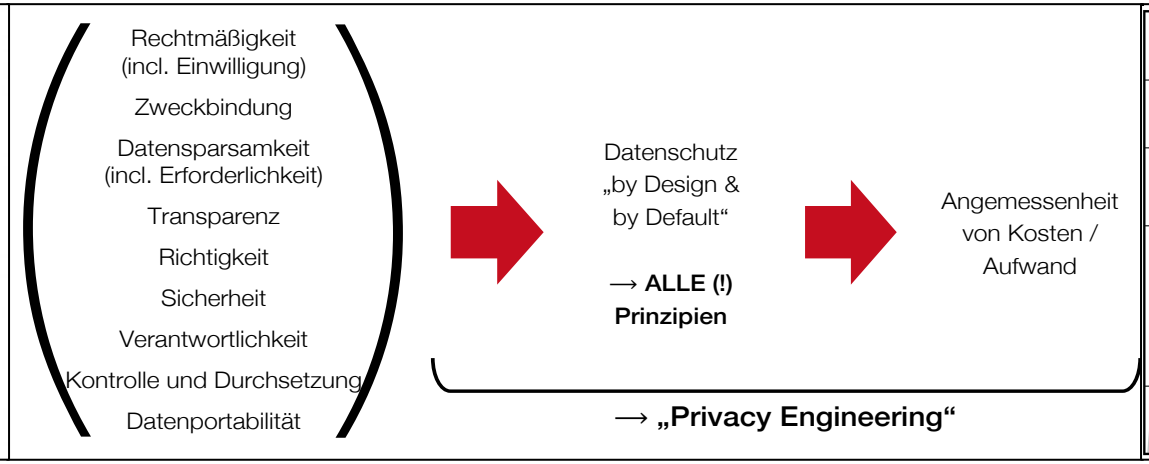
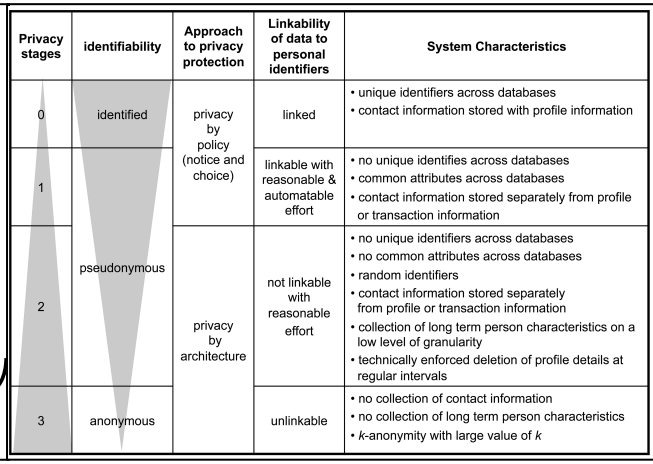
Information Governance
Lesson 09: Datenschutz 2 – Privacy Engineering





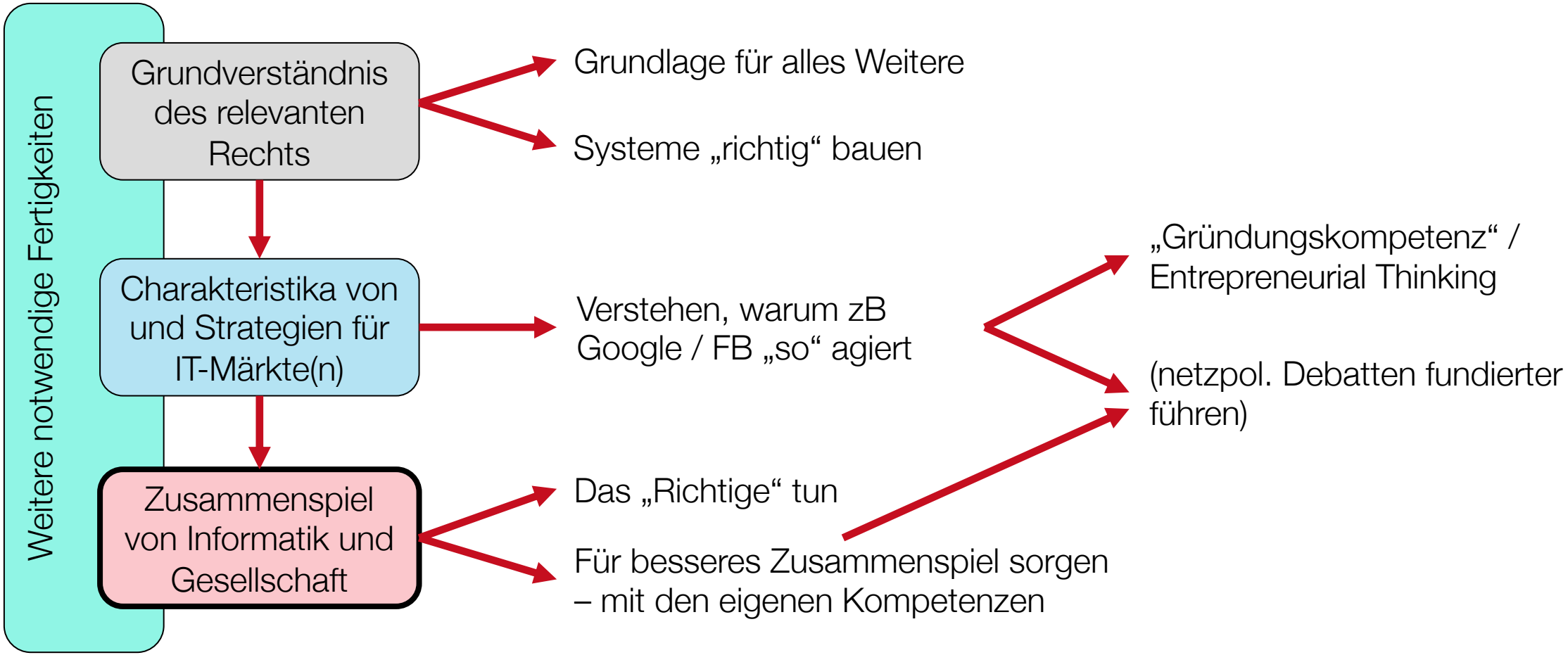
Privacy stages	identifiability	Approach to privacy protection	Linkability of data to personal identifiers	System Characteristics
0	identified	privacy by policy (notice and choice)	linked	• unique identifiers across databases • contact information stored with profile information
1			linkable with reasonable & automatable effort	• no unique identifies across databases • common attributes across databases • contact information stored separately from profile or transaction information
2	pseudonymous	privacy by architecture	not linkable with reasonable effort	• no unique identifiers across databases • no common attributes across databases • random identifiers • contact information stored separately from profile or transaction information • collection of long term person characteristics on a low level of granularity • technically enforced deletion of profile details at regular intervals
3	anonymous		unlinkable	• no collection of contact information • no collection of long term person characteristics • *k*-anonymity with large value of *k*

Frank Pallas

Information Systems Engineering
TU Berlin

ISEngineering

Information Governance – „Riding Skills“



Information Governance – Thematischer Überblick

Recht, Vertrag, Eigentum, Transaktion	Güterarten, „geist. Eigentum“, Lizenzmodelle	Wiss. Arbeiten	Wettbewerb, Preise, Transaktionsk.	Block A: Grundlagen
Informations-Ökonomie	Netzwerk- und Plattform-Ökon., Digital Commons	Digitale Transformation		Block B: Digitale Gesch.-modelle
Datenschutz: Rechtliche Grundlagen	Privacy Engineering	Konsultationen zu Semesterthemen	Surveillance	**Block C: Datenschutz & Privatheit**
Technikbasierte Regulierung, „Code as Law“	Konsultationen‘ + Postersessions (3 Wochen)	Technik- und informatische Berufsethik	Wunschthemen & Roundup	Block D: Gestaltungs-implikationen

Lesson 09: Datenschutz 2 – Privacy Engineering

Neun Prinzipien des Datenschutzrechts

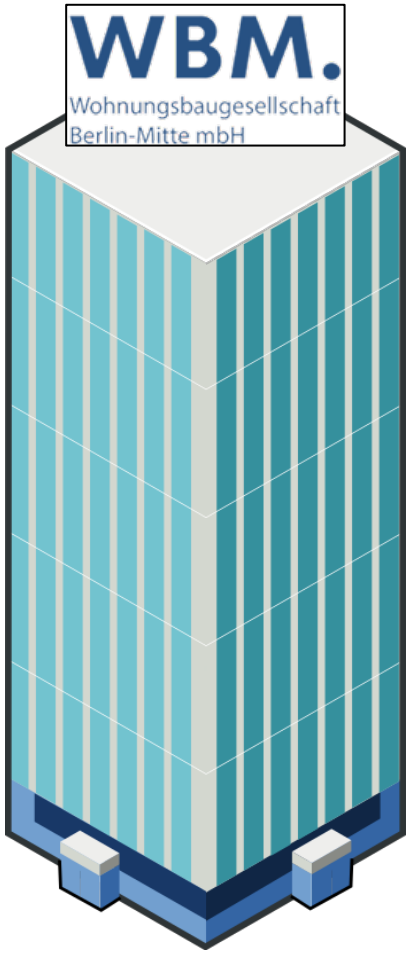
9+1 Prinzipien – „Privacy by Design" / „Privacy Engineering"

Technische Umsetzung von Datensparsamkeit

Wieviel Security ist angemessen? Experimente

Technische Ansätze für Transparenz und Zweckbindung

Recap: Stellen Sie sich vor...



„Entwerfen Sie ein tragfähiges Backend zur intelligenten Vernetzung des Gebäudebestands“

???

Grundlegende

Konzepte / Prinzipien

Recap

⟶ Ohne personenbezogene (oder „personenbeziehbare“) Daten keine Anwendbarkeit der GDPR

(und des Datenschutzrechts im Allgemeinen)

Recap: GDPR – Ziele

Natürliche Personen

Schutz von **Grundrechten**

Angemessenes Schutzniveu **und** freier Datenverkehr

Recap: 9 Kernprinzipien des Datenschutzes

Welche Konzepte kennen Sie aus der letzten Woche?

Recap: 4/9 Kernprinzipien des Datenschutzes

Rechtmäßigkeit (incl. Einwilligung)

Zweckbindung

Datensparsamkeit (incl. Erforderlichkeit)

Transparenz

9 Kernprinzipien des Datenschutzes

Rechtmäßigkeit (incl. Einwilligung)

Zweckbindung

Datensparsamkeit (incl. Erforderlichkeit)

Transparenz

Richtigkeit

Richtigkeit

z.B.

„Personenbezogene Daten müssen [...] sachlich richtig und erforderlichenfalls auf dem neuesten Stand sein; es sind alle angemessenen Maßnahmen zu treffen, damit personenbezogene Daten, die [...] unrichtig sind, unverzüglich gelöscht oder berichtigt werden“

Art. 5(1d) GDPR

→ Pflicht des „Controllers“ und Anrecht des „Data Subjects“!

9 Kernprinzipien des Datenschutzes

Rechtmäßigkeit (incl. Einwilligung)

Zweckbindung

Datensparsamkeit (incl. Erforderlichkeit)

Transparenz

Richtigkeit

Sicherheit

Sicherheit

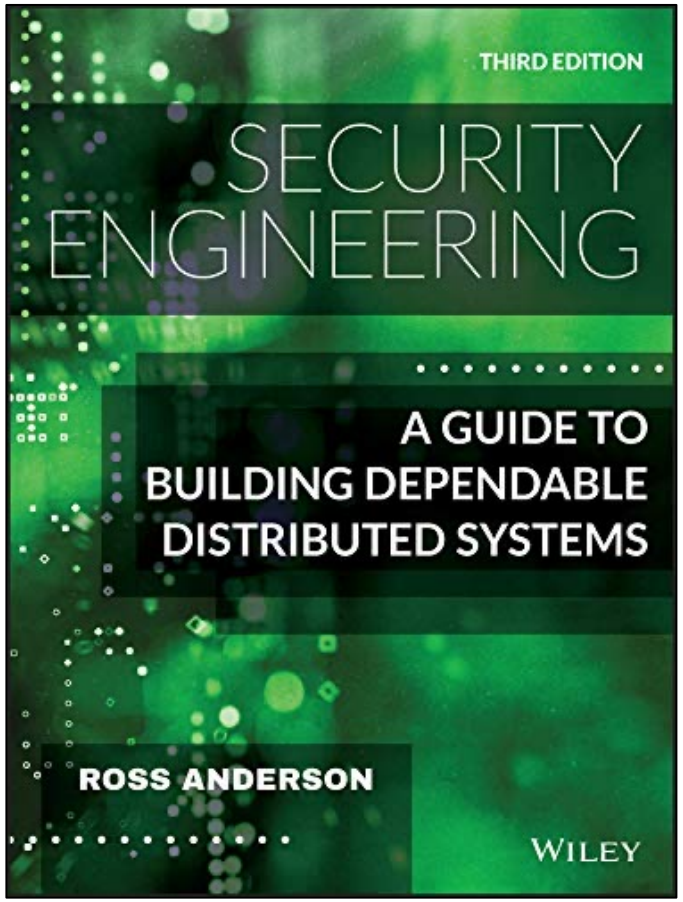
z.B.

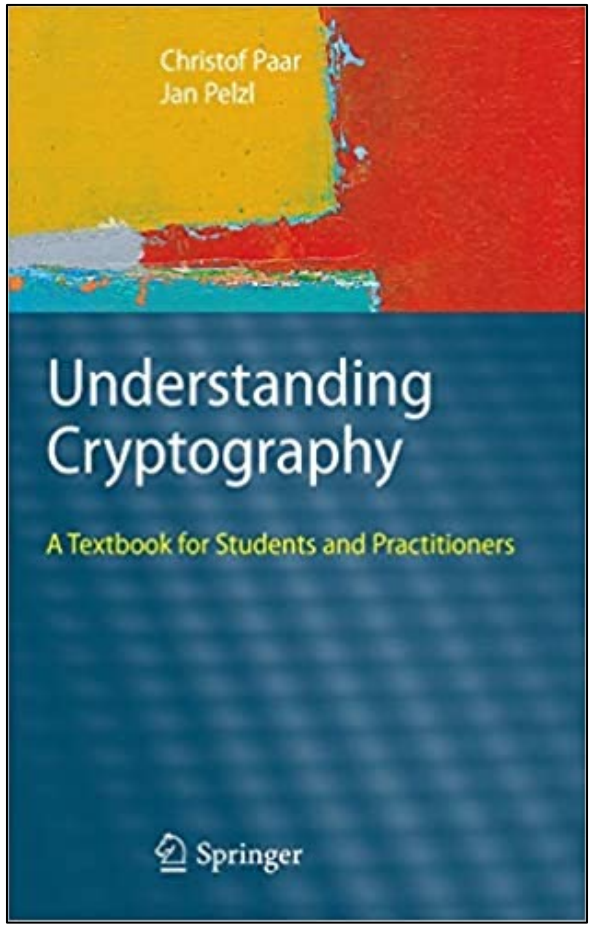
„Personenbezogene Daten müssen [...] in einer Weise verarbeitet werden, die eine angemessene Sicherheit der personenbezogenen Daten gewährleistet, einschließlich Schutz vor unbefugter oder unrechtmäßiger Verarbeitung und vor unbeabsichtigtem Verlust, unbeabsichtigter Zerstörung oder unbeabsichtigter Schädigung durch geeignete technische und organisatorische Maßnahmen“

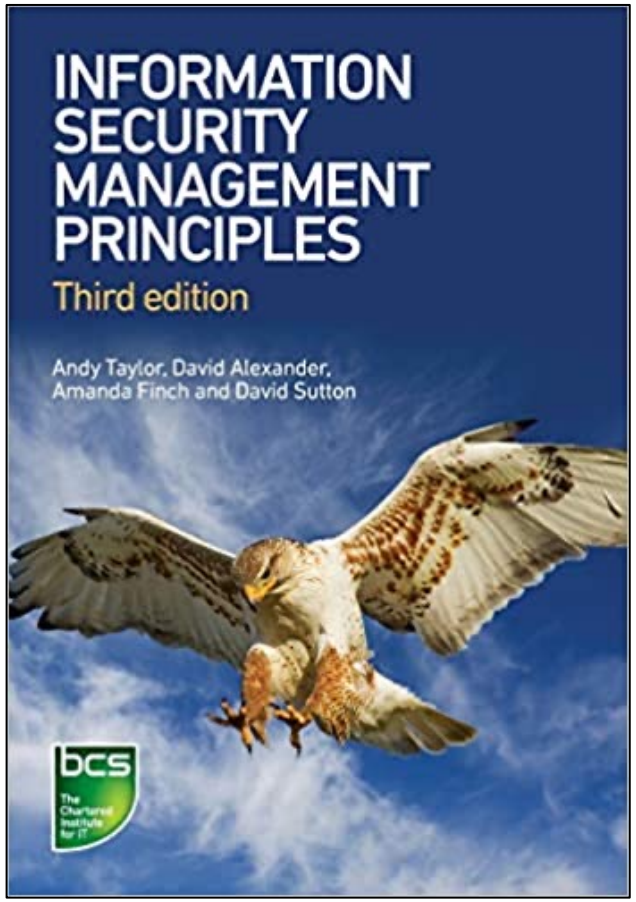
Art. 5(1f) GDPR

⟶ **Confidentiality** (Vertraulichkeit)**, Integrity** (Integrität)**, Availability** (Verfügbarkeit)
⟶ **C-I-A**

Sicherheit

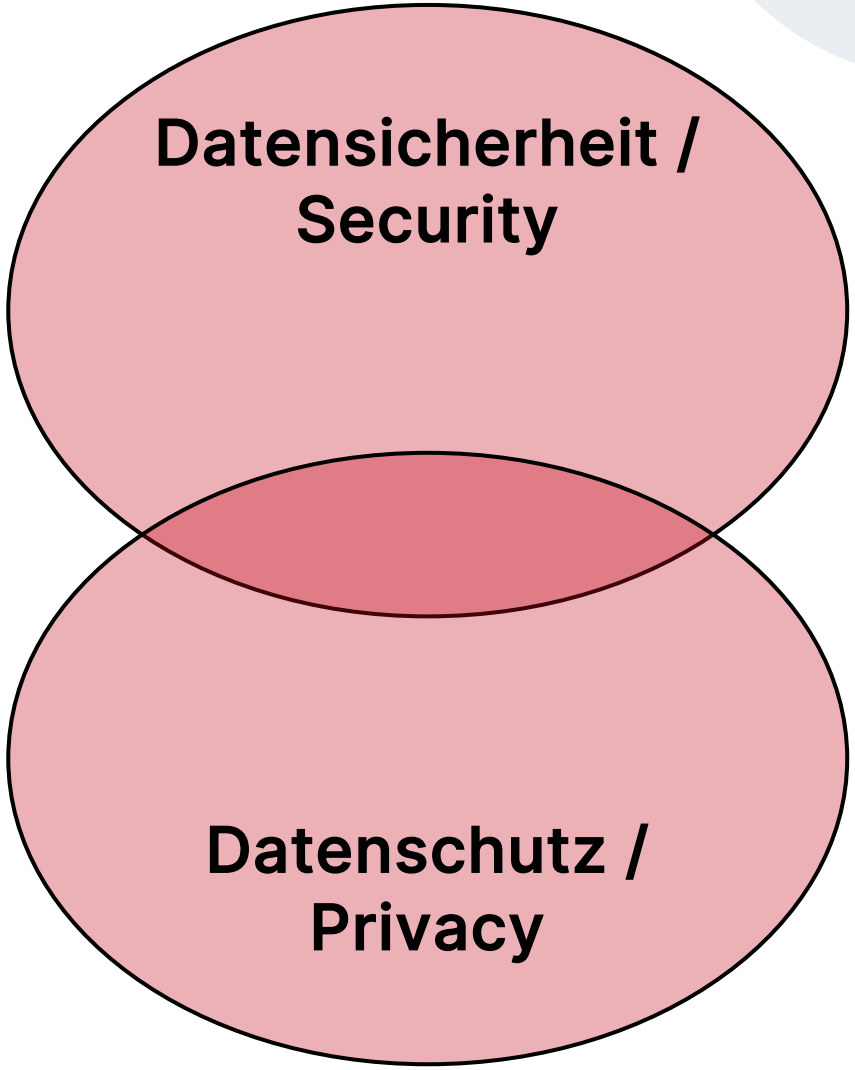

THIRD EDITION
SECURITY ENGINEERING
A GUIDE TO BUILDING DEPENDABLE DISTRIBUTED SYSTEMS
ROSS ANDERSON
WILEY


Christof Paar
Jan Pelzl
Understanding Cryptography
A Textbook for Students and Practitioners
Springer


INFORMATION SECURITY MANAGEMENT PRINCIPLES
Third edition
Andy Taylor, David Alexander, Amanda Finch and David Sutton
bcs

Datensicherheit / Datenschutz

- (Daten-) Sicherheit als **eines von mehreren** Kernprinzipien des Datenschutzes
- (Daten-) **Sicherheit auch jenseits von Datenschutz** auf vielfältige Weise relevant
- Ergo: Überschneidungen; Datenschutz und Datensicherheit sind **weder deckungsgleich noch** ist das Eine eine **Untermenge** des Anderen!



9 Kernprinzipien des Datenschutzes

Rechtmäßigkeit (incl. Einwilligung)

Zweckbindung

Datensparsamkeit (incl. Erforderlichkeit)

Transparenz

Richtigkeit

Sicherheit (insb. Vertraulichkeit, Integrität, Verfügbarkeit)

Verantwortlichkeit

Verantwortlichkeit

„Der Verantwortliche ist für die Einhaltung des Absatzes 1 verantwortlich und muss dessen Einhaltung nachweisen können“

Art. 5(2) GDPR

„Der Verantwortliche setzt [...] geeignete technische und organisatorische Maßnahmen um, um sicherzustellen und den Nachweis dafür erbringen zu können, dass die Verarbeitung gemäß dieser Verordnung erfolgt. “

Art. 24(1) GDPR

Belegen von „Rechtsbruch“ → Belegen von Konformität (vgl. „Beweislastumkehr“)

9 Kernprinzipien des Datenschutzes

Rechtmäßigkeit (incl. Einwilligung)

Zweckbindung

Datensparsamkeit (incl. Erforderlichkeit)

Transparenz

Richtigkeit

Sicherheit (insb. Vertraulichkeit, Integrität, Verfügbarkeit)

Verantwortlichkeit (incl. nachweisbare Rechtskonformität)

Kontrolle und Durchsetzung

Kontrolle und Durchsetzung

„Jeder Mitgliedstaat sieht vor, dass eine oder mehrere unabhängige Behörden für die Überwachung der Anwendung dieser Verordnung zuständig“

Art. 51(1) GDPR

„Der Europäische Datenschutzausschuss (im Folgenden „Ausschuss“) wird als Einrichtung der Union mit eigener Rechtspersönlichkeit eingerichtet. […] [Er] besteht aus dem Leiter einer Aufsichtsbehörde jedes Mitgliedstaats und […]“

Art. 68 GDPR

Kontrolle und Durchsetzung

BfDI Der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit

Berliner Beauftragte für Datenschutz und Informationsfreiheit

ULD Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein

Landesbeauftragte für Datenschutz und Akteneinsicht

. . .

