IT-sikkerhed: Privacy and Data protection GDPR
Privacy by Design

Privacy Technologies

Carsten Jørgensen
Department of Computer Science

### Forskellen

Fortsat fra forelæsningen d.25.oktober

Privacy vs. Data Protection





Sælg din egen data

"Forbyd 'cloud' " >< ingen privacy regler, lad virksomhederne og forbrugerne selv ordne det



## Is privacy dead?

### Facebook's Mark Zuckerberg Claims

by Terrence O'Brien on January 11, 2010 at 03:10 PM

FILED UNDER: celebrities, privacy, socialnetworking, web, facebook, google



## **Privacy is overrated**

265,188









# Google CEO: Secrets Are for Filthy People



Ryan Tate

Filed to: GOOGLEPLEX 12/04/09 4:48pm



Eric Schmidt suggests you alter your scandalous behavior before you complain about his company invading your privacy.

That's what the Google CEO told Maria Bartiromo during CNBC's big Google special last night, an extraordinary pronouncement for such a secretive guy.



Print | Font: A A + -

The generous explanation for Schmidt's

d, deal with it," Sun MicroSystems CEO

Scott McNealy is widely reported to have declared some time ago. Privacy in the digital age may not be as dead and buried as McNealy believes, but it's certainly on

life cumpout

## "Privacy er ikke naturligt for mennesker"





### Privacy er ikke noget nyt

Det er kulturelt hvor meget man deler

Men der har altid været ting man holdt for sig selv – Man fortalte ikke naboen hvor på marken man havde begravet sin guldskat, uanset hvor tæt man boede



### Privacy er ikke noget nyt

Det er - og har altid været - vigtigt, at have mulighed for at **styre** <u>hvad</u> man deler og <u>hvem</u> der får adgang til informationen.

Når man accepterede at naboen lyttede med var det også med den forståelse, at naboen ikke fortalte det videre til Kongen eller Lensgreven.



### Privacy er ikke noget nyt

Hvis man ønskede at holde noget hemmeligt var det muligt at finde et rum hvor det var svært for de andre beboer at lytte med.

### **DVS**

Man kendte risikoen – hvem havde mulighed for at lytte med.

Det var kun muligt at søge i mine breve, dagbøger osv hvis der var fysisk adgang til huset.

Loven og husets arkitektur beskyttede imod hvem som helst kunne overvåge mig

("Privacy drejer sig om kontrol over data")



### Privacy: tidselementet og hukommelsen



De ved du gik på gaden, men (hvis du havde tøj på) kan de ikke huske hvad dag, hvad tid eller hvad tøj.

Teknologi ændrer dette fundamentalt



- Mister kontrollen over "min" data
- Vi undervurderer antallet af modtagere af vores information



- Informationer sælges/misbruges/tages ud af konteksten: Mister kontrol, følte data var hemmeligt (i den nye situation)
- Vi glemmer hvor længe data eksisterer (Holder op med at tale når personen nærmer sig >< skriver på Facebook -> samtalen bliver indekseret en dag, eller personen bliver venner med en af forfatterne)



Overvågning og privacy

Ikke nødvendigvis dårligt:

Stor værdi i Big Data, i politiets og i efterretningstjenesternes arbejde



Registre

### Privacy er interessant

Vi taler om vidt forskellige vigtige issues, det forvirrer billedet hvis de blandes sammen i én gryde:

NSA m.fl
I DK og udenfor DK
Politi og efterretning

Kommerciel profilering

National overvågning

Offentlige- og virksomhedsregistre, Nationalt Genom Center Google
Facebook
Reklame-trackere
og anden kommerciel
profilering



### Privacy er interessant



Alle er vigtige - men de kræver forskellige hensyn og midler

- Lovgivning (DK, EU, US, andre)
- Krypto fra virksomheder (SSL, mail osv)
- Privacy beskyttende løsninger
- Bruge krypto selv (mail, diskskryptering, SMS osv)
- Ad-blockers, flere browsere, log-out, VPN



Privacy

Privacy drejer sig om kontrol over data:

Hvem får data, hvordan bliver det brugt, gemmer de det, hvem bliver data delt med og kan man få det slettet



### Trusler imod privacy

Uværdighed ved udstillelse Forlegenhed

Samkøring Bruger/kunde profilering Identitetstyveri Osv, osv, osv

Vil du accepterer pris differencering?

I'm Feeling Lucky



### Grundprincipper

Fair Information Practices (FIPs) - 1970ies (5 regler)

- 1.User should know about the data collection
- 2. User should be able to see collected data
- 3.Data should only be used for the purpose data was collected for
- 4. User can correct, amend or delete collected data
- 5. Data should be securely stored

(Ingen "collection limitations")



## Den Europæiske Unions Charter om Grundlæggende Rettigheder

### Artikel 7 Charter: Respekt for privatliv og familieliv

Enhver har ret til respekt for sit privatliv og familieliv, sit hjem og sin kommunikation.

### Artikel 8 Charter: Beskyttelse af personoplysninger

- 1. Enhver har ret til beskyttelse af personoplysninger, der vedrører den pågældende.
- Disse oplysninger skal behandles rimeligt, til udtrykkeligt angivne formål og på grundlag af de berørte personers samtykke eller på et andet berettiget ved lov fastsat grundlag. Enhver har ret til adgang til indsamlede oplysninger, der vedrører den pågældende, og til berigtigelse heraf.
- 3. Overholdelsen af disse regler er underlagt en uafhængig myndigheds kontrol.



### GDPR EU-lovgivning

En "forordning", dvs lovgivningen er "ens" i alle EUlande

Lovgivningen betyder bl.a.:

- Der kan udstedes bøder på op til 100 millioner euro eller op til 4 procent af en koncerns globale årlige omsætning
- Krav om underretning inden 72 timer ved sikkerhedsbrister (data breach notifikation) med datatab til Datatilsynet - og i nogle tilfælde alle de berørte kunder



### Nye GDPR EU-lovgivning

Mange virksomheder skal udnævne Databeskyttelses-medarbejdere (Data Protection Officers/Databeskyttelsesrådgivere)

- DPIA Privacy Impact Assessments
- Gruppesøgsmål
- Data protection by design and by default



## Hvad forstås ved personoplysninger?

Enhver form for information om en <u>identificeret</u> eller <u>identificerbar</u> fysisk person

Identificerbar: direkte eller indirekte identifikation gennem f.eks. navn, ID-nummer, journal nr. osv.

Alle rimelige hjælpemidler skal tages i betragtning

Både subjektive oplysninger (vurderinger) og objektive oplysninger

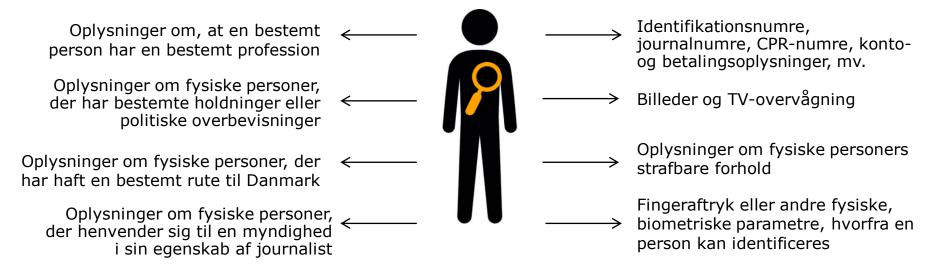






## Eksempler på personoplysninger

"Enhver form for information om en identificeret eller identificerbar fysisk person (den registrerede)"



Billeder og TV-overvågning i USA?



