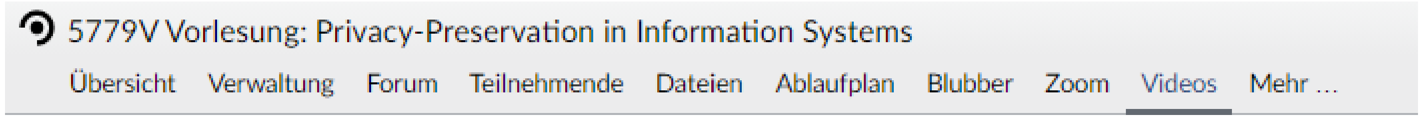
Privacy-Preservation Technologies in Information Systems Dr. Armin Gerl

WS 2021/2022

Organization

- Lecture will be conducted solely **virtual**
  - Every Tuesday: 14:15 to 15:45

- Lecture Recording
  - Lecture will be recorded; **but no guarantee!**
  - Recorded Lectures will be processed and uploaded via StudIP/Vimeo



5779V Vorlesung: Privacy-Preservation in Information Systems

- Exercise
  - Mix of synchron and asynchron Teaching
  - More Details in first exercise

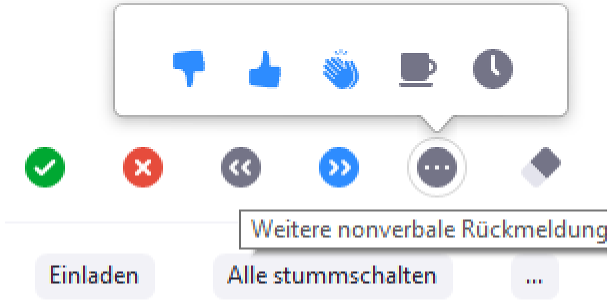
Virtual Lecture

- Video: I am looking for activated video screens ☺
  - Imagine you have to talk to a „black wall“
  - Please no inapproriate background pictures or „Zoom Bombing“ -> otherwise ban

- Audio: Please mute, except you have questions

- Use the Chat for Questions
  - I try to read/answer them in time
  - Please no spam

- Use the Zoom Icons for Feedback
  - Raise your Hand for Questions
  - Agree, Disagree, Faster, Slower, Short Break



Exams

- Plan
  - Regular Written Exam with a duration of 90 Minutes
  - 2 Exam Dates
    - 1. Exam: 22.02.2022; 13:00 – 15:00 HS 9 (AM)
    - 2. Exam: 21.04.2022; 14:00 – 16:00 HS 9 (AM)

To be confirmed by administration

- Backup Plan (if any COVID restrictions come)
  - Virtual Oral Exam or alternative exam type

Announcement

FIT Europe Seminar in Milan, Italy

Chapter 1: Introduction

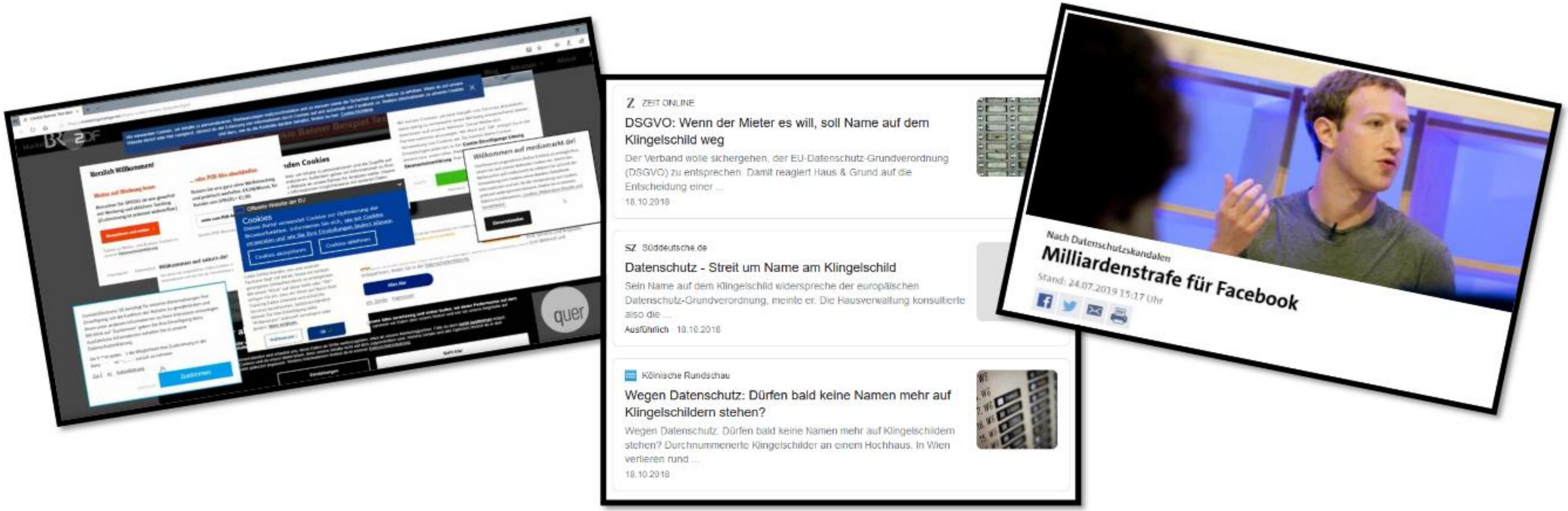
Privacy-Preservation Technologies in Information Systems

Dr. Armin Gerl

WS 2021/2022

Introduction – The Bad News

- Increased media presence of data protection issues since EU-wide regulations came into effect in 2018



ZEIT ONLINE

DSGVO: Wenn der Mieter es will, soll Name auf dem Klingelschild weg

Der Verband wolle sichergehen, der EU-Datenschutz-Grundverordnung (DSGVO) zu entsprechen. Damit reagiert Haus & Grund auf die Entscheidung einer ...

18.10.2018

SZ Süddeutsche.de

Datenschutz - Streit um Name am Klingelschild

Sein Name auf dem Klingelschild widerspreche der europäischen Datenschutz-Grundverordnung, meinte er. Die Hausverwaltung konsultierte also die ...

Ausführlich 18.10.2018

Kölnische Rundschau

Wegen Datenschutz: Dürfen bald keine Namen mehr auf Klingelschildern stehen?

Wegen Datenschutz: Dürfen bald keine Namen mehr auf Klingelschildern stehen? Durchnummerierte Klingelschilder an einem Hochhaus. In Wien verlieren rund ...

