

Big Data – personvernprinsipper under press



Datatilsynet

CIO Forum, Oslo 10.10.2013
Catharina Nes, seniorrådgiver

Big Data – personvernprinsipper under press

September 2013





Big Data er deg!





Neste pavevalg



Våre daglige gjøremål produserer store mengder data

Surfe på internett...



site tracking - clickstream - site analytics - email replication on devices via cloud infrastructure - social media trails

Ta en telefon...



location tracking (towers) - phone log files
data streams created/stored by apps

Kjøre bil...



location tracking at toll booths
instrumentation data (OnStar)

data streams initiated by ATM withdrawals
- transaction pattern logs
- audit/verification trails

Besøk i banken...



data streams from credit card transaction
product inventory changes
buying pattern logs (loyalty programs)

Ute å handle...

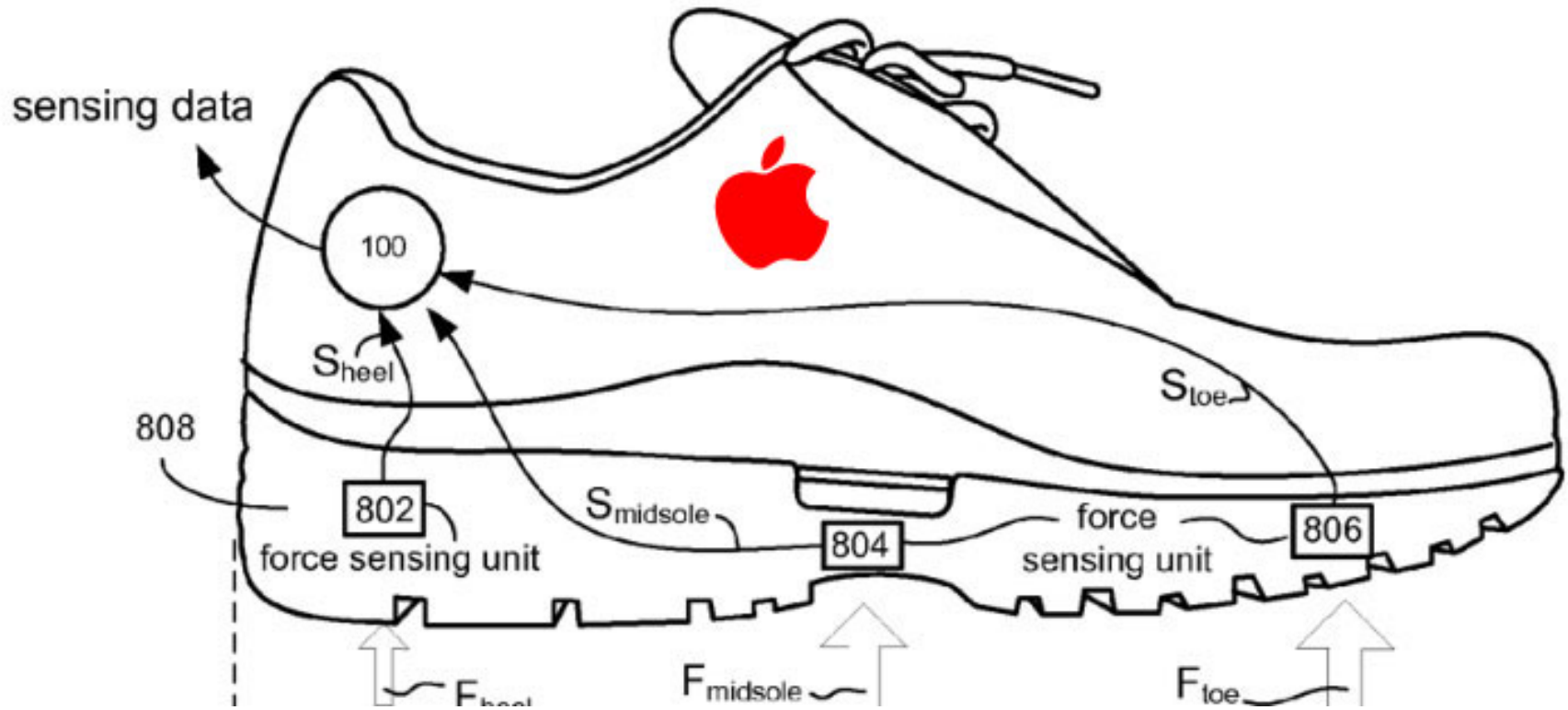


Data "eksos" produseres i vår omgang med moderne teknologi!

Slappe av hjemme...



smart meter
(electricity usage)



'Like' curly fries on Facebook? Then you're clever

'Like' curly fries? Then there's a good chance you've got a high IQ, according to a Cambridge University project to discover what we unwittingly reveal about ourselves on Facebook.



HVA PUTTER DU I POSEN? Er du Coop-medlem må du regne med at butikken har full kontroll på hva akkurat du handler. Informasjon om ditt handlemønster bruker kjeden for å lage tilbud som de tror at akkurat du vil bite på. (Foto: Kim Jansson)

Coop sender handlekurven din til analyse

Går gjennom alt medlemmene handler for å gi «relevante tilbud».



KREDITECHNOLOGY

Kreditech works like a big mosaic - Any online data that can be found about an individual will be used for fraud detection, identification, scoring. A decision /payout is made within seconds 100% automatically - no manual actions occur in the workflow. This generates customer value as unbanked individuals with no credit bureau history can get credit in 10min conveniently. Variable cost is 0 at unlimited scale thanks to fully automatized processes.



