„One more thing…“

Steve Jobs

Department für Rechtswissenschaften und Internationale Beziehungen

Donau-Universität Krems. Die Universität für Weiterbildung.

Technische Aspekte der Datenanonymisierung

Daniel Kissler MSc MA

mail@daniel-kissler.at

Prolog

Datenschutz

- DSG § 1 Abs. 1: *„[…] Anspruch auf Geheimhaltung der ihn betreffenden personenbezogenen Daten […]“*
- DSGVO Art 32 b): *„die Fähigkeit, die Vertraulichkeit, Integrität, Verfügbarkeit und Belastbarkeit […]“*

Security

Schutzziele

- Primäre: Vertraulichkeit, Integrität und Verfügbarkeit
- Sekundäre: Authentizität, Nichtabstreitbarkeit und Zurechenbarkeit

Technische und organisatorische Maßnahmen

Prolog

Organisatorische Maßnahmen

- Datenschutz-Managementsystem
- Datenschutzbeauftragter
- Richtlinien & Prozesse

Technische Maßnahmen

- Verschlüsselung
- Protokollierung
- BenutzerInnenmanagement

Bauliche Maßnahmen

- Einbruchssicherung
- Zutrittskontrollsystem
- Videoüberwachung

Anonymität

- Anonymität (altgr. ἀνώνυμος anónymos ‚nicht benannt‘)

- *„Ein Subjekt ist anonym, wenn das Subjekt in einer Menge von Subjekten (der Anonymitätsmenge) nicht hinreichend identifizieren werden kann.“*
  oder
- *„Anonymität bezeichnet das Fehlen der Zuordnung einer Person zu einer von ihr ausgeübten Handlung. Dies kann absichtlich (Geheimhaltung) oder unbeabsichtigt (Zufall) geschehen.“*

Anonymität: Bedrohungen

Aufdecken der Identität	Eine anonyme Person kann aufgrund eines veröffentlichten Datensatzes eindeutig identifiziert werden.
Aufdecken der Zugehörigkeit	Eine anonyme Person kann mit einer entsprechenden hohen Wahrscheinlichkeit einer Menge von Personen mit gewissen Attributen zugeordnet werden.
Aufdecken eines Attributs	Einer anonymen Person können sensible Attribute zugeordnet werden.
Attribution	Auf Basis von mehrere (abstrakten) Attributen, kann eine Identifikation durchgeführt werden.

Anonymität: Attribution

63,3% der U.S. Bevölkerung

„sind durch eine Attributkombination eindeutig identifizierbar (Geschlecht, Geburtsdatum, 5-stellige Postleitzahl)“ [1]

Anonymität: Generalisierung

- Generalisierung
  - Datenschutzrelevante Daten werden verallgemeinert, wodurch eine sog. K-Anonymität entsteht, die die betroffenen Personen schützt.
  - Je nach Verwendungszweck, variiert der Grad der Generalisierung bzw. ob eine Generalisierung überhaupt durchgeführt werden kann.
- Beispiel: med. Studie

Name	Geburtsjahr	PLZ	Geschlecht	Diagnose
John Doe	1982	33098	Männlich	Migräne
Thomas Muster	1982	33098	Männlich	Erkältung
Max Maier	1983	33098	Männlich	Rheuma
Otto Normal	1983	33098	Männlich	Depression
Jane Doe	1985	33100	Weiblich	Heuschnupfen
Lieschen Müller	1985	33100	Weiblich	Hypochondrie
Erika Musterfrau	1983	33098	Weiblich	Übergewicht
Jane Average	1983	33098	Weiblich	Migräne