### Personoplysninger

## **Opdeles** i to niveauer:

(Ikke personhenførbare oplysninger)

- Almindelige, ikke-følsomme personoplysninger
- Følsomme personoplysninger
   (f.eks. oplysninger om helbred, race, politisk baggrund, religiøs overbevisning mv.)
- Andre typer af følsomme personoplysninger (eksempelvis oplysninger om strafbare forhold, væsentlige sociale problemer mv.).
- Oplysninger til bedømmelse af økonomisk soliditet og kreditværdighed

### Personoplysninger

## Udgangspunktet er:

- Almindelige, ikke-følsomme personoplysninger må gerne behandles (udtrykkeligt angivet formål)
- Følsomme personoplysninger (f.eks. oplysninger om sundhed, race, politisk baggrund, religiøs overbevisning mv.) må IKKE behandles

"collected for specified, explicit and legitimate purposes"

#### Article 7

#### Respect for private and family life

Everyone has the right to respect for his or her private and family life, home and communications.

#### Article 8

#### Protection of personal data

- 1. Everyone has the right to the protection of personal data concerning him or her.
- 2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.
- 3. Compliance with these rules shall be subject to control by an independent authority.

### Samtykke

Et samtykke skal være en frivillig, specifik, informeret og utvetydig viljestilkendegivelse Dataansvarlige skal kunne bevise, at datasubjektet har samtykket til behandlingen Hvis samtykket gives skriftligt i en erklæring, der også vedrører andre forhold, skal en anmodning om samtykke forelægges på en måde, som klart kan skelnes fra de andre forhold

Samtykket kan gives mundtligt og skriftligt, herunder elektronisk

Afkrydsning af felt på hjemmeside – opt in, men ikke opt out (forudafkrydsede felter) Ikke stiltiende samtykke





### Undtagelserne – her gælder forordningen ikke

- Aktiviteter af ren privat eller familiemæssig karakter
  - F.eks. korrespondance, føring af adressefortegnelse eller sociale netværksaktiviteter og den type onlineaktiviteter
- Aktiviteter uden for EU-retten, f.eks. national sikkerhed
- Den fælles udenrigs- og sikkerhedspolitik
- Særlige regler for politi, anklagemyndighed og domstole
- Behandlinger i EU's institutioner



### Grundlæggende principper

### Personoplysninger skal behandles

- a) lovligt, rimeligt og på en gennemsigtig måde i forhold til den registrerede ("lovlighed, rimelighed og gennemsigtighed")
- b) indsamles til udtrykkeligt angivne og legitime formål og må ikke viderebehandles på en måde, der er uforenelig med disse formål ("formålsbegrænsning")
- c) være tilstrækkelige, relevante og begrænset til, hvad der er nødvendigt i forhold til de formål, hvortil de behandles ("dataminimering")
- d) være korrekte og om nødvendigt ajourførte ("rigtighed")
- e) opbevares på en sådan måde, at det ikke er muligt at identificere de registrerede i et længere tidsrum end det, der er nødvendigt til de formål, hvortil de pågældende personoplysninger behandles ("opbevaringsbegrænsning")
- f) behandles på en sikker måde ("integritet og fortrolighed").

## Artikel 8 Charter: Beskyttelse af personoplysninger

- Enhver har ret til beskyttelse af personoplysninger, der vedrører den pågældende.
- Disse oplysninger skal behandles rimeligt, til udtrykkeligt angivne formål og på grundlag af de berørte personers samtykke eller på et andet berettiget ved lov fastsat grundlag. Enhver har ret til adgang til indsamlede oplysninger, der vedrører den pågældende, og til berigtigelse heraf.
- Overholdelsen af disse regler er underlagt en uafhængig myndigheds kontrol.

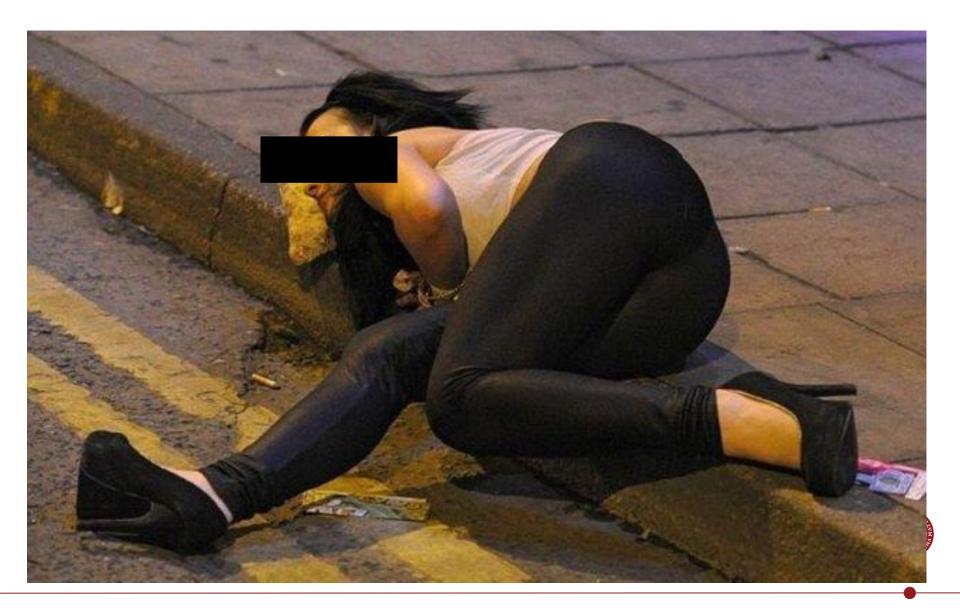
### "De registreredes" rettigheder

Adgang til data om sig selv Få data om sig selv rettet eller slettet

- Ikke uforholdsmæssig/overdreven i forhold til indsamlingsformålet
- Nyt samtykke ved ændring af formål



## Ret til at blive slettet?



### Persondata loven

# Samtykke er ikke altid nødvendigt Overførsel til udlandet "Behandling" af data

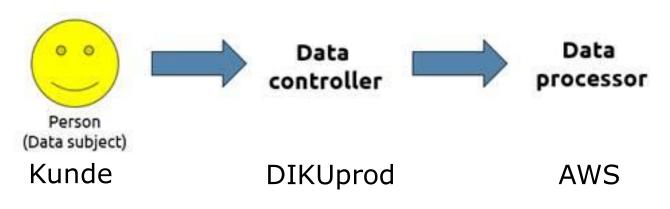
- Er en IP-adresse persondata?
- Hvorfor er et foto af en person jeg ikke kender persondata?
- Support i Indien?