→ Überwachen, fördern, Beschwerden entgegennehmen, ..., Ordnungsgelder verhängen

Kontrolle und Durchsetzung

„Bei Verstößen [...] Geldbußen von bis zu 20.000.000 EUR / 10.000.000 EUR oder im Fall eines Unternehmens von bis zu 4% / 2% seines gesamten weltweit erzielten Jahresumsatzes des vorangegangenen Geschäftsjahrs verhängt, je nachdem, welcher der Beträge höher ist.“

⟶ Abhängig vom Artikel, gegen den verstoßen wurde

Art. 83(2, 4, 5) GDPR

⟶ 1,2 Mrd EUR für Meta (2023)

⟶ bis zu 20/10 (>10/5) Mio EUR für WBM 2023
(60 / 30 Mio EUR für Deutsche Wohnen 2023)

9 Kernprinzipien des Datenschutzes

Rechtmäßigkeit (incl. Einwilligung)

Zweckbindung

Datensparsamkeit (incl. Erforderlichkeit)

Transparenz

Richtigkeit

Sicherheit (insb. Vertraulichkeit, Integrität, Verfügbarkeit)

Verantwortlichkeit (incl. nachweisbare Rechtskonformität)

Kontrolle und Durchsetzung

Datenportabilität

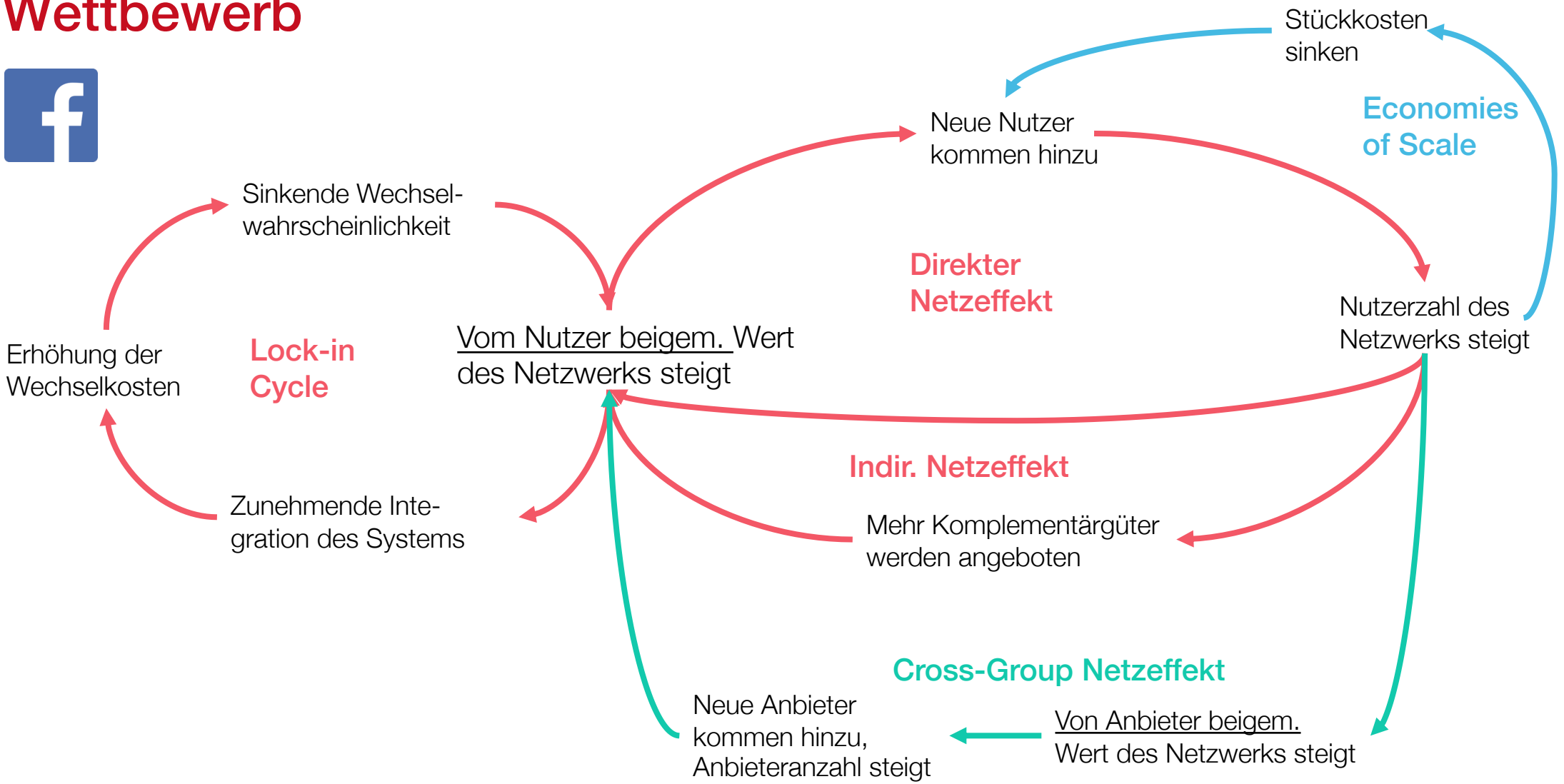
Datenportabilität

„1) Die betroffene Person hat das Recht, die sie betreffenden personenbezogenen Daten, die sie einem Verantwortlichen bereitgestellt hat, in einem strukturierten, gängigen und maschinenlesbaren Format zu erhalten, und sie hat das Recht, diese Daten einem anderen Verantwortlichen [...] zu übermitteln, sofern [...]

2) die betroffene Person das Recht, zu erwirken, dass die personenbezogenen Daten direkt von einem Verantwortlichen einem anderen Verantwortlichen übermittelt werden, soweit dies technisch machbar ist."

Art. 20 GDPR

Datenportabilität – Im Kern nicht Datenschutz sondern Wettbewerb



9 Kernprinzipien des Datenschutzes

Rechtmäßigkeit (incl. Einwilligung)

Zweckbindung

Datensparsamkeit (incl. Erforderlichkeit)

Transparenz

Richtigkeit

Sicherheit (insb. Vertraulichkeit, Integrität, Verfügbarkeit)

Verantwortlichkeit (incl. nachweisbare Rechtskonformität)

Kontrolle und Durchsetzung

Datenportabilität

→ ALLE Kernprinzipien müssen erfüllt werden!
(z.B. kein „Verzicht“ auf Datensparsamkeit möglich)

Lesson 09: Datenschutz 2 – Privacy Engineering

Neun Prinzipien des Datenschutzrechts

9+1 Prinzipien – „Privacy by Design" / „Privacy Engineering"

Technische Umsetzung von Datensparsamkeit

Wieviel Security ist angemessen? Experimente

Technische Ansätze für Transparenz und Zweckbindung

Kernprinzipien des Datenschutzes

9 **+ 1** Kernprinzipien

„Privacy / Data Protection by Design“

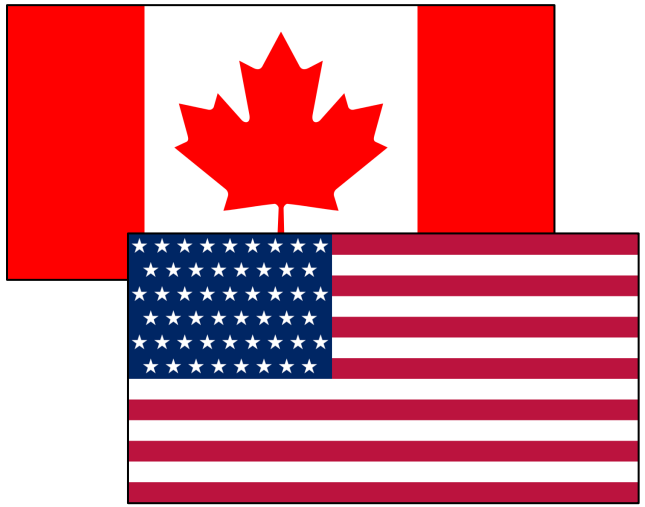
„Privacy“?

Privacy by Design:
The 7 Foundational Principles

1. ***Proactive*** not ***Reactive***:
   Preventative, not Remedial;
2. Privacy as the ***Default*** setting;
3. Privacy ***Embedded*** into Design;
4. ***Full*** Functionality:
   Positive-Sum, not Zero-Sum;
5. End-to-End **Security**:
   Full Lifecycle Protection;
6. Visibility **and** Transparency:
   Keep it **Open**;
7. Respect for User Privacy:
   Keep it **User-Centric**.

RYERSON UNIVERSITY

„Privacy“ vs. „Data Protection“: „Two Western Cultures“



„Privacy“

VS.

„Data Protection“

Article

The Two Western Cultures of Privacy: Dignity Versus Liberty

James Q. Whitman†

CONTENTS

† Ford Foundation Professor of Comparative and Foreign Law, Yale University. Earlier versions of this paper were presented to the helpful audiences at the Unidem Seminar on European and American Constitutionalism, the Institut des Hautes Études sur la Justice, the French-American Foundation of Paris, and Columbia and Yale Law Schools. The author would also like to thank Anita Allen, Friedrich Wenzel Bulst, Agnès Dunogué, Edward Eberle, William Edmundson, Christoph Paulus, and Jeffrey Rosen for their aid and advice. Unless otherwise noted, all translations are my own.

„Privacy“ vs. „Data Protection“: „Two Western Cultures“

„Continental privacy [is] a right to respect and personal dignity [...] rights to one‘s image, name, and reputation [...] what Germans call the right to informational self-determination [...].“

„[T]he American right to privacy [...] is the right to freedom from intrusions by the state, especially in one‘s own home [...] reasonable expectations of privacy“

Whitman (2004)

„Privacy“ vs. „Data Protection“

~~„Privacy“ vs. „Data Protection“~~

→ Die Unterscheidung zwischen „Privacy“ und „Data Protection“ ist zwar wichtig (insb. für juristische Argumentationen) – Hier reicht es aber, wenn Sie wissen, **dass** es einen Unterschied gibt.

„by Design & by Default“

„Unter Berücksichtigung des Stands der Technik, der Implementierungskosten und [...] trifft der Verantwortliche [...] geeignete technische und organisatorische Maßnahmen [...] die dafür ausgelegt sind, die Datenschutzgrundsätze [...] wirksam umzusetzen [...]“

„Der Verantwortliche trifft geeignete technische und organisatorische Maßnahmen, die sicherstellen, dass durch Voreinstellung grundsätzlich nur personenbezogene Daten, deren Verarbeitung [...] erforderlich ist, verarbeitet werden.“

Art. 25 GDPR

„by Design & by Default“

Rechtmäßigkeit (incl. Einwilligung)

Zweckbindung

Datensparsamkeit (incl. Erforderlichkeit)

Transparenz

Richtigkeit

Sicherheit

Verantwortlichkeit

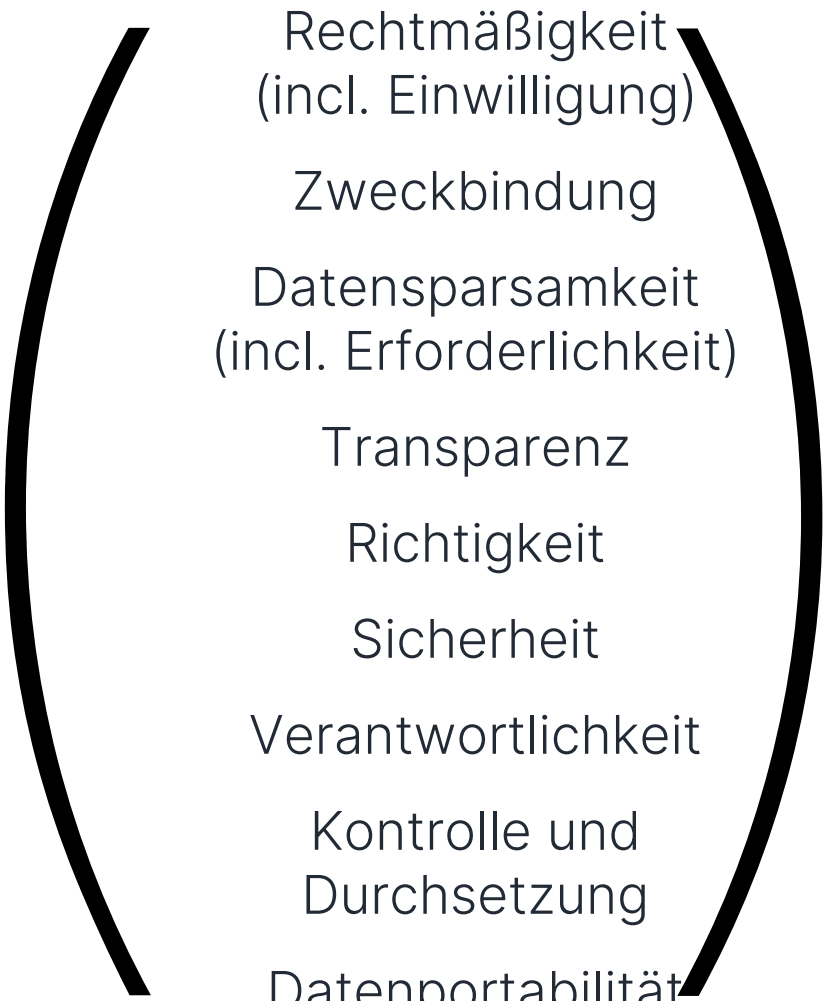
Kontrolle und Durchsetzung

Datenportabilität

Umsetzung durch technische und organisatorische Maßnahmen

→ **ALLE (!) Prinzipien**

„by Design & by Default“



Datenschutz „by Design & by Default“

→ **ALLE (!) Prinzipien**

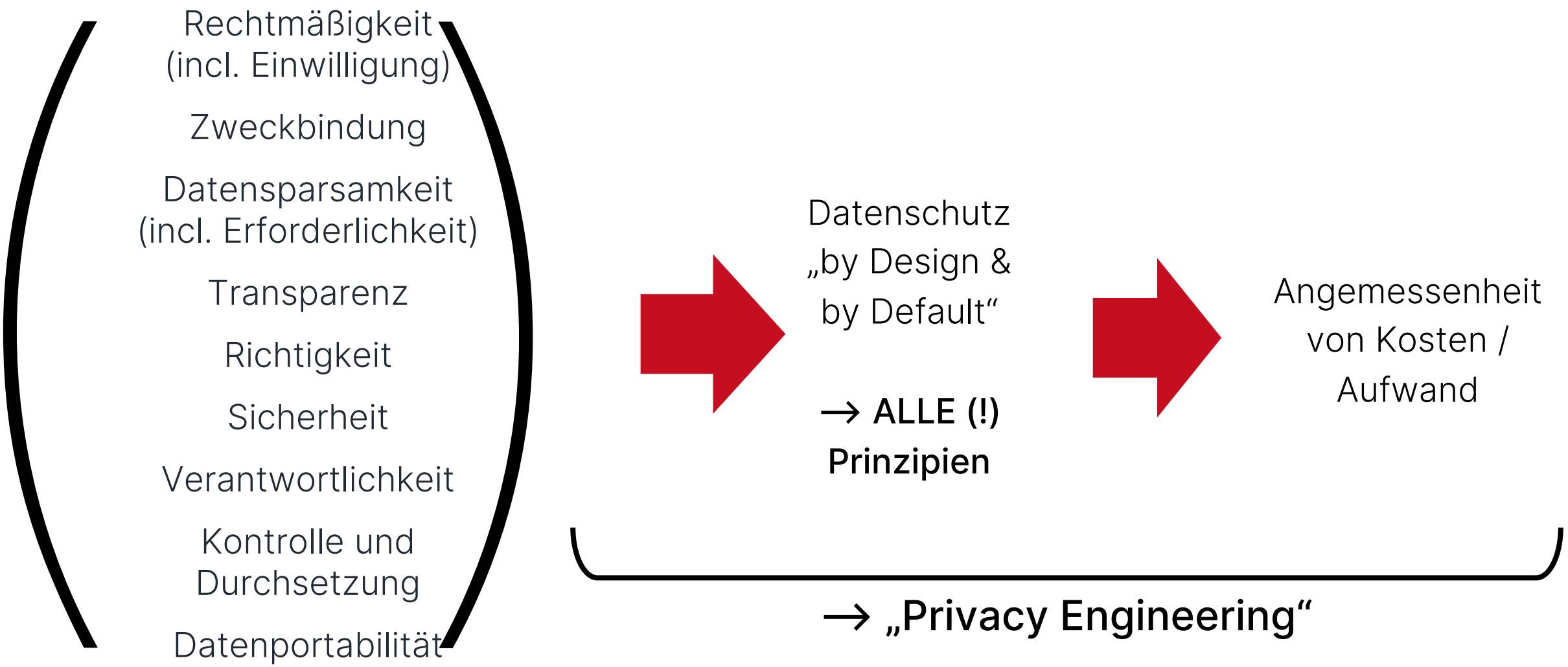
„by Design & by Default“

„Unter Berücksichtigung des Stands der Technik,
der Implementierungskosten und [...] trifft der Verantwortliche [...] geeignete
technische und organisatorische Maßnahmen [...] die dafür ausgelegt sind,
die Datenschutzgrundsätze [...] wirksam umzusetzen [...]“

„Der Verantwortliche trifft geeignete technische und organisatorische Maßnahmen, die sicherstellen, dass durch Voreinstellung grundsätzlich nur personenbezogene Daten, deren Verarbeitung [...] erforderlich ist, verarbeitet werden.“

Art. 25 GDPR

„by Design & by Default“



„Privacy Engineering“

Privacy stages	identifiability	Approach to privacy protection	Linkability of data to personal identifiers	System Characteristics
0	identified	privacy by policy (notice and choice)	linked	• unique identifiers across databases • contact information stored with profile information
1			linkable with reasonable & automatable effort	• no unique identifies across da • common attributes across data • contact information stored sep or transaction information
2	pseudonymous	privacy by architecture	not linkable with reasonable effort	• no unique identifiers across da • no common attributes across • random identifiers • contact information stored sep from profile or transaction info • collection of long term person low level of granularity • technically enforced deletion regular intervals
3	anonymous		unlinkable	• no collection of contact inform • no collection of long term pers • *k*-anonymity with large value

NISTIR 8062

An Introduction to Privacy Engineering and Risk Management in Federal Systems

DOI:10.1145/2633029

Privacy and Security
Can You Engineer Privacy?

The challenges and potential approaches to applying privacy research in engineering practice.

INDUSTRIAL-SIZE DATA SPILLS, leaks about large-scale secret surveillance programs, and personal tragedies due to inappropriate flows of information are

such as notice and choice, data retention limitation, and subject access rights. These principles are seen to be instrumental to making the collection and processing activities of organizations transparent. Although less amous, data protection principles need to be translated into technirequirements and are vulnerable arrow interpretations. Moreover, Ps fall short of mitigating all the acy concerns of users toward a n organization. They also do not ress privacy concerns users may

have with respect to other users, with people in their social environments, and toward a greater public.

Scholars from various fields have stepped up to the challenge of clearing the murky waters of privacy. Legal scholars and philosophers have proposed taxonomies of privacy violations[5] and a holistic framework for evaluating appropriate flows of information based on contextual social norms.[3] Social scientists and ethnographers have studied groups of people, online and offline, to develop better-informed understandings of users' needs. But how are engineers supposed to integrate

viewpoints

DOI:10.1145/3486631 Lea Kissner and Lorrie Cranor

Privacy
Privacy Engineering Superheroes

Privacy engineers are essential to both preventing and responding to organizational privacy problems.

DOES YOUR ORGANIZATION want to offer cookie choices without annoying popups? Do you want to share sensitive data in aggregate form without risking a privacy breach? Do you want to monitor data flows to ensure personal information does not end up

2022 International Workshop on Privacy Engineering – IWPE'22

CO-LOCATED WITH 7TH IEEE EUROPEAN SYMPOSIUM ON SECURITY AND PRIVACY

JUNE 6, 2022, GENOA (ITALY)

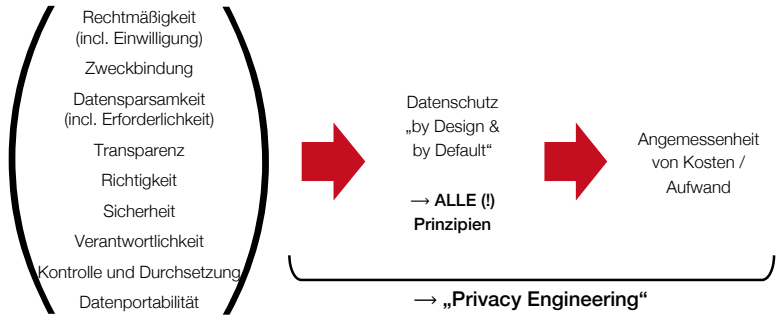
„Privacy Engineering“ – (Eine mögliche) Definition

„Privacy engineering [...] aims to provide methodologies, tools, and techniques to ensure systems provide acceptable levels of privacy. In [...] the EU [...] the General Data Protection Regulation sets the requirements that need to be fulfilled.“

Wikipedia, 28.11.2022

by Design & by Default – Privacy Engineering

- „Privacy“ vs. „Data Protection“: Ein wichtiger, aber nicht überall relevanter Unterschied
- Art. 25 GDPR verpflichtet zur Umsetzung **aller** Datenschutzprinzipien mittels technischer und organisatorischer Maßnahmen
- Allerdings sind nur solche Maßnahmen notwendig, deren **Kosten und Aufwand angemessen** sind
- Entwicklung, Einsatz und Aufwands-/Kostenbewertung entspr. technischer Mechanismen lassen sich unter dem Begriff **„Privacy Engineering“** zusammenfassen



Lesson 09: Datenschutz 2 – Privacy Engineering

Neun Prinzipien des Datenschutzrechts

9+1 Prinzipien – „Privacy by Design" / „Privacy Engineering"

Technische Umsetzung von Datensparsamkeit

Wieviel Security ist angemessen? Experimente

Technische Ansätze für Transparenz und Zweckbindung

9 Kernprinzipien des Datenschutzes

Rechtmäßigkeit (incl. Einwilligung)

Zweckbindung

Datensparsamkeit (incl. Erforderlichkeit)

Transparenz

Richtigkeit

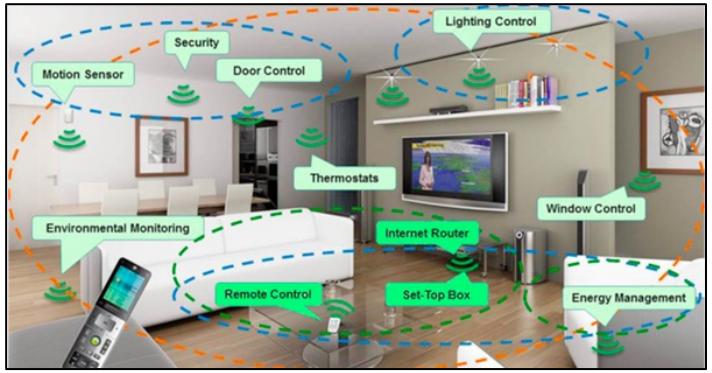
Sicherheit (insb. Vertraulichkeit, Integrität, Verfügbarkeit)

Verantwortlichkeit (incl. nachweisbare Rechtskonformität)

Kontrolle und Durchsetzung

Datenportabilität

Privacy Engineering: Datensparsamkeit

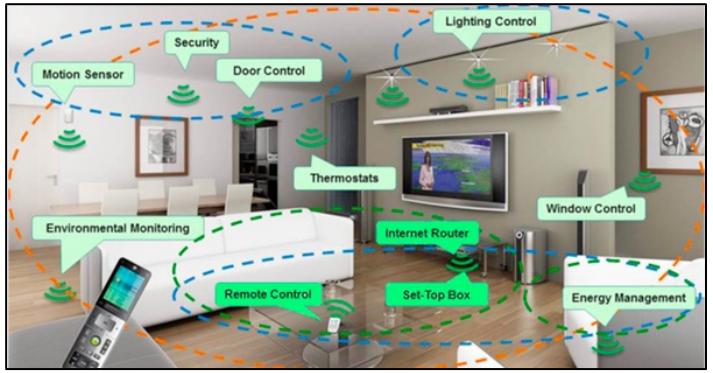


?

WBM.
Wohnungsbaugesellschaft
Berlin-Mitte mbH

→ Wie lässt sich Datensparsamkeit für den Anwendungsfall „Gebäudeautomation“ technisch umsetzen?

Privacy Engineering: Datensparsamkeit



?