8000
DATA POINTS
PER APPLICATION MAX.

24 / 7
OPERATING

0
CHANCE FOR
FRAUDSTERS :)



Utfordring # 1: Bruk av data til nye formål

- Big Data = *gjenbruk* av data
- Utfordrer prinsippet om **formålsbestemthet**
- Vil vi få en utvikling mot svært brede samtykker?



Utfordring # 2: Datamaksimalisering

- Big Data = nytt syn på data
- Verdien ligger i dataenes *fremtidige* bruksmuligheter
- Dette utfordrer prinsippet om **dataminimalisering**
- Men hvem vil ønske å slette penger?



Utfordring # 3: Mangel på åpenhet

- Hvilke opplysninger samles inn om meg, av hvem og til hvilket bruk?
- Tap av kontroll!
- Rett til innsyn er en grunnleggende rettighet, men blir vanskelig å praktisere.



Big Data



Flurry Is Big Data

Flurry is a big data company. We're already handling terabytes of total data. Next step: exabytes. As we've scaled our big data clusters in the world, running on Hadoop.

Since first offering Flurry Analytics in January 2009, the volume of data we handle has grown consistently by 20% month-over-month. At our peak, we handle more than 25,000 session reports per second: that is over 3 times the total number of tweets per second managed by Twitter (source: @twitterglobalpr). Flurry tracks over 250 billion total data points per month, including any unique event or action tracked within any app sessions.

Big Data Builds Big Apps

Usage Statistics
Category Benchmarks
Custom Events & Parameters
Audience Segmentation
Custom Segmentation
More Power Features

←
• • •
→

IN-APP PURCHASES All Versions ▾ Across All Time ▾

Event Statistics
User Paths
Event Parameters
Event Logs

Level Purchase Made ▾

More Parameters. More Insight

Beyond measuring what consumers purchase, this developer also tracks at which level consumers purchase. Doing so, they can further optimize revenue. For example, most users in this game buy at level 1, but not again until level 8. Among this developer's goals should be to increase purchases before level 8 and/or to maximize engagement until players reach level 8, since that's when they'll likely purchase again.