### Privacy / Databeskyttelse (Data Protection)

Firmaet DIKUprod indsamler forskellig persondata om sine kunder bl.a. navn, adresse, mail og telefonnummer

DIKUprod er begyndt at bruge Amazon Web Services (AWS) cloud løsninger. Persondata overføres til AWS





Privacy / Databeskyttelse (Data Protection)

## Hvem ejer den persondata DIKUprod overfører til AWS?

Firmaet DIKUprod indsamler forskellig persondata om sine kunder bl.a. navn, adresse, mail og telefonnummer

DIKUprod er begyndt at bruge Amazon Web Services (AWS) cloud løsninger. Persondata overføres til AWS





Privacy / Databeskyttelse (Data Protection)

Privacy drejer sig om (din) kontrol over data:

Hvem får data, hvordan bliver det brugt, gemmer de det, hvem bliver data delt med og kan man få det slettet

Privacy != Secrecy

Man kan ikke skjule noget for big data, men man kan lave regler - og der kan være konsekvens hvis nogen bryder reglerne



### Privacy / Databeskyttelse er ikke secrecy

the grugq Retweeted





Jessy Irwin 🦙 🥏 @jessysaurusrex

privacy isn't | about hiding, | it's sharing | on ur terms |

(\\_\_/) || (• 人•) || / づ

22/12/2016, 19.46



# Mit nye firma Privacy i en cloud-verden?



# Mit nye firma

Rådmand vil registrere nattens

ballademagere

Rådmand i Arhus vil - som et forsøg - op

Rådmand Gert Bjerregaard (V) fremsen forhører sig om mulighederne for på fors natteliv.

 I dag kan en ballademager, der allered fortsætte "festen" på naborestauranten.

Har politiet har givet et tilhold, skal pe er et kraftigt signal og en sanktion, der

Det forudsætter dog et fælles register kunne godt tænke mig at indføre en fors Espersen om mulighederne for at starte Ritt: Ballademagere skal registreres 08. jan. 2008 18.14 Indland

Køhenhavns overhorgmester Ritt Bjerregaard er nu parat til ystem, der kan udelukke voldelige ker i hovedstadens natteliv.

> al kunne registrere og afvise ilhold af politiet, oplyser ariat. Initiativet kommer efter den oa knivstikkerier.

anisation, Horesta, jubler over Commune.

ringssystem på landsplan, men det som har været fremsynet, set taget det seriøst, siger Gry Asnæs for Horesta Natteliv.

aurationer ikke registrere e, der har solgt narko eller ingssystemet kræver derfor en



SKREVET AF MORTEN ESPERSEN

### BRØNDBY: GIV OS LOV TIL AT REGISTRERE BALLADEMAGERE

Brøndby beder politikerne om lov til at registrere ballademagerne til klubbens kampe.

Synes godt om 5 personer synes godt om dette.

Brøndby har de seneste dage fået skudt i skoene, at de ikke gør nok for at forhindre ballademagere til klubbens kampe. Brøndby vil ikke registerere klubbens udebanefans, men vil gerne registrere ballademagerne, hvilket de dog ikke må. Direktør Ole Palmå gør opmærksom på, at Brøndby altså bruger mange resurser på fan-arbejde.

 Dette er DBU vidende om og det burde alle politikere være orienteret omkring både i folketinget og i

Superligaen Matches

Odense BK - Brøndby 06/11 17:00

2.00 3.35 3.65



Københavns Borgerrepræsentation! Hvis de er i tvivl om vores indædte lyst til og etora arhaida for at minimara hallada i forhindalea mad fodholdkamna (vorae

## Hvad var problemet?



FORSIDEN	TESTRESULTATER	NYHEDER	GODE RÅD		VÆR /
Hvad søger du?:	Søg	Genveje: K	øb adgang	Klager A	Aftaler og køb

Presse » Bølleregister skal være mere sikkert

### Bølleregister skal være mere sikkert

Vi er alle interesserede i et sikkert natteliv. Men det skal ikke ske på bekostning af vores privatliv, skriver Forbrugerrådet i et høringssvar til Justitsministeriet.

Af Charlotte Friis Eriksen

Sikkerheden skal være i top, når danskerne går i byen. Og det kommende registreringssystem, også kaldet bølleregistret, kan holde ballademagerne ude af landets diskoteker og barer.

Men sikkerheden i systemet skal forbedres. Det skriver Forbrugerrådet i et høringssvar til bekendtgørelsen om bølleregistret, der har det formelle navn "Bekendtgørelse om videregivelse og behandling af oplysninger om restaurationsforbud".

Læs hele høringssvaret som pdf

#### Oplysningerne kan udnyttes

Bøllerne er ikke de eneste, der registreres. Alle gæster vil ifølge diskoteks- og restaurationsbranchen kunne blive bedt om at oplyse navn, cpr-nummer, fingeraftryk, e-mail og foto.



### LÆS OGSÅ OM

Private registre ude af kontrol Forbrugere fanges i databaser

#### NDRE HAR SET

Virksomheder har et ansvar

Derfor er det meget vigtigt, at alle gæsternes oplysninger behandles på den rigtige måde, skriver Forbrugerrådet i høringssvaret til Justitsministeriet. Ellers kan oplysningerne udnyttes kommercielt eller personligt.

#### Kunderne bør anonymiseres

"Det er fuldt forståeligt, at restaurationsbranchen vil holde bøllerne ude. Men man kan sagtens indsamle oplysningerne og samtidig kryptere dem, så hver kunde ikke kan identificeres med navn og cpr-nummer," siger Anette Høyrup, jurist i Forbrugerrådet.

Anonymiseringen kan blandt andet ske ved at tildele gæsterne numre, så de bliver uidentificerbare for andre end diskoteket eller baren. Anette Høyrup henviser til IT- og Telestyrelsens anbefalinger om nye digitale sikkerhedsmodeller.

Læs IT- og Telestyrelsens anbefalinger om nye digitale sikkerhedsmodeller (pdf)

#### Sikkerheden skal følge tiden

Mit nye firma tilbyder:

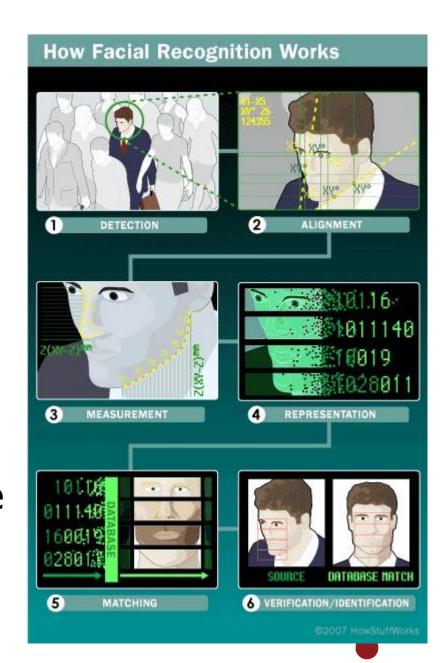
- Gratis kameraer med ansigtsgenkendelse til alle danske barer og diskoteker!
- Ingen registrering af navn, CPR eller anden personlig information kun *foto og køn*





Mit nye firma tilbyder:

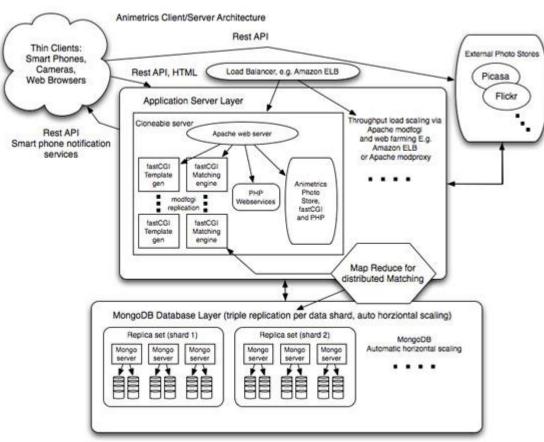
- Hurtig genkendelse, efter registrering.
- ■Ingen kø!
- Gratis udstyr <u>og</u> gratis deltagelse for alle barer og diskoteker



Mit nye firma tilbyder:

Cloud løsning, Serverless og webservices, alt

over internettet



Hvad får de – og hvad får jeg?



