

Big data – privacy og vækst

Man siger, at vi med internettet lever i en global landsby, hvor alle ved alt om alle – men forskellen er en langt højere grad af ulighed i adgangen til oplysningerne. Alle ved **ikke** alt om alle. Nogle få ved til gengæld rigtig, rigtig meget om rigtig mange.

Den ulige adgang til data

Staten, kommunerne og rigtig mange virksomheder opbevarer data. Data om dig og mig og os alle sammen og det er rigtig mange data. Det har vi ikke altid givet lov til. Vi har ikke bevidst og aktivt givet de data fra os. Og har vi for eksempel været ved lægen eller indlagt på hospitalet, har det været en nødvendighed uden alternativer. Det er uafklaret, rent juridisk, hvem der ejer disse data. P.t. opfører staten sig som om den ejer de data, der samles ind gennem den offentlige sektor og virksomhederne samler data og holder på dem. Det kan være dine indkøb, elforbrug, dine kondital og løbetider eller hvor du er og har været henne i de sidste par uger. EU kommissionen anslår, at værdien af europæernes persondata har en potentiel vækstmulighed på tæt på 1 mia. euro frem til 2020.¹ Set fra virksomhedernes side gradueres data på deres værdi. Datamængder som kunne være til gavn for samfundet via forskning eller udvikling af programmer eller nye apps, bliver således ikke frit tilgængelige, så længe de har en værdi for virksomheden. Virksomhederne er her for at tjene penge. Et nyere men efterhånden godt brugt ordsprog hedder "Når produktet er gratis – er det dig, der er varen". Spørgsmålet er, hvornår vi som dataudbydere begynder at tage penge for at være produktet? Og kender vi overhovedet vores markedsværdi?

Hvad Danmark ved om dig

En mulighed er at kræve at offentlige virksomheder viser alt den data de har til rådighed om en, når man går ind på deres

¹ Factsheet: Data protection: Progress on EU reform now irreversible after European Parliament vote, 12. Marts 2014: http://ec.europa.eu/justice/data-protection/law/index_en.htm

hjemmeside. Det kunne gøres ved en knap på alle hjemmesider, der giver adgang til at se al info de har om dig. Formålet er, at man skal kunne se mængden af data, der findes og man skal kunne se sine egne data. Princippet er, at data skal være tilgængelige for den, der har leveret dem.

Hvorfor klarere lovgivning om databeskyttelse?

Kun 26% af de europæiske brugere af sociale netværk og 18% af brugere, der handler på nettet føler at de har fuldstændig kontrol over deres persondata på nettet.

43% af europæiske internetbrugere mener, at de er blevet spurgt om flere informationer end nødvendigt.

58% af europæiske forbrugere føler, at der ikke er et alternativ til at oplyse krævede personlige oplysninger ved køb eller service på nettet.

90% af europæerne ønsker samme lovgivning over hele Europa.

Kilde: Factsheet "Why do we need an EU data protection reform?" http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm

Privacy by default

Når man besøger private virksomheders hjemmesider bliver der opsamlet en ukendt mængde data. Ikke bare om, hvad du foretager dig lige nu på denne specifikke hjemmeside, men også via cookies om, hvad du ellers foretager dig på nettet. Løsningen på det er privacy by default. Dvs. et software design koncept, der forhindrer opsamling, fremvisning eller videregivelse af enhver form for personlige data uden eksplicit tilladelse fra brugeren. En af sidegevinsterne ved

privacy by default er, at mange brugere ikke ved eller kan gennemskue, hvilke og hvor mange oplysninger om dem, de videregiver ved at besøge en hjemmeside eller være aktive på de sociale medier. Udfordringen er at finde en standard, der tager højde for forskellige typer software, forskellige brugerbehov og som virker, hvor den skal. Men uden at blive et forstyrrende irritationsmoment ved for eksempel konstant at præsentere brugeren for advarsler og krav om godkendelse, før man kan komme videre.

