

Big Data

Hva er Big Data, og hva betyr Big Data for deg?

PwC Consulting



Introduksjon

Big Data er et begrep som har gått fra å være et buzzword, til å bli noe norske virksomheter nå aktivt må forholde seg til. Det er ingen tvil om at 2014 var året hvor Big Data gikk fra å være en hype som et fåtall norske virksomheter hadde implementert som en del av sin IT-satsing, til å komme høyt opp på agendaen også utenfor IT-direktørens kontor. Big Data er, forenklet sagt, ny teknologi som gjør det mulig å analysere større og mer komplekse datamengder hurtigere og mer nøyaktig enn tidligere.

Mange selskaper sitter på enorme datamengder om kunder, produkter, transaksjoner, osv., men mangler en plan for hvordan dataene kan benyttes til å få ny innsikt som grunnlag for beslutningsstøtte. Samtidig genereres det enorme mengder data fra ulike eksterne kilder, som kombinert med interne data kan gi et bredere analysegrunnlag og ny informasjon. Det er i mange virksomheter et stort behov for å etablere en god plan for hvordan de kan hente ut verdi fra disse dataene gjennom analyse og hvordan de kan utnytte den nye innsikten i forretningsprosessene.

I en fersk nordisk undersøkelse om Big Data fant SAS Institute og Intel at 50 % av selskapene i undersøkelsen enda ikke har etablert en IT-arkitektur som er i stand til å håndtere de enorme datamengdene som nå er tilgjengelige. Samtidig svarte 80 % selskapene at de ser et økende behov for å samle inn og analysere mer data og 92 % svarte at de ser på analyse av disse dataene som en mulighet for å gi selskapet et konkurransefortrinn i forhold til konkurrentene ¹.

Det er et gap mellom hva selskaper i det nordiske markedet ønsker å gjøre for å sikre økt innsikt og bedre beslutningsstøtte og hva de faktisk gjør for å bli bedre på dette. Denne artikkelen vil forklare hva Big Data er og hvilke muligheter som ligger i en strukturert tilnærming til Big Data og analyse.

Hva er Big Data?

Hva er egentlig Big Data? Det første forsøket på en definisjon av Big Data ble lansert i 2001: Big Data er data som kan kjennetegnes ved tre V'er, Volume (volum), Variety (variasjon) og Velocity (hastighet).

- Volume: Tilgjengelig data øker i omfang – IBM estimerer at det innen 2020 kommer til å genereres 2,3 trilljoner gigabyte data hver dag, som tilsvarer 107 millioner blue-ray disk eller 4,2 milliarder fulle CD'er
- Variety: Ulike typer data er tilgjengelig fra en rekke ulike datakilder - 400 millioner Twitter-meldinger sendes hver dag, og stadig flere datakilder kobles opp mot internett og genererer informasjon
- Velocity: Data er tilgjengelig raskere enn før og vi ønsker at dataene er tilgjengelige i tilnærmet sanntid

IBM har valgt å ta med enda en V i sin definisjon av Big Data, Veracity (pålitelighet) – usikkerhet knyttet til kvalitet i dataene, og viktigheten av å kunne stole på dataene ².

I 2012 spisset Gartner denne definisjonen ytterligere: Big Data er informasjon som kjennetegnes ved høyt volum, høy hastighet og/eller høy grad av variasjon, og som krever nye metoder for prosessering og tilrettelegging for å kunne fungere som grunnlag for forbedrede beslutninger, økt innsikt og optimalisering av prosesser ³.



Foruten mengden data som nå genereres, er Big Data egentlig noe nytt? Historien er full av eksempler på hvordan mennesker har strukturert og satt sammen mengder av informasjon for å ta bedre beslutninger. 3000 år f.Kr ble det i Egypt og Mesopotamia innført metoder for å loggføre alt fra handelstransaksjoner til bruk av mål og materialer i byggkonstruksjoner – det første kjente eksempelet på strukturert datainnsamling.