Hvad får jeg – egentlig?

- Liste over kendte ballademagere (kan sælges)
- Hvor er pigerne lige nu
- Hvilke barer og andre steder er populære

# Og hvad så?



Hvad får jeg – egentlig?

- Hvor skal der reklameres i byen
- •"Få en gratis drink hvis du ..."
- •Gå i Byen App
- •Få en gratis XZY, når du alligevel skal forbi Salon 39



Hvad får jeg – egentlig?

- Map til Facebook…
  - Gratis information, begrænset aldersrum,
  - Skal kun bruge DK-profiler, køn og byer er kendt

Teknisk muligt – men er det også lovligt i EU? hvad med USA?

En løsning med bedre data beskyttelse?



#### Privacy overvejelser

Designvalg Privacy vurderinger

## Privacy by Design Privacy Enhancing Technologies

## Model for privacy





#### Privacy by Policy

# Pause



"Before I write my name on the board, I'll need to know how you're planning to use that data."



## **Privacy forbedringer**



#### Privacy model

# Model for privacy



Hvor vil i ligge – for den specifikke løsning?



## Privacy by Design

## **Usability**



## Usability and Psychology

Brug af viden om menneskelige tankegang til at få mennesker til at tage "det rigtige" valg

Mennesker vælger ofte det, der er nemmest og det, der ligge lige for

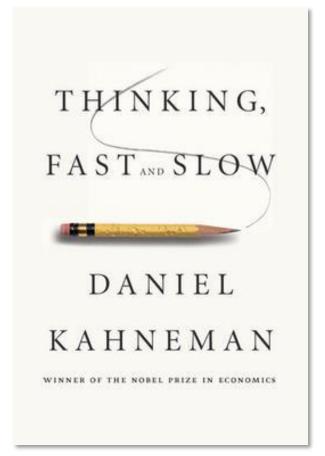


Usability and Psychology - mentale modeller

Mennesker har to adskilte systemer til at behandle information

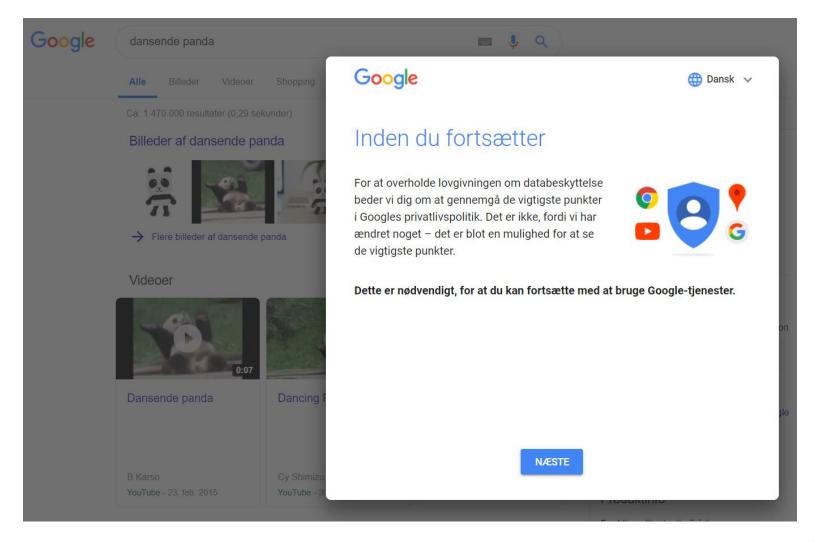
**System 1:** er hurtig automatisk og meget påvirkelig (doven autopilot)

System 2: behandler information langsomt, velovervejet og inddrager specifikke mål (logik og rationel tankegang)





## Brugerens mentale model





## Usability and Psychology



## Usability and Psychology - nudging

Sikkerheds nudging er f.eks.

Sikre defaults

Path of least resistance

Just in time

Feedback

OSV, OSV



## Usability / cookies

Users mental model når de går ind på en hjemmeside er ikke privacy eller sikkerhed

Overvej hvad brugerens forventninger er når de går ind på hjemmesiden eller bruger jeres app

Privacy er en "secondary task" for brugeren, det er ikke derfor de er på hjemmesiden - de vil forsøge at undgår/omgå secondary tasks

Vis ikke mere info end er brugbar for brugeren – dvs ikke alt cookie info, kun relevant eller selvvalgt...



Privacy by Design

## **Privacy Enhancing Technologies (PETs)**



## Privacy by Design - PETs







3 forskellige nøgler:

Er det ikke lettere at have den samme nøgle til hoveddøren som til pengeskabet og til cykellåsen ?

Risiko og forskellige sikkerhedsniveauer



Digital identitet - Trusler imod privacy

Forskel på kontanter og Dankort?



#### Digital identitet

## Diagnose baseret på blodprøve:

- a) CPR = samme nøgle
- Identificeret overfor lægen
- Identificeret overfor laboratoriet
- b) Ingen sammenkædning
- Ikke identificeret overfor laboratoriet
- Ikke identificeret overfor lægen



## Privacy by Design – Privacy Enhancing Technologies

Transaktionsisolering/transaktionsanonymisering

- Anonymous credentials ("Kontekstafhængige akkreditiver")
- Formålsspecifikke nøgler

Formålet er ofte ikke "identifikation af en bruger" men at få valideret et sikkerhedsaspekt ("er over 18", "kvinde", "er aktiv studerende ved DIKU", "er dansk statsborger")

Model for privacy

#### Identity og privacy

Selektiv afsløring af attributter fra akkreditiv

CPR: v ■

Kommune:

Køn: v

Alder: >18 år



# Bruger kan ikke direkte sammenkædes på tværs af serviceudbydere

## Serviceudbyder 1

CPR: \

Kommune:

Køn: v

Alder: >18 år

## Serviceudbyder 2

CPR: v

Kommune: KBH

Køn: v

Alder:

## Serviceudbyder 3

CPR: v

Kommune:

Køn: v

Alder: >18 år

Transaktionsanonymisering vs. kommunikationsanonymisering

## Kommunikationsanonymisering:

Bevidst: Ingen registrering af IP-adresser, mail, telefonnumre, cookies, device-ids, mac-adresser osv., osv

Øger sandsynligheden for virksomhed ikke kan identificerer parter

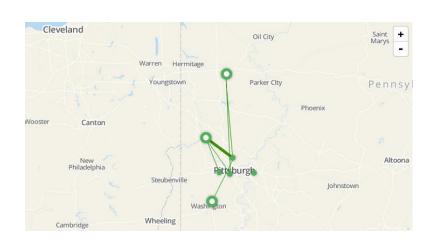
Se f.eks. Whistleblower systemer



#### Privacy by Design - dataminimering

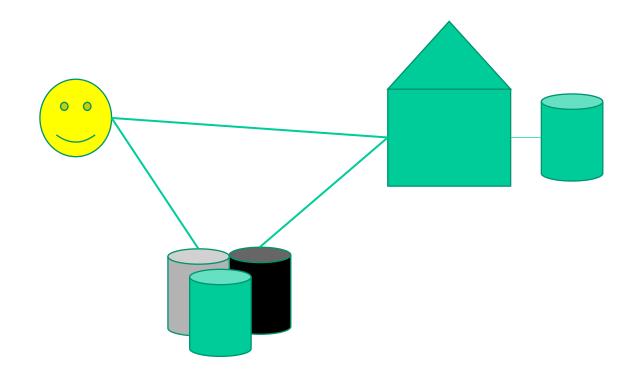
Save/compute: On device vs in the network

Undgå unikke identifiers





## Teknikker til vurdering af privacy





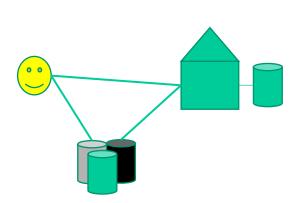
## Teknikker til vurdering af privacy

User Sphere: Device ID, SMS, mikrofon,contacts1. What user data do we collect?