Big data CSR

En anden mulighed er at skabe et valg for brugeren. Når vi køber dagligvarer som kaffe kan vi vælge bæredygtige mærker. Når vi køber ind på nettet kan vi vælge web-butikker med e-mærket. Det har en omkostning, ofte på prisen, alligevel er det et marked, der er vokset voldsomt de seneste 10 år. På samme måde burde man kunne vælge hjemmesider, der tager hensyn til dit behov for privacy. En privacy-venlig hjemmeside kunne give mulighed for at vælge en "valgt fra" knap. Det giver mulighed for at vælge information eller handlinger fra, hvis du synes, at du i samme omgang afgiver for mange informationer. Omvendt kan man også give mulighed for at afgive information, hvis det er en fordel. Nogle kan for eksempel se det som en kæmpe fordel, at man ved at afgive informationer via GPS om, hvor man kører, kan modtage gode råd om trafikforhold, benzinforbrug, servicestationer og kørselsøkonomi. Andre vil gerne kunne køre uden at der bliver registreret andet end for eksempel vejtype eller hvor meget man har kørt indenfor et større område.

Konkret kan det gøres ved at klassificere ved hjælp af vejledning og standardisering.

Privacy

Privacy handler ikke om at skjule oplysninger, men om retten til at kunne bestemme over informationer om os selv.

Privacy by design

Når hensyn til privatlivet tages med i opbygningen af systemet. Privacy by design er kendetegnet ved 7 fundamentale principper:

1. Det er proaktiv beskyttelse,
2. privacy by default, dvs. automatisk beskyttelse som standardindstilling,
3. privacy er indlejret i it-design og ikke noget man vælger til,
4. beskyttelse ses som plussum og ikke som et trade off med f.eks. sikkerhed,
5. fuld –livscyklus beskyttelse, hvor data bliver sikkert opbevaret og sikkert destrueret,
6. uanset hvilken praksis, der bruges skal det være åbent og verificerbart for alle og endelig
7. brugercentreret, dvs. brugervenligt og med fokus på brugerens privatliv.

Privacy by default

En standardindstilling, der forhindrer indsamling, visning eller videregivelse af personlige data.

Big Privacy

Er baseret på Privacy by design men udvidet til også at omfatte netværk, værdikæder og økosystemer ved brug og produktion af Big data.

Standarder for privacy

Der findes følgende standarder til beskyttelse af privatlivets fred på internettet:

- ISO 27001

Tidssvarende lovgivning er det nødvendige udgangspunkt

Den nuværende lov om persondatabeskyttelse er 19 år gammel, forældet og yder for svag beskyttelse. Det er afgørende vigtigt med lovgivning – virksomhederne gør det ikke af sig selv. Lovgivning kan være med til at sikre motivation for at der investeres i kryptering. Og samtidig er netop klar lovgivning en af forudsætningerne for at virksomheder tør binde an med Big data. Selv med god velvilje kan det være svært at gennemskue, hvor grænserne går, når det gælder personfølsomme oplysninger. Jura er derfor et vigtigt pejlemærke for, hvad der er i orden og hvad der ikke er acceptabelt.

Men samtidig er udfordringen: Hvad hjælper lovgivning, hvis vi ikke ved, at den er brudt? Dataopsamling sker "bagved". Der er behov for meget skarpere datarevision, så virksomhederne ikke overtræder reglerne. Man kan sammenligne med karteldannelse, hvor den største udfordring er opdage, hvad der faktisk foregår.

Big data og persondata har tre brudflader, der skal tages vare på:

Data høst: Når der opsamles data i dag er det langt fra altid, at de enkelte borgere aktivt har givet tilsagn om at deres private data må opsamles og hvad de må bruges til. Og ofte er folk slet ikke klar over, hvad der registreres.

Data mining: Her renses data og analyseres i forhold til andre data med det formål at skabe sammenhæng mellem aktuelle forhold og dermed give os ny viden. Det kan være i forskningssammenhænge, men det kan også være til rent kommercielle forhold.

Applikationsfasen: Datamining resulterer i en algoritme, der er en generaliseret form af den nye viden. Værdien af den ligger i koblingen til den virkelige verden, f.eks. et datasæt af skoleelever eller kunder. Afhængig af algoritmen kan man herved for eksempel forudsige handlinger eller præferencer. Et kendt eksempel er "Dem, der lånte denne bog, kunne også lide xxx". Mere bekymrende kan det blive, når man forsøger at differentiere priser på eksempelvis rejser alt efter, hvilke præferencer man i øvrigt

ved at kunden har. Et eksempel på det er Safari-sagen.²

En lovgivning på området skal tage hensyn til disse tre brudflader og til at der i dag er tale om langt større muligheder for opsamling af data og samkøring af data meget hurtigt og meget billigere end tidligere.

