

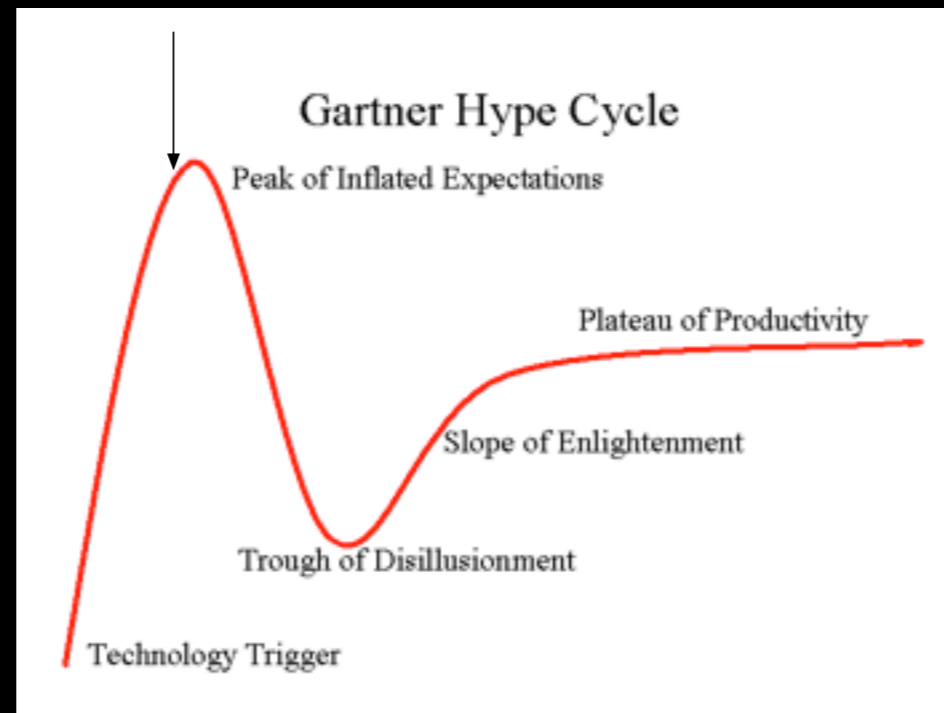


Big data – en balansegang mellom nytteverdi og personvern.

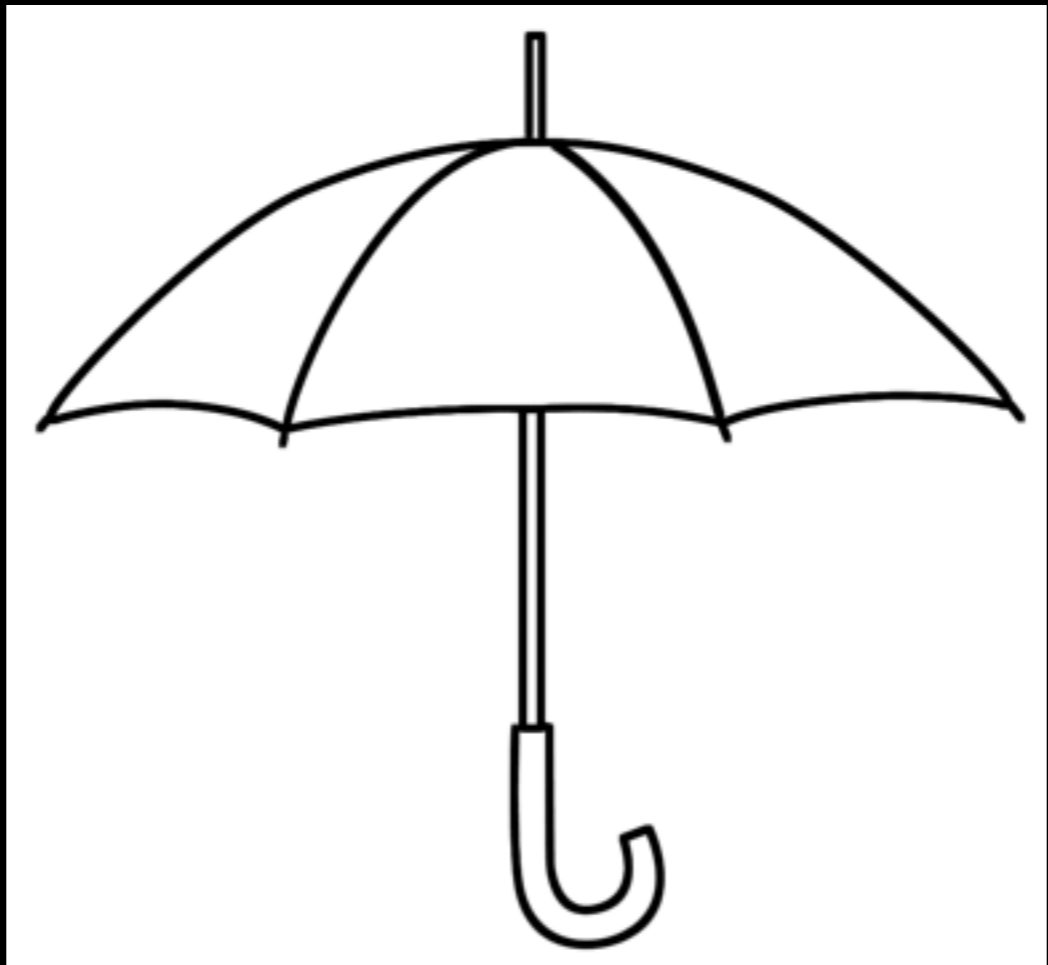
Hyggelig at så mange kunne komme. Big data vil kunne hjelpe oss å utvinne store verdier, gi stor samfunnsnytte men det er fallgruver knyttet til disse verdiene. Jeg kommer ikke til å snakke mye om hva big data er, mer om hva det grunner i og hva det kan bety for oss.