18.10.2018

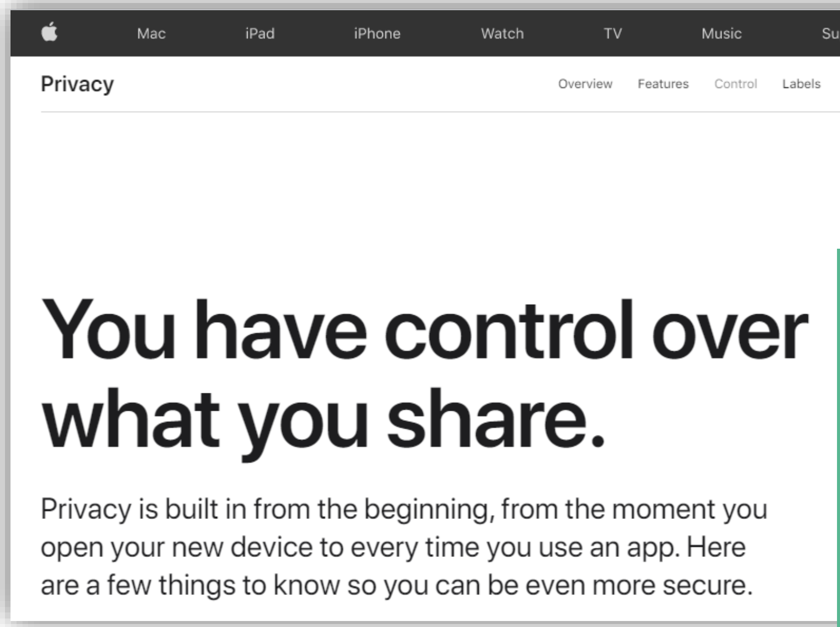
Nach Datenschutzskandalen

Milliardenstrafe für Facebook

Stand: 24.07.2019 15:17 Uhr

Introduction – The Good News

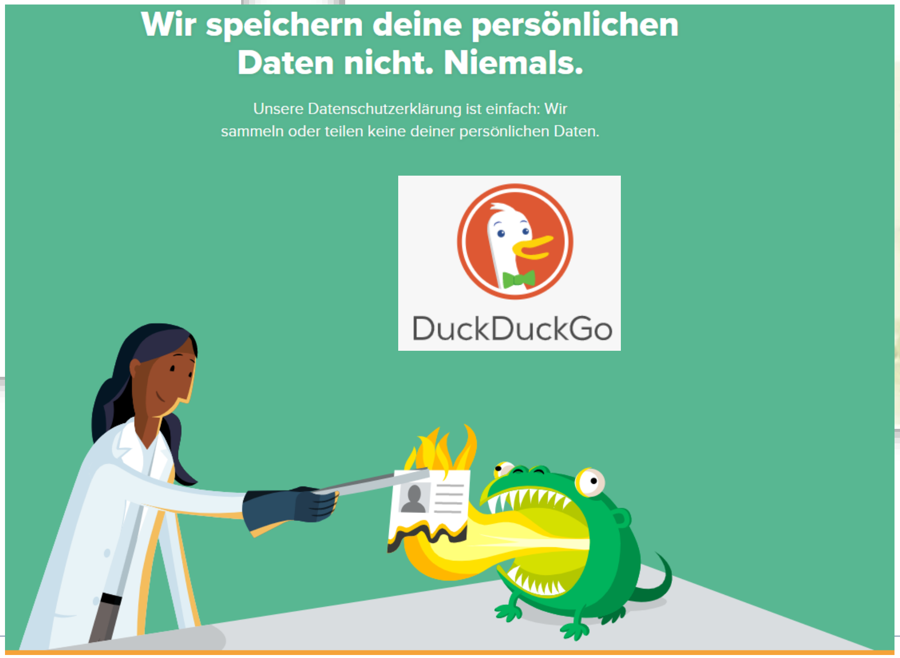
- Privacy for Marketing or as a unique „Feature“



You have control over what you share.

Privacy is built in from the beginning, from the moment you open your new device to every time you use an app. Here are a few things to know so you can be even more secure.

Android 12 Beta 3 with scrolling screen-shots, privacy indicator APIs, enhanced auto-rotate and more released



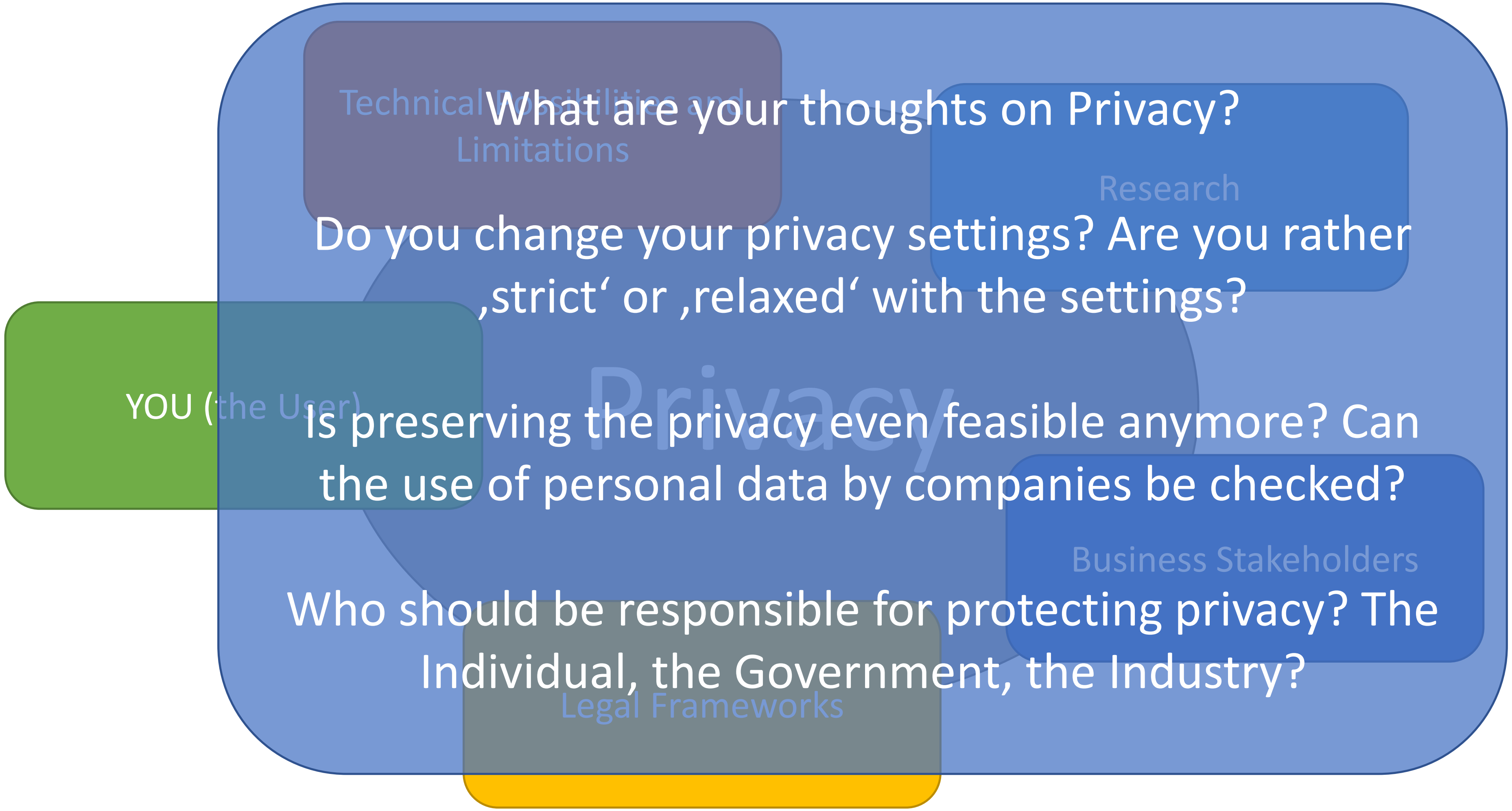
Wir speichern deine persönlichen Daten nicht. Niemals.