Skadevolder betaler

Sikkerhed kan også motiveres ved at betale for omkostningerne ved fejl. Eksempelvis ved identitetstyveri, der har både store menneskelige, men også økonomiske konsekvenser for den, der rammes. Firmaer bør modsat i dag have pligt til at oplyse, når der sker datalæk og de skal bære de økonomiske omkostninger ved "oprydningen" ved f.eks. identitetstyveri. En metode er altså at kunne gøre sikringen af brugernes privatliv til et problem for de udbydere, der indsamler data.

Hvem holder øje?

Det er Datatilsynet, der fører tilsyn med om reglerne overholdes i Danmark. Men antallet af inspektioner er ikke vokset i takt med udviklingen på internettet³. Det er afgørende, at Datatilsynet har ressourcerne og uafhængigheden til at kunne løfte overvågningsopgaverne i takt med at digitaliseringen stiger både private og i det offentlige. Både fordi vi som privatpersoner bruger nettet mere og mere, men også som en logisk konsekvens af den stigende digitalisering af den offentlige sektor og frigørelse af offentlige data. Presset på sikkerheden stiger med digitaliseringen og det stiller krav til vogterne. Et bud kan også være at supplere med et skarpere tilsyn med reglerne ved hjælp af datarevisorer som vi kender det fra regnskabsrevisorer.

² Læs mere på conferences.sigcomm.org/con-next/2013/program/p1.pdf

³ Se <http://www.datatilsynet.dk/publikationer/datatilsynets-aarsberetninger/>

EU kommissionens Data Protection Reform er vejen frem

EU kommissionens Data Protection Reform også kendt som persondataforordningen, er en vigtig milesten på vejen til en mere fair og tillidsvækkende brug af de muligheder Big data giver. EU er på dette område en vigtig frontløber. Desværre støtter Danmark ikke forordningen pga hensyn til forvaltningstraditionen, modsat de øvrige lande. En lang række interesseorganisationer, herunder DI, PROSA, Forbrugerrådet Tænk m.fl. støtter reformen. Hovedpointerne heri er bl.a.:

Retten til at blive glemt: Bedre adgang til at få ens data slettet, når man vil det.

Dataportabilitet: Herunder at flytte egne oplysninger fra én udbyder til en anden.

Kontrol over egne data: Retten til aktivt at sige ja til opsamling af persondata og virksomheders og organisationers pligt til at advare ved brud på datasikkerhed.

Databeskyttelse først: Privacy by design og Privacy by default indgår som krav i EU's persondataforordning, også ved f.eks. sociale netværk.

FN's Verdenserklæring om Menneskerettigheder 1948, artikel 12:
"No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks".

Den Europæiske Menneskerettighedskonvention, 1950, artikel 8, stk.1:
"Enhver har ret til respekt for sit privatliv og familieliv, sit hjem og sin korrespondance".

Uddannelse og dannelse

Big data åbner en verden af nye muligheder. Det stiller krav og modkrav. Hvordan får vi mest muligt ud af de nye muligheder og hvordan lærer vi at begå os i en verden, der ikke glemmer. Beskyttelse af privatlivets fred og sikring af it-sikkerhed i det hele taget, kræver en indsats på flere fronter. Både når det gælder uddannelsen af de it-specialister, der udvikler systemerne, og når det handler om at lære folkeskolebørnene at gebærde sig i det digitale samfund.

Uddannelse af ingeniører og teknikere

Erhvervsstyrelsen har i deres analyser konstateret en tendens til stigende efterspørgsel efter analytikere med stor indsigt i håndtering, strukturering, analyse og visualisering af data. Kernegruppen af eftertragtede analytikere består af matematikere, statistikere, dataloger, ingeniører og økonomer med de rette kompetencer.⁴ Danmark er i front på områder som befolkningens IT-parathed, men udviklingen går langsomt, når det gælder Big data. Der er derfor behov for en målrettet indsats i forhold til at styrke uddannelserne og efteruddannelserne på netop de kompetencer der kræves ved håndtering af Big data, hvis ikke vi skal falde helt bagud.

Samtidig er det vigtigt, at uddannelse og efteruddannelserne også får mere fokus på risici ved IT-sikkerhed og brud på privacy. Teknologi er hverken mere eller mindre end det, den bliver brugt til. Men i takt med at teknologien kan mere og mere, bliver risikoen for misbrug og konsekvenserne ved fejl som f.eks. lækager langt større. Skal vi bevare tilliden til fordelene ved Big data, skal bevidstheden om, hvordan IT-professionelle bruger data skærpes.