I 1839 begynte den amerikanske marineoffiseren Matthew Fontaine Maury å samle inn informasjon fra store mengder gamle sjøkart, skipslogger og bøker med informasjon om strømninger, vind og værforhold på spesifikke datoer. Hver for seg ga «datakildene» ikke noe mer informasjon enn det man kunne lese fra kilden selv, men da han satte dette sammen klarte han å lage et komplett kart over Atlanterhavet, hvor informasjon om vær- og vindforhold ble fordelt ut på de korrekte koordinatene i kartet. I 1855 resulterte denne datainnsamlingen i boken «The

Physical Geography of the Sea», med mer enn 1,2 millioner datapunkter, som revolusjonerte navigasjon i skipsfarten ⁴.

I dag er bøkene og loggfilene til Maury supplert med data som vi genererer hver eneste dag fra Twitter, Facebook, verdipapirtransaksjoner, sensorer osv. Vi gjør akkurat det samme som Matthew Fountaine Maury gjorde for over 150 år siden, vi setter sammen og analyserer strukturert og ustrukturert informasjon for å lære mer om verden rundt oss.



Hva er driverne bak Big Data?

Big Data kan sies å være summen av store mengder strukturerte og ustrukturerte data.

- Strukturerte data: Transaksjonsdata som vi typisk finner i tradisjonelle datavarehus, økonomisystemer (ERP), kundedatabaser (CRM), osv.
- Ustrukturerte data: Data som typisk ikke kan struktureres i tabeller – Twitter-meldinger, bilder, tekst, osv.

I tillegg til å være en kombinasjon av strukturerte og ustrukturerte data, så kommer data i enorme mengder fra en rekke ulike kilder - Internet of Things. Dette begrepet er sentralt når vi snakker om Big Data – alt er på internett og kan knyttes sammen ved hjelp av data. Nye biler er koblet mot internett og genererer enorme mengder informasjon om alt fra kjøremønstre til når sentrale komponenter i bilen må

repareres eller skiftes ut. Telefonene våre er konstant på nett og bruk av apper generer data om våre nettvaner, kjøpsvaner, hvor vi befinner oss og private interesser og meninger. Smart TV'en din er koblet opp mot internett slik at Netflix kan lage analyser basert på dine TV-vaner. Til og med huset eller leiligheten din er på internett – du kan bruke egne apper til å styre strømforbruket i boligen din ⁵. Slik kan vi fortsette.

En vanlig misforståelse er at Big Data kun handler om ustrukturerte data, som det genereres trillioner av gigabyte av hver eneste dag, og som er majoriteten av all tilgjengelig informasjon. Big Data handler om hvordan vi setter sammen informasjon fra interne og strukturerte datakilder med ustrukturert informasjon for å skape ny innsikt.