Unsere Datenschutzerklärung ist einfach: Wir sammeln oder teilen keine deiner persönlichen Daten.

Data Protection Challenges

- Technical progress provides the means to look deeper into the private lives of citizens than ever before
- Even anonymized data sets can be mined for useful information with advanced analysis tools in data warehouses
- Actors in the data business are primarily objected to obey the legal restriction imposed upon them
- Moral concerns are only starting to be acknowledged in recent years due to public backlash after data protection scandals and leaks
- A framework suitable for examining large and complex data sets on privacy issues would solve many problems for data processors and sources alike
- The academic field that tries to tackle this problem is still expanding
- A variety of privacy models for different areas of applicability already exist, each with their own benefits and limitations

What has to be considered?



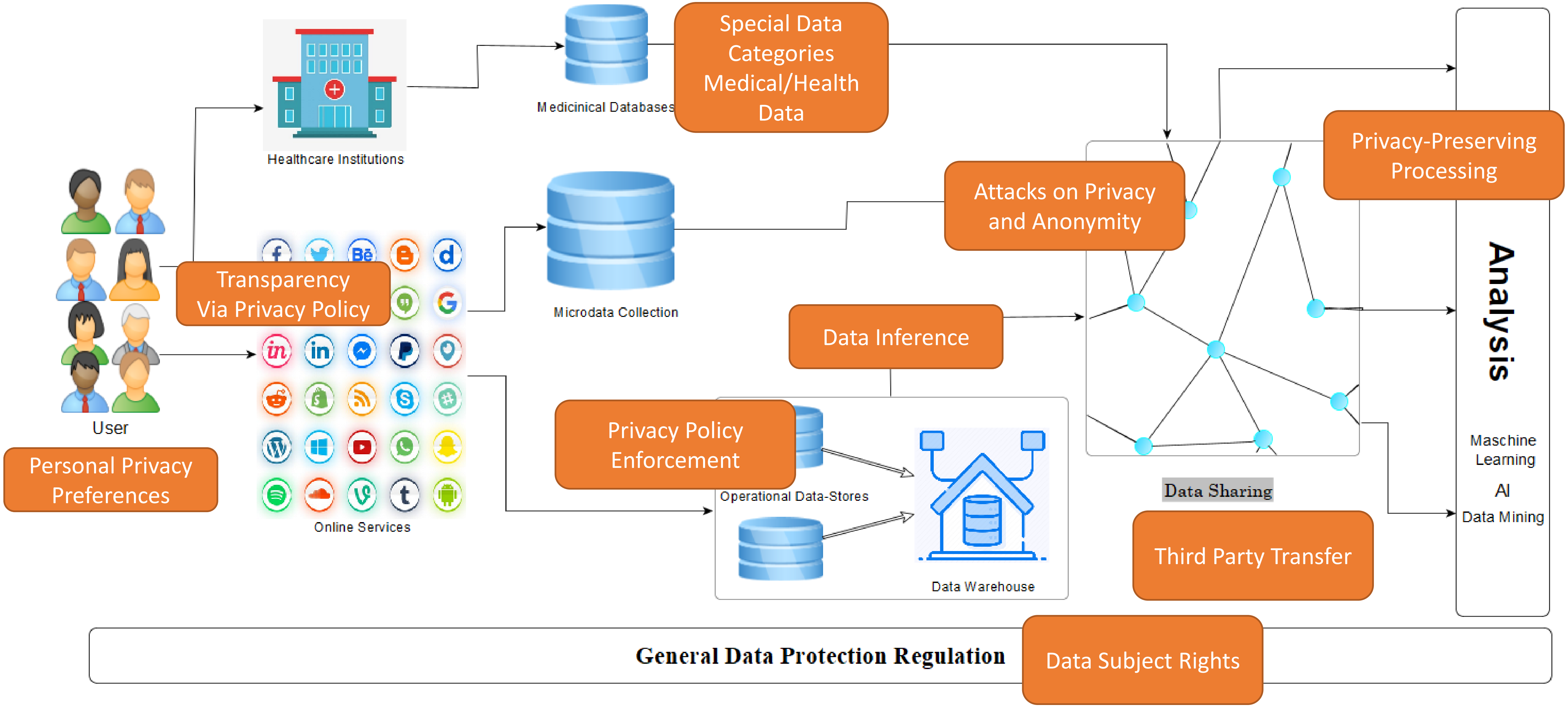
What are your thoughts on Privacy?

Do you change your privacy settings? Are you rather ‚strict‘ or ‚relaxed‘ with the settings?

Is preserving the privacy even feasible anymore? Can the use of personal data by companies be checked?

Who should be responsible for protecting privacy? The Individual, the Government, the Industry?

What has to be considered?



Healthcare Institutions
Medicinical Databases
Special Data Categories Medical/Health Data
Microdata Collection
Attacks on Privacy and Anonymity
Privacy-Preserving Processing
User
Transparency Via Privacy Policy
Personal Privacy Preferences
Online Services
Data Inference
Privacy Policy Enforcement
Operational Data-Stores
Data Warehouse
Data Sharing
Third Party Transfer
Analysis
Maschine Learning
AI
Data Mining
General Data Protection Regulation
Data Subject Rights

1.1

Data Protection: Developments and Problems

Privacy-Preservation Technologies in Information Systems

Dr. Armin Gerl

WS 2021/2022

Data Protection is a fundamental Right

- In Germany, the so called „Recht auf informationelle Selbstbestimmung“ was derived from other fundamental rights on 15.12.1983 by the BVerfG
- Context for this judgment was the attempted connection between collected data from different government branches and agencies into a single database
- After discussing the issue, the court came to a conclusion called „Volkszählungsurteil“:
  - People in a free democracy only retain their capacity to partake and act freely in said democracy, if they can decide how their data is collected, processed and stored for themselves
  - If an individual's preferences are disclosed without consent, said individual won‘t be able to freely interact within society
  - Instead the individual will subordinate their personal preferences to those of its surroundings, inhibiting the potential to enrich the society with diversity and open discourse
  - The last point opposes fundamental rights of the „Grundgesetz“ and led the court to their final decision