⁴ "Big data som vækstfaktor i dansk erhvervsliv", Erhvervsstyrelsen, Irisgroup, december 2013.

Big data kan skabe

- Ny viden
- Optimering af ressourcer
- Forudsigelser
- Overvågning

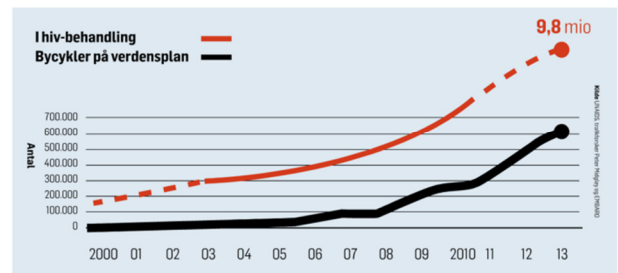
Om resultatet bliver til noget positivt eller negativt afgøres af måden vi bruger data på.

Human sensors og naturkatastrofer

Blandt de mere dramatiske, men også mest avancerede eksempler på brug af Big data er et forskningsprojekt på University of Illinois, USA, hvor professor Tarek Abdelzaher har fokus på social sensing og hvordan man får pålidelige informationer ud af "upålidelige" data som eksempelvis tweets under begivenheder som orkaner, jordskælv eller civile uroligheder. Kan man opsamle informationer fra menneskelige sensorer kan det være et afgørende parameter for at klarlægge omfanget af en katastrofe og sende den rigtige hjælp ud allerede mens det står på. Her er det ikke selve høstningen af data, der er problemet, da tweets netop er helt åbne og offentligt tilgængelige informationer givet bevidst af brugeren. Brugt på den rigtige måde kan det til gengæld effektivisere livsvigtigt rednings- og hjælpearbejde.

Skal forudsigelser have konsekvenser, så er det afgørende at være på det rene med, at statistisk signifikant sammenhæng ikke nødvendigvis er det samme som en videnskabeligt bevist sammenhæng. Korrelation medfører ikke kausalitet. Udfordringen blandt de kommende IT-professionelle er at sikre en bevidsthed omkring, hvordan vi sikrer at brug af data ikke ender i brud på sikkerheden eller ulovlig overvågning. Ligesom man skal være opmærksom på, om de nye

sammenhænge som Big data afdækker, faktisk er sande, så vi ikke risikerer falske forudsigelser:



Datajournalistik: Bycykler sikrer hiv-behandling

Skal vi sikre privatlivets fred skal privacy tænkes ind i og i langt højere grad integreres i vores software, som det eksempelvis gøre i privacy by design-tankegangen – det skal ingeniørerne opdrages til.

"Hvis det er gratis, er det dig, der er varen"

Brugerne skal opdrages til, at der ikke er noget, der hedder gratis. Virksomhedernes forretningsmodel er, at de ganske vist giver dig gratis adgang til at bruge f.eks. et spil, en app eller et socialt medie. Til gengæld får de masser af forbrugerdata tilbage, som de kan bruge til markedsføring, videresalg mv.

Hvis man kan gøre forbrugerne mere bevidste vil det også ændre virksomhedernes fokus, så tilbud om øget privacy bliver et konkurrenceparameter.

Bekymringen i befolkningen er stigende, men holdningen til, hvor grænserne går, er vidt forskellige. Der er endnu ikke en fasttømret kultur for, hvad der er acceptabelt og hvor vi sætter grænserne. Der er omvendt heller ikke endnu en bred erfaring med, hvad der er konsekvenserne af at udlevere oplysninger, ligesom det langt fra altid er klart for folk, hvor mange oplysninger man afgiver, eksempelvis ved at downloade en app som spillet "Angry birds".

Samtidig er forholdet til, hvad der er en del af privatlivet forskelligt i forskellige lande. I Sverige og Norge er alle selvangivelser eksempelvis offentlige. I England er kameraovervågning i det offentlige rum acceptabelt i langt højere grad end

i Danmark, til gengæld er der en lettere forbløffet skepsis overfor vores CPR-nr. system. Et projekt om road pricing i Nordjylland viste, at nogle af bilisterne ser en klar fordel i at kunne overvåge børnene, når de låner bilen. Og man kunne nemt forestille sig, at nogle gerne bytte private data for bilkørsel for mere og bedre information om kørsel i egen bil.