Hva er Big data?

Skal ikke gå veldig mye inn på det, da dette seminaret ikke handler om hva big data er, men hva vi skal gjøre med det. Det er veldig vanskelig å definere big data. Mange liker å definere det rundt dataen, at den er variabel, volumøs eller har høy hastighet. Jeg liker å tenke på det som en felles forståelse av mulighetene som kommer fra all dataen som nå er og gjøres tilgjengelig.



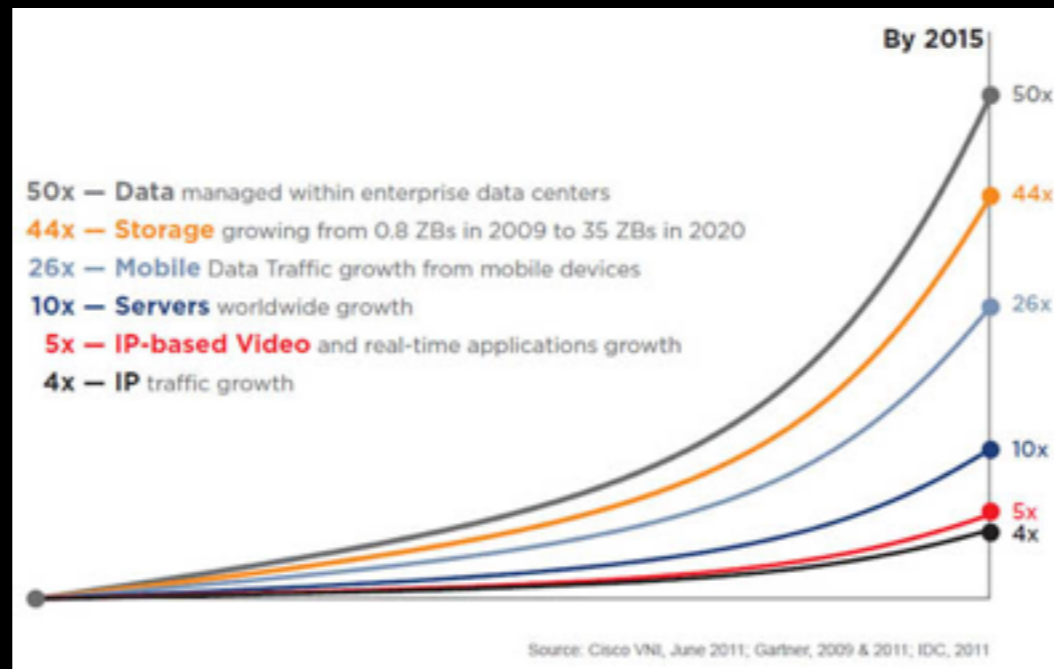
Big data er en hype sier mange, i denne presentasjonen vil jeg peke på at det ikke er en hype, heller et misbrukt uttrykk.



paraplybegrep der flere søger bort I fra dårlig vær og nærmere finansiering..