WBM.
Wohnungsbaugesellschaft Berlin-Mitte mbH

→ Frühes Filtern / Aussortieren

→ Lokale vs. zentrale Verarbeitung

→ Pseudonymisierung / Anonymisierung / Aggregation

Privacy Engineering: Datensparsamkeit

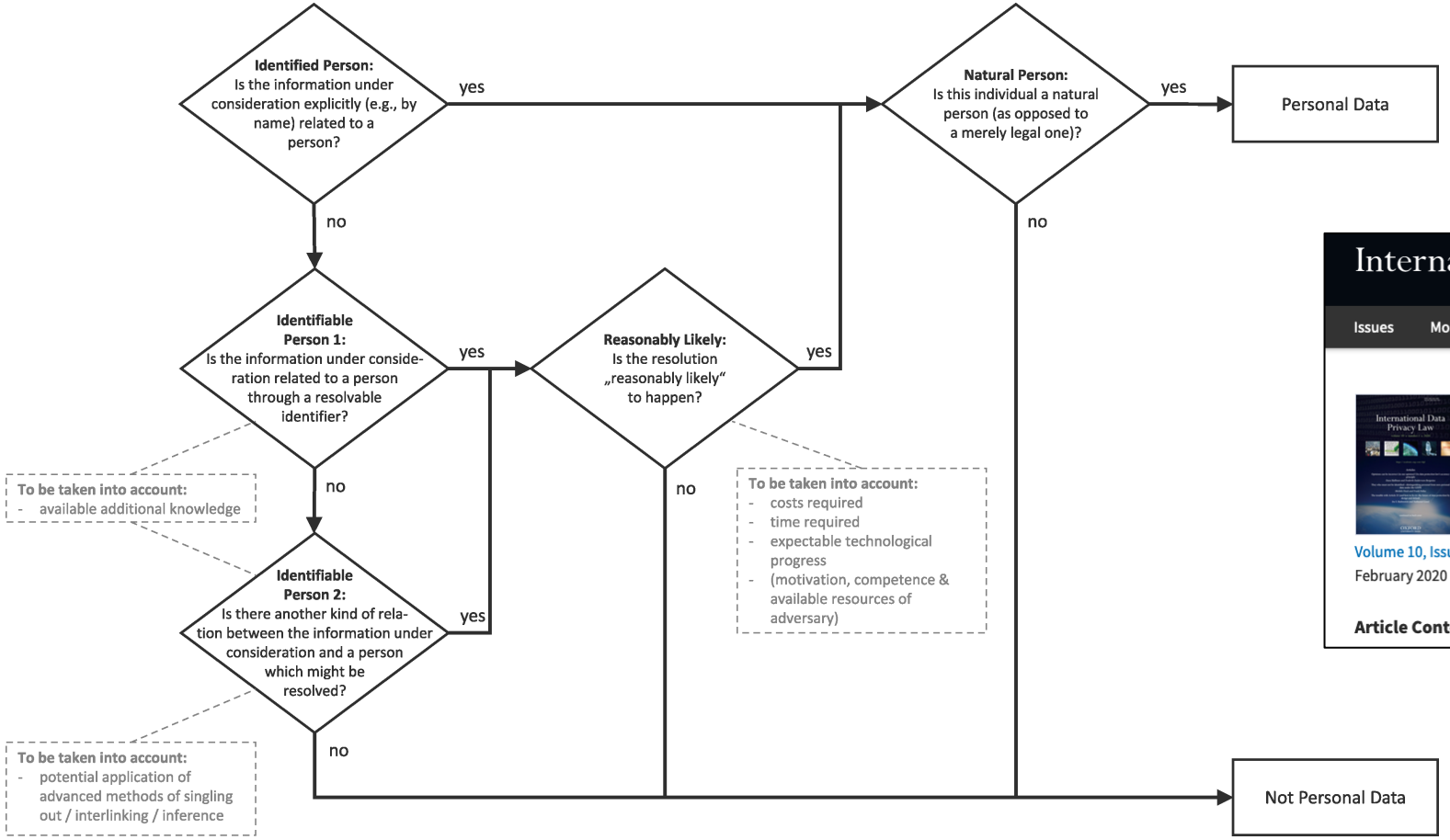
Privacy stages	identifiability	Approach to privacy protection	Linkability of data to personal identifiers	System Characteristics
0	identified	privacy by policy (notice and choice)	linked	• unique identifiers across databases • contact information stored with profile information
1	pseudonymous		linkable with reasonable & automatable effort	• no unique identifies across databases • common attributes across databases • contact information stored separately from profile or transaction information
2		privacy by architecture	not linkable with reasonable effort	• no unique identifiers across databases • no common attributes across databases • random identifiers • contact information stored separately from profile or transaction information • collection of long term person characteristics on a low level of granularity • technically enforced deletion of profile details at regular intervals
3	anonymous		unlinkable	• no collection of contact information • no collection of long term person characteristics • *k*-anonymity with large value of *k*

Spiekermann, S., & Cranor, L. F. (2009). Engineering Privacy. IEEE Transactions on Software Engineering, 35(1), 67-82.

Spiekermann & Cranor unterscheiden zwischen:

- **„identified data“:** Daten weisen einen expliziten Personenbezug auf
- **„pseudonymous data (a)“:** Daten sind pseudonymisiert, Personenbezug kann (für Partei X) mit „reasonable and automatable effort“ wiederhergestellt werden
- **„pseudonymous data (b)“:** Daten sind pseudonymisiert, Personenbezug kann (für Partei X) nicht mit „reasonable and automatable effort“ wiederhergestellt werden
- **„anonymous data“:** Personenbezug ist “unwiederbringlich“ entfernt

Privacy Engineering: Datensparsamkeit



Datensparsamkeit

Zeitstempel	Wohnung	Device	Status
… 18:33:26	1101 (F. Pallas)	002-Bew. Flur	an
… 18:41:18	1101 (F. Pallas)	002-Bew. Flur	aus
…	…	…	…
… 19:30:48	1101 (F. Pallas)	016-Balkontür	offen
… 19:30:53	1101 (F. Pallas)	016-Balkontür	zu
… 19:35:31	1101 (F. Pallas)	016-Balkontür	offen
… 19:35:34	1101 (F. Pallas)	016-Balkontür	zu
…	…	…	…
… 00:12:14	1101 (F. Pallas)	016-Balkontür	offen
… 00:12:17	1101 (F. Pallas)	016-Balkontür	zu
… 00:16:48	1101 (F. Pallas)	016-Balkontür	offen
… 00:16:52	1101 (F. Pallas)	016-Balkontür	zu
… 00:17:02	1101 (F. Pallas)	002-Bew. Flur	an
…	…	…	…

Datensparsamkeit

In explizit personenbezogener Rohform können Daten potentiell tiefgreifende Einblicke in Lebensgewohnheiten geben

Pseudonymisierung

Zeitstempel	Wohnungspseudonym	Device	Status
… 18:33:26	0x37a8cf204b1	002-Bew. Flur	an
… 18:41:18	0x37a8cf204b1	002-Bew. Flur	aus
…	…	…	…
… 19:30:48	0x37a8cf204b1	016-Balkontür	offen
… 19:30:53	0x37a8cf204b1	016-Balkontür	zu
… 19:35:31	0x37a8cf204b1	016-Balkontür	offen
… 19:35:34	0x37a8cf204b1	016-Balkontür	zu
…	…	…	…
… 00:12:14	0x37a8cf204b1	016-Balkontür	offen
… 00:12:17	0x37a8cf204b1	016-Balkontür	zu
… 00:16:48	0x37a8cf204b1	016-Balkontür	offen
… 00:16:52	0x37a8cf204b1	016-Balkontür	zu
… 00:17:02	0x37a8cf204b1	002-Bew. Flur	an
…	…	…	…

Datensparsamkeit

Einfache Pseudonymisierung (z.B. Ersetzen durch Zufallszahl und parallel gehaltene Zuordnungstabelle) minimiert Risiko

Allerdings: Wer z.B. Zugriff auf Zuordnungstabelle hat, hat Einblicke weiterhin

Anonymisierung

Zeitstempel		Device	Status
… 18:33:26		002-Bew. Flur	an
… 18:41:18		002-Bew. Flur	aus
…		…	…
… 19:30:48		016-Balkontür	offen
… 19:30:53		016-Balkontür	zu
… 19:35:31		016-Balkontür	offen
… 19:35:34		016-Balkontür	zu
…		…	…
… 00:12:14		016-Balkontür	offen
… 00:12:17		016-Balkontür	zu
… 00:16:48		016-Balkontür	offen
… 00:16:52		016-Balkontür	zu
… 00:17:02		002-Bew. Flur	an
…		…	…

Datensparsamkeit

Ohne Identifier sind Rückschlüsse ungleich schwerer – grundsätzlich aber nicht ausgeschlossen (z.B. durch Verschränken mit anderen Datensätzen, Erkennung innerer Zusammenhänge, ...)

Aggregation

Zeitstempel	Gebäude	Kategorie	Anzahl im gesamten Gebäude
… 18:33:26	11	Aktivität	5
… 18:41:18	11	Aktivität	4
…		…	…
… 19:30:48	11	Außenfenster offen	12
… 19:30:53	11	Außenfenster offen	11
… 19:35:31	11	Außenfenster offen	12
… 19:35:34	11	Außenfenster offen	10
…		…	…
… 00:12:14	11	Außenfenster offen	4
… 00:12:17	11	Außenfenster offen	3
… 00:16:48	11	Außenfenster offen	4
… 00:16:52	11	Außenfenster offen	3
… 00:17:02	11	Aktivität	7
…		…	…

Datensparsamkeit

Aggregation lässt „individuelle“ Werte in Summen etc. „untergehen“

Aggregation: Insbesondere

K-Anonymität:

Daten werden so depersonalisiert / generalisiert, dass immer mindestens *k* Personen nicht voneinander unterschieden werden können

L-Diversität:

Daten werden so depersonalisiert / generalisiert, dass jede Gruppe mindestens *L* unterschiedliche „sensible“ Werte beinhaltet

Beispiel *k*-Anonymität, *k*=3

Matrikelnr.	Name	Fachsemester	Note
--	--	<= 5	2.3
--	--	6	2.3
--	--	7-10	2.0
--	--	<= 5	1.7
--	--	<= 5	3.0
--	--	>= 11	1.0
--	--	>= 11	1.3
--	--	<=5	2.3
--	--	7-10	1.7
--	--	6	3.3
--	--	<= 5	1.0
--	--	6	2.0
--	--	>=11	1.3
--	--	7-10	3.3

Beispiel *k*-Anonymität, *k*=3

Matrikelnr.	Name	Fachsemester	Note
--	--	<= 5	2.3
--	--	6	2.3
--	--	7-10	2.0
--	--	<= 5	1.7
--	--	<= 5	3.0
--	--	>= 11	1.0
--	--	>= 11	1.3
--	--	<=5	2.3
--	--	7-10	1.7
--	--	6	3.3
--	--	<= 5	1.0
--	--	6	2.0
--	--	>=11	1.3
--	--	7-10	3.3

k-Anonymität erlaubt z.B. statistische Erkenntnisse, ohne einzelne identifizierbare Personen

Aber: wenn sensible Werte für alle Mitglieder einer Gruppe gleich/ähnlich sind, lassen sich aus bekannter Zugehörigkeit zu einer Gruppe Rückschlüsse ziehen

Beispiel *l*-Diversität, *l*=3

MatrikelNr.	Name	Fachsemester	Note
--	--	<= 6	2
--	--	<= 6	2
--	--	>= 7	2
--	--	<= 6	2
--	--	<= 6	3
--	--	>= 7	1
--	--	>= 7	1
--	--	<= 6	2
--	--	>= 7	2
--	--	<= 6	3
--	--	<= 6	1
--	--	<= 6	2
--	--	>= 7	1
--	--	>= 7	3

Datensparsamkeit

They who must not be identified—distinguishing personal from non-personal data under the GDPR

Michèle P

Internati

https://d

Publishe

	1. Start from known person	2. Start from content
A. ID-based re-identification	'Find all transactions that John Smith was involved in, based on his known ID'	'Find the persons involved in transaction *X*, based on known IDs of all persons to be considered'
B. Content-based re-identification	'Find all transactions that John Smith was involved in through matching transaction data with his known bank account history'	'Find the persons involved in transaction *X* through matching transaction data with bank account histories of all persons to be considered'

„Differential Privacy“

Privacy Engineering: Datensparsamkeit

- Ein und dasselbe Datum kann gleichzeitig für Partei A personenbezogen / -beziehbar sein und für Partei B nicht – etwas wenn Partei A Zugriff auf eine Zuordnungsliste oder „Zusatzwissen“ hat
- Datensparsamkeit (oder Datenminimierung) meint Minimierung **personenbezogener** Daten (für Partei X)
- Dies kann durch Minimierung der Daten oder durch „Minimierung“ / Entfernung / Erschwerung des Personenbezugs geschehen
- Vielzahl von Techniken / Verfahren zur Datenminimierung / Anonymisierung

Privacy stages	identifiability	Approach to privacy protection	Linkability of data to personal identifiers	System Characteristics
0	identified	privacy by policy (notice and choice)	linked	• unique identifiers across databases • contact information stored with profile information
1	pseudonymous		linkable with reasonable & automatable effort	• no unique identifies across databases • common attributes across databases • contact information stored separately from profile or transaction information
2		privacy by architecture	not linkable with reasonable effort	• no unique identifiers across databases • no common attributes across databases • random identifiers • contact information stored separately from profile or transaction information • collection of long term person characteristics on a low level of granularity • technically enforced deletion of profile details at regular intervals
3	anonymous		unlinkable	• no collection of contact information • no collection of long term person characteristics

Matrikelnr.	Name	Fachsemester	Note
--	--	<= 5	2.3
--	--	6	2.3
--	--	7-10	2.0
--	--	<= 5	1.7
--	--	<= 5	3.0
--	--	>= 11	1.0
--	--	>= 11	1.3
--	--	<=5	2.3
--	--	7-10	1.7
--	--	6	3.3
--	--	<= 5	1.0
--	--	6	2.0
--	--	>=11	1.3
--	--	7-10	3.3

Lesson 09: Datenschutz 2 – Privacy Engineering

Neun Prinzipien des Datenschutzrechts

9+1 Prinzipien – „Privacy by Design" / „Privacy Engineering"

Technische Umsetzung von Datensparsamkeit

Wieviel Security ist angemessen? Experimente

Technische Ansätze für Transparenz und Zweckbindung

Recap: „by Design & by Default“

Rechtmäßigkeit (incl. Einwilligung)

Zweckbindung

Datensparsamkeit (incl. Erforderlichkeit)

Transparenz

Richtigkeit

Sicherheit

Verantwortlichkeit

Kontrolle und Durchsetzung

Datenportabilität

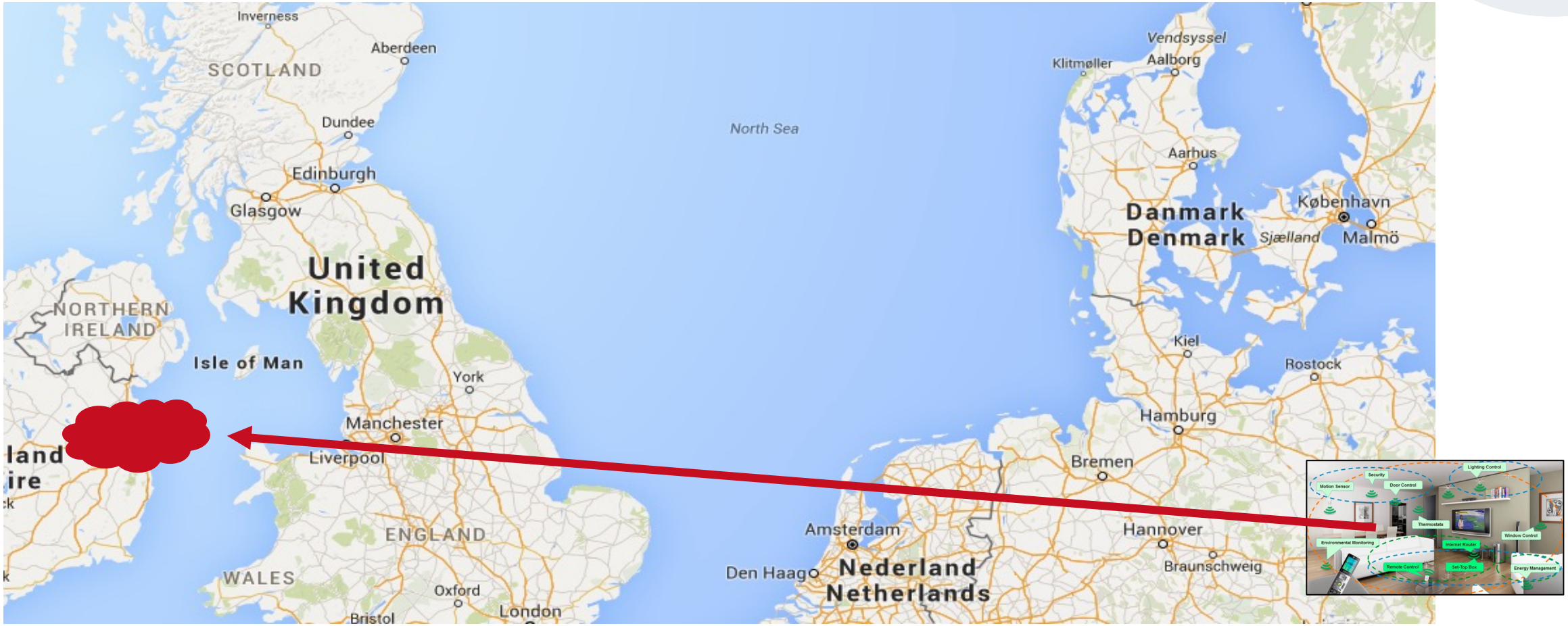
→ Datenschutz „by Design & by Default“

→ ALLE (!) Prinzipien

→ Angemessenheit von Kosten / Aufwand

→ „Privacy Engineering“

Sicherheit: Beispiel



Sicherheit: Beispiel

„Data in Transit Encryption!“

„Fully Homomorphic Encryption!“

Sicherheit: Beispiel

„Data in Transit Encryption!“

„Fully Homomorphic Encryption!“

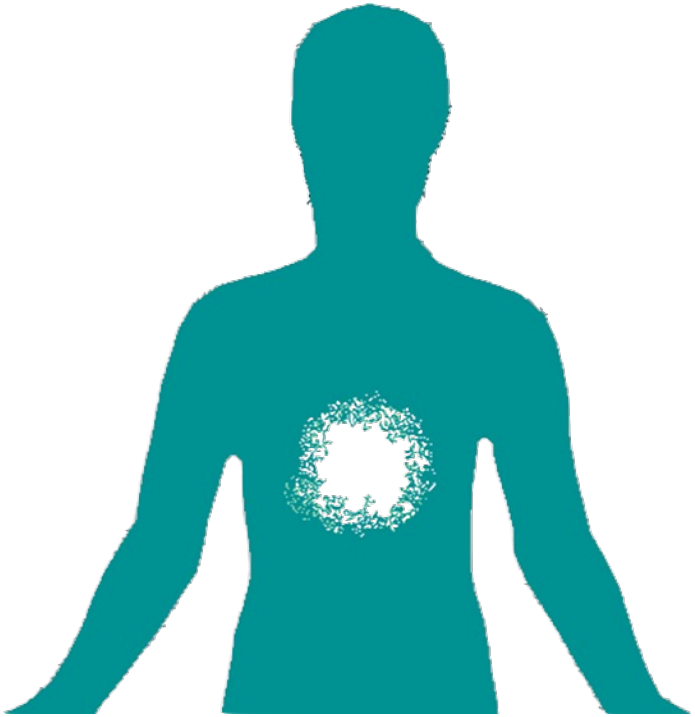
→ Welche? In welcher Konfiguration?

Aufwandsangemessenheit im rechtlichen Diskurs

„Es kommt drauf an…“

„Lässt sich immer nur am Einzelfall entscheiden…“

…



Recap: „by Design & by Default“

Rechtmäßigkeit (incl. Einwilligung)

Zweckbindung

Datensparsamkeit (incl. Erforderlichkeit)

Transparenz

Richtigkeit

Sicherheit

Verantwortlichkeit

Kontrolle und Durchsetzung

Datenportabilität

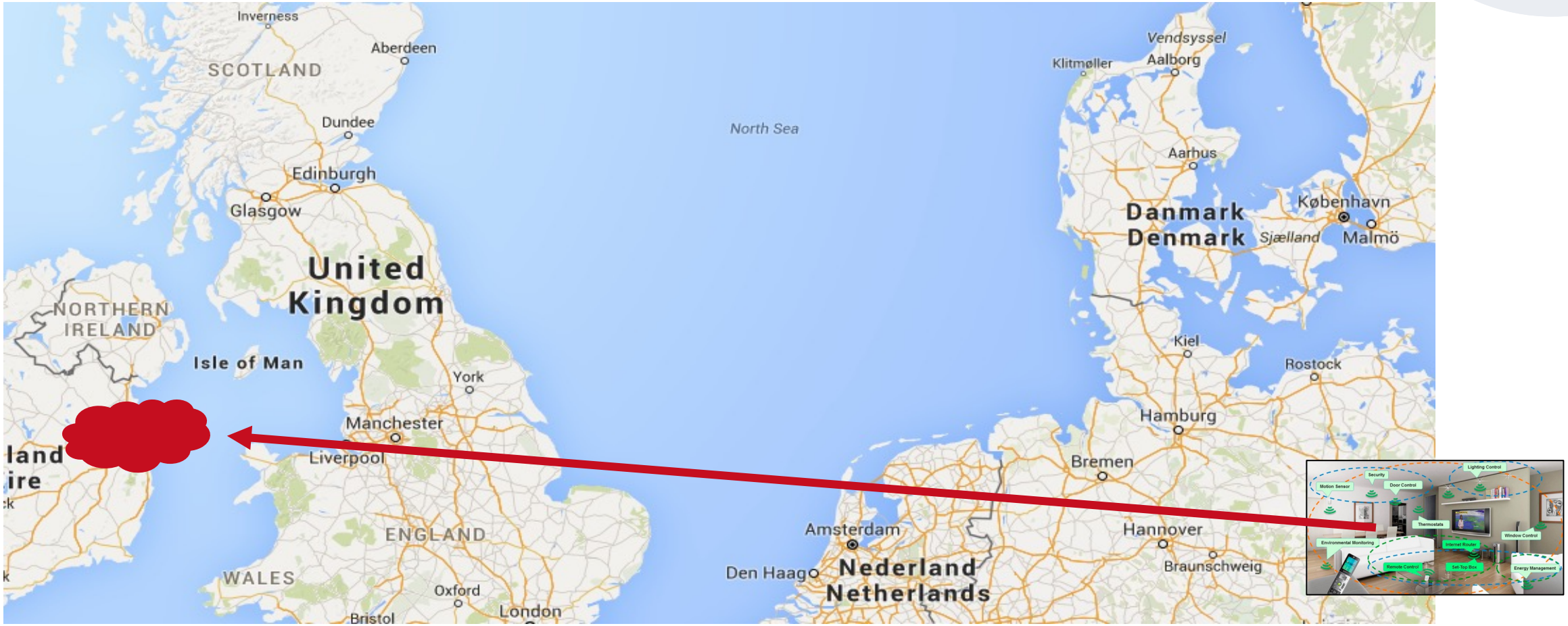
→ Datenschutz „by Design & by Default“

→ ALLE (!) Prinzipien

→ Angemessenheit von Kosten / Aufwand

→ „Privacy Engineering“

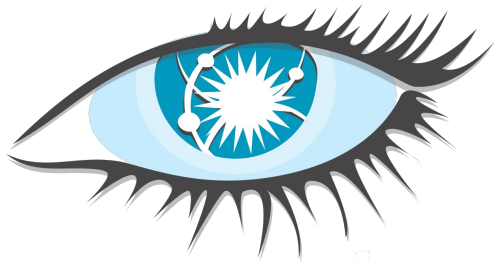
Beispiel: Aufwandsangemessenheit bei Datentransfers