Parameter Distribution Download CSV

Level	Count
Other	2.48M
15	301.56K
9	304.56K
8	410.66K
1	501.62K
10	284.34K
16	240.64K
11	230.14K
2	199.53K
12	196.53K

Utfordring # 4: Farvel anonymitet?

- Gjennom sammenstilling av data kan det oppstå risiko for reidentifisering
- Er anonyme data mulig i Big Data-alderen?

THREAT LEVEL

privacy

Netflix Spilled Your *Brokeback Mountain* Secret, Lawsuit Claims

BY RYAN SINGEL 12.17.09 4:29 PM

Follow

Like 177

Tweet 18

+1 0

5



Utfordring # 5: Sammenstilling kan avsløre sensitive data

- Sammenstilling av ikke-sensitive data kan gi sensitivt resultat
- Kan frembringe informasjon om:
 - seksuell legning
 - sykdommer
 - politisk overbevisning

The Creepiness Factor: How Obama and Romney Are Getting to Know You

APR 10 2012, 12:45 PM ET  Share 30  Tweet 233  +1 28

The presidential campaigns have the technology to know more about voters than any other bids in history.



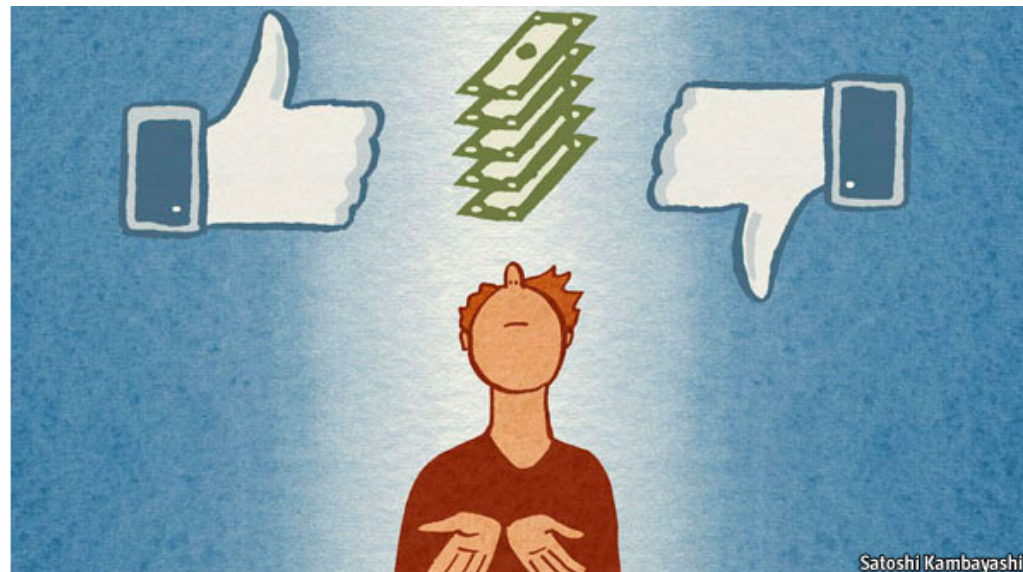
TECH | 2/16/2012 @ 11:02AM | 1,885,680 views

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

GreenCarReports.com

Utfordring # 6: "The dictatorship of data"

- Vi blir summen av våre digitale spor
- Kan befeste fordommer og sosial ekskludering
- Big Data glemmer kontekst: beslutninger treffes på uriktige data





Anbefalinger

for personvernvennlig bruk av Big Data



skaff samtykke

eller

anonymiser



Husk risikovurdering

Gode rutiner for anonymisering viktig

Kartlegg sannsynligheten for at dataene kan reidentifiseres



Vær åpen!

Hvilke data samles inn, fra hvor og til hvilket bruk?

Hvordan fungerer algoritmen?



Innebygd personvern skaper tillitt

Les om sju steg for innebygd personvern
på våre nettsider

“With Big Data comes big responsibilities”

danah boyd, Microsoft Research

Takk for meg!
Catharina Nes
cane@datatilsynet.no