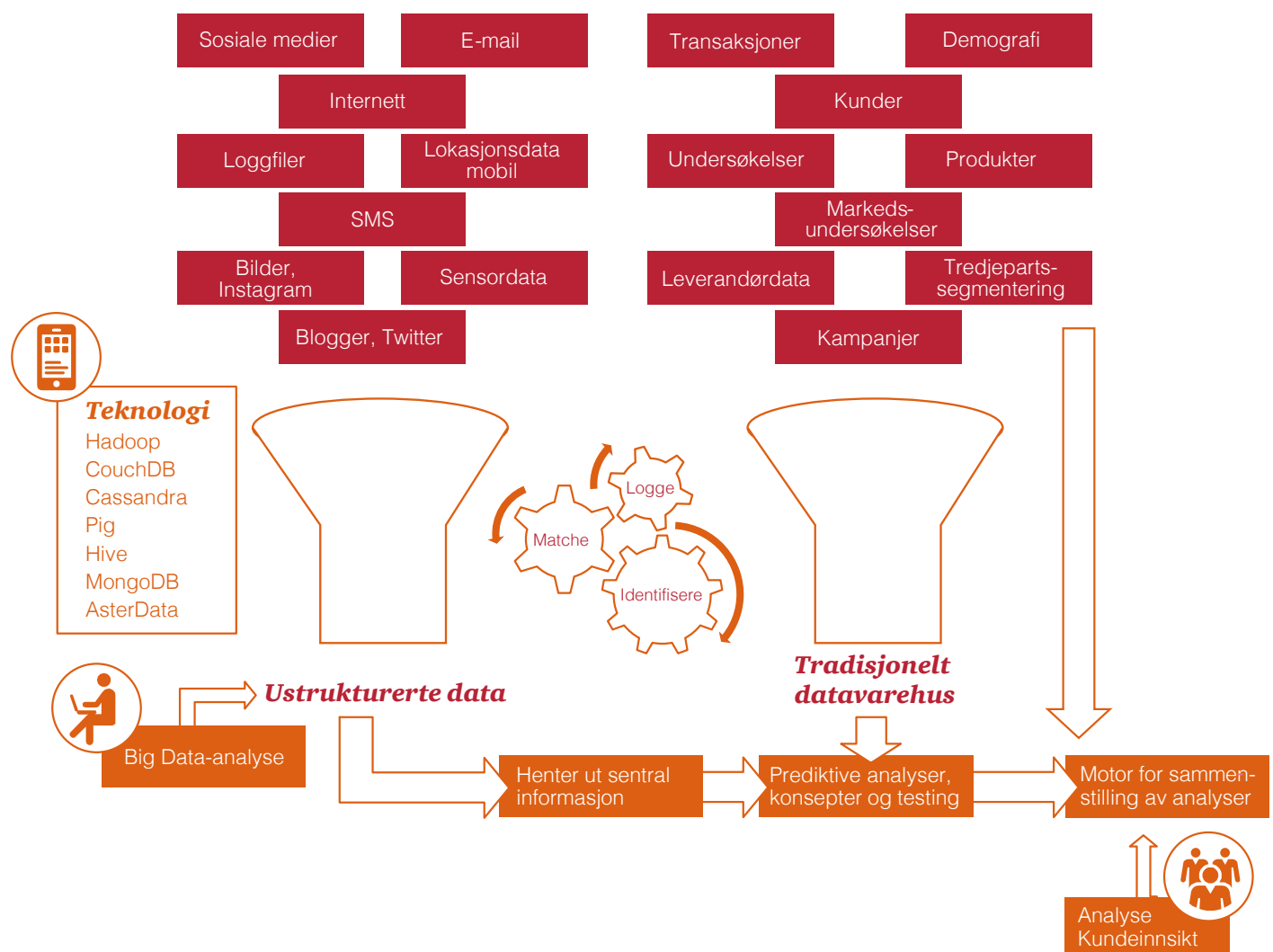




Den største og viktigste årsaken til at vi nå kan snakke om Big Data som et viktig beslutningsgrunnlag er at det nå finnes tilgjengelig teknologi som gjør det mulig å samle inn og lagre enorme datamengder raskt og kostnadseffektivt. Vi kan sette sammen strukturert og ustrukturert informasjon og bruke denne informasjonen til å se sammenhenger, lage analyser og generere innsikt som vi tidligere ikke trodde var mulig.

Figuren gir et forenklet bilde av hvordan man kan samle sammen informasjon fra ustrukturerte og strukturerte datakilder og hente ut informasjon som grunnlag for relevante analyser. Big Data i seg selv gir ingen verdi. Verdien genereres når vi klarer å analysere de riktige dataene og gi ny innsikt og bedre beslutningsgrunnlag for virksomheten. Den største verdien fra en Big Data-plattform oppnås når man klarer å lage gode visualiseringer og presentasjoner basert på komplekse analyser på relevant informasjon.