Hvorfor nå?

Hvorfor har big data blitt så stort nå? Jeg skal peke på 3 (4) ting: tilgjengeligheten på virkelige, pålitelige data har økt dramatisk, og ikke minst; datamengden har blitt mange ganger større, de inneholder mer kunnskap.



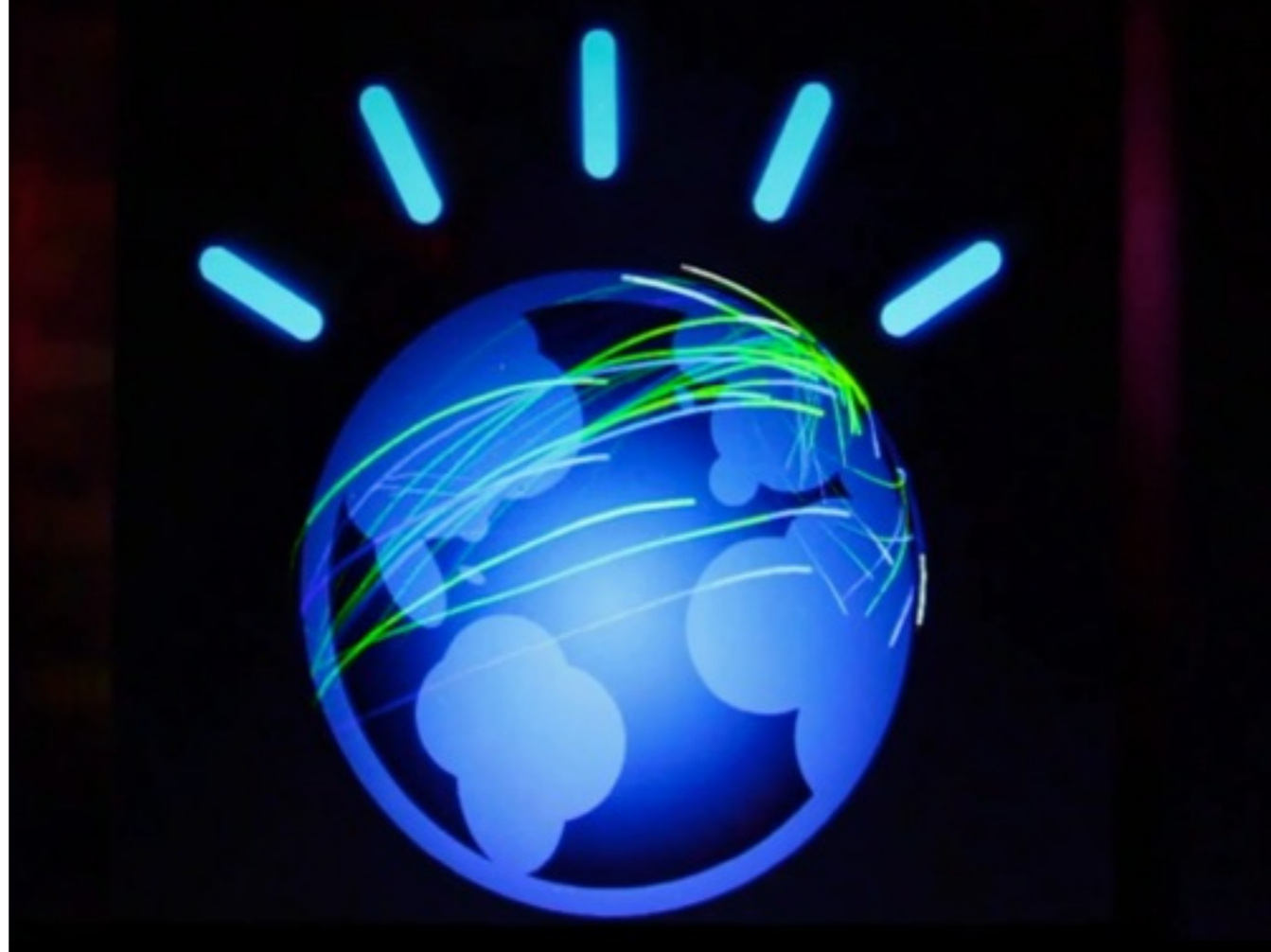
+

[Jeffrey Dean and Sanjay Ghemawat 2004]

Og de blir større!
Samtidig har verktøyene for å behandle store datamenger blitt tilgjengeliggjort, før var de interne verktøy hos google ()

Hva kan Big Data gjøre?