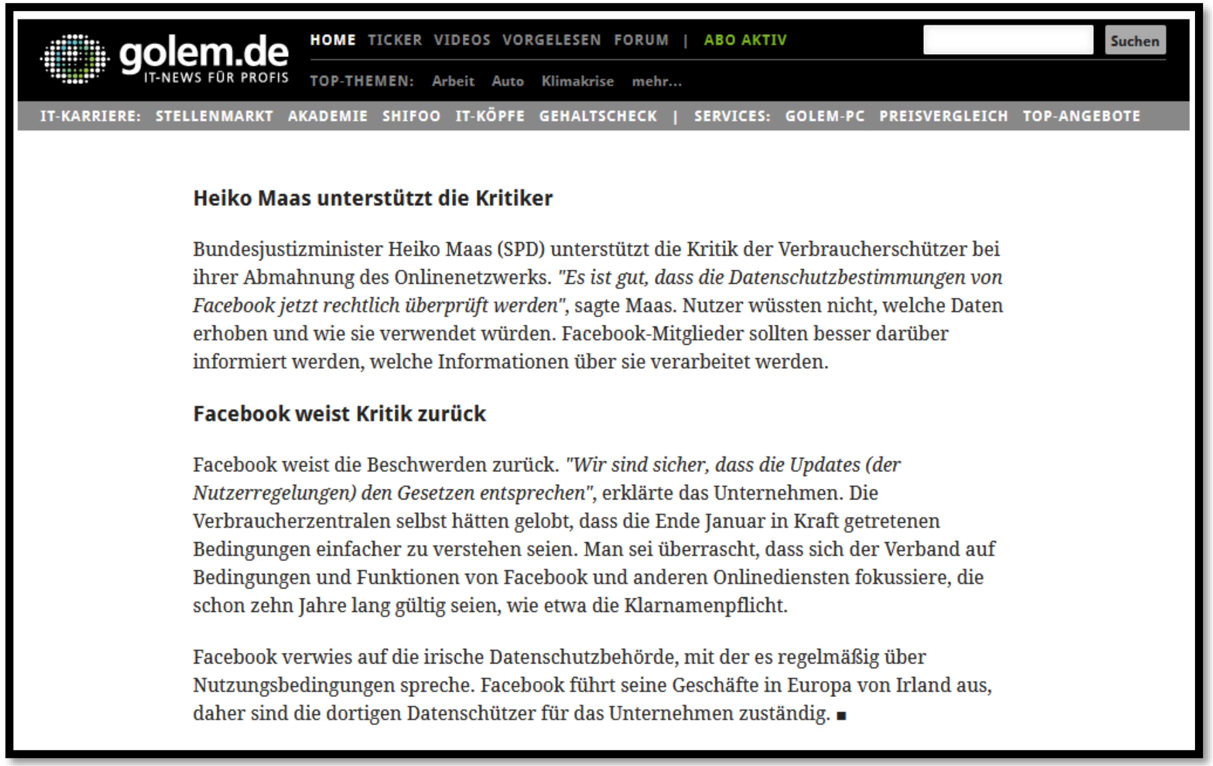
Data Collection: Then vs. Now

- In the year 1983 the amount of data collected was minimal
- Small scale customer data collected by singular businesses and disconnected state agencies were the main actors
- Data trade (or data breaches) was conducted in paper form
- Comparing different data sets and aligning their contents (eliminating duplicates and creating profiles) was mostly done by hand
- Needless to say, dealing with a large data collection was time consuming and in general didn‘t pay off for most actors

- Today, data is collected in some form on almost every interaction we take part in
- Even when we‘re not actively engaging with services, our data is collected and processed (CCTV, E-Health, E-Government)
- Most businesses expanded from a local to a global stage
- State agencies cooperate more intensely and collected data crosses borders
- The internet allows for attacks on databases on a daily basis
- Analysis of collected data is done with exponentially more powerful tools, using A.I. and high performance computing

Internationalization and Data Protection

- Global actors circumventing local (German) regulations through lobbying, relocation and loopholes


golem.de
IT-NEWS FÜR PROFIS
HOME TICKER VIDEOS VORGELESEN FORUM | ABO AKTIV
Suchen
TOP-THEMEN: Arbeit Auto Klimakrise mehr...
IT-KARRIERE: STELLENMARKT AKADEMIE SHIFOO IT-KÖPFE GEHALTSCHECK | SERVICES: GOLEM-PC PREISVERGLEICH TOP-ANGEBOTE

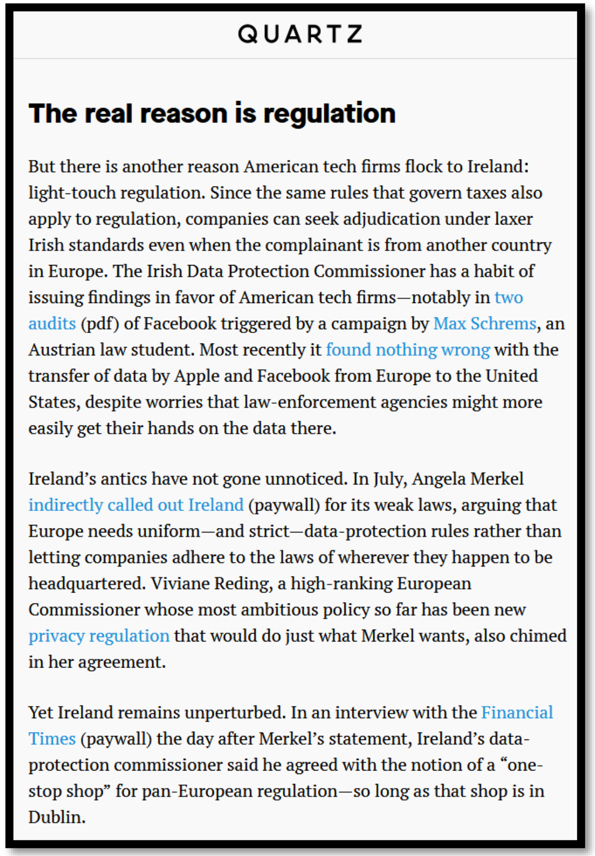
Heiko Maas unterstützt die Kritiker

Bundesjustizminister Heiko Maas (SPD) unterstützt die Kritik der Verbraucherschützer bei ihrer Abmahnung des Onlinenetzwerks. *"Es ist gut, dass die Datenschutzbestimmungen von Facebook jetzt rechtlich überprüft werden"*, sagte Maas. Nutzer wüssten nicht, welche Daten erhoben und wie sie verwendet würden. Facebook-Mitglieder sollten besser darüber informiert werden, welche Informationen über sie verarbeitet werden.