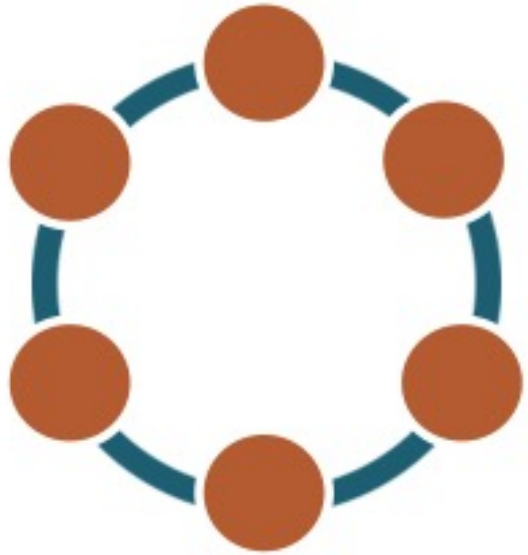


Beispiel: Cassandra TLS Performance

Generell:

- Verteilte, hochskalierbare Datenbank
- Verwendet von Apple (10PB), Netflix (420TB), ebay (250TB), ...
- Typische Anwendung: Selbst verwaltetes Deployment auf x Cloud-Instanzen

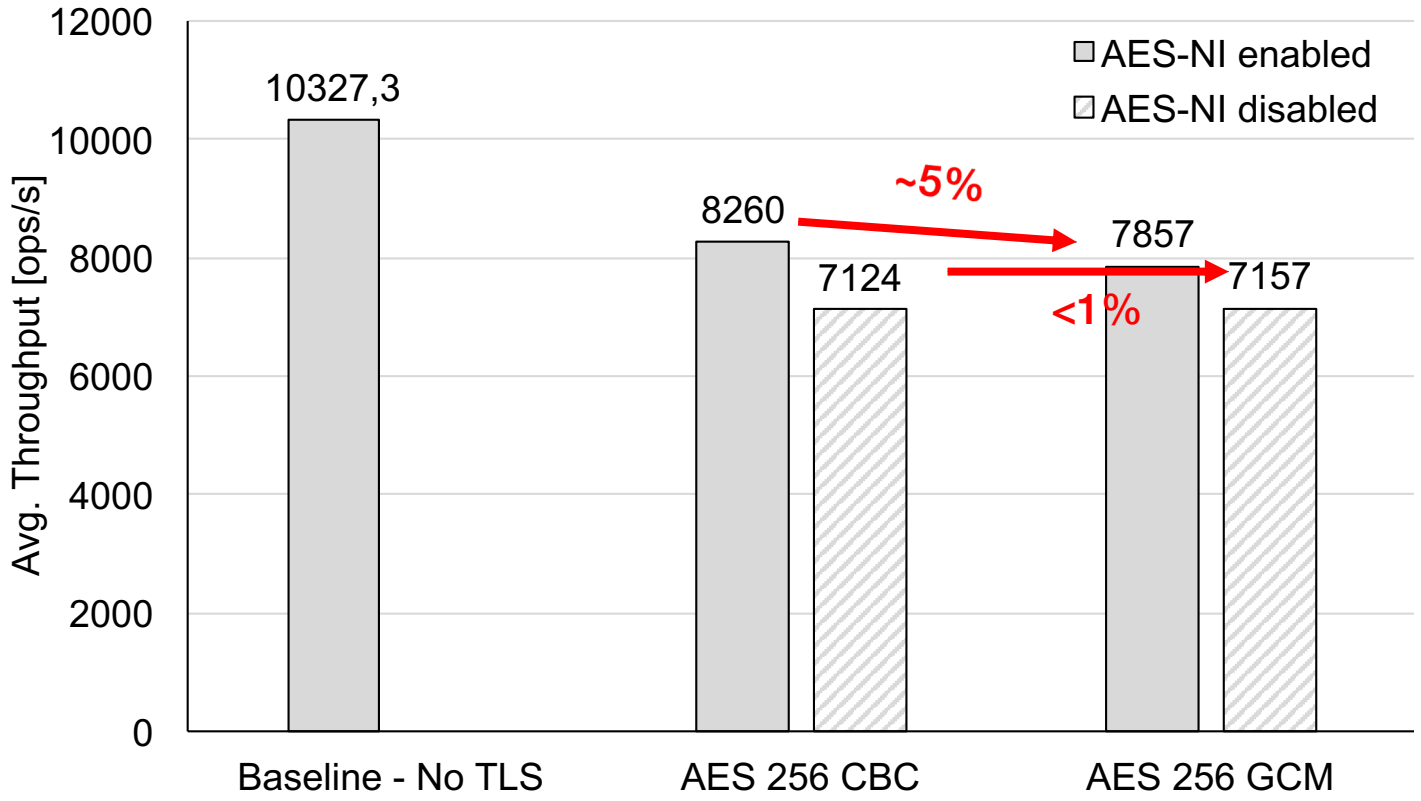




Data in transit encryption:

- TLS
- Alle üblichen Konfigurationsoptionen (Schlüssellänge, Betriebsmodus, HW-support, ...)

Cassandra TLS Optionen, 3 Nodes, 50/50 load



→ Wenn AES-NI verfügbar, höhere Sicherheit durch GCM „die 5% wert“?

→ Wenn AES-NI nicht verfügbar, kein Grund, GCM nicht zu nutzen.

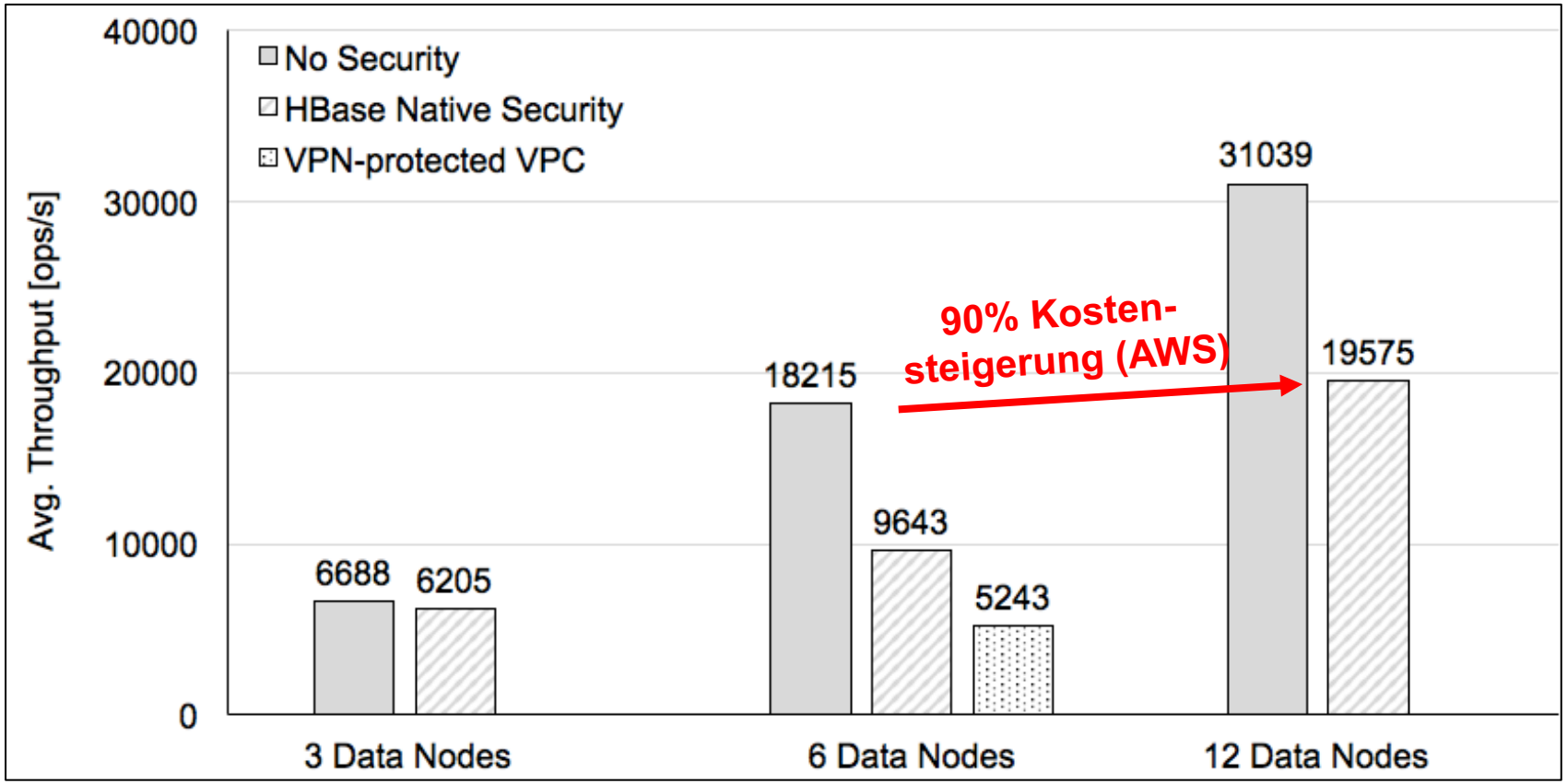
Beispiel: HBase native encryption

Kernkomponente im Hadoop BigData Ökosystem

Native encryption – An/Aus auf 2 “Layern”

Zu erwartender Performance Impact laut Doku: ~10%

Beispiel: HBase native encryption



Pick Your Choice in HBase: Security or Performance

Frank Pallas, Johannes Günther, David Bermbach
Information Systems Engineering Research Group
TU Berlin
Berlin, Germany
Email: {fp,jg,db}@ise.tu-berlin.de

***Abstract*—When analyzing sensitive data in a cloud-deployed Hadoop stack, data-in-transit security needs to be enabled, especially in the underlying storage tier. This, however, will affect the performance of the system and may partially offset the cost benefits of the cloud.**

In this paper, we discuss two strategies for securing HBase deployments in the cloud. For both, we present benchmarking results which show performance impacts that significantly exceed the suggested 10% from the official documentation. These results demonstrate (i) that security configurations should follow a rational decision process based on benchmarking results and (ii) that the security architecture of HBase/HDFS should be redesigned with an emphasis on performance.

***Keywords*-Benchmarking; HBase; Performance; Security**

I. Introduction

Big data technology can provide key business insights to enterprises of all sizes. However, until a few years ago, this was only available to big corporations that could afford the necessary infrastructure for a data warehouse or big data cluster in their local data center.

For smaller players, (public) cloud-based deployments can serve as low risk door openers to the world of big data. However, even big businesses can benefit from cloud-based deployments: big data use cases which need compute resources periodically (e.g., a bank running complex risk analyses once a month) or for variably-sized data sets (e.g., a retail company analyzing sales details right before Christmas and a week later) are a natural fit for cloud deployments that offer affordable, pay-as-you-go resources with instant scalability and high availability wherever and whenever needed.

At the same time, though, cloud-based deployments require additional security precautions due to data privacy regulations or simply for protecting core business interests. However, the performance impact of securing the big data cluster may offset the original cloud benefits – an aspect that has not been focused on in big data research yet. For instance, the official Apache HBase documentation only mentions an estimated performance impact of approximately 10% arising from the activation of built-in native security mechanisms [1, Section 58.3] which, as we will later see, is usually incorrect and also considers HBase native security as the only option. At the same time, cloud computing research, e.g., [2], has found interesting effects when enabling transport security – comprising a broad spectrum of both negative *and positive* performance impacts depending on the specific configuration.

In this paper, we focus on HBase, – the NoSQL system underlying the Hadoop ecosystem – and the performance impact of enabling security features therein, specifically of enabling transport encryption to ensure confidentiality of data in transit. We explicitly consider the big data engines on top of HBase, e.g., Hadoop or Spark, beyond the scope of this paper. We, therefore, present the following contributions:

- A thorough discussion of different strategies for ensuring data confidentiality for cloud-based HBase deployments
- A comprehensive experimental evaluation of the performance impact incurred by following a specific strategy
- A discussion of lessons learned and recommendations for big data deployments in public clouds

This paper is structured as follows: Based on a realistic application scenario and some foundations on cloud security and the HBase architecture provided in section II, we present two possible approaches for securing data in transit to, from, and within an HBase cluster deployed in a public cloud in section III. Our experiments for determining the performance impact of these approaches and the respective results are presented in section IV and discussed in section V. We close with related work (section VI) and a conclusion (section VII).

II. Foundations

In this section, we will give an overview of the application scenario behind our work and resulting security requirements in cloud environments, before presenting a brief introduction to HBase and its relevant communication channels.

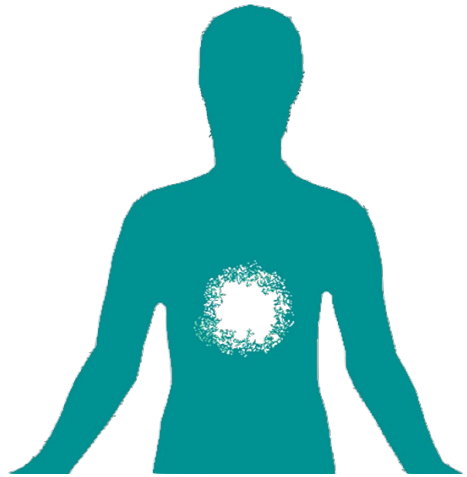
A. Application Scenario

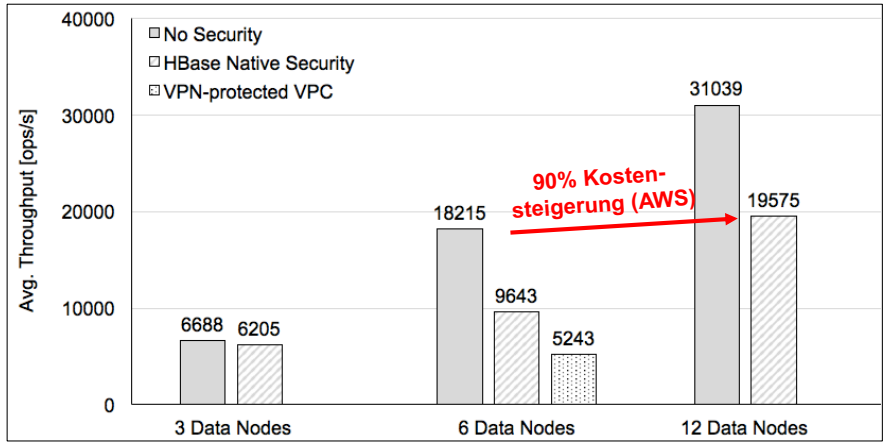
For our considerations, we assume the scenario of a mid-sized DIY store chain that already operates a comparably small centralized data center that serves both the ERP system as well as data analysis tasks. As storage backend, the store uses HBase. Due to recent growth in business, the ERP is affected by increased latency leading to delays at checkout and thus to decreased customer satisfaction. Furthermore, the insufficient capacity of the storage backend also impairs business intelligence through significantly longer data analysis runtimes.

Pallas, F., Bermbach, D., Müller, S., & Tai, S. (2017, April). Evidence-based security configurations for cloud datastores. In Proceedings of the Symposium on Applied Computing (pp. 424-430). ACM.

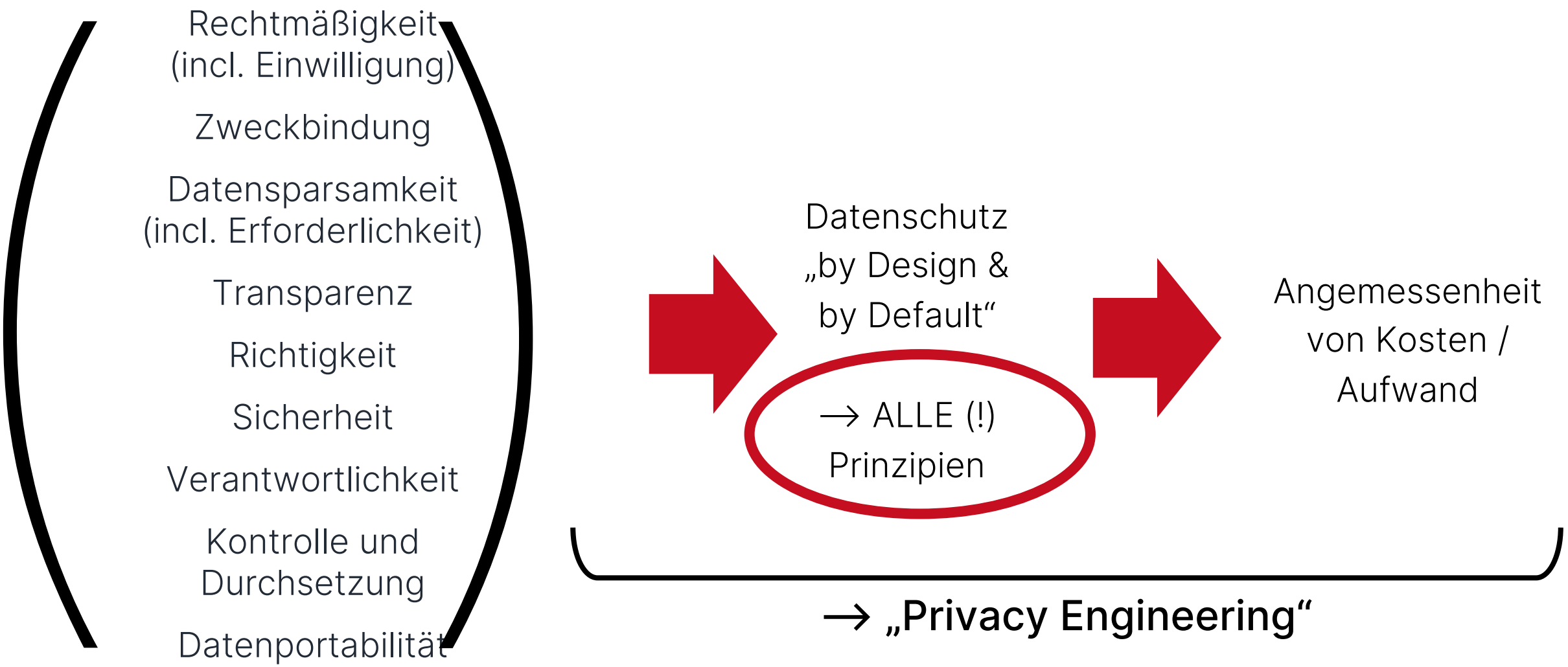
Privacy Engineering: Aufwandsangemessenheit

- Der Aufwand und die Kosten, die ein technischer Mechanismus zur Umsetzung eines Datenschutz-prinzips nach sich zieht, sind entscheidend für die rechtliche Verpflichtung zu dessen Anwendung
- Gleichzeitig sind die tatsächlichen Kosten (unterschiedlicher Optionen) im technisch-rechtlichen Diskurs oftmals nicht ausreichend bekannt
- Rationalität und Nachvollziehbarkeit entsprechender Abwägungen sind daher eingeschränkt
- Hilfreicher Ansatz: Experimentelle Bestimmung tatsächlicher Overheads im konkreten Fall





„by Design & by Default“



Lesson 09: Datenschutz 2 – Privacy Engineering

Neun Prinzipien des Datenschutzrechts

9+1 Prinzipien – „Privacy by Design" / „Privacy Engineering"

Technische Umsetzung von Datensparsamkeit

Wieviel Security ist angemessen? Experimente

Technische Ansätze für Transparenz und Zweckbindung

Technische Datenschutz-Mechanismen existieren (insb. in der wissenschaftlichen Diskussion) vor allem für die Prinzipien Sicherheit und Datensparsamkeit („Anonymisierung“ etc.)

Tatsächlich sind aber technische Ansätze **für alle Prinzipien** notwendig

→ „Privacy Engineering Beyond Anonymization and Security“

9 Kernprinzipien des Datenschutzes

Rechtmäßigkeit (incl. Einwilligung)

Zweckbindung

Datensparsamkeit (incl. Erforderlichkeit)

Transparenz

Richtigkeit

Sicherheit (insb. Vertraulichkeit, Integrität, Verfügbarkeit)

Verantwortlichkeit (incl. nachweisbare Rechtskonformität)

Kontrolle und Durchsetzung

Datenportabilität

GDPR: „Transparenz“

Werden personenbezogene Daten bei der betroffenen Person erhoben, so teilt der Verantwortliche der betroffenen Person zum Zeitpunkt der Erhebung dieser Daten Folgendes mit:

a) den Namen und die Kontaktdaten des Verantwortlichen [...]

b) gegebenenfalls die Kontaktdaten des Datenschutzbeauftragten;

c) die Zwecke, für die die personenbezogenen Daten verarbeitet werden sollen, sowie die Rechtsgrundlage für die Verarbeitung;

d) [...]

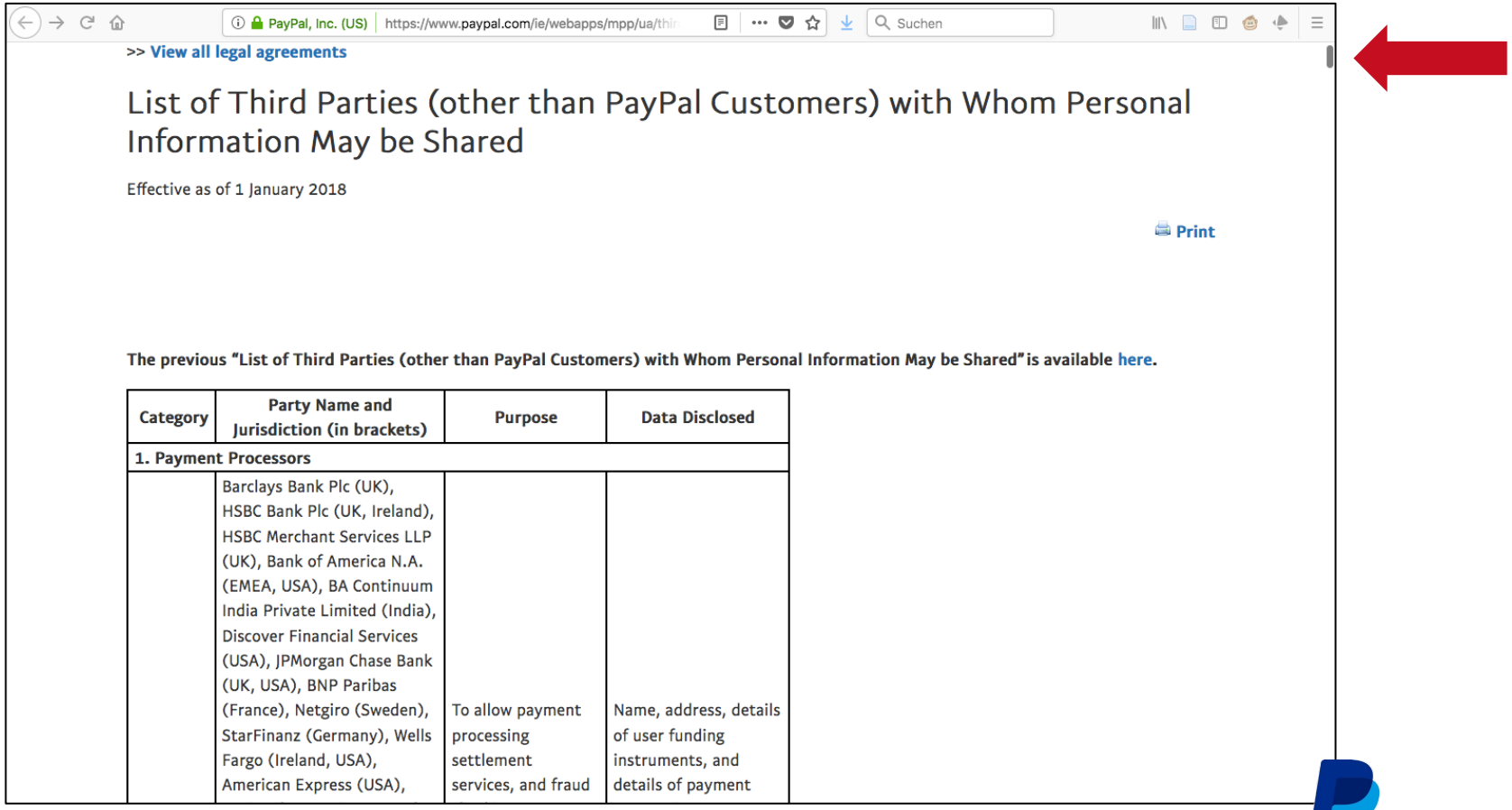
e) gegebenenfalls die Empfänger oder Kategorien von Empfängern der personenbezogenen Daten und

f) gegebenenfalls die Absicht des Verantwortlichen, die personenbezogenen Daten an ein Drittland oder eine internationale Organisation zu übermitteln [...]