Jeg skal vise dere to eksempler, og



Big data, i min forstand, kan hjelpe å dianostisere kreft: I følge en test har leger 50% treff, mens IBMs Watson når 90%. Dette gjøres hovedsaklig via data innsakning: 600,000 medisinske bevis, mer enn to millioner sider fra medisinske journaler and og en mulighet til å søke igjennom 1,5 millioner pasient journaler.



LAGARBEID: De sju gutta foran dataskjermene skal ha vel så mye av æren for Red Bulls suksess som bilen. Foto: Red Bull Racing

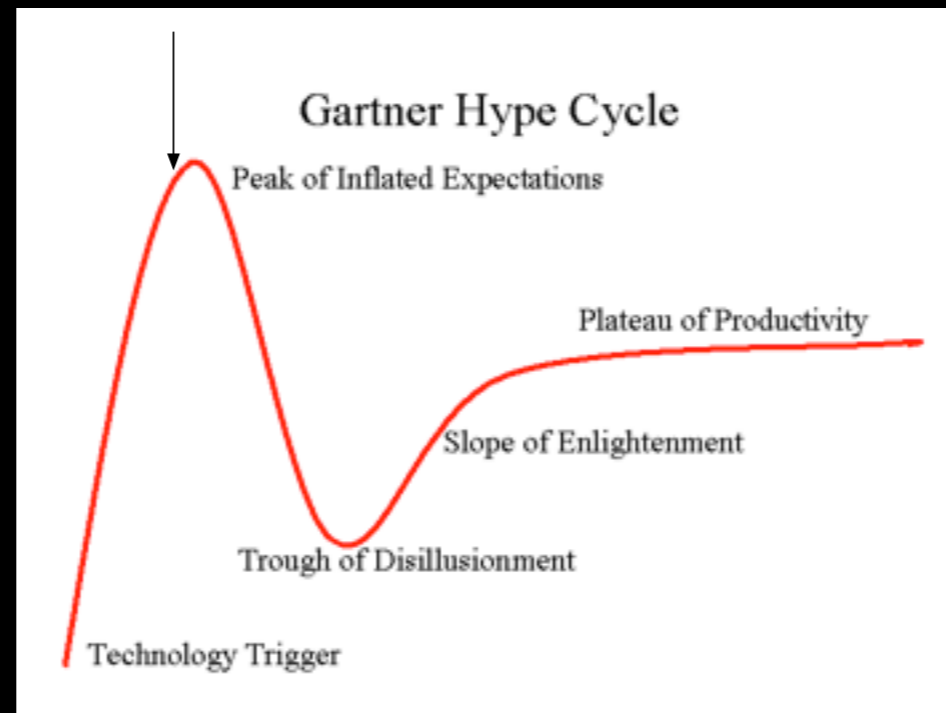
24 menn i et vindusløst rom er hemmeligheten bak Red Bulls Formel 1-suksess