Figur 1 - Sammensetning av strukturert og ustrukturert informasjon som analysegrunnlag for økt kundeinnsikt



Hvorfor skal du bry deg om Big Data?

Big Data byr på muligheter norske virksomheter må møte med optimisme og nysgjerrighet, men også med realisme og bevissthet rundt utfordringene dette medfører. Større, bedre og raskere tilgang til nye data, kombinert med relevante

analyser og innsikt, skaper innovasjon, konkurransekraft og økt produktivitet. Samtidig skal man ha et bevisst forhold til personvern og sikker håndtering av sensitiv informasjon.

Selskaper i en rekke bransjer henter ut forbedrede resultater som følge av riktig bruk av Big Data

Figur 2 - Big Data sikrer kontinuerlig og forbedring ⁶

Media

Big Data har gjort det mulig for medieselskaper å **få ett bilde på kundene** sine - hva leser du, når leser du det, hvor lenge leser du. Dette skaper fokusert produktutvikling og markedsføring, og generer ny innsikt for spesialtilpasning av innhold på nett og i aviser

Forsikring

Forsikringsselskaper tar i bruk Big Data for å lage bedre **prediktive modeller som beslutningsgrunnlag** for prising, markedsføring, salg og sammensetning av forsikringsporteføljen. Bruk av sensorer i biler gjør det mulig å prise bilforsikringen basert på ditt kjøremønster, og ikke alder og kjønn

Helse

Helsesektoren bruker Big Data til å analysere **effektivitet og kvalitet i behandlingen**. Big Data gjør det mulig for en lege å hente inn riktig informasjon om mulige diagnoser fra alle verdens artikler og bøker i løpet av sekunder, basert på informasjon om symptomer kombinert med erfaringsdata

Telekom

Sofistikerte metoder for analyse av kundeatferd har gjort selskaper i stand til å identifisere tidligere ukjente årsaker til redusert **kundetilfredshet og høy churn-rate**. Big Data gjør det mulig å drive målrettet produktforbedring og proaktiv kundebehandling mye mer effektivt enn før

Hva gjøres globalt?

Mange land og virksomheter har kommet svært langt når det gjelder å ta i bruk de mulighetene Big Data representerer. I førersetet finner man selskaper som Amazon, Apple, Netflix, Facebook, Twitter, Google, Yahoo og IBM. I et globalisert marked hvor norske virksomheter konkurrerer med stadig flere utenlandske aktører blir det enda viktigere å kunne konkurrere på teknologi og innovasjon. Hvorfor er Big Data viktig i denne sammenheng? Fordi virksomheter som er gode på Big Data klarer å identifisere nye satsingsområder og forbedringsområder gjennom økt innsikt. Fordi virksomheter som er gode på Big Data får et konkurransefortrinn i forhold til andre gjennom å ta riktige beslutninger basert på riktig informasjon, informasjon som i prinsippet er tilgjengelig for alle, men som kun de som satser på strukturert dataanalyse klarer å utnytte. Dette skaper igjen nye muligheter for kontinuerlig forbedring av virksomheten ⁷.

Netflix er et velkjent eksempel på et selskap som kan betraktes som en innovator innen Big Data og dataanalyse. Selskapet ble stiftet i 1997, og de neste ti årene satset de primært på distribusjon av DVD'er til abonnenter. Deres suksess innen DVD-segmentet la grunnlaget for deres satsing på streamingtjenester, som de tidlig forstod ville bli den viktigste konkurrenten til filmleie på DVD. De viste tidlig vilje til innovasjon ved å være i forkant av teknologiutviklingen. Allerede i 2010 var de den største kilden til internettrafikk i Nord-Amerika gjennom sine streamingtjenester. Året etter opplevde de en markant nedgang i antall DVD-abonnenter, men dette ble langt på vei utjevnet av den enorme økningen i abonnenter på deres streamingtjenester.