Men konsekvenserne er begyndt at vise sig og det har en effekt. På EU niveau siger 70 % af adspurgte europæere, at de er bekymrede. Flere lækager og en stigning på 77 % i datakriminalitet betyder, at der nu er en tendens til at det begynder at vende i befolkningernes holdninger.

Konklusion: Forebyggelse er billigere end...

Der er store fordele ved brug af Big data i forskning og i den offentlige sektor. Der er også meget, der tyder på, at der er en erhvervsfordel for Danmark ved at lade virksomheder få et eller andet niveau af adgang til de mange oplysninger om danskernes helbred, læge- og medicinforbrug lagret i cpr-registret i gennem snart en generation. Men en forudsætning for at vi kan dette er tillid til, at vi ikke pludselig står overfor negative konsekvenser i form af dyrere priser på eksempelvis rejser eller forsikringer, identitetstyveri eller offentliggørelse af følsomme oplysninger, for eksempel om vores helbred.

En måde at sikre fordelene og undgå ulemperne er at integrere data efter principperne i Privacy by Design. En anden er at skabe et etisk kodeks for, hvad der er god praksis for virksomheder på nettet. En tredje er at skabe efterspørgsel efter privacy niveauer, altså, at man som bruger kan vælge det niveau af åbenhed man er klar til. Formålet er ganske enkelt, at man som bruger er bevidst om, at man er en del af en handel og derfor selvfølgelig bør forhandle sig til den bedst mulige pris. Men også, at man giver rum for, at folk, der er villige til at afgive oplysninger også kan få fordelene af eksempelvis bedre trafikinformation eller mere målrettede reklamer.

Et dansk etisk moralkodeks?

Den engelske Information Commissioner's Office (ICO) har publiceret "Anonymisation: managing data protection risk", november 2012. Denne code handler om risikohåndtering i forbindelse med personfølsomme data og er et led i den britiske "open data agenda".

Se mere på:

<http://search.ico.org.uk/ico/search?q=anonymisation%3A+managing+data>

I Danmark har Datatilsynet en række vejledninger på deres hjemmeside, se www.datatilsynet.dk

IDA anbefaler

- Private og offentlige virksomheder bør på en let tilgængelig måde oprette adgang til alle de oplysninger som den pågældende virksomhed har om dig og hvad der ikke bliver slettet, når du forlader siden, f.eks. ved at man kan trykke på et ikon/knap.
- Privacy bør være en standardindstilling og ikke noget man skal gøre aktivt for at sikre. Der bør i højere grad udvikles pc'ere med indbygget software så fx browsere allerede har default indstillinger med højt privacy niveau, ligesom man allerede nu kan købe tablets med børnesikkerhedsindstillinger
- Kommunikation med det offentlige skal krypteres efter principperne om anbefalet post.
- Uddannelserne skal have fokus på privacy, sikkerhed og databeskyttelse.
- Etik og moral-parametre baseret på konsekvensrisiko skal fremmes på uddannelserne, så de unge kan sige fra, når kommende arbejdsgivere kræver uetiske ting af dem.
- IDA støtter den nye EU forordning på niveau med DI, PROSA og Forbrugerrådet Tænk, Institut for Menneskerettigheder m.fl.
- Forureneren betaler-princippet bør indføres. Omkostninger ved IT brud på sikkerheden og identitetstyveri betales af udbyder, fx bank eller virksomhed.
- Nødvendigt, at der sker en opdatering af lovgivningen, så den modsvarer det teknologiske kvantespring som Big data giver.
- Datatilsynet styrkes til at kunne rådgive om Big data, udstikke retningslinjer på niveau med ombudsmanden og gøres uafhængigt.
- Der skal startes pilotprojekter op med øget rådgivning til virksomheder om krav til privacy og mulighederne for at markedsføre sig på god skik.

- Det bør overvejes om der er behov for at supplere med krav om IT-revision.
- IDA bør udvikle et etisk kodeks som medlemmer skal overholde.

Vil du vide mere?

Kontakt

Chefkonsulent Grit Munk
gmu@ida.dk, tlf: 30596596

Chef for Politik, Presse og Analyse Søren Lauridsen,
sla@ida.dk, tlf: 30637723