Dagbladet 5.Okt

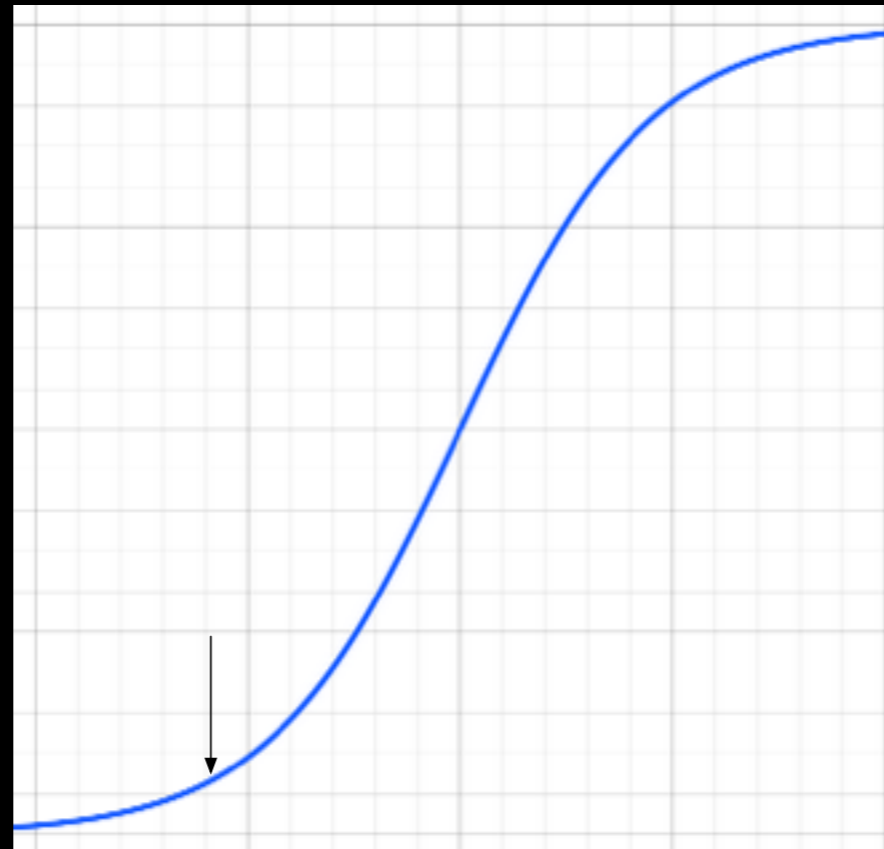
Hjelp deg å vinne formel 1 løp: "100 gigabytes" data for hvert løp, 100 vis av sensorer; temperature, g-krefter, spin etc. Blir sendt til England der 24 analytikere sitter klar for å analysere. Disse kan da bidra til å avgjøre om man må inn til pitstop, eller minst like viktig; sanke data for å bevise at konkurrenten har brutt reglene.

Mye, mye mer.

- Twitter -> influensa utbrudd
- Google trends -> Markedet
- Mange "open data" initiativ.
- Data i dvale.



Så ja, når det gjelder termen "big data" er vi kanskje der.



Men når det kommer til verdier, er det flere indikasjoner på at vi er her. For å sette dette i kontekst; McKinsey gjorde i 2011 en studie der de kom frem til at big data kunne føre til: **300 milliarder dollar besparelser i helsevesenet i USA**

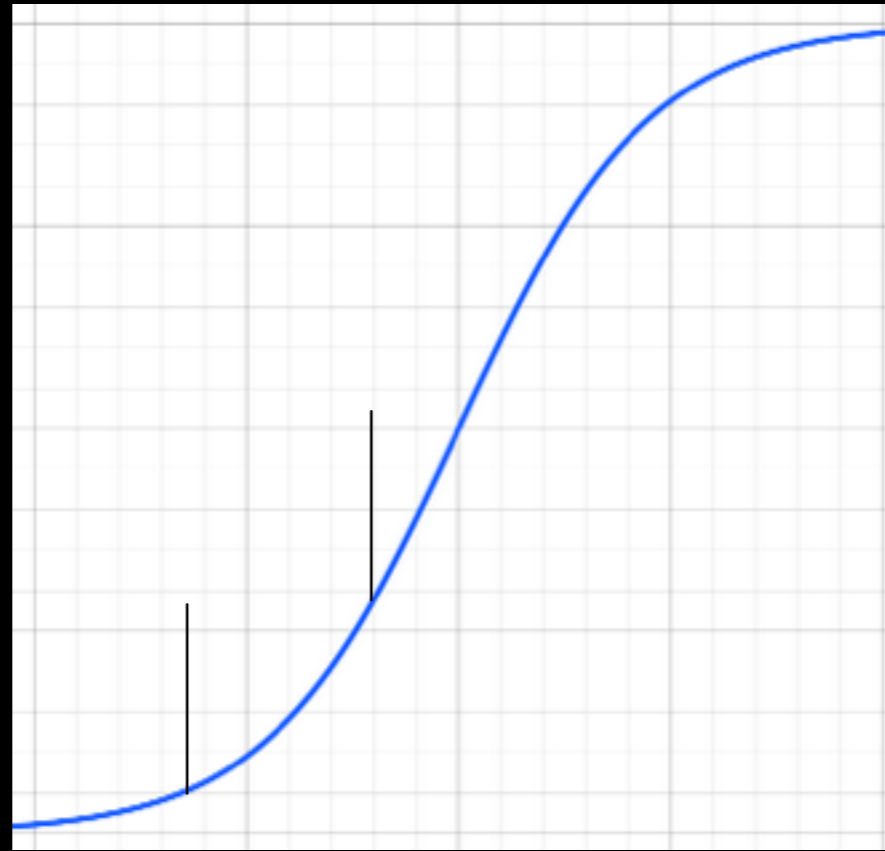
Hvordan?