Netflix har gjennom sine abonnenter tilgang til informasjon om deres seervaner; hva ser de på, når ser de på det, når trykker de på pause, hvilke filmer og skuespillere søker de etter? Dette datagrunnlaget, kombinert med sofistikerte analyseverktøy, gjorde Netflix i stand til å si at en serie regissert av David Fincher, med Kevin Spacey i hovedrollen, basert på det britiske dramaet «House of Cards» ville slå an blant deres abonnenter. Og de fikk helt rett.

Netflix så muligheter der andre så begrensninger. De så at med den teknologiske innovasjonen streamingtjenestene representerte lå det uforløste muligheter. Fra å være en DVD-distributør gikk Netflix til å bli et ledende selskap innen streaming og ikke minst filmproduksjon. De var innovative nok til å forstå at man ved å ta i bruk dataanalyser av abonnentenes seervaner kunne lage noen av de mest populære TV-seriene de siste årene – «House of Cards» og «Orange is the New Black».

Et annet eksempel er Macy's, som sammen med IBM har implementert en løsning for å samle inn og analysere Big Data for å få enda bedre innsikt i kundene sine. Hensikten med dette er å få en enda mer målrettet, priseffektiv og relevant markedsføring gjennom sammenstilling og analyse av sanntidsdata fra deres nettsider med andre relevante kundeparametere som lokasjonsdata og informasjon om kundene. Målet med dette er å kunne skape en skreddersydd og bedre kundeopplevelse for hver eneste kunde ⁸.

Hva gjøres i Norge?

I det norske markedet har aktører som Telenor og Schibsted over tid satset på Big Data og analyse som har gitt dem unik innsikt i kunder, kundeatferd og dermed økt kvalitet i kundedialog og markedsføring. Telenor har et eget forskningsmiljø (Telenor Research) som samarbeider med fagmiljøer innen Big Data på Harvard, MIT, Cambridge og Northeastern University. De har gjennomført forsøk på effekten av tradisjonell direkte markedsføring, sammenlignet med markedsføringskampanjer basert på Big Data-analyse. Det viste seg at kampanjene var 13 ganger mer effektive enn tradisjonelle markedsføringskampanjer når mottagerne ble valgt basert på avanserte algoritmer og analyse av enorme mengder attributter for å identifisere den optimale målgruppen for kampanjen ⁹.

Schibsted har også jobbet strukturert med Big Data over flere år, og har de siste årene sagt at de skal konkurrere med Google, LinkedIn, Facebook og Amazon, noen av de ledende aktørene innen dataanalyse. Gjennom sin betalingsløsning, SPID, får Schibsted tilgang til verdifull informasjon om kundene sine. Som kunde benytter man seg av SPID både når man skal logge seg på Finn.no og når man skal lese Aftenposten på nett. Dette gjør at Schibsted er i stand til å lære mye om vår atferd som forbrukere – hva vi leser, når vi leser det, hvilke artikler vi synes er interessante. Ved å samle inn dataene og ha en strukturert tilnærming til analyse kan Schibsted bli enda tryggere på å ta nye produkter som lanseres treffer det riktige segmentet ¹⁰.

Både Telenor og Schibsted har skaffet seg et verdifullt datagrunnlag og har en strukturert tilnærming til hvordan de skal bruke dette datagrunnlaget til å ta riktige strategiske beslutninger. Ved å være tidlig ute med analyse av Big Data er sannsynligheten stor for at de har etablert et konkurransefortrinn i forhold til mange av konkurrentene sine.

Hvordan går du frem for å utnytte Big Data-analyse?

Viktige beslutninger kan ikke lenger baseres på magesfølelsen til sentrale beslutningstakere. Strategisk viktige beslutninger vil tas i et høyere tempo enn før, med krav om tilgang til gode analyser basert på data som er relevante for virksomheten – Big Data¹¹. De virksomhetene som ikke tar dette innover seg vil risikere å havne langt bak de som har tilgang ny innsikt om kunder og markeder.