[...]

Art. 13 GDPR

GDPR: „Transparenz“



PayPal, Inc. (US) https://www.paypal.com/ie/webapps/mpp/ua/thir

Suchen

>> View all legal agreements

List of Third Parties (other than PayPal Customers) with Whom Personal Information May be Shared

Effective as of 1 January 2018

Print

The previous "List of Third Parties (other than PayPal Customers) with Whom Personal Information May be Shared" is available here.

Category	Party Name and Jurisdiction (in brackets)	Purpose	Data Disclosed
1. Payment Processors			
	Barclays Bank Plc (UK), HSBC Bank Plc (UK, Ireland), HSBC Merchant Services LLP (UK), Bank of America N.A. (EMEA, USA), BA Continuum India Private Limited (India), Discover Financial Services (USA), JPMorgan Chase Bank (UK, USA), BNP Paribas (France), Netgiro (Sweden), StarFinanz (Germany), Wells Fargo (Ireland, USA), American Express (USA),	To allow payment processing settlement services, and fraud	Name, address, details of user funding instruments, and details of payment

PayPal

GDPR: „Transparenz“

Opinion | THE PRIVACY PROJECT

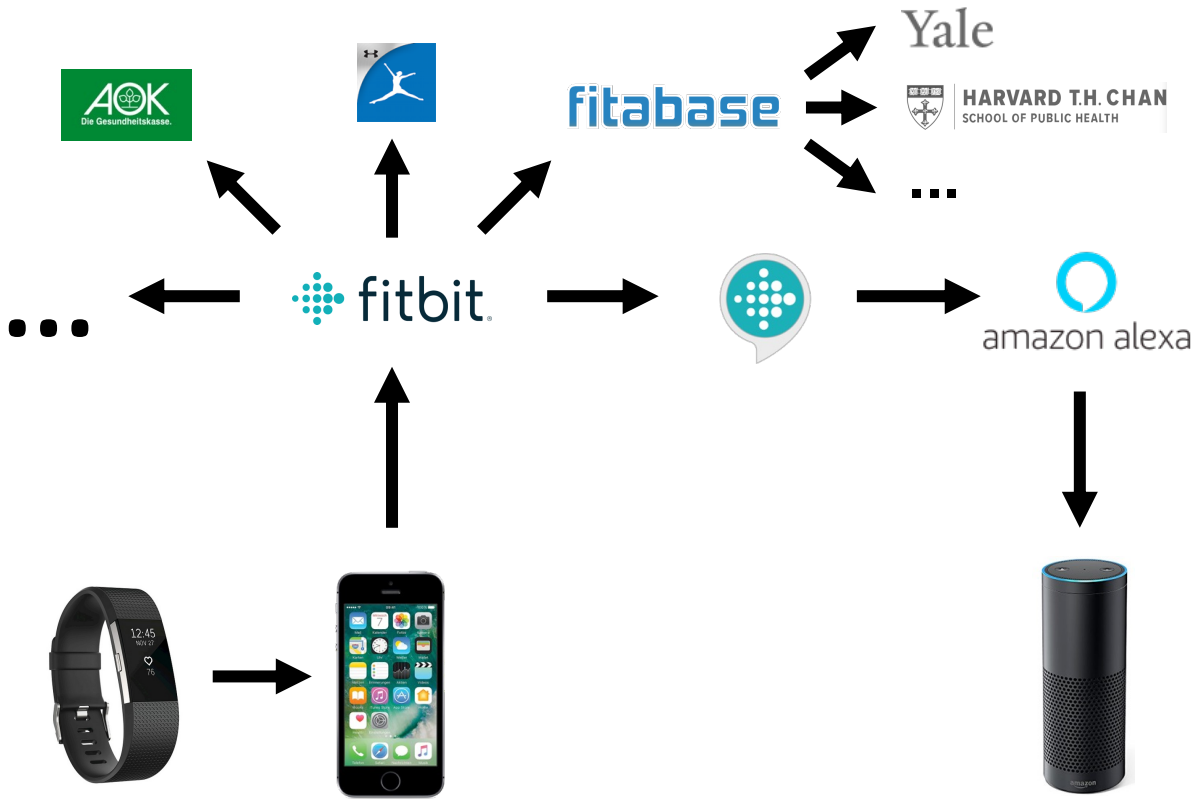
We Read 150 Privacy Policies. They Were an Incomprehensible Disaster.

By Kevin Litman-Navarro

In the background here are several privacy policies from major tech and media platforms. Like most privacy policies, they're verbose and full of legal jargon — and opaquely establish companies' justifications for collecting and selling your data.

The data market has become the engine of the internet, and these privacy policies we agree to but don't fully understand help fuel it.

Transparenz in aktuellen Szenarien



Wer bekommt welche Daten auf welcher Rechtsgrundlage?

Auch hier:
Aufwände für Individuum zu hoch (und Darstellung zu kompliziert/unverständlich), als dass Informationen in der Realität tatsächlich rezipiert würden

→ Technische Repräsentation von Transparenzinformationen?

GDPR: „Transparenz“

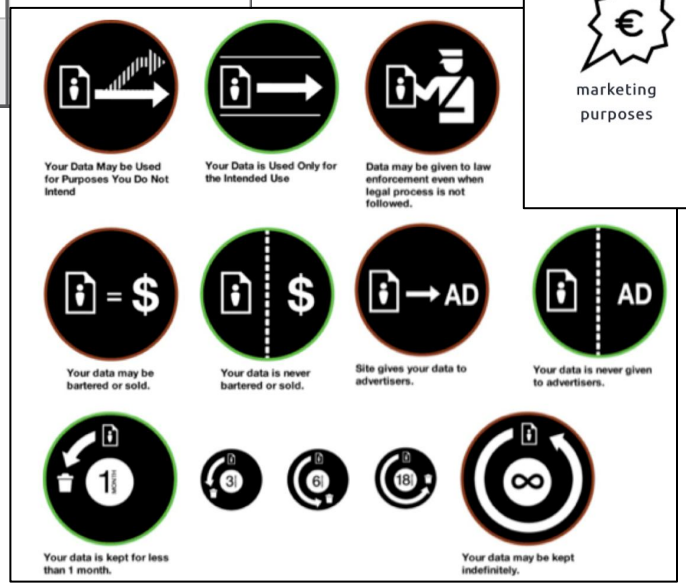
Transparent information, communication and modalities for the exercise of the rights of the data subject

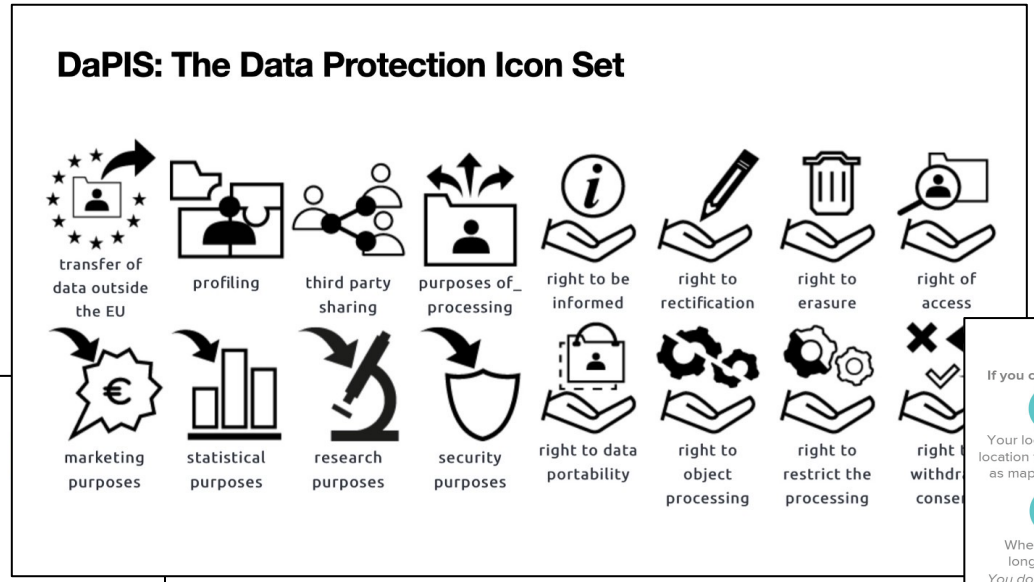
1. The controller shall take appropriate measures to provide any information referred to in Articles 13 and 14 and any communication under Articles 15 to 22 and 34 relating to processing to the data subject in a concise, transparent, intelligible and easily accessible form, using clear and plain language, in particular for any information addressed specifically to a child. The information shall be provided in writing, or by other means, including, where appropriate, by electronic means. When requested by the data subject, the information may be provided orally, provided that the identity of the data subject is proven by other means.

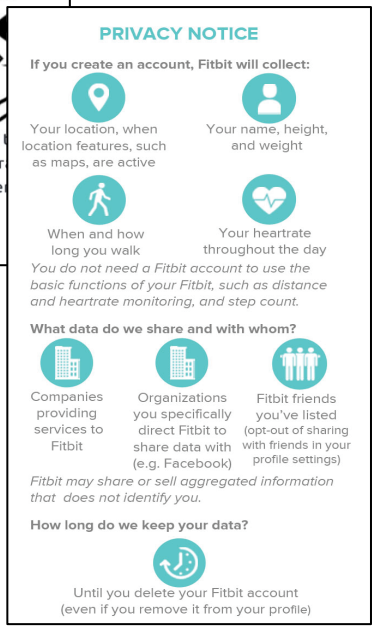
7. The information to be provided to data subjects pursuant to Articles 13 and 14 may be provided in combination with standardised icons in order to give in an easily visible, intelligible and clearly legible manner a meaningful overview of the intended processing. Where the icons are presented electronically they shall be machine-readable.

GDPR: Transparenz und Icons

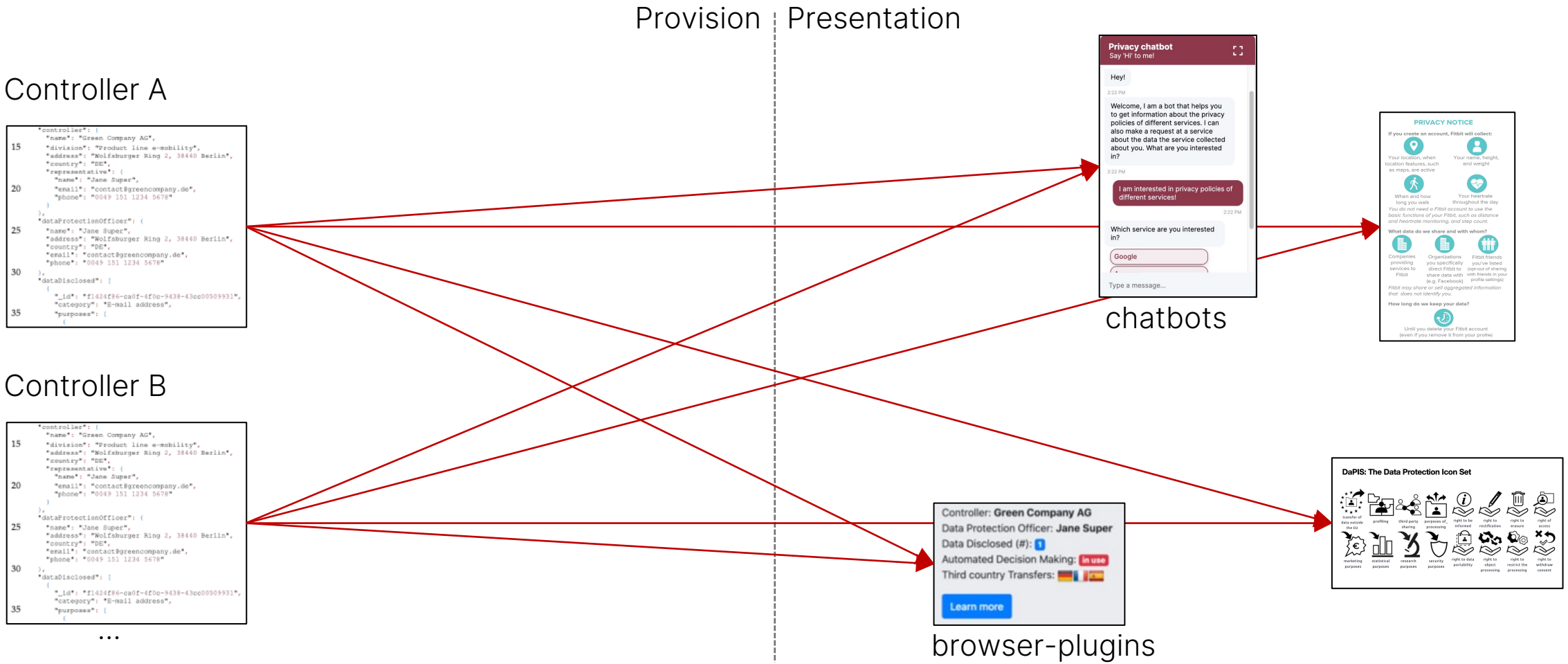
TYPE OF DATA COLLECTED	GENERAL DATA PRACTICES	DATA SHARING
contact: name, mailing address, email, or phone number	**ad customization:** user data may be used for the purpose of customizing advertising	**affiliates:** affiliates and subsidiaries bound by the same privacy practices
computer: IP address, browser type, or operating system	**third party tracking:** site allows third parties to place advertisements that may track user behavior	**contractors:** third party contractors bound by the same privacy practices
interactive: browsing behavior or search history	**public display:** service allows users to contribute information which may be displayed publicly	**third parties:** third parties not subject to same data practices
financial: account status or activity, credit information, or purchase history	**user control:** users allowed to access and correct personal data collected	
content: contents of personal communications, stored documents or media	**data retention:** explicitly stated duration of retention for personal data collected	







Transparenz: Trennen von Bereitstellung und Darstellung (Provision vs. Presentation)



Transparenz: Bereitstellung in maschinenlesbarer Form

Provision | Presentation

Controller A

Controller B

...

Transparenz: Bereitstellung in maschinenlesbarer Form

Transparency and modalities

Article 12

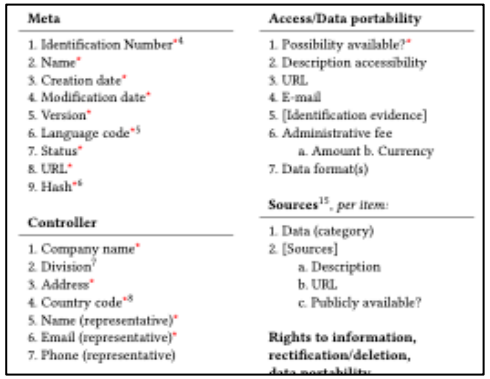
Transparent information, communication and modalities for the exercise of the rights of the data subject

1. The controller shall take appropriate measures to provide any information referred to in Articles 13 and 14 and any communication under Articles 15 to 22 and 34 relating to processing to the data subject in a concise, transparent, intelligible and easily accessible form, using clear and plain language, in particular for any information addressed specifically to a child. The information shall be provided in writing, or by other means, including, where appropriate, by electronic means. When requested by the data subject, the information may be provided orally, provided that the identity of the data subject is proven by other means.

Rechtliche Anforderungen (e.g., GDPR)

Reference(s)				Transparency information
13 (1a)	14 (1a)		30 (1a)	Controller
13 (1b)	14 (1b)		30 (1a)	Data protection officer
13 (1c)	14 (1c)	15 (1a)	30 (1b)	Purposes
13 (1c)	14 (1c)			Legal basis
13 (1d)	14 (2b)			Legitimate interests
13 (1e)	14 (1e)	15 (1c)	30 (1d)	Recipient (categories)
13 (1f)	14 (1f)	15 (1c)	30 (1e)	Third country transfer
13 (1f)	14 (1f)	15 (2)	30 (1e)	Adequacy (third country)
13 (1f)	14 (1f)	15 (2)	30 (1e)	Access and Data portability
13 (2a)	14 (2a)	15 (1d)	30 (1f)	Retention or storage criteria

Kategorisierung



Meta

1. Identification Number*[4]
2. Name*
3. Creation date*
4. Modification date*
5. Version*
6. Language code*[5]
7. Status*
8. URL*
9. Hash*[6]

Controller

1. Company name*
2. Division?
3. Address*
4. Country code*[8]
5. Name (representative)*
6. Email (representative)*
7. Phone (representative)

Access/Data portability

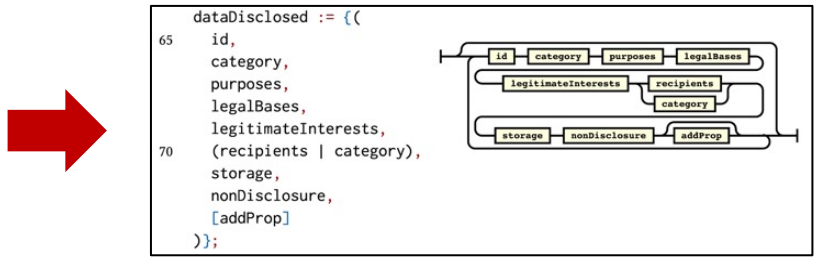
1. Possibility available?*
2. Description accessibility
3. URL
4. E-mail
5. [Identification evidence]
6. Administrative fee
   a. Amount b. Currency
7. Data format(s)

Sources[15], *per item:*

1. Data (category)
2. [Sources]
   a. Description
   b. URL
   c. Publicly available?

Rights to information, rectification/deletion,

Detaillierte Analyse benötigter Ausdrucksmächtigkeit



```
dataDisclosed := {(
  id,
  category,
  purposes,
  legalBases,
  legitimateInterests,
  (recipients | category),
  storage,
  nonDisclosure,
  [addProp]
)};
```


Formale Sprachspezifikation

```
"controller": {
  "name": "Green Company AG",
  "division": "Product line e-mobility",
  "address": "Wolfsburger Ring 2, 38440 Berlin",
  "country": "DE",
  "representative": {
    "name": "Jane Super",
    "email": "contact@greencompany.de",
    "phone": "0049 151 1234 5678"
  }
},
"dataProtectionOfficer": {
  "name": "Jane Super",
  "address": "Wolfsburger Ring 2, 38440 Berlin",
  "country": "DE",
  "email": "contact@greencompany.de",
  "phone": "0049 151 1234 5678"
},
"dataDisclosed": [
  {
    "_id": "f1424f86-ca0f-4f0c-9438-43cc00509931",
    "category": "E-mail address",
    "purposes": [
      {
```

Rechtlich abgeglichene, maschinenlesbare Repräsentation

Technische Repräsentation von Transparenzinformationen im Einklang mit GDPR-Anforderungen

TILT: A GDPR-Aligned Transparency Information Language and Toolkit for Practical Privacy Engineering

Anonymous Author(s)

ABSTRACT

In this paper, we present TILT, a transparency information language and toolkit explicitly designed to represent and process transparency information in line with the requirements of the GDPR and allowing for a more automated and adaptive use of such information than established, legalese data protection policies do.

We provide a detailed analysis of transparency obligations from the GDPR to identify the expressiveness required for a formal transparency language intended to meet respective legal requirements. In addition, we identify a set of further, non-functional requirements that need to be met to foster practical adoption in real-world (web) information systems engineering. On this basis, we specify our formal language and present a respective, fully implemented toolkit around it. We then evaluate the practical applicability of our language and toolkit and demonstrate the additional prospects it unlocks through two different use cases: *a)* the inter-organizational analysis of personal data-related practices allowing, for instance, to uncover data sharing networks based on explicitly announced transparency information and *b)* the presentation of formally represented transparency information to users through novel, more comprehensible, and potentially adaptive user interfaces, heightening data subjects' actual informedness about data-related practices and, thus, their sovereignty.

Altogether, our transparency information language and toolkit allow – differently from previous work – to express transparency information in line with actual legal requirements and practices of modern (web) information systems engineering and thereby pave the way for a multitude of novel possibilities to heighten transparency and user sovereignty in practice.