Recipient Sphere: Virksomheden: statistik, intern profilering, info til 3.party (Y/N)

2. What do we do with your data?

3.party Sphere: Profiling, trackers, data mining



3. Who do we share your data with?

Do we allow 3.party access in any way? (directly, or do we share data later?)

## Et par privacy by design eksempler



#### Movia



# Slut med ove billetkontroll

Fra i dag er Movias billetk på uniformen.

## FAKTA Sådan beha og slag

Billetkontrollører får kameraer på uniformen efter flere trusler, spark og slag

Fra i dag bliver ansatte i Movia udstyret med bodycams.



#### Må man det?

#### Straffelovens § 263, stk. 1

"Med bøde eller fængsel indtil 6 måneder straffes den, som uberettiget

. . .

3) ved hjælp af et apparat <u>hemmeligt</u> aflytter eller optager udtalelser fremsat i enrum, telefonsamtaler eller anden samtale mellem andre eller forhandlinger i lukket møde, <u>som han ikke selv deltager i</u>, eller hvortil han uberettiget har skaffet sig adgang."

#### Argumenter

- Ikke "hemmelig" optagelse
- Kontrolløren deltager i samtalen
- Ikke forsæt til at optage bi-personer



#### Må man det?

#### Straffelovens § 263, stk. 1

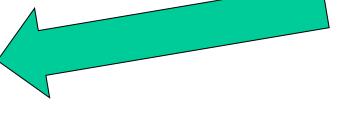
"Med bøde eller fængsel indtil 6 måneder straffes den, som uberettiget

...

3) ved hjælp af et apparat <u>hemmeligt</u> aflytter eller optager udtalelser fremsat i enrum, telefonsamtaler eller anden samtale mellem andre eller forhandlinger i lukket møde, <u>som han ikke selv deltager i</u>, eller hvortil han uberettiget har skaffet sig adgang."

#### Argumenter

- Ikke "hemmelig" optagelse
- Kontrolløren deltager i samtalen
- Ikke forsæt til at optage bi-personer





#### Hvad skal der til?

## Sådan informerer vi

- Skilt på uniform
- Mundtlig information
- Folder
- Information på website
- Pressemeddelelse
- Interne retningslinjer

## Indsigtsret

#### Personer på optagelsen har indsigtsret;

- Kunder der kontrolleres
- Kontrolløren selv
- Andre (bi-personer)

#### Hvordan gennemføres indsigtsretten?

- Optagelsen sikres fra sletning
- Teknisk sløring af billede og lyd
- Sikker identifikation af kunden
- Sikker udlevering



#### Hvad skal der til?

#### Sikkerhed

- · Udleveres kun til politiet
- Skærm er deaktiveret og password-beskyttet
- Kontrolløren kan ikke se optagelsen
- Få medarbejdere har adgang
  - Ved udlevering til politiet
  - Ved håndtering af indsigtsbegæringer
- Automatisk sletning efter 14 dage

- Tag informationsdelen alvorligt
- Begræns lagring af optagelser
- Vær klar til at håndtere indsigtsbegæringer



## Privacy by Design - dataminimering



## Eksempler på dataminimering:

- Kun indsamle efter en hændelse
- Kun indsamle med mellemrum
- Ikke gemme data hvis alt ok
- Slette data efter 2 timer eller gemme 30 dage
- Ikke mere detaljeret information end nødvendigt, f.eks. generelt område >< detaljeret info</li>
- Benzin ok ja/nej?

"collected for specified,
explicit
and legitimate purposes"



## PIA og DPIA

(Privacy Impact Assessments og Data Protection Impact Assessments)



#### Privacy – risiko vurderingen er stadig vigtig

#### Threat model:

Skal der beskyttes imod uheld eller imod bevidste angreb imod privacy?

Hvem skal der beskyttes imod: Global overvågning, virksomheden selv, 3.part, delt it-udstyr hos brugerne, eller imod andre brugere af tjenesten?



#### PIA - Privacy Impact Assessment

## Overvej:

#### Den data der indgår

Følsomhed, hvad bruges data til, potential konsekvens ved brud osv

#### Teknologierne der bruges

Nye/ukendte, "intrusive" osv

#### **Data-flows**

Hvordan data indsamles, opbevares, bruges, slettes -

I organisationen, udenfor organisationen, udenfor EU osv



#### Start med persondata-flow analyse

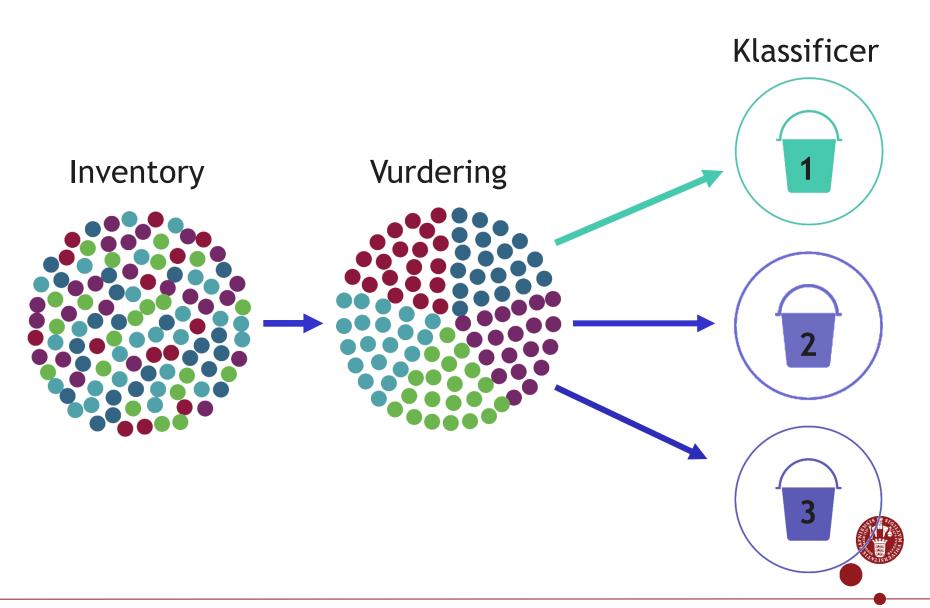
#### Skab overblik:

Dataflow analyse skal afklare:

- Hvor kommer persondata ind?
- Hvor går data hen?
- Hvor får 3.parties adgang til persondata?
- Er databehandler aftaler mm på plads?
- Hvad er hjemmelsgrundlag?
- Osv