Facebook weist Kritik zurück

Facebook weist die Beschwerden zurück. *"Wir sind sicher, dass die Updates (der Nutzerregelungen) den Gesetzen entsprechen"*, erklärte das Unternehmen. Die Verbraucherzentralen selbst hätten gelobt, dass die Ende Januar in Kraft getretenen Bedingungen einfacher zu verstehen seien. Man sei überrascht, dass sich der Verband auf Bedingungen und Funktionen von Facebook und anderen Onlinediensten fokussiere, die schon zehn Jahre lang gültig seien, wie etwa die Klarnamenpflicht.

Facebook verwies auf die irische Datenschutzbehörde, mit der es regelmäßig über Nutzungsbedingungen spreche. Facebook führt seine Geschäfte in Europa von Irland aus, daher sind die dortigen Datenschützer für das Unternehmen zuständig. ■



The real reason is regulation

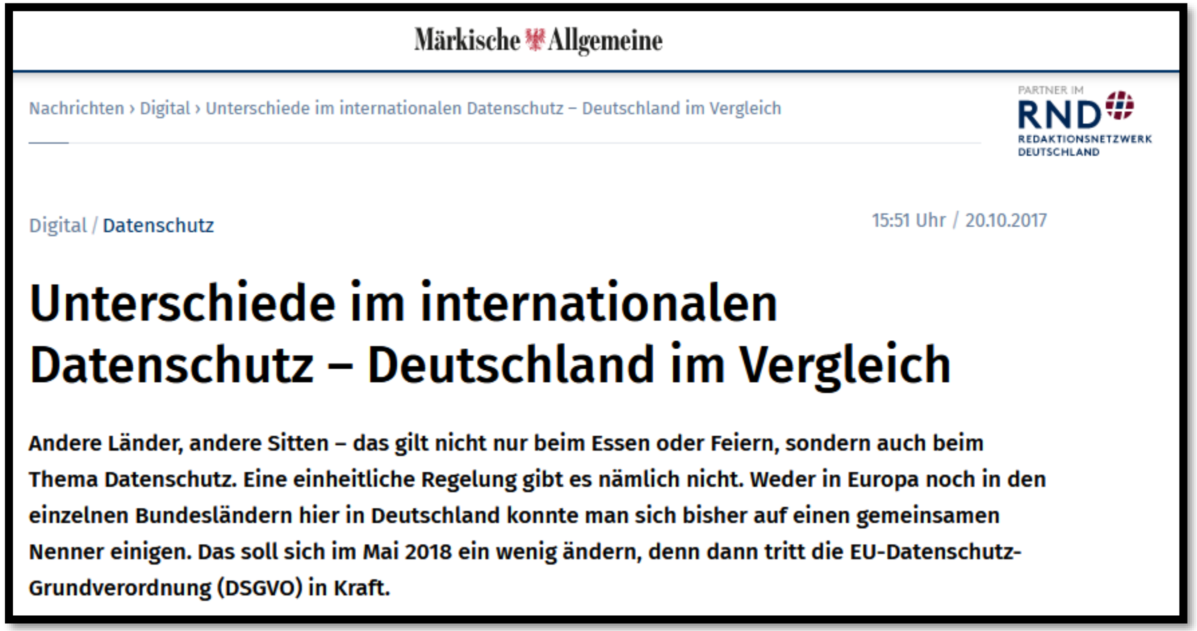
But there is another reason American tech firms flock to Ireland: light-touch regulation. Since the same rules that govern taxes also apply to regulation, companies can seek adjudication under laxer Irish standards even when the complainant is from another country in Europe. The Irish Data Protection Commissioner has a habit of issuing findings in favor of American tech firms—notably in two audits (pdf) of Facebook triggered by a campaign by Max Schrems, an Austrian law student. Most recently it found nothing wrong with the transfer of data by Apple and Facebook from Europe to the United States, despite worries that law-enforcement agencies might more easily get their hands on the data there.

Ireland's antics have not gone unnoticed. In July, Angela Merkel indirectly called out Ireland (paywall) for its weak laws, arguing that Europe needs uniform—and strict—data-protection rules rather than letting companies adhere to the laws of wherever they happen to be headquartered. Viviane Reding, a high-ranking European Commissioner whose most ambitious policy so far has been new privacy regulation that would do just what Merkel wants, also chimed in her agreement.

Yet Ireland remains unperturbed. In an interview with the Financial Times (paywall) the day after Merkel's statement, Ireland's data-protection commissioner said he agreed with the notion of a "one-stop shop" for pan-European regulation—so long as that shop is in Dublin.

European Countermeasures

- Large differences in data protection regulations between states and even regions lead to a European standard with the GDPR



Märkische Allgemeine

Nachrichten › Digital › Unterschiede im internationalen Datenschutz – Deutschland im Vergleich

PARTNER IM RND REDAKTIONSNETZWERK DEUTSCHLAND

Digital / Datenschutz

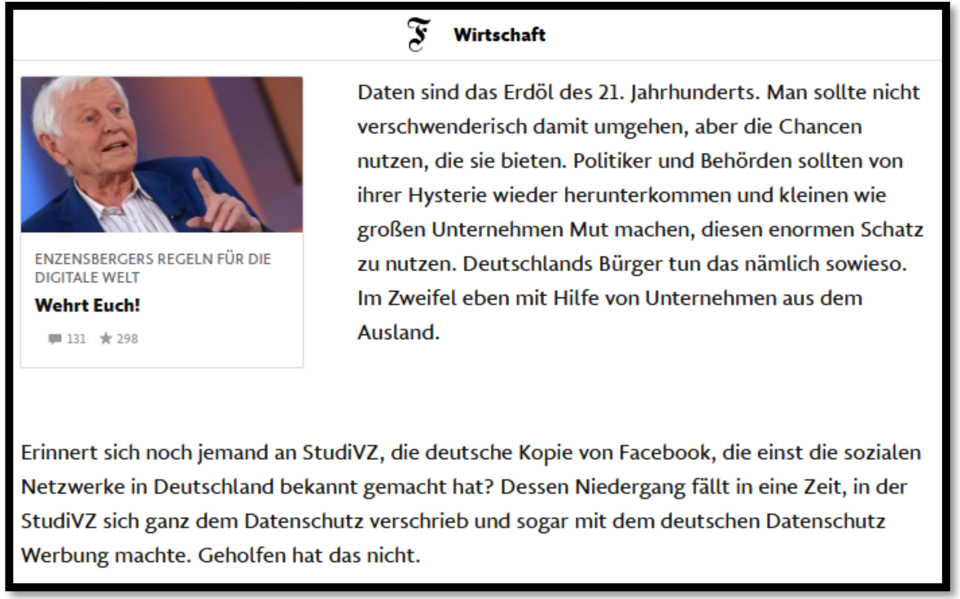
15:51 Uhr / 20.10.2017

Unterschiede im internationalen Datenschutz – Deutschland im Vergleich

Andere Länder, andere Sitten – das gilt nicht nur beim Essen oder Feiern, sondern auch beim Thema Datenschutz. Eine einheitliche Regelung gibt es nämlich nicht. Weder in Europa noch in den einzelnen Bundesländern hier in Deutschland konnte man sich bisher auf einen gemeinsamen Nenner einigen. Das soll sich im Mai 2018 ein wenig ändern, denn dann tritt die EU-Datenschutz-Grundverordnung (DSGVO) in Kraft.