CCS CONCEPTS

• **Applied computing** → **Law**; • **Information systems** → **Information systems applications**; **Web data description languages**; • **Software and its engineering** → **Formal language definitions**; **Context specific languages**; • **Security and privacy** → **Privacy protections**;

KEYWORDS

Data transparency, GDPR, data protection, privacy by design, legal informatics, privacy engineering

FAccT, 2021, Toronto, Canada

ACM ISBN 978-x-xxxx-xxxx-x/x/xx…$xx.00
https://doi.org/xx.xxx/xxx_x

ACM Reference Format:
Anonymous Author(s). 2021. TILT: tion Language and Toolkit for Prac *of ACM FAccT Conference (FAccT* 11 pages. https://doi.org/xx.xxx/x

1 INTRODUCTION

Transparency has been a core technical deliberations aroun larly applies to data privacy in in a sovereign and self-determi choices, individuals need to h actual facts and givens regard including, e.g., which party co purposes, how long this data i protection laws and regulatio European General Data Prote California Consumer Privacy A include transparency rules fo them to reveal respective info

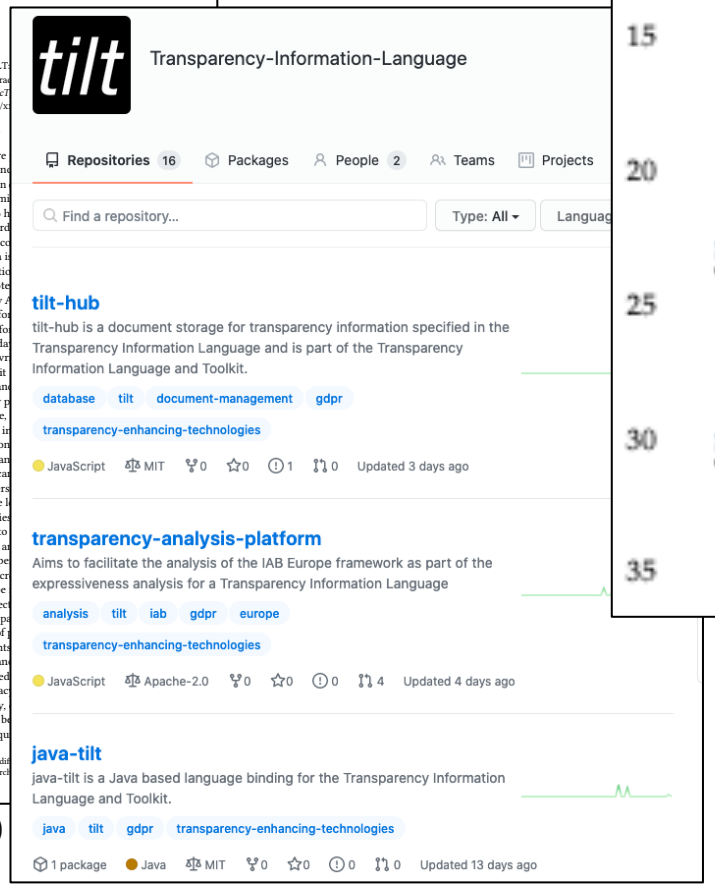
These obligations are toda tive transparency parts of wr policies do, however, exhibit limit their actual reception an subjects: First of all, privacy p written in legalese language, locate transparency-related i them correctly [24, 33]. Secon different logical structures an statements, causing significa every new policy to be unders discriminates people who are l transfers towards other partie tice. Data subjects wanting to texts therefore have to read a establish the logical interdepe

Together with the ever-incr thus, of privacy policies to be vacy policies, including respec read anymore before using a pa a certain collection and use of tions, transparency statements self-serving formal complian tering data subjects' informed conduct with regard to privac

For other areas of privacy, edly been demonstrated to be and administrative effort requ

[1]Being well aware of the slightly dif Protection", we use these terms interch

Grünewald und Pallas (2021)



tilt Transparency-Information-Language

Repositories 16 | Packages | People 2 | Teams | Projects

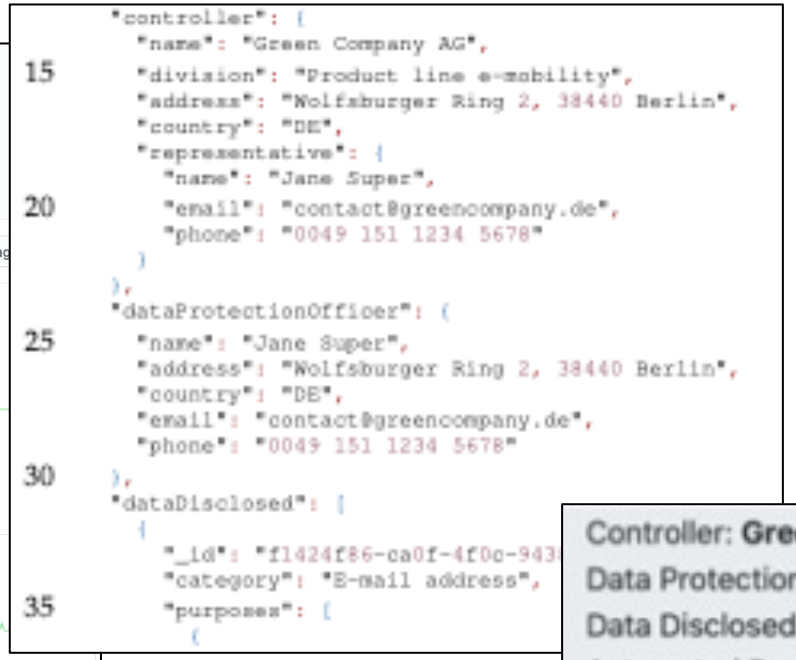
Find a repository... | Type: All | Languag

tilt-hub
tilt-hub is a document storage for transparency information specified in the Transparency Information Language and is part of the Transparency Information Language and Toolkit.
database | tilt | document-management | gdpr | transparency-enhancing-technologies
JavaScript | MIT | 0 | 0 | 1 | 0 | Updated 3 days ago

transparency-analysis-platform
Aims to facilitate the analysis of the IAB Europe framework as part of the expressiveness analysis for a Transparency Information Language
analysis | tilt | iab | gdpr | europe | transparency-enhancing-technologies
JavaScript | Apache-2.0 | 0 | 0 | 0 | 4 | Updated 4 days ago

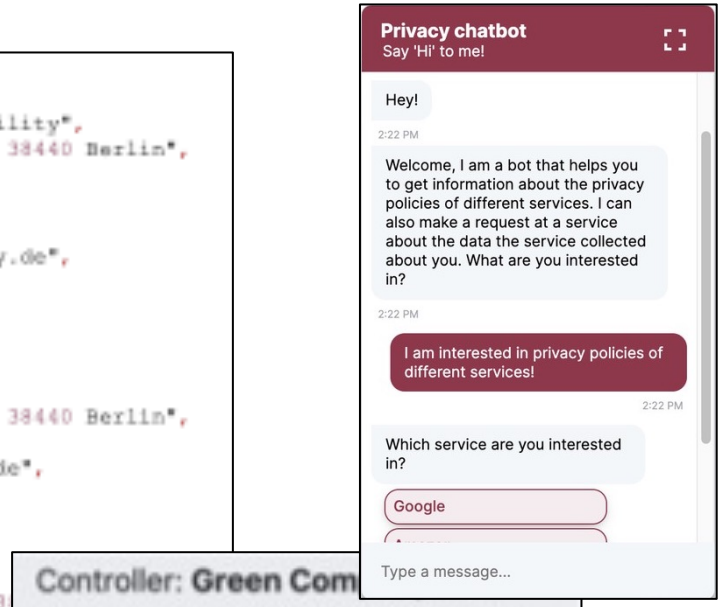
java-tilt
java-tilt is a Java based language binding for the Transparency Information Language and Toolkit.
java | tilt | gdpr | transparency-enhancing-technologies
1 package | Java | MIT | 0 | 0 | 0 | 0 | Updated 13 days ago

https://github.com/Transparency-Information-Language



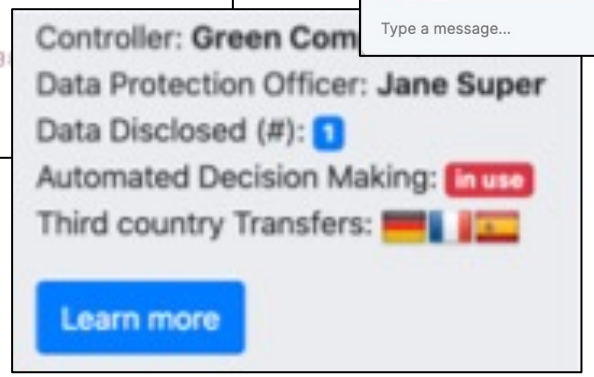
```
"controller": {
  "name": "Green Company AG",
  "division": "Product line e-mobility",
  "address": "Wolfsburger Ring 2, 38440 Berlin",
  "country": "DE",
  "representative": {
    "name": "Jane Super",
    "email": "contact@greencompany.de",
    "phone": "0049 151 1234 5678"
  }
},
"dataProtectionOfficer": {
  "name": "Jane Super",
  "address": "Wolfsburger Ring 2, 38440 Berlin",
  "country": "DE",
  "email": "contact@greencompany.de",
  "phone": "0049 151 1234 5678"
},
"dataDisclosed": [
  {
    "_id": "f1424f86-ca0f-4f0c-943
    "category": "E-mail address",
    "purposes": [
      {
```

Formale, maschinenlesbare Repräsentation



Privacy chatbot
Say 'Hi' to me!

Hey!
2:22 PM
Welcome, I am a bot that helps you to get information about the privacy policies of different services. I can also make a request at a service about the data the service collected about you. What are you interested in?
2:22 PM
I am interested in privacy policies of different services!
2:22 PM
Which service are you interested in?
Google
Type a message...



Controller: **Green Com**
Data Protection Officer: **Jane Super**
Data Disclosed (#): 1
Automated Decision Making: in use
Third country Transfers:
Learn more

Neue Darstellungsmöglichkeiten

9 Kernprinzipien des Datenschutzes

Rechtmäßigkeit (incl. Einwilligung)

Zweckbindung

Datensparsamkeit (incl. Erforderlichkeit)

Transparenz

Richtigkeit

Sicherheit (insb. Vertraulichkeit, Integrität, Verfügbarkeit)

Verantwortlichkeit (incl. nachweisbare Rechtskonformität)

Kontrolle und Durchsetzung

Datenportabilität

Zweckbindung: Relevanz

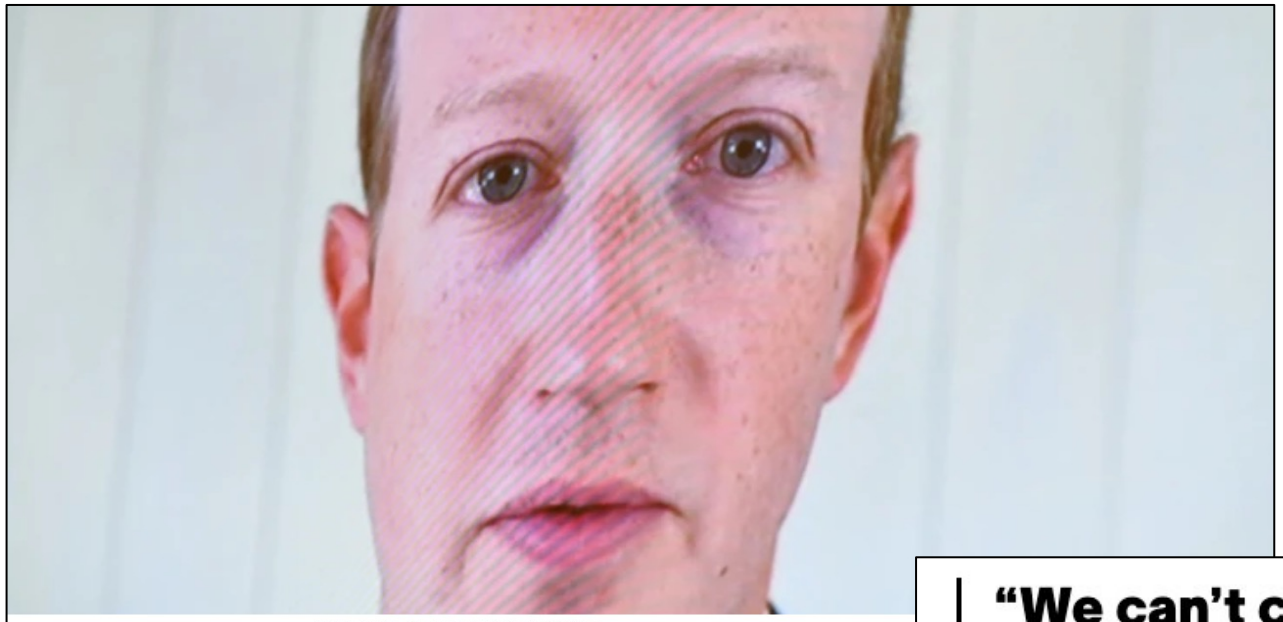


IMAGE: MANDEL NGAN/POOL/AFP VIA GETTY IMAGES

MOTHERBOARD
TECH BY VICE

Facebook Doesn't Know What It Does With Your Data, Or Where It Goes: Leaked Document

https://www.vice.com/en/article/akvmke/facebook-doesnt-know-what-it-does-with-your-data-or-where-it-goes

> **"We can't confidently make controlled policy changes or external commitments such as 'we will not use X data for Y purpose.' And yet, this is exactly what regulators expect us to do"**

Zweckbindung: Technische Umsetzung?

→ z.B. „Purpose-Based Access Control (PBAC)“: Zugriffsberechtigungen auch auf Basis verfolgter Zwecke

Bisher (fast) ausschließlich als Erweiterungen für einzelne DBs (z.B. IBM DB2)
→ „Hippocratic Databases“

Aber: Passt nicht zu aktuellen Entwicklungspraktiken und -modellen
(z.B. DB-agnostic ORMs, PubSub-/Streaming-Systeme, FaaS/CloudNative, ...)

→ „Purpose-Awareness“ für aktuelle System- und Programmierparadigmen und Komponenten (ORM, MQTT, Kafka, FaaS environments, ...)

Zweckbindung: Technische Umsetzung

Towards Application-Layer Purpose-Based Access Control

— Preprint ACM SAC 2020 —

Frank Pallas
TU Berlin
Information Systems Engineering Research Group
Berlin, Germany
fp@ise.tu-berlin.de

Max-R. Ulbricht
TU Berlin
Information Systems Engineering Research Group
Berlin, Germany
mu@ise.tu-berlin.de

Stefan Tai
TU Berlin
Information Systems En
Research Group
Berlin, German
st@ise.tu-berlin.

Thomas Peikert
TU Berlin
thomas.peikert@campus.tu-berlin.de

Marcel Reppenhagen
TU Berlin
marcel.reppenhagen@campus.tu-berlin.de

Daniel Wenze
TU Berlin
daniel.wenzel@campus.tu

Paul Wille
TU Berlin
paul.wille@campus.tu-berlin.de

Karl Wolf
TU Berlin
karl.wolf@campus.tu-berlin.de

Preprint, to appear in: The 35th ACM/SIGAPP Symposium On Applied Computing (ACM SAC 2020), Brno, Czech Republic, March 30-April 3, 2020.
Final version availible at:
https://doi.org/10.1145/3341105.3375764

ABSTRACT

In this paper, we propose an architecturally novel approach to implementing purpose-based access control in practice. Different from previous proposals, our approach resides on the application instead of the data(base) layer. This allows for significantly better integration with established architectures and practices of real-world application engineering and to achieve database independence.

To validate practical applicability, we provide two exemplary implementations and briefly assess the introduced overhead in matters of achievable throughputs. Results significantly depend on data and query type but basically suggest bearable overheads for realistic applications even though possible performance optimizations have not been implemented in our proofs-of-concept yet. Our approach thus proposes significantly better practical feasibility than previous ones and exhibits reasonable overheads. It therefore paves the way for purpose-based access control to be actually adopted in practice.

KEYWORDS

Privacy, data protection, purpose limitation, access control, PBAC, privacy by design, privacy engineering, web engineering

1 INTRODUCTION

Privacy by Design[1] (PbD) is one of the core concepts of modern privacy legislation, aiming at the effective implementation of privacy principles through concrete technologies and their design. For instance, the European General Data Protection Regulation (GDPR) requires data controllers to "implement appropriate technical and organisational measures [..] designed to implement principles [...] in an effective manner and to integrate safeguards into the processing" [10, Art. 25 (1)].

Noteworthily, this obligation refers to "data pr ples" in general. Even though (academic) discussion Design largely revolve around data minimization (pseudonymization, etc.) and security, other prin also need to be addressed properly.

Of the various established privacy principles ar those codified in Art. 5 of the GDPR, we herein co principle of purpose limitation and the possibilitie it technologically, "by Design". Basically, purpose legislative obligation is codified in Art. 5 of the GD

> Personal data shall be [...] collected for spe plicit and legitimate purposes and not fur cessed in a manner that is incompatible w purposes [10, Art. 5 (1 b)]

Other examples vividly illustrating the particu purpose limitation throughout the regulatory regi include the limitation of individually provided co more specific purposes" (Art. 6 (1)), the rigorous pur for processing special categories of personal data l 9, or the obligations to inform data subjects about which data is collected and processed given in Ar track of the purposes that certain pieces of perso used for and ensuring these limitations to actually b is thus a core obligation for any party collecting personal data under the regime of the GDPR.

While the principle of purpose limitation has so implemented through rather administrative proce by policy", according to Spiekermann and Cran proach proves increasingly inappropriate in the ligh foreseeable givens. With omnipresent collection and its use for novel, initially unforeseen use-cases

[1] Being well aware of the slightly different notions between "Privacy" and "Data Protection", we use these terms interchangeably herein.

2021 IEEE 25th International Enterprise Distributed Object Computing Conference (EDOC)

Messaging with Purpose Limitation – Privacy-Compliant Publish-Subscribe Systems

Karl Wolf
Information Systems Engineering
TU Berlin
Berlin, Germany
kw@ise.tu-berlin.de

Frank Pallas
Information Systems Engineering
TU Berlin
Berlin, Germany
fp@ise.tu-berlin.de

Stefan Tai
Information Systems Engineering
TU Berlin
Berlin, Germany
st@ise.tu-berlin.de

***Abstract*—Purpose limitation is an important privacy principle to ensure that personal data may only be used for the declared purposes it was originally collected for. Ensuring compliance with respective privacy regulations like the GDPR, which codify purpose limitation as an obligation, consequently, is a major challenge in real-world enterprise systems. Technical solutions under the umbrella of purpose-based access control (PBAC), however, focus mostly on data being held at-rest in databases, while PBAC for communication and publish-subscribe messaging in particular has received only little attention. In this paper, we argue for PBAC to be also applied to data-in-transit and introduce and study a concrete proof-of-concept implementation, which extends a popular MQTT message broker with purpose limitation. On this basis, purpose limitation as a core privacy principle can be addressed in enterprise IoT and message-driven integration architectures that do not focus on databases but event-driven communication and integration instead.**

***Index Terms*—purpose limitation, publish-subscribe, messaging, GDPR, privacy, privacy engineering**

I. Introduction

Privacy is of key importance to any enterprise. Regulations like the GDPR [1] in the EU specifically prescribe how personal information is (not) to be used by organizations. Among the core principles of regulations like the GDPR is the principle of purpose limitation. This privacy principle requires purposes for the processing of data to be specified and that compliance to declared purposes must be ensured.

Related work on purpose limitation, however, has mostly focused at data-at-rest, that is, data persisted in a database and accessed (for a particular purpose) later on. Respective technical mechanisms are typically subsumed under the term "purpose-based access control (PBAC)". Concrete implementations proposed include low-level database extensions often referred to as "Hippocratic Databases" [2]–[4] as well as higher-level approaches integrating PBAC into established programming abstractions for database access such as object-relational-mappers (ORMs) [5].

Surprisingly, little attention has been paid on purpose limitation and PBAC for data-in-transit, that is, data traveling through the network by means of communication middleware. With event-driven architectures and stream-based processing, data may not necessarily be persisted anymore, but runs through processing pipelines in the form of messages. Especially in enterprise IoT settings, with large-scale integration of multiple data sources and use cases involving real-time monitoring, such message-oriented processing patterns are increasingly prevalent.

To close this gap, we herein propose PBAC for data being in-transit in message-oriented architectures. In this regard, we aim to answer the following research questions:

- How can the privacy principle of purpose limitation be technically implemented in real-world message-oriented architectures?
- What design options need to be considered in doing so?
- What is the performance impact to be expected from the different design options?

To address these questions and the aforementioned gap, we introduce and study a concrete implementation, HivePBAC, an extension to one of the most widely used MQTT message brokers – HiveMQ. Our solution brings purpose-awareness to the publication and consumption of personal data in use cases following the publish-subscribe pattern. As such, our technical solution complements any administrative and organizational privacy measures common in practice.

In particular, our contributions are:

- An in-depth analysis of requirements and prerequisites to be taken into account when introducing purpose-awareness to publish-subscribe systems,
- A model and design to introduce purpose-related functionalities – particularly allowing the publication of data to be bound to sets of hierarchically structured allowed (AIP) and prohibited intended purposes (PIP) and subscriptions to be made for explicitly specified access purposes (APs) – to existing MQTT brokers while still providing backward compatibility for clients without dedicated purpose-related functionality,
- A proof-of-concept implementation for a purpose-aware MQTT broker, enforcing said purpose restrictions against APs at different filtering points in time (on publish, on subscribe, etc.),
- An easy-to use Python client library allowing to integrate respective purpose-related capabilities into sending and receiving clients with low effort, and

978-1-6654-3579-6/21/$31.00 ©2021 IEEE
DOI 10.1109/EDOC52215.2021.00027

Technische Universität Berlin
Faculty IV – Electrical Engineering and Computer Science
Information Systems Engineering Chair
Prof. Dr.-Ing. Stefan Tai

Thesis
in pursuit of the degree of Master of Science
at Technische Universität Berlin

In-Transit Purpose-Based Access Control for Scalable Content-Based Publish/Subscribe Systems

Submitted by

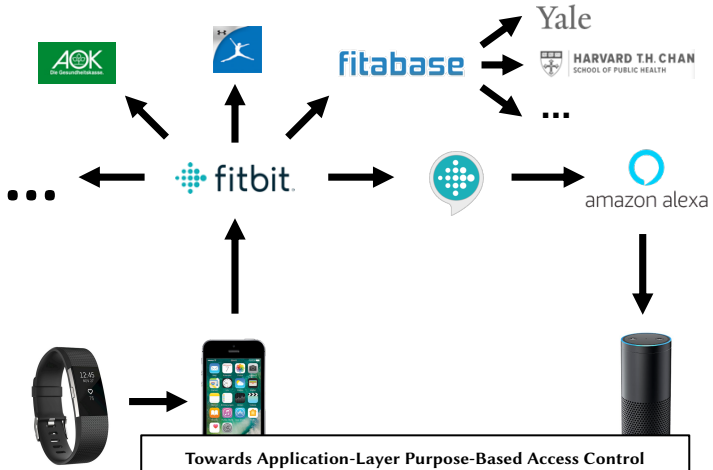
Examiners: Prof. Dr.-Ing. Stefan Tai, TU Berlin
Prof. Dr.-Ing. David Bermbach, TU Berlin

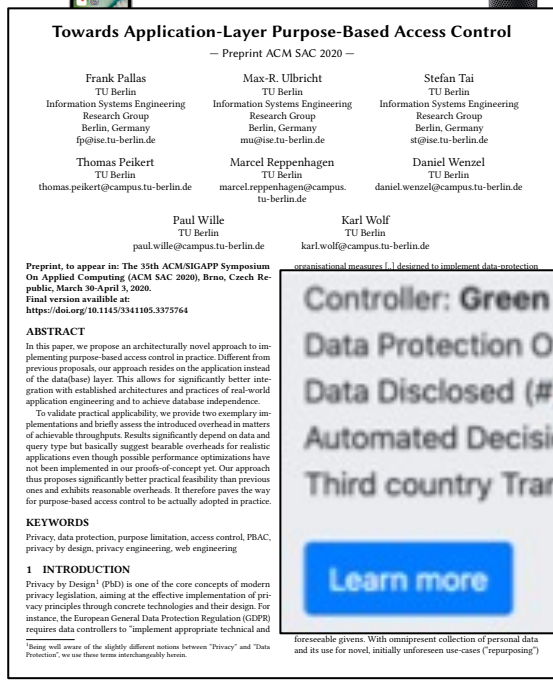
Advisors: Dr.-Ing. Frank Pallas, TU Berlin
M. Sc. Karl Wolf, TU Berlin

2022-09-20

Privacy Engineering: Jenseits von Sicherheit und Minimierung

- Technische Mechanismen für Privacy / Data Protection by Design derzeit vor allem für Datensparsamkeit / -minimierung und Sicherheit
- Tatsächlich sind aber technische Ansätze für **alle Prinzipien** notwendig
- Auch Prinzipien wie Transparenz oder Zweckbindung lassen sich technisch abbilden
- Wichtig dabei: **Tatsächliche** rechtliche Anforderungen und **Implementierungsaufwand** in aktuellen Technologiestacks etc. von Beginn an mitdenken!





Towards Application-Layer Purpose-Based Access Control

— Preprint ACM SAC 2020 —

Frank Pallas, TU Berlin, Information Systems Engineering Research Group, Berlin, Germany, fp@ise.tu-berlin.de
Max-R. Ulbricht, TU Berlin, Information Systems Engineering Research Group, Berlin, Germany, mu@ise.tu-berlin.de
Stefan Tai, TU Berlin, Information Systems Engineering Research Group, Berlin, Germany, st@ise.tu-berlin.de
Thomas Peikert, TU Berlin, thomas.peikert@campus.tu-berlin.de
Marcel Reppenhagen, TU Berlin, marcel.reppenhagen@campus.tu-berlin.de
Daniel Wenzel, TU Berlin, daniel.wenzel@campus.tu-berlin.de
Paul Wille, TU Berlin, paul.wille@campus.tu-berlin.de
Karl Wolf, TU Berlin, karl.wolf@campus.tu-berlin.de

Preprint, to appear in: The 35th ACM/SIGAPP Symposium On Applied Computing (ACM SAC 2020), Brno, Czech Republic, March 30-April 3, 2020.
Final version available at:
https://doi.org/10.1145/3341105.3375764

ABSTRACT

In this paper, we propose an architecturally novel approach to implementing purpose-based access control in practice. Different from previous proposals, our approach resides on the application instead of the data(base) layer. This allows for significantly better integration with established architectures and practices of real-world application engineering and to achieve database independence.