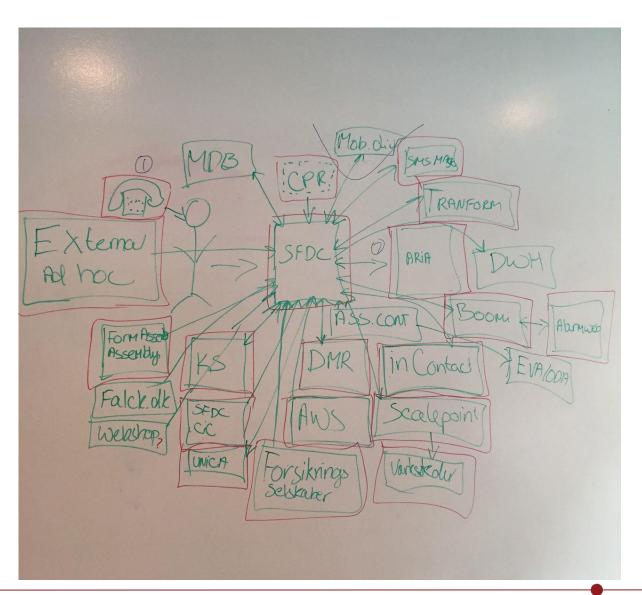


#### Data klassifikation



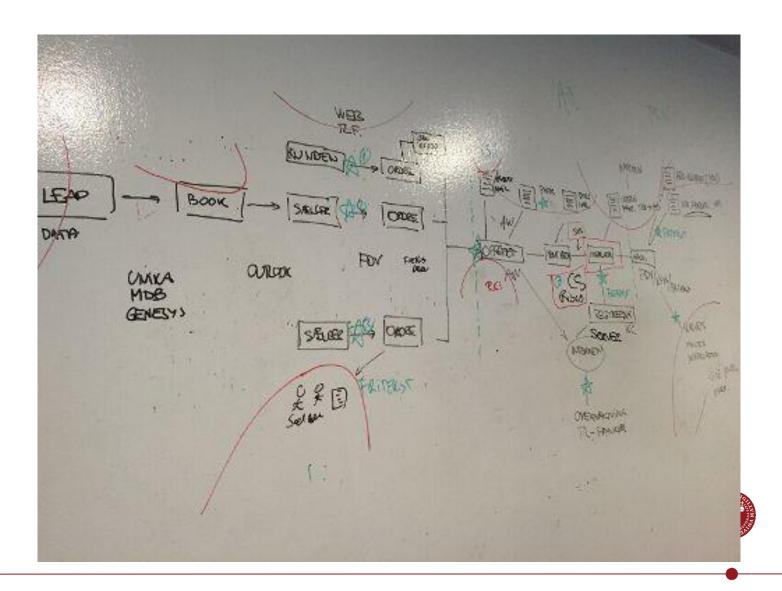
#### Sample "Level 1" test-flow

#### **Analyse 1**



### Sample "Level 1" test-flow

#### Analyse 2



#### Persondata - hjælp

#### Persondatalovens sikkerhedskrav

Udgangspunktet for loven er, at en løsning der behandler persondata ALTID skal overholde de nedenstående 12 grundkrav:

#### Grundlæggende krav

#### 1. Er der lavet en risikovurdering?

Kravene til en løsning skal altid vurderes ud fra de specifikke omstændigheder. Man skal derfor ALTID lave en risikovurdering og en Privacy Impact Assessment, der beskriver hvilke ekstra krav, udover de 12 grundkrav, den specifikke løsning kræver.

#### 2. Bliver persondata behandlet sikkert?

Der skal være sikkerhedspolitikker, procedurer og andre retningsliner i virksomheden, der beskriver, hvordan it-sikkerheden konkret er etableret i organisationen.

Lokale sikkerhedspolitikker/procedurer kan være nødvendige.

Politikker, procedurer og retningslinier skal efterleves, overholdelsen kan revideres af ekstern itrevision.

#### 3. Ved medarbejderne hvordan de skal behandle persondata sikkert?

Medarbejdere, der behandler persondata skal instrueres i hvordan de skal behandle data. (Der skal være awareness aktiviteter i forhold til sikkerhedspolitikker, procedure og persondataloven)

#### 4. Er kontrakter på plads med underleverandører?

Der skal være etableret skriftlige aftaler med databehandlere til sikring af, at datasikkerheden lever op til persondataloven, samt at den dataansvarlige påser dette.

Dvs er databehandler aftaler og underdatabehandler aftaler på plads?

Kray kan ligge både i den underliggende infrastruktur og i den daglige brug af en applikation/løsning:

Infrastruktur		Applikationen	
5.	Fysisk adgangskontrol	5.	Fysisk adgangskontrol
	Uautoriseret adgang skal kontrolleres.		Fysisk adgang til applikationen/anden
	Dette dækker fra begrænset adgang til		persondata skal beskyttes.



## **Data protection technologies**



#### Anonymisering/pseudonymisering

## **Anonymisering** >< Pseudonymisering

'Pseudonymous data' means personal data that cannot be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organizational measures to ensure non-attribution;

Anonym data er ikke persondata, pseudonymiseret data er persondata



#### Anonymisering/pseudonymisering

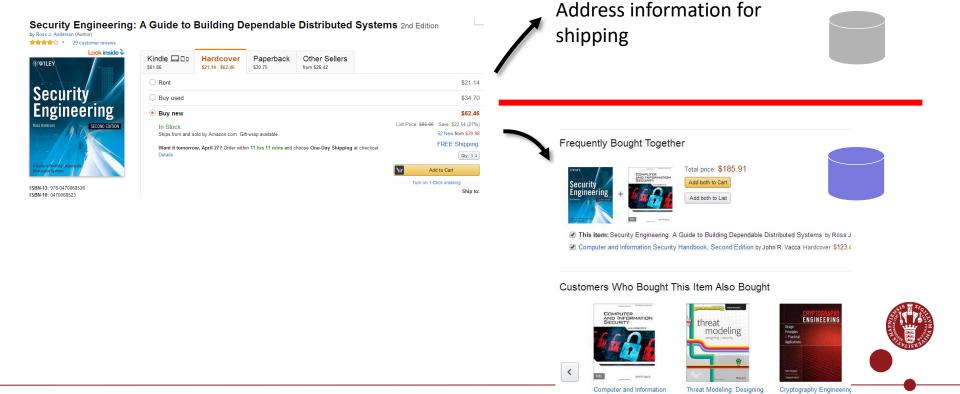
## **Anonymisering** >< Pseudonymisering

Begrænse risiko ved datatab Dele data internt/eksternt Forskning osv...



# Eksempel på pseudonymisering: Kun adgang til persondata med specifikt arbejdsbehov



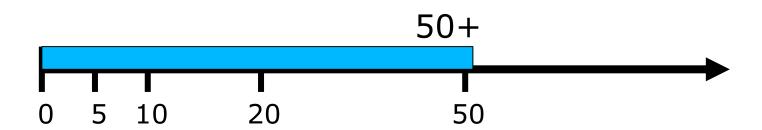


Privacy by Design – dataminimering igennem pseudonymisering

## Hvor mange gange har en bruger brugt "Hjælp":



Tydeligt tegn på usability problemer, bør måske overveje vores løsning







#### 1b. Pseudonymise personal data as soon as possible

"Pseudonymisering": udvalgte kategorier af personpdata erstattes med "koder" så data ikke kan henføres til en specific person uden nøglen kendes. F.eks. kan CPR-numre erstattes af en kode, der opbevares et andet sted. (cpr 123456-1234 = ID 12345abc)

Pseudonymiseret data er stadig persondata fordi det stadig er <u>muligt</u> at koble oplysningerne sammen.