Global Differences

- The notion of ownership of ones data is a very European thought process
- The USA for example still sees collected data as property of the collector, resulting in a competitive advantage for companies outside the EU



Wirtschaft

ENZENSBERGERS REGELN FÜR DIE DIGITALE WELT

Wehrt Euch!

131 ★ 298

Daten sind das Erdöl des 21. Jahrhunderts. Man sollte nicht verschwenderisch damit umgehen, aber die Chancen nutzen, die sie bieten. Politiker und Behörden sollten von ihrer Hysterie wieder herunterkommen und kleinen wie großen Unternehmen Mut machen, diesen enormen Schatz zu nutzen. Deutschlands Bürger tun das nämlich sowieso. Im Zweifel eben mit Hilfe von Unternehmen aus dem Ausland.

Erinnert sich noch jemand an StudiVZ, die deutsche Kopie von Facebook, die einst die sozialen Netzwerke in Deutschland bekannt gemacht hat? Dessen Niedergang fällt in eine Zeit, in der StudiVZ sich ganz dem Datenschutz verschrieb und sogar mit dem deutschen Datenschutz Werbung machte. Geholfen hat das nicht.

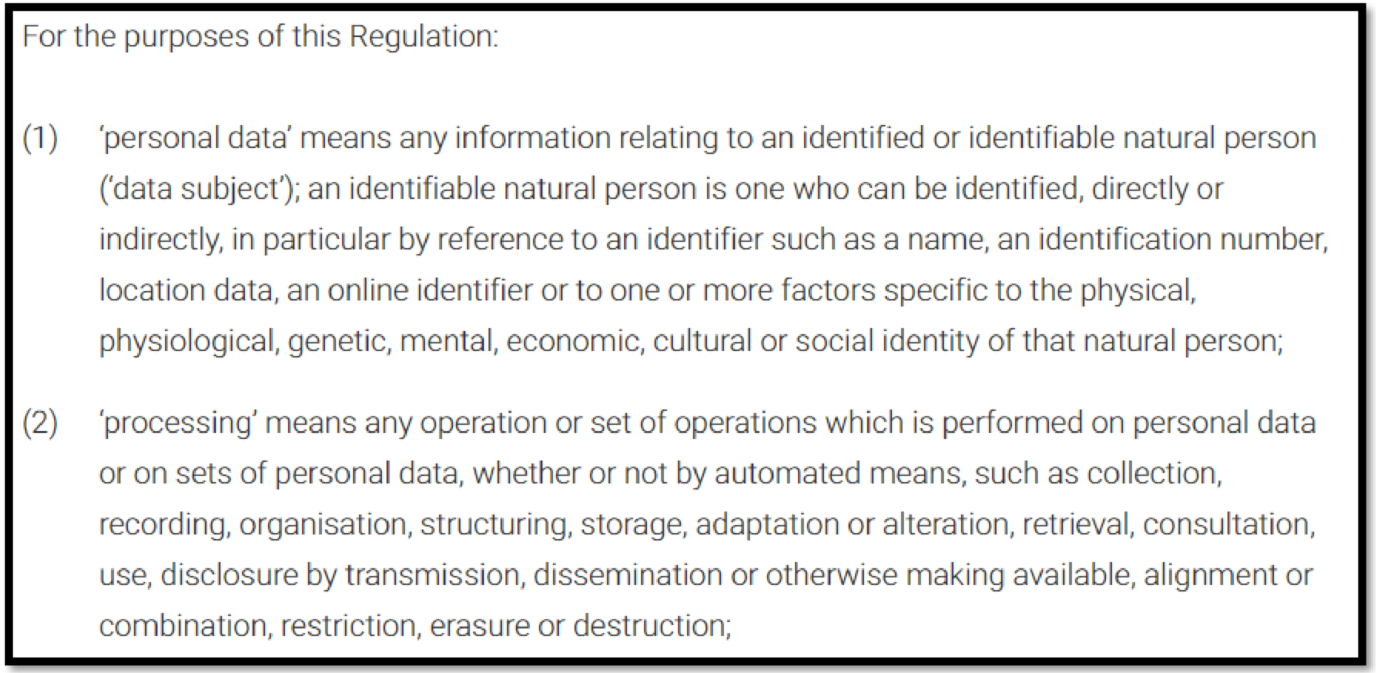
- The European standardization of data protection laws deters new investors and inhibits the development of the IT sector in exchange for protecting the privacy of EU citizens

The Impact of the GDPR

- Relocating their headquarters to circumvent local regulations isn‘t a loophole for companies anymore
- Also foreign businesses collecting data on EU soil are now restricted by the new law, with violations resulting in huge fines
- Fining is done with respect to a companies revenue, preventing the biggest global players from throwing money at the problem
- The law spearheaded a new standard for data protection laws globally
- Subsequently, the design of the GDPR was copied in many government legislations (e.g. California and Brazil)
- Companies are now forced to deal with the issue of data protection much more intensely than before

Contents of the GDPR

- Excerpt from definitions of key words from the GDPR (Art. 4):



For the purposes of this Regulation:

(1) 'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

(2) 'processing' means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;

- Building on these pretty broad definitions, the GDPR generally prohibits processing of personal data without legal basis or consent and imposes extensive requirements on non-personal data collected

From legal to technical Models

- Question: How to enforce GDPR effective and efficient?
- Enforcement by “privacy officers” does not seem efficient and effective
- Often rather more legal then technical solutions for privacy

- Legal definitions are often very “vague”
- Techniques trying to satisfy these legal definitions have to be „state of the art“
- This state is continually driven forward by new technologies being conceived by researchers and implemented by companies
- The academic community demands a mechanism that ensures legal conformity of data processing and that gives the individual more control over the whole process