To validate practical applicability, we provide two exemplary implementations and briefly assess the introduced overhead in matters of achievable throughputs. Results significantly depend on data and query type but basically suggest bearable overheads for realistic applications even though possible performance optimizations have not been implemented in our proofs-of-concept yet. Our approach thus proposes significantly better practical feasibility than previous ones and exhibits reasonable overheads. It therefore paves the way for purpose-based access control to be actually adopted in practice.

KEYWORDS

Privacy, data protection, purpose limitation, access control, PBAC, privacy by design, privacy engineering, web engineering

1 INTRODUCTION

Privacy by Design[1] (PbD) is one of the core concepts of modern privacy legislation, aiming at the effective implementation of privacy principles through concrete technologies and their design. For instance, the European General Data Protection Regulation (GDPR) requires data controllers to "implement appropriate technical and

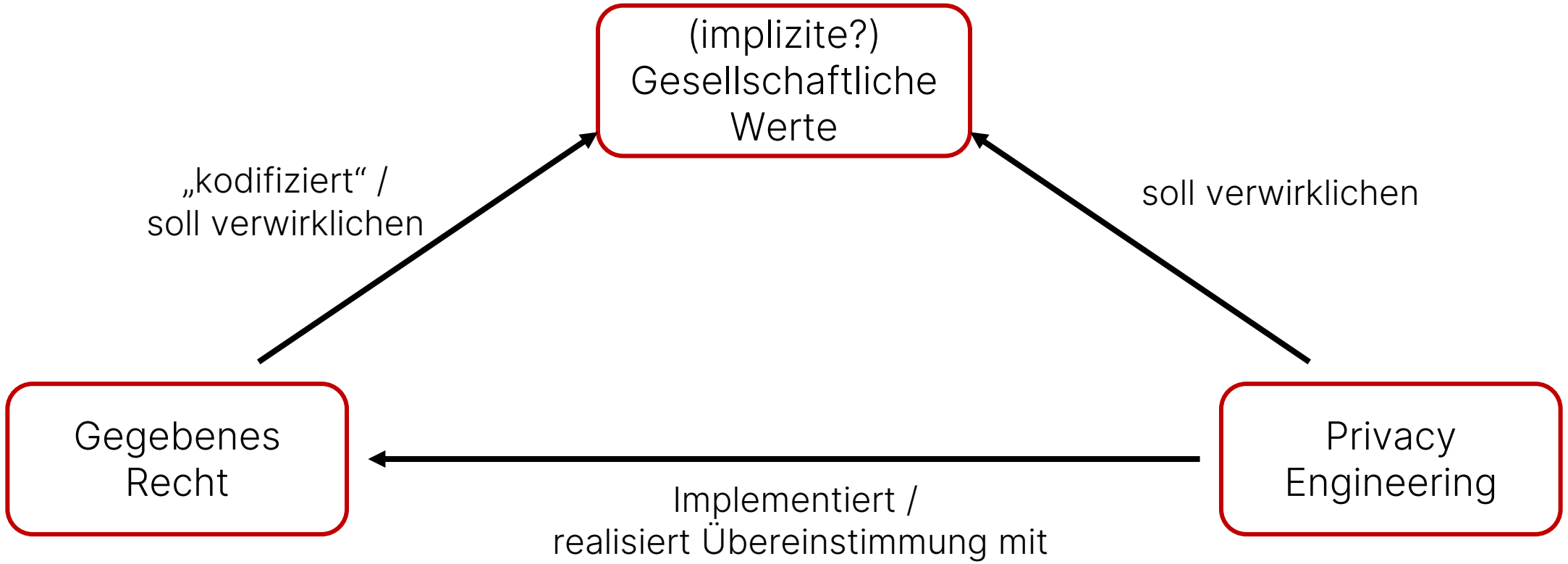
[1]Being well aware of the slightly different notions between "Privacy" and "Data Protection", we use these terms interchangeably herein.

organisational measures [...] designed to implement data-protection

foreseeable givens. With omnipresent collection of personal data and its use for novel, initially unforeseen use-cases ("repurposing")

„Privacy Engineering“

„Privacy Engineering“?



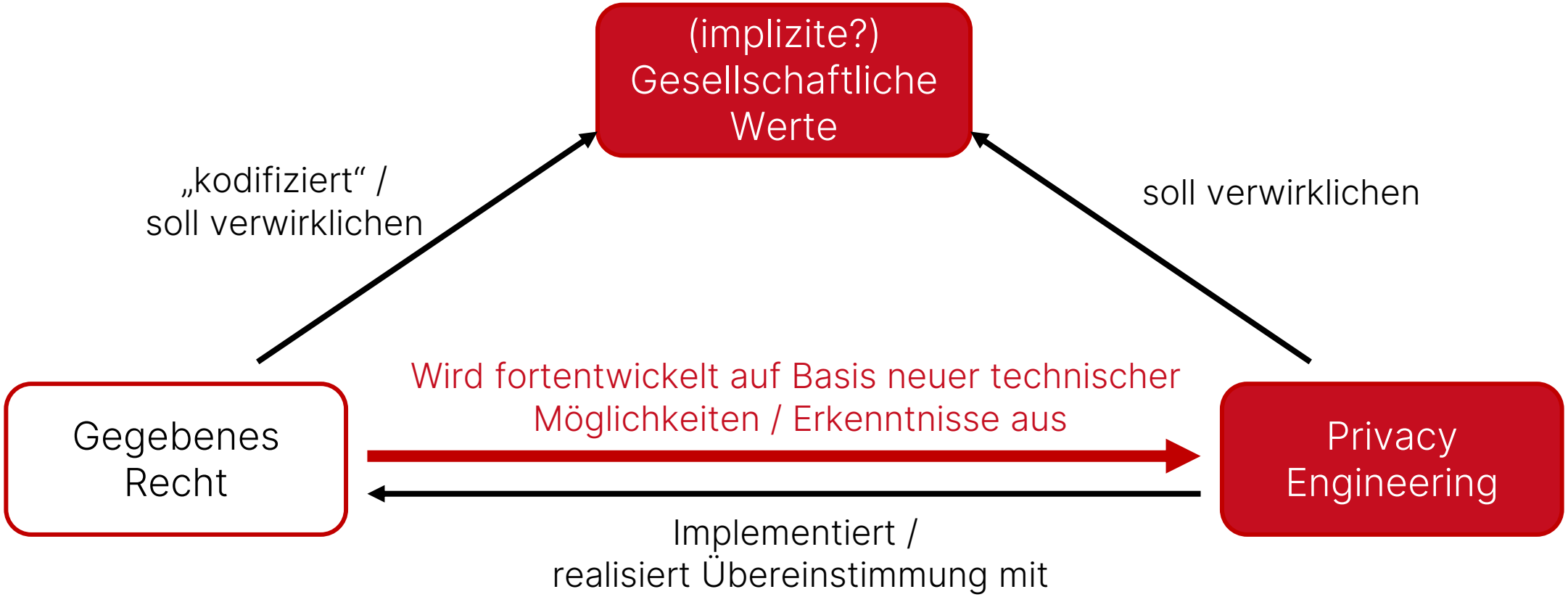
GDPR: „Transparenz“

Transparent information, communication and modalities for the exercise of the rights of the data subject

1. The controller shall take appropriate measures to provide any information referred to in Articles 13 and 14 and any communication under Articles 15 to 22 and 34 relating to processing to the data subject in a concise, transparent, intelligible and easily accessible form, using clear and plain language, in particular for any information addressed specifically to a child. The information shall be provided in writing, or by other means, including, where appropriate, by electronic means. When requested by the data subject, the information may be provided orally, provided that the identity of the data subject is proven by other means.

7. The information to be provided to data subjects pursuant to Articles 13 and 14 may be provided in combination with standardised icons in order to give in an easily visible, intelligible and clearly legible manner a meaningful overview of the intended processing. Where the icons are presented electronically they shall be machine-readable.

„Privacy Engineering“?



Information Governance – The Horizontal Policy Elevator

→ Technische Fortentwicklung des „Stands der Technik“ und des „mit angemessenem Aufwand Umsetzbaren“

→ Neue explizite und implizite Verpflichtungen zur Umsetzung

Privacy Engineering – Want more?

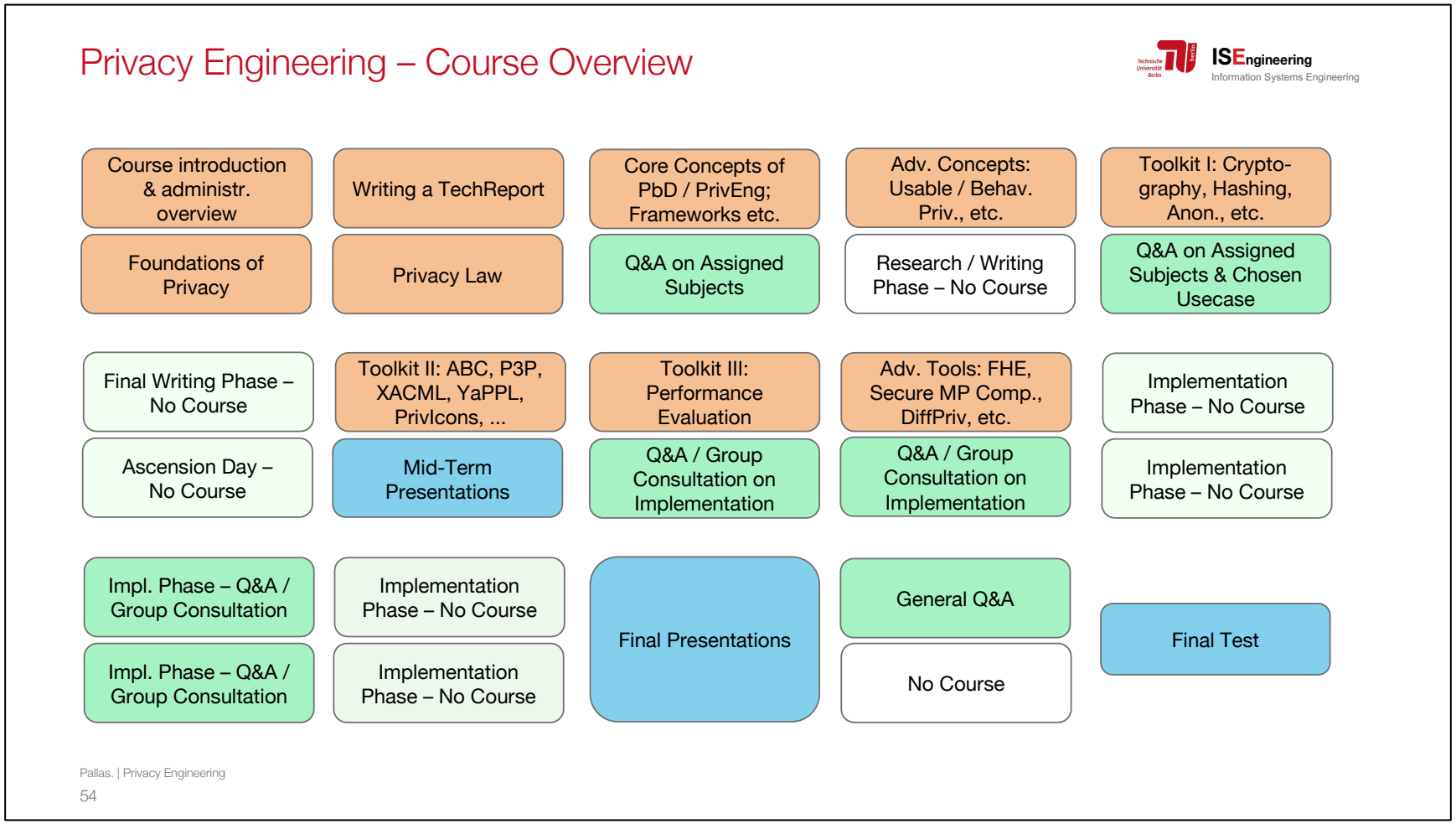
More Privacy Engineering?

Privacy Engineering

ISEngineering
Information Systems Engineering

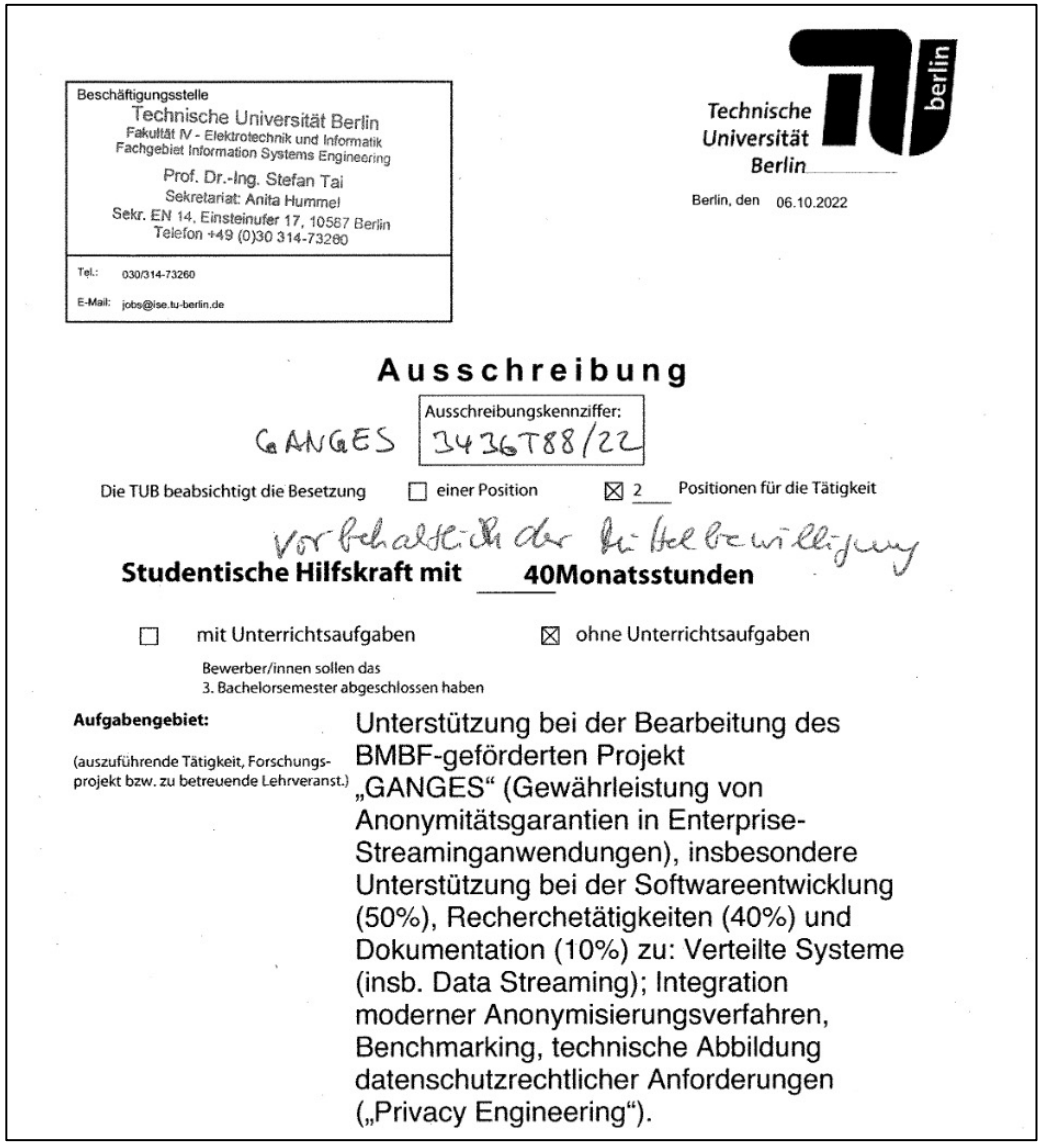
Lesson 1: Kick-Off / Introduction / Domain Overview “Privacy Engineering”

More Privacy Engineering?



→ Inter-/transdisziplinär, Implementierungsfokus, SoSe, Master

More Privacy Engineering?



Beschäftigungsstelle
Technische Universität Berlin
Fakultät IV - Elektrotechnik und Informatik
Fachgebiet Information Systems Engineering
Prof. Dr.-Ing. Stefan Tai
Sekretariat: Anita Hummel
Sekr. EN 14, Einsteinufer 17, 10587 Berlin
Telefon +49 (0)30 314-73260

Tel.: 030/314-73260
E-Mail: jobs@ise.tu-berlin.de

Technische Universität Berlin

Berlin, den 06.10.2022

Ausschreibung

GANGES | Ausschreibungskennziffer: 3436T88/22

Die TUB beabsichtigt die Besetzung ☐ einer Position ☒ 2 Positionen für die Tätigkeit

vorbehaltlich der Mittelbewilligung

Studentische Hilfskraft mit 40Monatsstunden

☐ mit Unterrichtsaufgaben ☒ ohne Unterrichtsaufgaben

Bewerber/innen sollen das
3. Bachelorsemester abgeschlossen haben

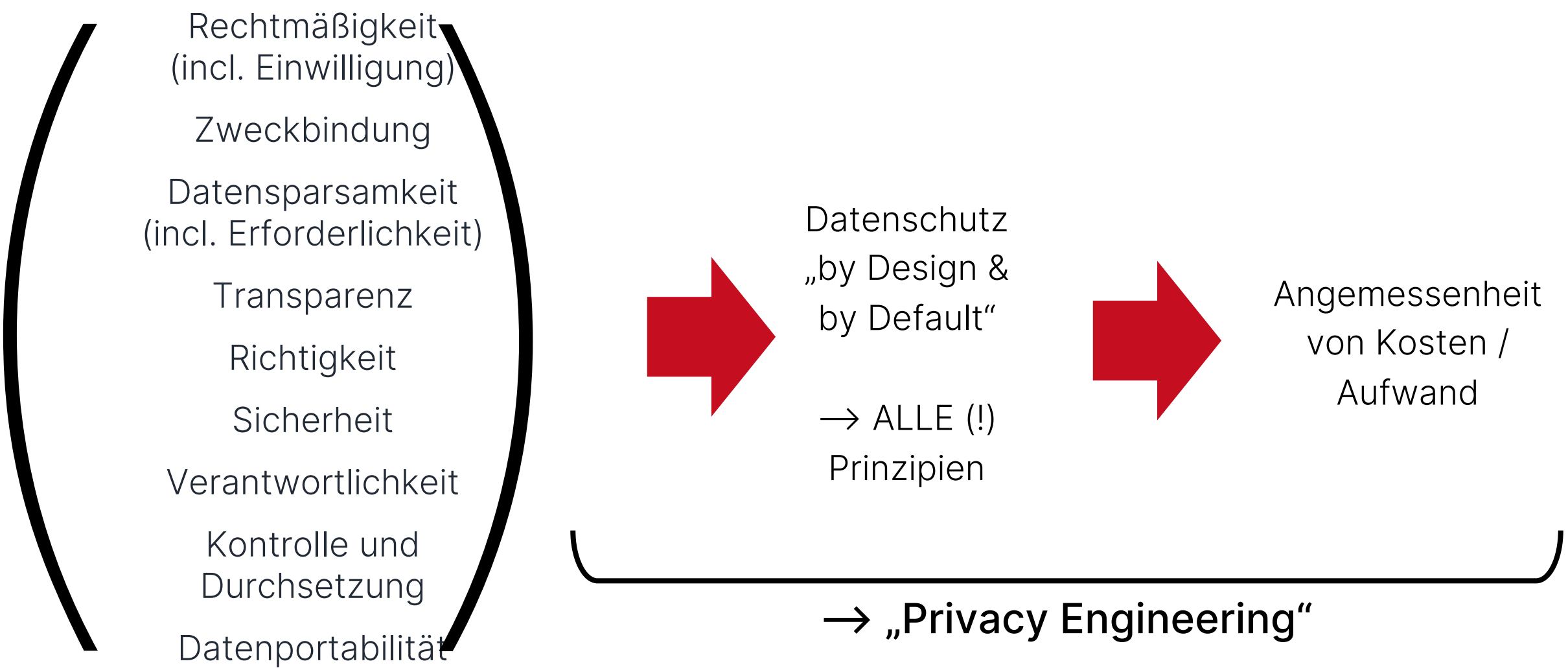
Aufgabengebiet:

(auszuführende Tätigkeit, Forschungsprojekt bzw. zu betreuende Lehrveranst.)

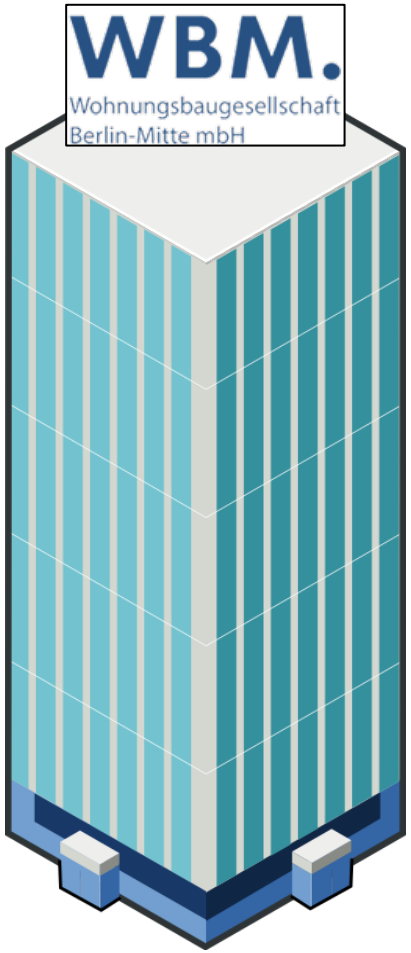
Unterstützung bei der Bearbeitung des BMBF-geförderten Projekt „GANGES“ (Gewährleistung von Anonymitätsgarantien in Enterprise-Streaminganwendungen), insbesondere Unterstützung bei der Softwareentwicklung (50%), Recherchetätigkeiten (40%) und Dokumentation (10%) zu: Verteilte Systeme (insb. Data Streaming); Integration moderner Anonymisierungsverfahren, Benchmarking, technische Abbildung datenschutzrechtlicher Anforderungen („Privacy Engineering“).

Get in touch!

„by Design & by Default“



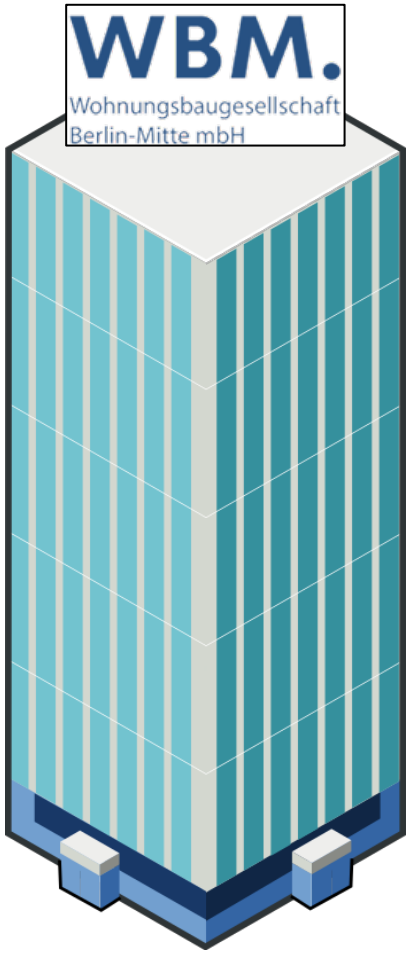
Stellen Sie sich vor...



„Entwerfen Sie ein tragfähiges Backend zur intelligenten Vernetzung des Gebäudebestands“

???

Stellen Sie sich vor...



„Entwerfen Sie ein tragfähiges Backend zur intelligenten Vernetzung des Gebäudebestands“

§§

!