Pseudonymiseringen kan give en bedre beskyttelse fordi det ikke umiddelbart er muligt at genkende individer.

#### Pseudonymisering

**Vi indsamler:** CPR, mail, købshistorik, alder. Vi har tre kunder:

CPR	Mail	Købshistorik	Alder
112233-1122	aaa@mail.com	Bog	18
445566-1234	bbb@gmail.com	TV	55
321321-6543	cccc@live.dk	Avis	55

#### Pseudonymisering af kunde data:

CPR	Mail	Købshistorik	Alder
а	1	Bog	18
b	2	TV	55
С	3	Avis	55

Vi kan nu lave big data:

18 årige fortrækker bøger, 55 årige foretrækker tv og aviser



Men pas på

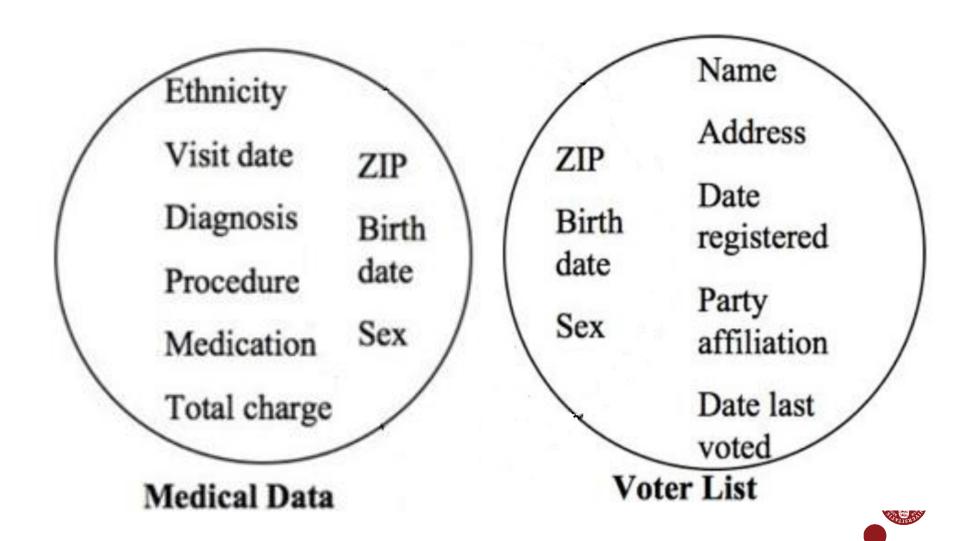
**Singling out** er muligheden for at isolerer data om et individ

Linkability er muligheden for at sammenkæde to eller flere records i en eller flere databaser
Hvis man kan identificerer at to records matcher same gruppe individer, men ikke bestemme individet i gruppen beskytter teknikken imod "singling out", men ikke imod linkability

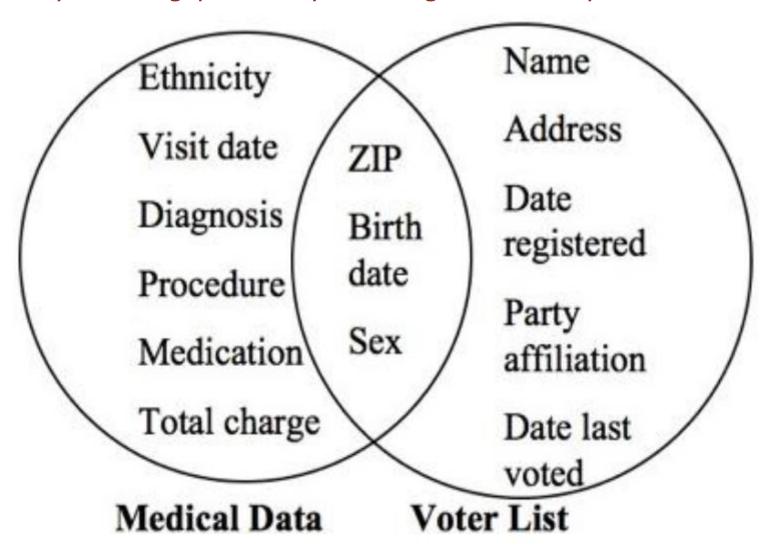
**Inference** er muligheden for at udlede en værdi udfra en anden værdi.



#### Anonymisering/pseudonymisering - linkability



#### Anonymisering/pseudonymisering - linkability





#### Pseudonymisering

Vi indsamler: CPR, mail, købshistorik, alder.

Vi har tre kunder:

CPR	Mail	Købshistorik	Alder
112233-1122	aaa@mail.com	Bog	18
445566-1234	b@gmail.com	TV	55
321321-6543	cccc@live.dk	Avis	55

#### Pseudonymisering af kunde data:

CPR	Mail	Købshistorik	Alder
а	1	Bog	18
b	2	TV	55
С	3	Avis	55



#### Pseudonymisering

CPR	Mail	Købshistorik	Alder
112233-1122	aaa@mail.com	Bog	18
445566-1234	b@gmail.com	TV	55
321321-6543	cccc@live.dk	Avis	55

CPR	Mail	Købshistorik	Alder
а	1	Bog	18
b	2	TV	55
С	3	Avis	55

#### Vi ved

Alice er kunde og er 18 år: Alice har købt en bog, hendes cpr er derfor 112233-1122 – Singling out Bob har købt en avis, Bobs CPR er 321321-6543 – Inference





Effektiv anonymisering forhindre, at et individ bliver identificeret i et dataset, at 2 entries/datasæt sammenkædes og fra at udlede information

Afhængig af kontekst og formål for anonymiserede data vil der ofte være brug for forskellige teknikker for at forhinder identifikation af personen

Typisk vil det ikke være nok bare at fjerne direkte identificerbare elementer

Anonymisering er **en vidtgående teknik** for en virksomhed



# Data generalisering

(Modificering af skala eller magnitude (dvs en region i stedet for en by, en måned i stedet for en uge).

# Data randomisering

(Noice – random ændring af værdier, f.eks. højde +-10cm),

Permutation – blanding af værdier så nogle er kunstigt linket til andre data subjects)

Bør altid kombineres med fjernelse af tydelige attributer/quasi-identifiers

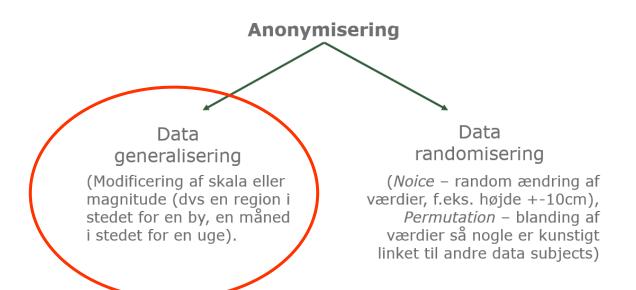


#### Anonymisering - Generalisering

#### Generalization

Generalisering kan effektivt beskytte imod Singling out, men beskytter ikke i alle tilfælde

Kræver specifikke ekstra overvejelser for at forhindre Linkability og Inference





#### **Data Generalization**

The main issue with k-anonymity is that it does not prevent inference attacks. In the following example, if the attacker knows that a specific individual is in the data set and was born in 1964, he also knows that the individual had a heart attack Furthermore, if we know that this dataset was obtained from a French organisation then each individual resides in Paris as the first three digits of Parisian postcodes are 750\*).