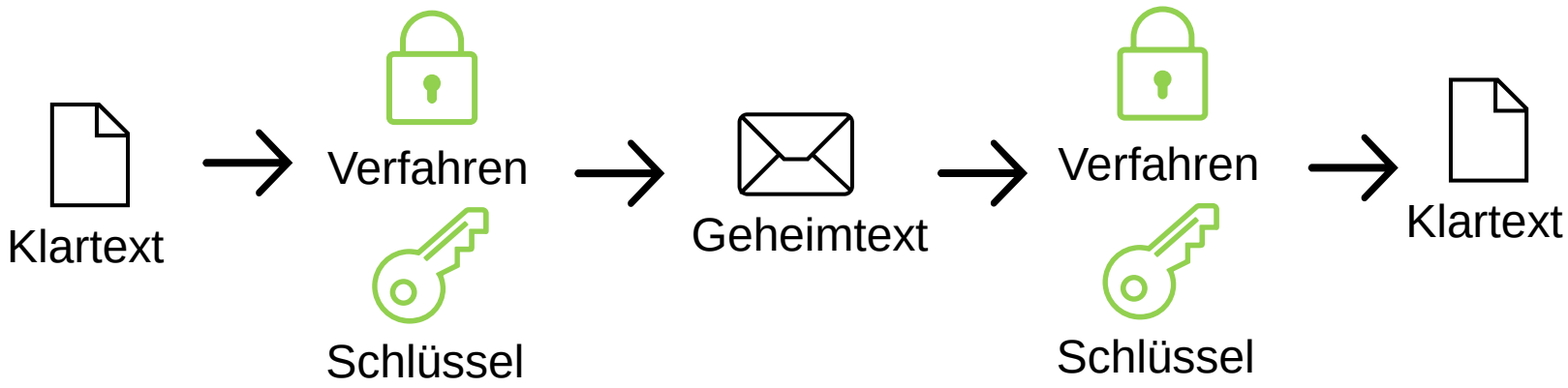
Geburtsjahr	PLZ	Geschlecht	Diagnose
1982–1983	33098	Männlich	Migräne
1982–1983	33098	Männlich	Erkältung
1982–1983	33098	Männlich	Rheuma
1982–1983	33098	Männlich	Depression
1983–1985	33*	Weiblich	Heuschnupfen
1983–1985	33*	Weiblich	Hypochondrie
1983–1985	33*	Weiblich	Übergewicht
1983–1985	33*	Weiblich	Migräne

Anonymität: Generalisierung

- Das Konzept der Generalisierung verfügt über Schwächen:
  - Unsorted Matching-Angriff
    Datenbankeinträge werden oftmals sortiert abgelegt, wodurch die Reihenfolge der Einträge determiniert sind.
  - Komplementärveröffentlichung
    Veröffentlichte (generalisierte) Daten lassen sich aufgrund von gemeinsamen Nennern zusammenfügen.
  - Homogenitätsangriff
    Eine Gruppe teilt das selbe Attribut (z.B. tödl. Krankheit), wodurch eine eindeutige Identifikation nicht mehr zwingend notwendig ist.
  - Angriff mit Hintergrundwissen
    Gewisse Attribute lassen auf weitere Informationen schließen (z.B. Schwangerschaftsbeschwerden, Hodenkrebs)
  - Oftmals ist eine Generalisierung von pers. bez. Daten nicht möglich, da der Anwendungsfall entscheidend ist.

Anonymität: Verschlüsselung

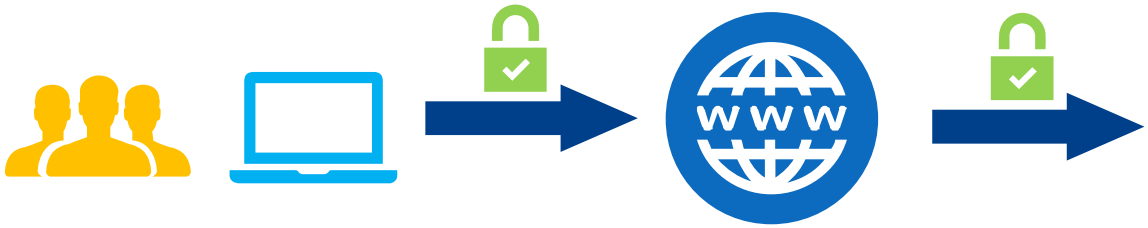
- Verschlüsselung
  Grundsätzlich funktioniert jedes Verschlüsselungsverfahren nach dem selben Prinzip:



Anonymität: Verschlüsselung

- Im Alltag ist die Verschlüsselung oftmals gar nicht mehr sichtbar (beispielsweise im Browser oder am Smartphone).
- Im Bereich Datenschutz wird durch Verschlüsselung die Vertraulichkeit und Integrität gewährleistet aber die Verfügbarkeit eingeschränkt.
- Durch die Verschlüsselung gehen keine Informationen verloren (wie z.B. bei der Generalisierung).
- In der Praxis ist die Verschlüsselung von Daten mit erheblichem Aufwand verbunden, sowohl in der Konzeptionierung als auch in der tatsächlichen Umsetzung.
- Des Weiteren können die Daten auf unterschiedlichen Ebenen geschützt werden.