Rechtmäßigkeit (incl. Einwilligung)
Zweckbindung
Datensparsamkeit (incl. Erforderlichkeit)
Transparenz
Richtigkeit
Sicherheit
Verantwortlichkeit
Kontrolle und Durchsetzung
Datenportabilität

Datenschutz „by Design & by Default“

→ **ALLE (!) Prinzipien**

Angemessenheit von Kosten / Aufwand

→ **„Privacy Engineering“**

Nächste Woche(n)

7	04.12.23	Datenschutz 2: Privacy Engineering [FP] **(in Präsenz)**	1. 9+1 Prinzipien: „Privacy by Design" 2. Technische Umsetzung von Datensparsamkeit (incl. *k*-anonymity etc.) 3. „Wieviel Security ist angemessen?", Experimente 4. Praktische Beispiele für „Systems-Oriented Privacy Engineering"
	07.12.23	Gruppenkonsultation Poster/Essay **(via Zoom)**	Gruppenindividuelle Betreuung (nach Absprache) [Tutor:Innen]
8	11.12.23	Datenschutz 3: Surveillance **(in Präsenz)** [FP]	1. Datenschutz vs. Überwachung 2. Benthams Panopticon 3. Gesamtgesellschaftliche Bedeutung 4. Freiheit vs. Sicherheit? Zum Umgang mit sich widersprechenden Grundrechten
	14.12.23	Q&A zu Block C **(via Zoom)**	

fin.

Bonusfolien: Technische Mechanismen zur Einwilligung

9 Kernprinzipien des Datenschutzes

Rechtmäßigkeit (incl. Einwilligung)

Zweckbindung

Datensparsamkeit (incl. Erforderlichkeit)

Transparenz

Richtigkeit

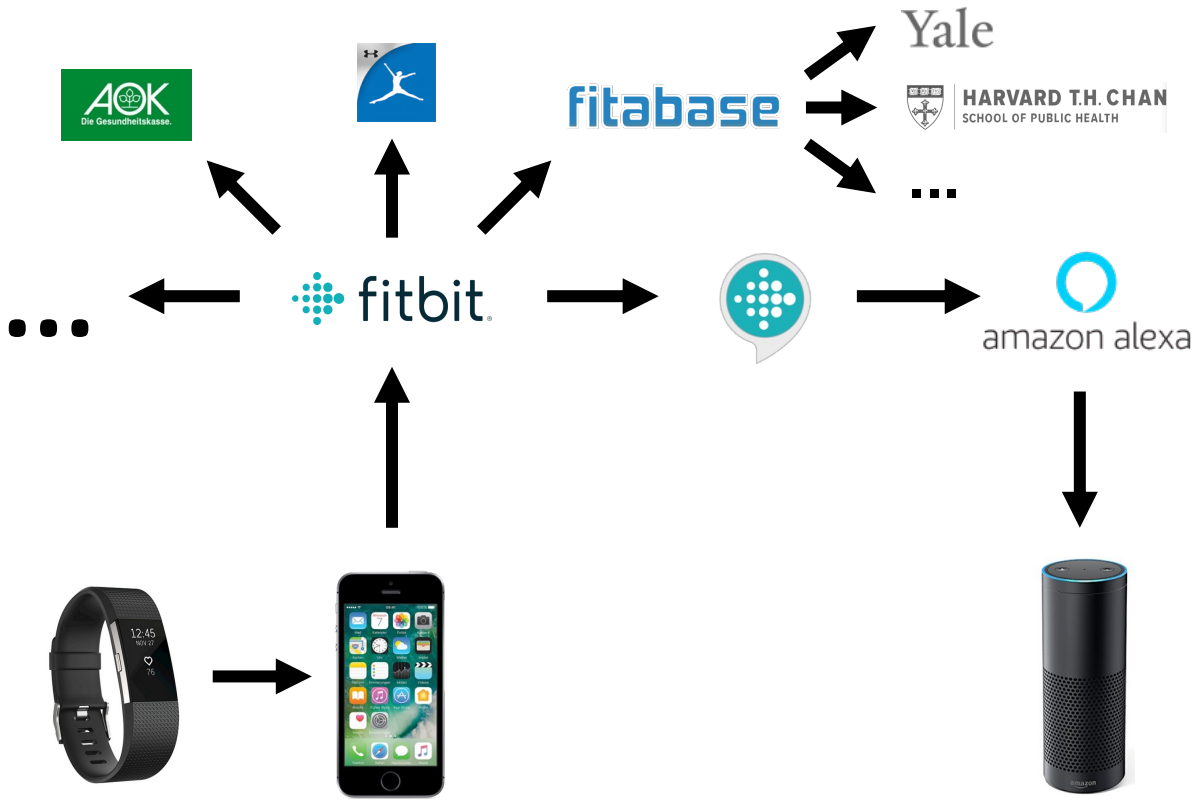
Sicherheit (insb. Vertraulichkeit, Integrität, Verfügbarkeit)

Verantwortlichkeit (incl. nachweisbare Rechtskonformität)

Kontrolle und Durchsetzung

Datenportabilität

Einwilligung in aktuellen Szenarien



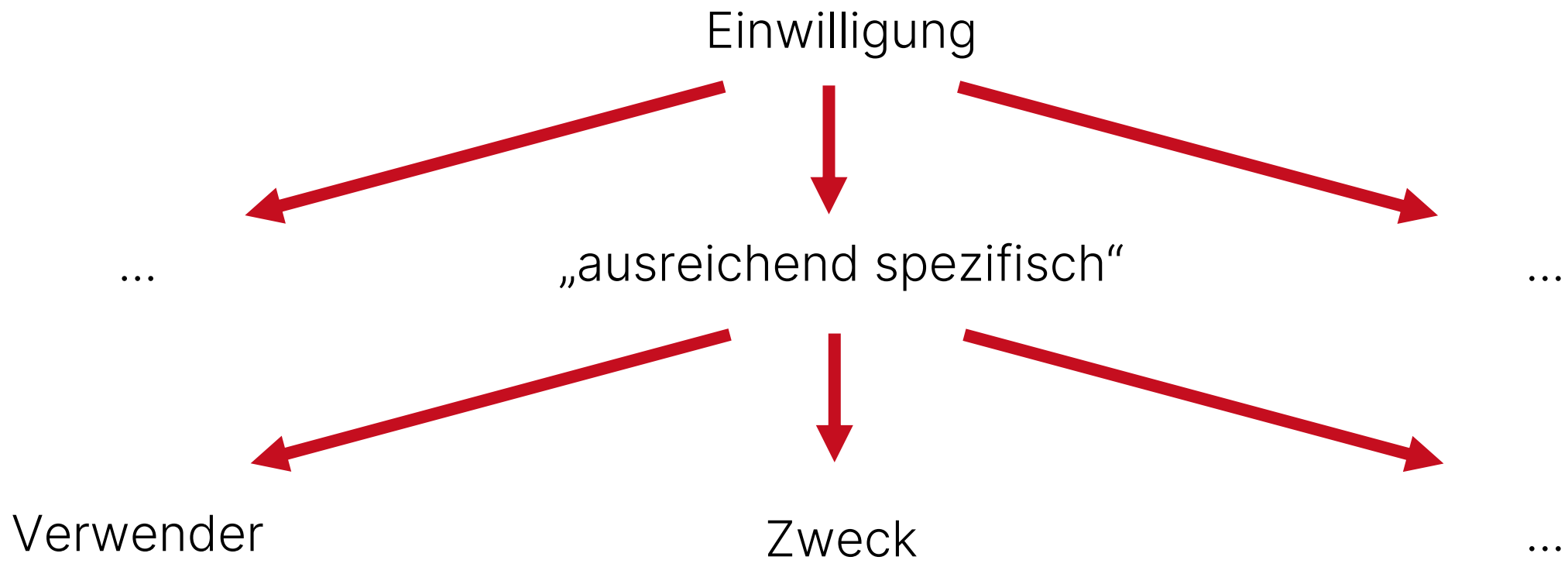
Wer soll welche Daten zu welchem Zweck bekommen?

Aktuelle Szenarien sind durch viele, hochvernetzte Akteure gekennzeichnet

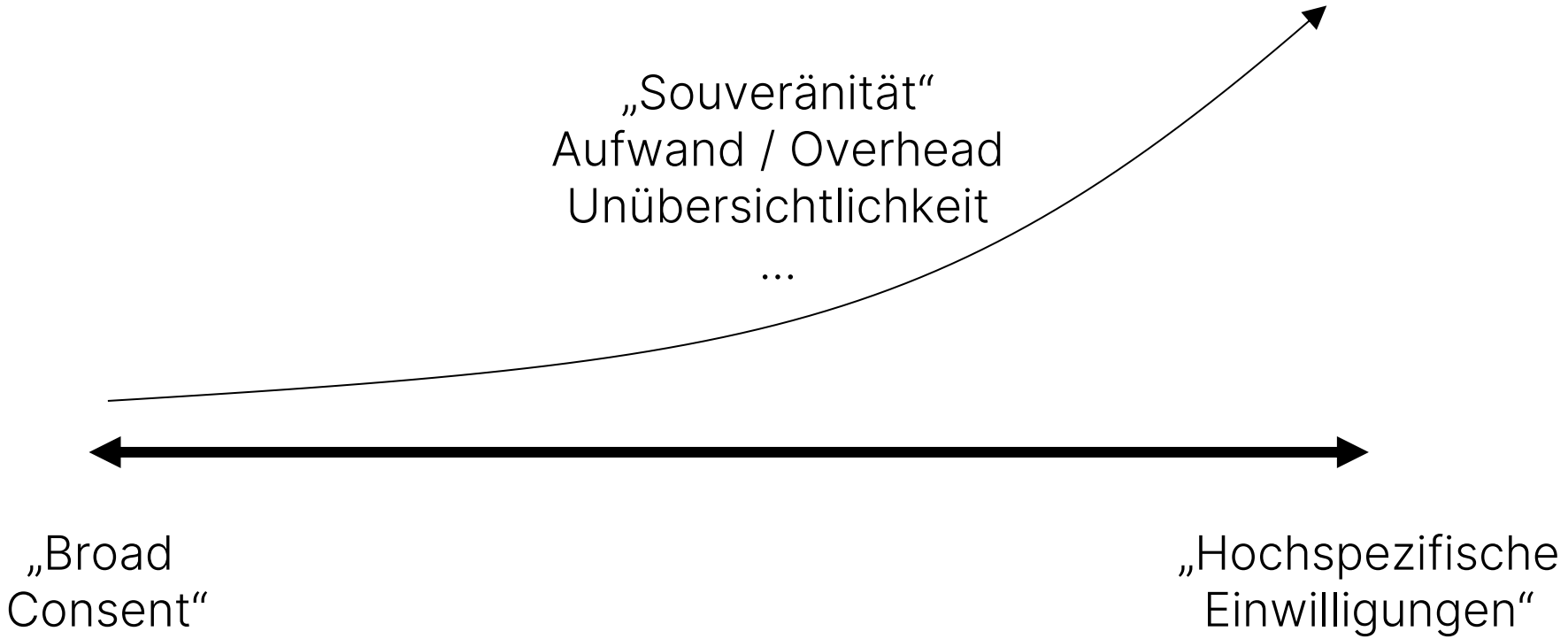
„Informierte Einwilligungen“ würden pro Akteur signifikante Aufwände für data subject erfordern (lesen, für unterschiedliche Zwecke entscheiden / anpassen, ja klicken, ...)

In ioT-Szenarien fehlen zudem nötige Interaktionsmöglichkeiten

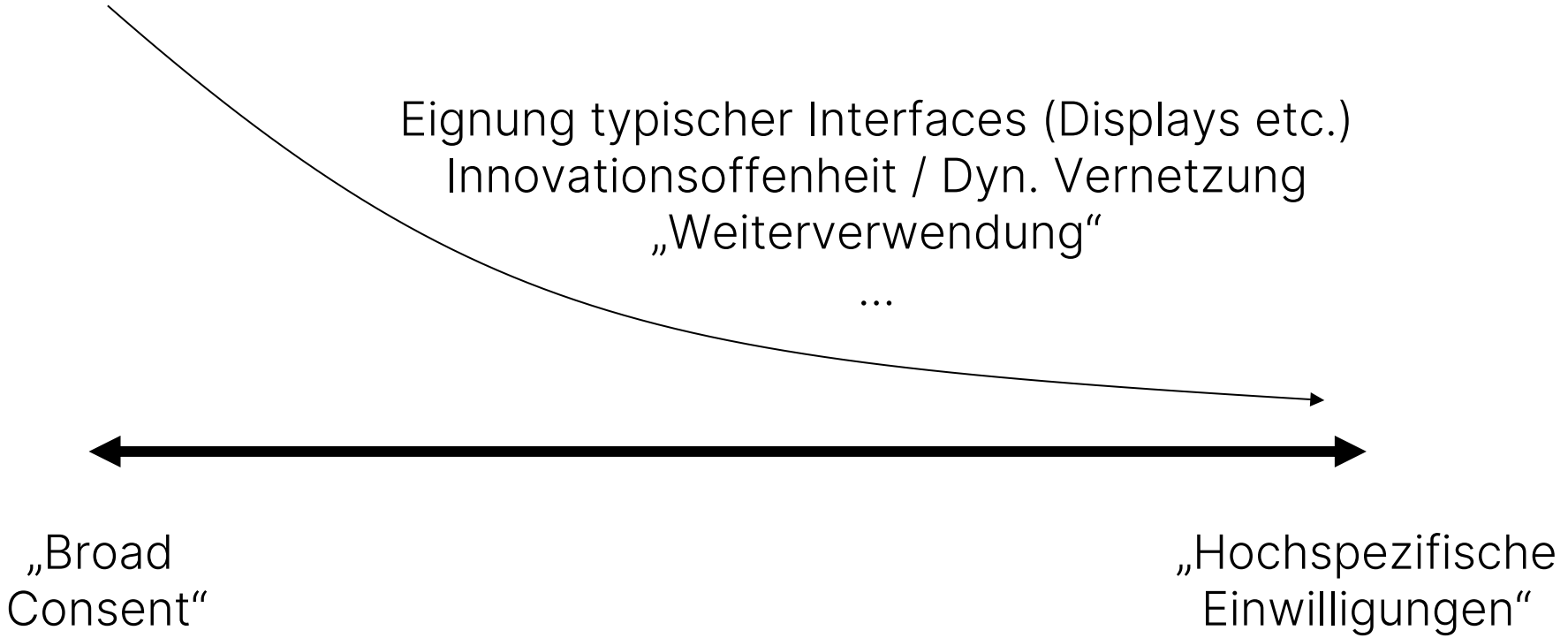
Beispiel: Einwilligung im Internet der Dinge



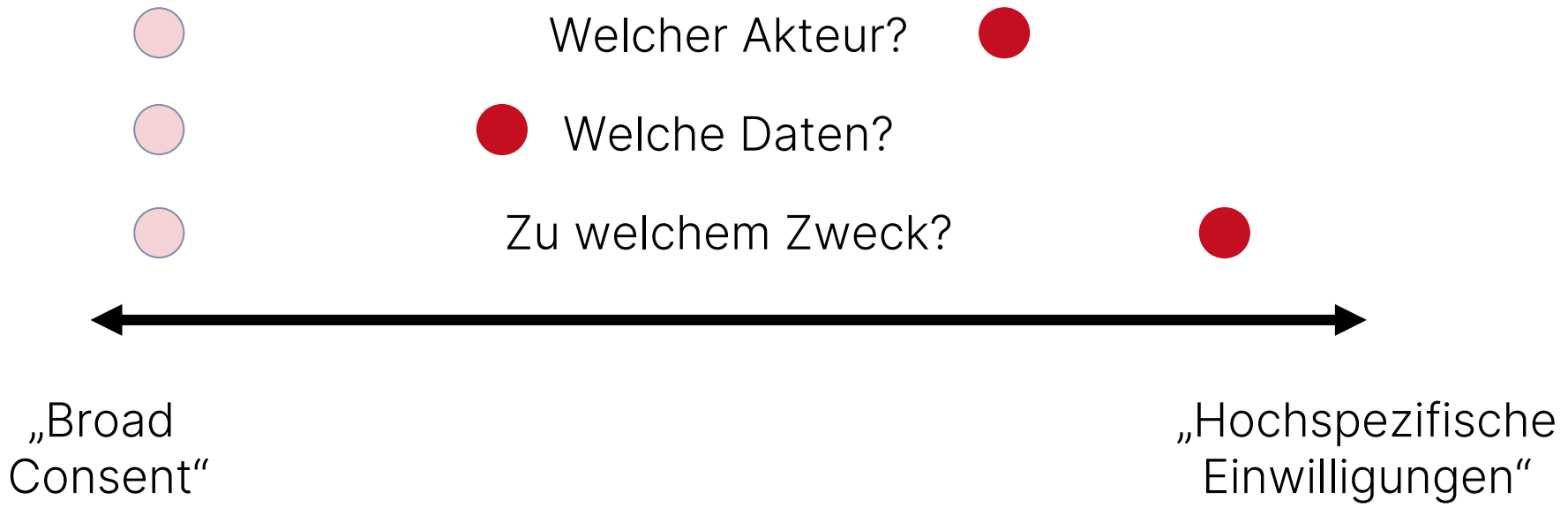
Beispiel: Einwilligung im Internet der Dinge



Beispiel: Einwilligung im Internet der Dinge

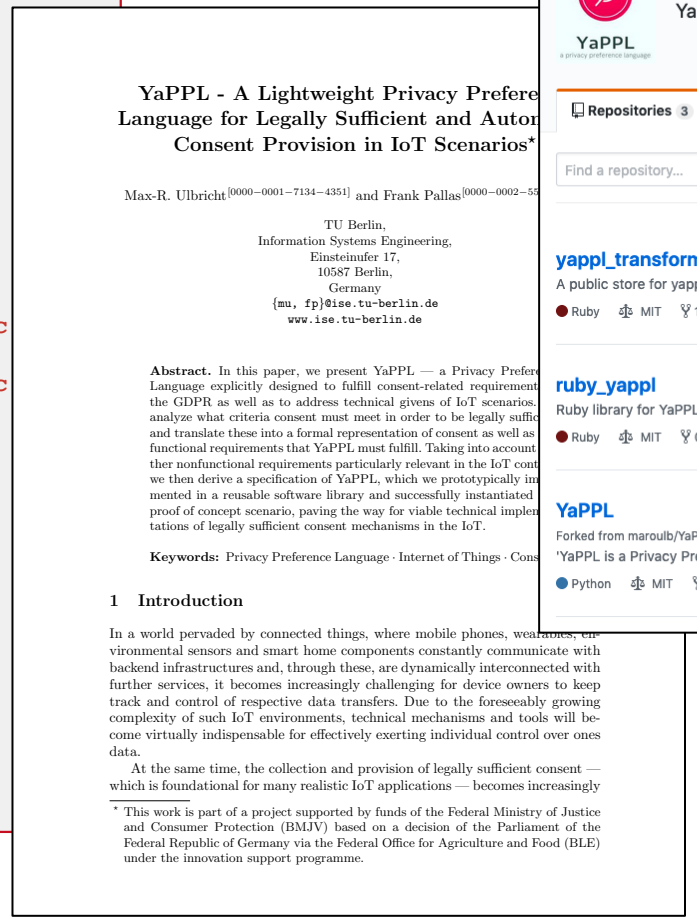


Beispiel: Einwilligung im Internet der Dinge



Technische Umsetzung: Formale Repräsentation von Einwilligungen nach GDPR-Anforderungen

```
{
"_id": 4477,
"preference": [{
            "rule": {
                        "purpose": {
                                    "permitted": [],
                                    "excluded": []
                        },
                        "utilizer": {
                                    "permitted": [],
                                    "excluded": []
                        },
                        "transformation": [
                                    {"attribute": "tr_func
                                    ...
                                    {"attribute": "tr_func
                        ],
                        ”valid_from": "date-time”,
                        "exp_date": "date-time"
            }
},
{
            "rule": {
                        "purpose": {
                        ...
                        "exp_date": "date-time"
            }
}
]
}
```



YaPPL - A Lightweight Privacy Prefere
Language for Legally Sufficient and Autor
Consent Provision in IoT Scenarios*

Max-R. Ulbricht[0000−0001−7134−4351] and Frank Pallas[0000−0002−55

TU Berlin,
Information Systems Engineering,
Einsteinufer 17,
10587 Berlin,
Germany
{mu, fp}@ise.tu-berlin.de
www.ise.tu-berlin.de

Abstract. In this paper, we present YaPPL — a Privacy Prefere
Language explicitly designed to fulfill consent-related requirement
the GDPR as well as to address technical givens of IoT scenarios.
analyze what criteria consent must meet in order to be legally suffic
and translate these into a formal representation of consent as well as
functional requirements that YaPPL must fulfill. Taking into account
ther nonfunctional requirements particularly relevant in the IoT cont
we then derive a specification of YaPPL, which we prototypically im
mented in a reusable software library and successfully instantiated
proof of concept scenario, paving the way for viable technical implem
tations of legally sufficient consent mechanisms in the IoT.

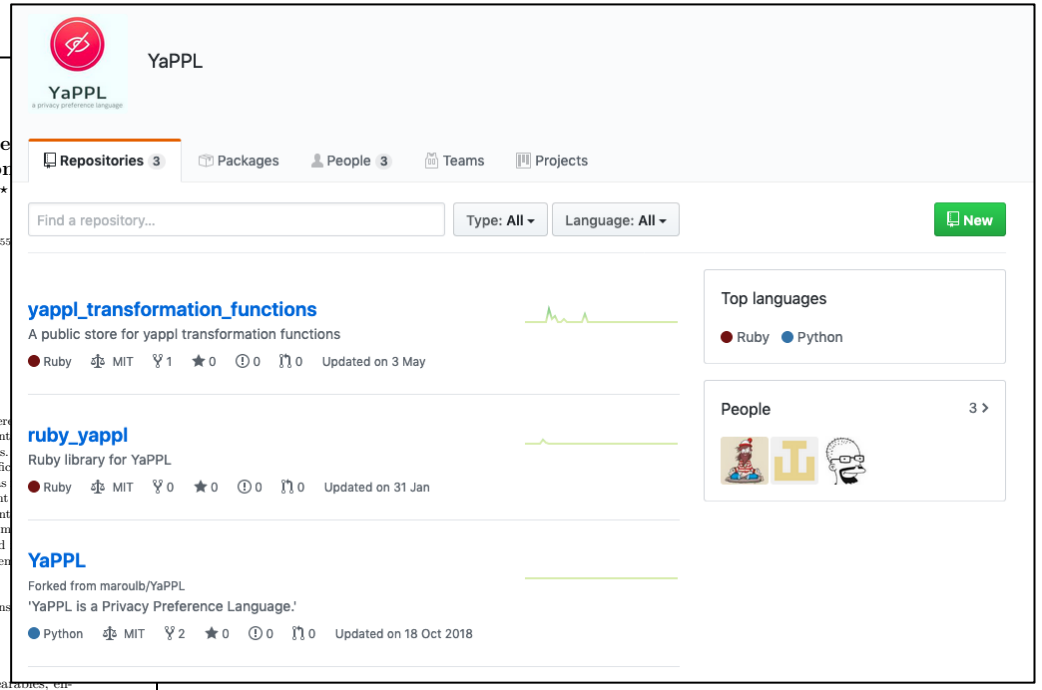
Keywords: Privacy Preference Language · Internet of Things · Cons

1 Introduction

In a world pervaded by connected things, where mobile phones, wearables, environmental sensors and smart home components constantly communicate with backend infrastructures and, through these, are dynamically interconnected with further services, it becomes increasingly challenging for device owners to keep track and control of respective data transfers. Due to the foreseeably growing complexity of such IoT environments, technical mechanisms and tools will become virtually indispensable for effectively exerting individual control over ones data.

At the same time, the collection and provision of legally sufficient consent — which is foundational for many realistic IoT applications — becomes increasingly

* This work is part of a project supported by funds of the Federal Ministry of Justice and Consumer Protection (BMJV) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support programme.



https://github.com/yappl