Year	Gender	ZIP	Diagnosis
1957	M	750*	Heart attack
1957	M	750*	Cholesterol
1957	M	750*	Cholesterol
1964	M	750*	Heart attack
1964	M	750*	Heart attack

Table 2. An example of poorly engineered k-anonymisation

#### Anonymisering er svært (inferens)



To find Tony Blair's record, for example, you'd look for all patients who underwent cardioversion and catheter ablation at Hammersmith Hospital on October 19th 2003

#### Anonymisering er svært







#### Data Generalization: K-anonymity

K-anonymity teknikker grupperer attributter med mindst K andre for at forhindre én person bliver identificeret

Attributter generaliseres så individerne deler samme værdi

Hvis k er for lille vil inference-angreb selvfølgelig være mere effektive:

K=2 eller K>10

Alder	,	Alder
18		<20
55	K=2	>20
55	1 2	>20

#### Data Generalization – I-divertity

## **L-diversity**

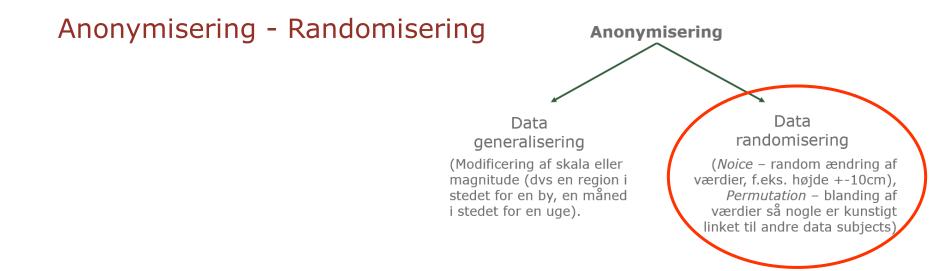
L-diversity udvider k-anonymity til at sikre, at <u>alle</u> <u>klasse</u> af attributter har mindst L forskellige værdier

L=10 i "Diagnosis", "Zip" og "Year"

Year	Gender	ZIP	Diagnosis
1957	M	750*	Heart attack
1957	M	750*	Cholesterol
1957	M	750*	Cholesterol
1964	M	750*	Heart attack
1964	M	750*	Heart attack

Alder
18
55
55





**Randomization** er grupperings-teknikker. Fjerner sammenkædningen mellem data og individet

Data gøres tilstrækkeligt upræcist til det ikke kan henføres til et specifikt individ

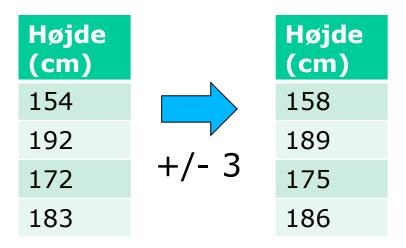
Noice/permutation



#### Anonymisering – Noice (Randomisering)

**Noice** ændrer atributter i et datasæt så de er mindre nøjagtige, men distributionen beholdes

Hvis højde oprindeligt blev målt til nærmeste centimeter kan anonymiseret datasæt f.eks. være præcist til +-10cm (random)





#### Anonymisering – Permutering (Randomisering)

**Permutering** blander værdier i en table så de bliver kunstigt linket til et andet data subject Nyttig når det er vigtigt at beholde distribution af attributer i datasættet

Kan være random værdier (kan være svært, "ændre alle 2'er til 8'er" er ikke nok fordi mønstret kan genkendes)

Alternativt flyttes værdier fra en record til en anden -> range og distribution er uændret, men korrelationen mellem værdier og individ er ødelagt

### Anonymisering – svag permutering

Year	Gender	Job	Income (permuted)
1957	M	Engineer	70k
1957	M	CEO	5k
1957	M	Unemployed	43k
1964	M	Engineer	100k
1964	M	Manager	45k

Table 1. An ineffective example of anonymisation by permutation of correlated attributes



Both generalisering- og randomiserings teknikker har svagheder men begge kan have privacy værdi - afhængig af formål og situationen

"Identifikation" er ikke kun at identificerer en persons navn eller adresse. Husk også potential identifikation via:

> singling out, linkability og inference.



Fuld anonymisering er svært i praksis

Location data: nok med 4 location datapoint til entydig at identificere en person (1.5 mio personer i undersøgelsen) (hjem, arbejde, træning, ven, eller lign)

=> Svært at fjerne data nok til at anonymisere i praksis. Data skal stadig være brugbart og jo mere coarse, jo mindre brugbart i praksis

## Netflix long tail problem:

http://www.cs.utexas.edu/~shmat/shmat\_oak08ne tflix.pdf

Anonymiseringsteknikker kan give fuld databeskyttelse – men kun hvis det gøres rigtigt

Dvs kontekst (krav) og formål skal være helt klare for at kunne opnå rette niveau af anonymisering

Yderligere teknikker kan være nødvendige for at sikre imod data kan unikt identificerer et individ



#### Singling out

Areas of interest to identify a specific person

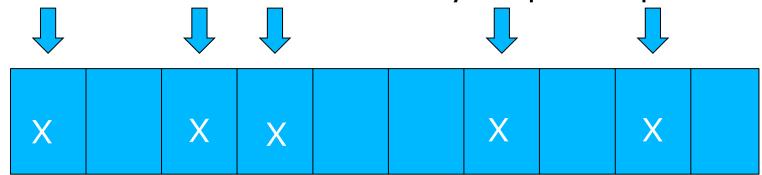


Potential areas of tracking:



#### Singling out

Areas of interest to identify a specific person

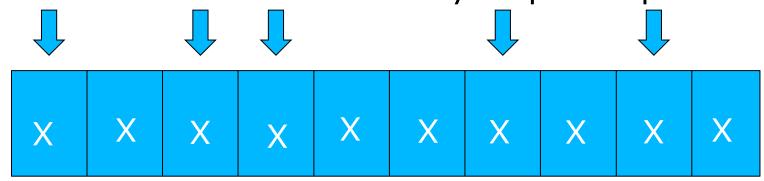


Potential areas of tracking: Match



#### Singling out

Areas of interest to identify a specific person



Potential areas of tracking: Still Matches

Adding noice does not protect the person against de-anonymization (but could help others)



### Hvor unik er jeg?





#### Differencial privacy

Kan f.eks. bruges til at finde

- Gennemsnit
- Om bestemte atributter er til stede
- Frekvens

"using hashing, subsampling and noise injection to enable... crowdsourced learning while keeping the data of individual users completely private."

Interesseret i gennemsnit? Check med 0,1%, eller er 10% tilstrækkeligt = Sampling



#### Privacy er interessant (og vigtigt)



- Din kontrol over din data
- Ikke samme nøgle til hoveddøren som til pengeskabet og til cykellåsen
- Tænk data beskyttelse med ind i arbejdet



## Spørgsmål