Anonymität: Verschlüsselung



1. Transportverschlüsselung
2. Verschlüsselung auf Applikations-Ebene
3. Verschlüsselung durch die Datenbank bzw. das Datenbankmanagementsystem
4. Verschlüsselung auf Dateiebene
5. Verschlüsselung auf HW-Ebene

Pseudonymität

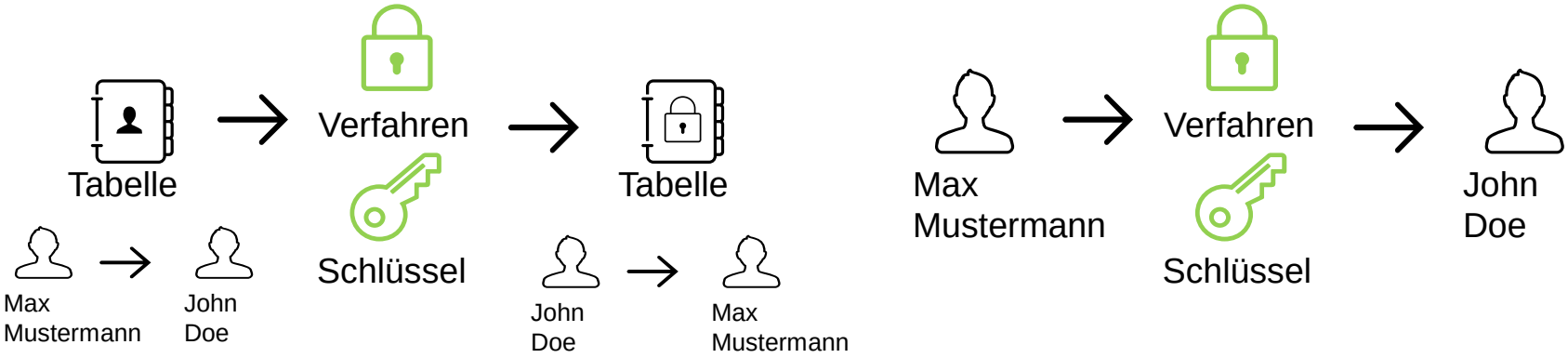
- In vielen Bereichen ist Anonymität nicht möglich, da z.B. die Zuordnung von Kundendaten notwendig ist.
- Bei der Pseudonymisierung werden ein oder mehrere Identifikationsmerkmale, auf dessen Basis eine eindeutige Identifizierung möglich ist, durch sog. Pseudonyme ersetzt, um die Feststellung der Identität des Betroffenen auszuschließen oder wesentlich zu erschweren.
- Die Pseudonymisierung spielt im Bereich der Nachvollziehbarkeit, Dokumentation und Zuordnung von Handlungen eine wesentliche Rolle:
  „*Wer hat wann, was gemacht?*“

Pseudonymität

- In der Praxis werden Pseudonyme oftmals selbst gewählt (insb. im Internet): Forum, Verkaufsplattformen, E-Mailadresse, etc.
- Im Bereich der Dokumentation und Nachweiserbringung (Stichwort: Nichtabstreitbarkeit) erfolgt die Pseudonymisierung oftmals durch das System selbst.
- Hierfür gibt es zwei Möglichkeiten
  - Das System selbst verwaltet die Pseudonyme
  - Vor oder nach der Verarbeitung werden die Identifikationsmerkmale pseudonymisiert.

Pseudonymität

- Die Pseudonymisierung kann mittels Zuordnung (z.B. Max Musterman -> John Doe) oder durch direkte Verschlüsselung erfolgen.
- Um die Tabelle der Zuordnungen zu schützen, bietet es sich an diese ebenfalls zu verschlüsseln.



Pseudonymität

- Durch die Absicherung der Pseudonyme mittels Verschlüsselung kann ein Mehraugenprinzip und eine Funktionstrennung (Seperation of Duty) erfolgen:
  - Dies kann einerseits mittel dig. Zertifikate gewährleistet werden oder
  - Andererseits durch Aufteilung d. Schlüssels auf den jeweiligen Adressatenkreis:

Schlüssel	
b3BlbnNzaC1rZXktdjEAAAAABG5vbmUAAAA Ebm9uZQAAAAAAAABAAACFwAAAAdzc2gtcn NhAAAAAwEAAQAAAgEA0I5sVxcSgnSaijdQs /A6BGuyOOs7kUEBqkPQMooEZRnwOoxdbJW6	Betriebsrat
NJUnmZrf9gwaKr4/4IIS+ullR38MiFcLYMT NGJzkJLWrJfF12vZoifsNYNN5CicQMGEdnu	Datenschutzbeuaftragte
b7EhYeuyJoe+P4eVPbCYaIRKobmbhAi7UBt KCEVQgo4tSLN9O7ZvOTHcrgYxT1TThBFbMY 9RhZ27Hg8saazCd2kWId2yb6gE1wkVtsTX9 cQd/rhSgisNbLUpZet3rrVyi+WuGm1xWovC	CIO
BnBvcDUQLb9bQPV8iFGGE5+RC8Og6qGy2Xe PLpiI+0tnX08R48zW9zkDgr5Y8nlPj9WT6O	CISO
.....	