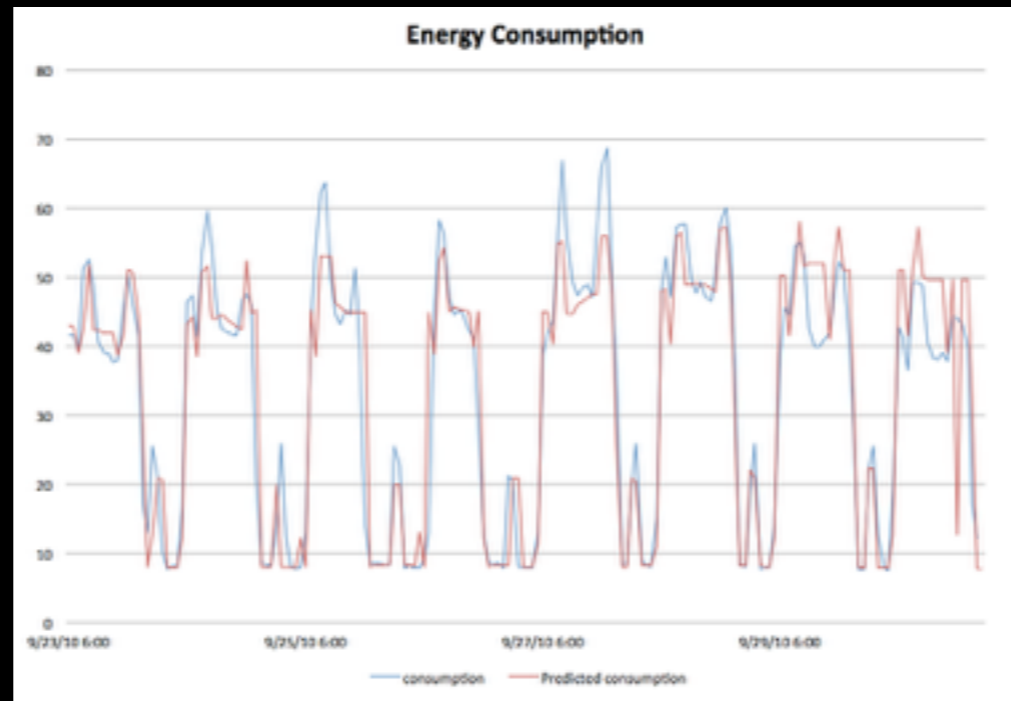
Twitter Storm



- Grunnarbeidet har blitt gjort (yahoo -> hadoop og lignende (mahout etc.)),
Dette vil forsette men .. og de store jobber med det!



nå som vi kan jobbe på store mengder data vil arbeidet skru seg mot å bruke moderne analyse metoder, og pga mengden data; hovedsaklig AI/ML metoder for å forstå dataen bedre. (se hvem google ansetter). Mye kan hentes ut via eksisterende metoder anvendt på nye matter, på større mengder data. Den største verdien ligger dog igjen; vi trenger nye metoder.



Grok solutions, Jeff Hawkins

Et lite eksempel på noen som har tenkt litt anderledes i analyse... Cortical Learning Algorithm, inspirert av menneske hjernen.

Men

Privat?

- Kupong for kreftmedisin?
- Nytte vs Privatliv
- Automatisk dataverge (2020)
- Utvinn det trygge, la resten ligge..
- Hvor skal grensen gå? og hvem setter den?

Trygt, med privatlivet i behold? Hvordan skal vi dra nytte av dette og samtidig ikke utsette privatlivet for innskrenkninger?

Skal vi lage et stort komplisert system som spør hver borger om hva han vil dele? ("vil du dele din medisinske data, du er på vei inn i en risikogruppe..", et slikt system er trolig litt frem i tid..)

I nærmeste fremtid er det antagelig lurt å se på det strategisk "rather be safe than..". Dra nytte av "trygg" data, og la resten ligge (intill videre) litt som barentshavet..



bjornmagnus.mathisen@sintef.no