1.2
Data Publishing Problems

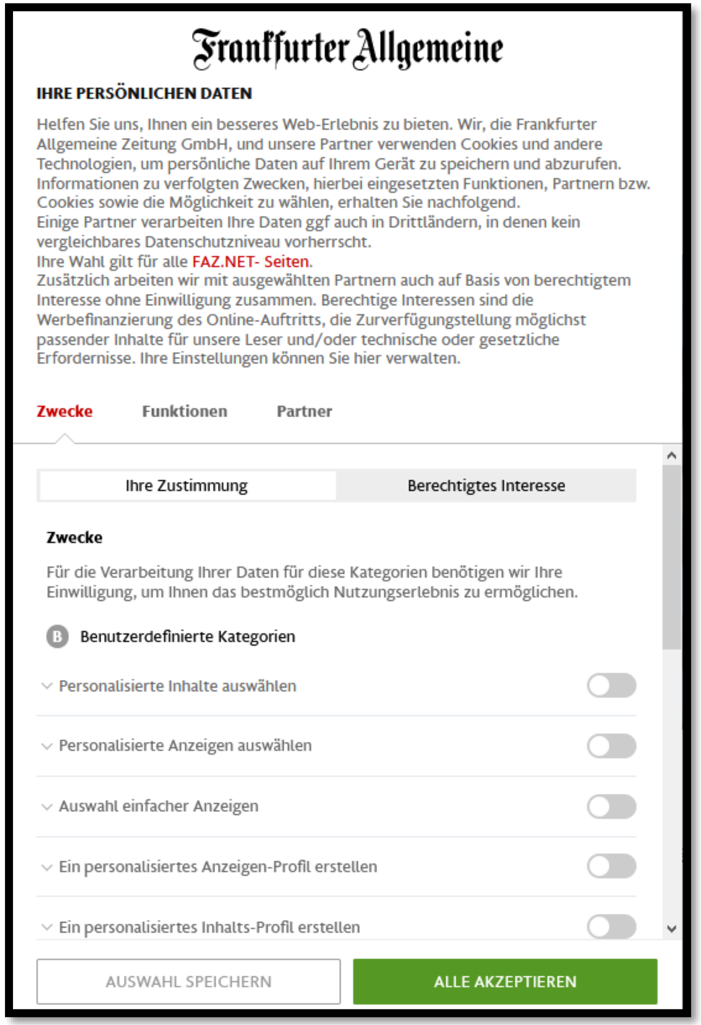
Privacy-Preservation Technologies
in Information Systems

Dr. Armin Gerl

WS 2021/2022

Data Trade Requirements

- Almost all actors in the data collecting business also participate in data trade to gain the maximum value from the information available
- The GDPR and similar frameworks impose regulations on publishing and trading data
- With new legal frameworks, every person (data subject) now has a variety of rights and choices concerning the processing and collection of their data
- Depending on the choices of the data subject , collectors have to take different measures to ensure the processing and privacy protection of data is conducted accordingly



Frankfurter Allgemeine

IHRE PERSÖNLICHEN DATEN

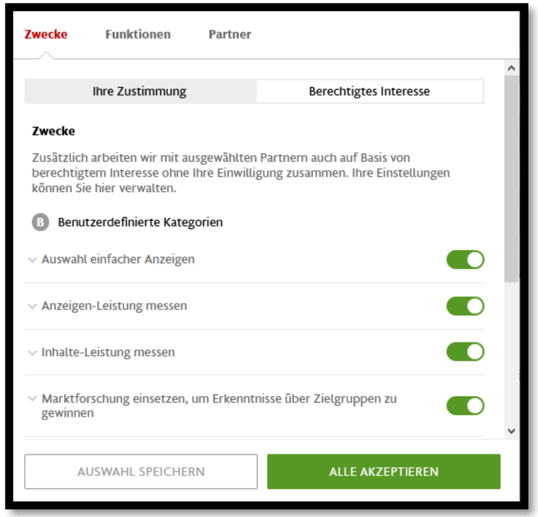
Helfen Sie uns, Ihnen ein besseres Web-Erlebnis zu bieten. Wir, die Frankfurter Allgemeine Zeitung GmbH, und unsere Partner verwenden Cookies und andere Technologien, um persönliche Daten auf Ihrem Gerät zu speichern und abzurufen. Informationen zu verfolgten Zwecken, hierbei eingesetzten Funktionen, Partnern bzw. Cookies sowie die Möglichkeit zu wählen, erhalten Sie nachfolgend.
Einige Partner verarbeiten Ihre Daten ggf auch in Drittländern, in denen kein vergleichbares Datenschutzniveau vorherrscht.
Ihre Wahl gilt für alle FAZ.NET- Seiten.
Zusätzlich arbeiten wir mit ausgewählten Partnern auch auf Basis von berechtigtem Interesse ohne Einwilligung zusammen. Berechtige Interessen sind die Werbefinanzierung des Online-Auftritts, die Zurverfügungstellung möglichst passender Inhalte für unsere Leser und/oder technische oder gesetzliche Erfordernisse. Ihre Einstellungen können Sie hier verwalten.

Zwecke

Für die Verarbeitung Ihrer Daten für diese Kategorien benötigen wir Ihre Einwilligung, um Ihnen das bestmöglich Nutzungserlebnis zu ermöglichen.

B Benutzerdefinierte Kategorien

- Personalisierte Inhalte auswählen ○
- Personalisierte Anzeigen auswählen ○
- Auswahl einfacher Anzeigen ○
- Ein personalisiertes Anzeigen-Profil erstellen ○
- Ein personalisiertes Inhalts-Profil erstellen ○



Zwecke

Zusätzlich arbeiten wir mit ausgewählten Partnern auch auf Basis von berechtigtem Interesse ohne Ihre Einwilligung zusammen. Ihre Einstellungen können Sie hier verwalten.

B Benutzerdefinierte Kategorien

- Auswahl einfacher Anzeigen ●
- Anzeigen-Leistung messen ●
- Inhalte-Leistung messen ●
- Marktforschung einsetzen, um Erkenntnisse über Zielgruppen zu gewinnen ●

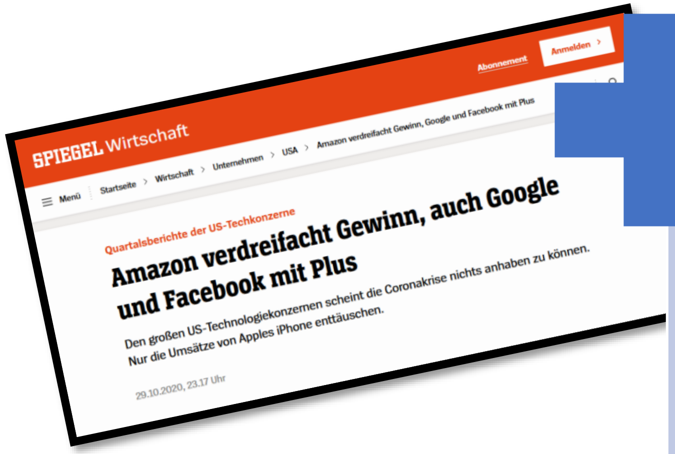
Data Processing Requirements

- To avoid the chaos and complexity that a strict tracking and handling of every possible combination of choices for each subject throughout the analysis process would bring, companies often anonymize all data before analysing
- If anonymization isn‘t done in advance, statistical analysis is usually performed on the basis of predefined condition that permit the processing of personal data:
    - Legitimate Interest
    - Permission by contract, if the person is the client bound by said contract
    - Collectors having to obey legal obligations (e.g. criminal prosecution)
    - Life-threatening contexts (e.g. rescue services)
    - Exercise of delegated official authority (executive power) or public interest
    - On the basis of a conflict between fundamental rights and safeguarding responsibilities (e.g. childcare)

Legal Basis
or
Consent

Knowledge is Power (and Money)

- Data often compared to have similar significance in the 21st century as the discovery of gold had in the 19th century, sparking a gold rush



SPIEGEL Wirtschaft
Abonnement
Anmelden ›
☰ Menü
Startseite › Wirtschaft › Unternehmen › USA › Amazon verdreifacht Gewinn, Google und Facebook mit Plus

Quartalsberichte der US-Techkonzerne

Amazon verdreifacht Gewinn, auch Google und Facebook mit Plus

Den großen US-Technologiekonzernen scheint die Coronakrise nichts anhaben zu können. Nur die Umsätze von Apples iPhone enttäuschen.