Pseudonymität

- In bestimmten Fällen ist selbst die Pseudonymisierung von Identifikationsmerkmalen nicht zielführend.
- IP-Adressen
  - Jedes Gerät, welches an ein lokales Netzwerk angeschlossen wird verfügt über eine IP-Adresse (statisch) bzw. bekommt es eine zugewiesen (dynamisch).
  - Unter bestimmten Umständen ist das Gerät und der Benutzer durch die IP-Adresse eindeutig identifizierbar. -> IP-Adressen pseudonymisieren?
  - IP-Adressen (IPv4) haben 32 Bit (ca. 4,3 Mrd. Adressen), welche in vier Oktetts (Blöcke zu je 8 Bits) aufgeteilt werden und üblicherweise dezimal dargestellt werden: 000.000.000.000 – 255.255.255.255
  - Pseudonymisierung v. öffentlichen IP-Adressen ist nicht möglich, da alle vergeben sind (kein Pool).

Pseudonymität

- Viele IP- Adressen sind reserviert bzw. gesperrt (special purpose)
- Pseudonymisierung v. lokalen IP-Adressen ist technisch möglich, birgt Herausforderungen iZm. der autom. Verarbeitung (z.B. Log-Management)
- Direkter Bezug zwischen BenutzerIn und IP-Adresse kann nur äußerst selten hergestellt werden.
- Gerade im Bereich Security (z.B. forensische Analyse, Malware-Analyse, Threat Intelligence) sind IP-Adressen eine essentielle Informationsquelle um Zusammenhänge, Bigger Pictures, etc. besser verstehen zu können.
- Präferenz: Pseudonymisierung des Benutzers

IANA IPv4 Special-Purpose Address Registry [2]

Address Block	Name
0.0.0.0/8	"This network"
0.0.0.0/32	"This host on this network"
10.0.0.0/8	Private-Use
100.64.0.0/10	Shared Address Space
127.0.0.0/8	Loopback
169.254.0.0/16	Link Local
172.16.0.0/12	Private-Use
192.0.0.0/24 [2]	IETF Protocol Assignments
192.0.0.0/29	IPv4 Service Continuity Prefix
192.0.0.8/32	IPv4 dummy address
192.0.0.9/32	Port Control Protocol Anycast
192.0.0.10/32	Traversal Using Relays around NAT Anycast
192.0.0.170/32	192.0.0.171/32, NAT64/DNS64 Discovery
192.0.2.0/24	Documentation (TEST-NET-1)
192.31.196.0/24	AS112-v4
192.52.193.0/24	AMT
192.88.99.0/24	Deprecated (6to4 Relay Anycast)
192.168.0.0/16	Private-Use
192.175.48.0/24	Direct Delegation AS112 Service
198.18.0.0/15	Benchmarking
198.51.100.0/24	Documentation (TEST-NET-2)
203.0.113.0/24	Documentation (TEST-NET-3)
240.0.0.0/4	Reserved
255.255.255.255/32	Limited Broadcast

Technische Aspekte

- Privacy/Security by Design!
- Es ist zwingend erforderlich sich bereits in der Konzeptionierungsphase dem Thema Datenschutz (und Security) zu widmen (Lastenheft!). Nachträgliche Adaption sind möglich aber meist immens ressourcenintensiv.
- Verschlüsselung ist ein probates Mittel um Schutz und/oder Anonymität/Pseudonymität zu gewährleisten. Erfordert jedoch ein entsprechendes Schlüsselmanagement und kann dadurch die Verfügbarkeit gefährden (Schlüsselverlust).
- Die Verschlüsselung selbst erhöht den Ressourcenbedarf (je nach System/Implementierung)

Fazit

- Gewährleistung von Anonymität ist in der Praxis selten möglich, falls doch geht damit immer ein Informationsverlust einher.
- Durch die Pseudonymisierung sind Handlungen nachvollziehbar und Zusammenhänge erkennbar.
- Datenschutz und Security verhalten sich diametral zueinander.
- Technische Umsetzung von datenschutzrechtlichen Anforderung ist oftmals mit erheblichen Aufwand verbunden.
- Technische Maßnahmen sind lediglich ein Teil der Lösung.

Vielen Dank!

Fragen?

Kontaktdaten

Daniel Kissler, MSc MA
mail@daniel-kissler.at
+4369910990800

Quellen

[1] Philippe Golle. Revisiting the uniqueness of simple demographics in the US population. In Proceedings of the 5th ACM workshop on Privacy in electronic society, pages 77–80. ACM, 2006.

[2] https://www.iana.org/assignments/iana-ipv4-special-registry/iana-ipv4-special-registry.xhtml