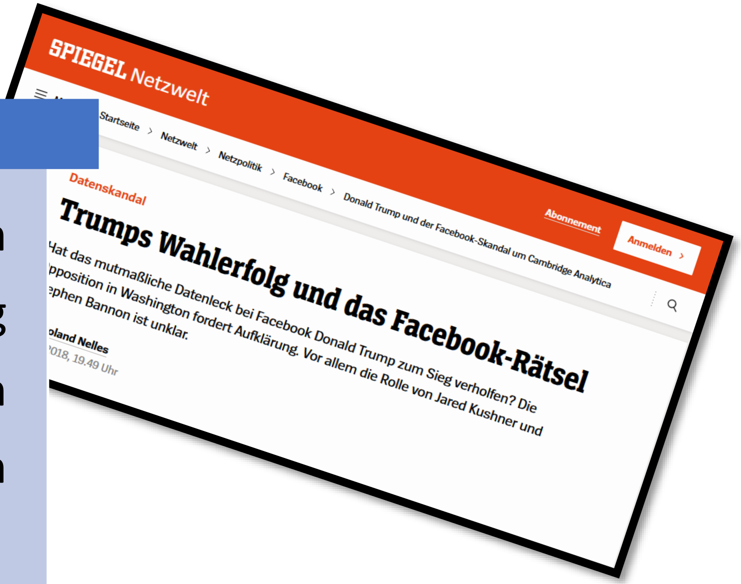
29.10.2020, 23.17 Uhr

+

- Advertising industry
- Strategic company development
- Demographic and administrative decisions
- Law enforcement

−

- Targeted manipulation
- Blackmailing
- Intimidation
- Loss of freedom



SPIEGEL Netzwelt
Abonnement
Anmelden ›
Startseite › Netzwelt › Netzpolitik › Facebook › Donald Trump und der Facebook-Skandal um Cambridge Analytica

Datenskandal

Trumps Wahlerfolg und das Facebook-Rätsel

Hat das mutmaßliche Datenleck bei Facebook Donald Trump zum Sieg verholfen? Die pposition in Washington fordert Aufklärung. Vor allem die Rolle von Jared Kushner und ephen Bannon ist unklar.

oland Nelles

018, 19.49 Uhr

Setting the Stage

- The bigger companies nowadays make a profit by collecting and analysing data from their current and potential future customers themselves
- An entire ecosystem of businesses that specialize on data collection and analysis for other companies and interest groups has emerged in recent years
- The number of services accessible on the internet for „free“ increased drastically, because the service providers are able to make money from information collected
- Yet, costumers are becoming more and more aware of this new form of payment and the legal regulations have tightened
- An exchange of data is desirable for all profiting actors in the information business, but trading or publishing requires compliance to legal regulations
- The goal is therefore: Maximizing utility while minimizing individual identifiability in a data set

External Matching

- Problem: Even if data has been anonymized, matching with background information might still lead to re-identification
- Publically available census data or publications from other sources are good candidates for conducting an external matching attempt
- For illustration purposes, lets transform the following secret data set:

Surname	Name	1st Treatment	2nd Treatment	Diagnose
Thomas	Meier	01.10.2020	13.10.2020	Lung cancer
Beate	Wimmer	01.10.2020	09.10.2020	Pelvic fractur
Maximilian	Huber	05.10.2020	07.10.2020	HIV

External Matching (cont.)

- Lets say that the hypothetical anonymisation process replaces the Names with codes and censors the Diagnoses, resulting in the sanitized version of the set, that is then published

Surname	Name	1st Treatment	2nd Treatment	Diagnose
Akdmclkm	Indncdc	01.10.2020	13.10.2020	*
Bkjdnckjn	Dgdhcd	02.10.2020	09.10.2020	*
Kodcndcd	Ldcdoo	05.10.2020	07.10.2020	*

- In addition to the sanitized set, an attacker also got the following information on the opening hours of treatment centres from public sources:

Monday	Tuesday	Wednesday	Thursday	Friday
HIV-Consulting	Irradiation	HIV-Consulting	Irradiation	Surgery

- For a potential employer, it would be easy to infer Mr. Hubers HIV-diagnose, if he has seen him on the 5th (Monday) or 7 (Wednesday) October at the clinic

Analysing the Data Set

- This shows that it isn‘t always sufficient to only look at the data that is to be protected
- Recombination attacks, like the one we just illustrated, also need to be taken into account -> privacy models
- In order to better distinguish the risks the attributes of tabular data pose to the individuals they represent, many models are based on a classification into 4 subject-dependent categories
- These classifications aren‘t absolute, the same column in different data sets aren‘t always sorted into the same categories -> context is important

1.3
Summary

Privacy-Preservation Technologies in Information Systems

Dr. Armin Gerl

WS 2021/2022

Recap of Chapter

- Privacy is a topic that is influenced by many Stakeholders
- Data collection on individuals is performed on a large scale by different actors, who also trade data amongst one another
- This data is used to support buisness and adminstrative decisions or perform targeted advertising, making it more and more valuable
- On the other hand, the right to data protection is slowly being established as a fundamental right across the globe, pioneered by the GDPR
- Companies face a trade-off between the usefullness of their collection and sufficient privacy of their customers being violated through re-identification
- First, an overview of this lectures topics is given...

Overview of Lecture Topics

We are here

Chapter	Est. Extent	Est. Dates
Chapter 1: Introduction	~1 Lecture	19.10.21
Chapter 2: From GDPR to Privacy Languages	~3 Lecture	26.10., 02.11., 09.11.21
Chapter 3: Basics on Data Anonymization in IS	~1 Lecture	16.11.2021
Chapter 4: Privacy Risks and Anonymization Techniques	~4 Lectures	23.11., 30.11.21, 07.12., 14.12.21
Chapter 5: Privacy in Health-Care	~2 Lectures	21.12., 11.01.21
Chapter 6: Privacy in Data Warehouses	~2 Lectures	18.01., 25.01.22
Chapter 7: Privacy in Social Networks	~2 Lectures	01.02., 08.02.22
Chapter 8: Current Research and Outlook	~2 Lectures	
Exam Preparation Lecture	1 Lecture	