Analyse av Big Data betyr ikke at man skal analysere alle de data som finnes der ute for å se om man kanskje finner trender og sammenhenger ingen har tenkt på før. Analyse av Big Data handler om å ha en tydelig formening om hvilke data som er viktige, hva man skal bruke disse dataene til, samt gode verktøy og kompetansen til å skape analyser og innsikt gjennom disse dataene.

Netflix er en av flere suksesshistorier på selskaper som har vokst gjennom en bevisst strategi knyttet til teknologisk innovasjon og bruk av Big Data og dataanalyse som sitt

viktigste beslutningsgrunnlag. Big Data er noe alle virksomheter nå bør ha et bevisst forhold til og vår anbefaling er å først etablere en plan for hvordan virksomheten skal forholde seg til de enorme informasjonsmengdene som genereres hver eneste dag, og hvordan man skal få innsikt fra disse dataene.

I PwC anbefaler vi at man har en strukturert tilnærming til hvordan virksomheten skal identifisere de mulighetene som finnes i analyse av Big Data. Det første, og viktigste, er å skape en kulturendring i organisasjonen hvor man i større grad tar beslutninger basert på prediksjoner om fremtiden, prediktive analyser av strukturerte og ustrukturerte data, som supplement til den tradisjonelle historiske rapporteringen. Videre er det viktig å være tydelig på hva man skal bruke den nye informasjonen til, og gjennom dette definere hvilken informasjon som er viktig for nettopp din virksomhet.

Figur 3 - PwC, fem steg til en suksessfull Big Data-implementering



Leveranser

- | | | | | |
|--|--|--|--|--|
| <ul style="list-style-type: none">• Innsyn i begrepet og identifisering av potensielle bruksområder• Vurdering av Big Data's relevans for virksomheten• Etablert hensikt og formål med et Big Data-initiativ | <ul style="list-style-type: none">• Initielt business case for Big Data• Prioriterte bruksområder• Overordnet kildeanalyse (interne og eksterne datakilder)• Forslag til løsningskonsepter og pilot | <ul style="list-style-type: none">• Komplett målbilde for Big Data (Big Data-strategi)• Business Case• Juridiske og etiske retningslinjer for håndtering av data• Oppdatert IT-arkitektur – hvordan håndtere Big Data sammen med eksisterende systemer? | <ul style="list-style-type: none">• Implementering av en pilotløsning• Evaluering av pilot – fokus på kvalitet på visualisering og analyse av data• Vurdering av videre utrulling av løsningen | <ul style="list-style-type: none">• Konstant videreutvikling av løsningen• Oversikt over trender innen teknologi og bransje |
|--|--|--|--|--|

I PwC ser vi at de mest suksessrike Big Data-satsingene i Europa og USA er basert på en stegvis implementering gjennom fem steg:

- 1** Definere hvordan Big Data kan svare på sentrale problemstillinger i en eller flere av forretningsområdene i virksomheten – kan Big Data gi oss ny innsikt?
- 2** Sikre forankring hos sentrale beslutningstakere og utarbeide et konsept for videre satsing på Big Data-analyse – velg utgangspunkt for en Big Data-løsning
- 3** Analysere og evaluere eksisterende IT-arkitektur og avdekke behov for investeringer i hardware, software og kompetanse for å kunne analysere ny informasjon effektivt – velg tilnærming til en Big Data-implementering
- 4** Gjennomføring av pilotprosjekter og evalueringer av disse som grunnlag for beslutning om videre satsing – pilot, vurdering og operasjonalisering
- 5** Stegvis implementering av nye bruksområder og kontinuerlig utvikling av analysegrunnlag og IT-arkitektur – evaluering og videreutvikling

Ta gjerne kontakt med oss for å diskutere hvordan din virksomhet kan skape verdi fra Big Data og dataanalyse!

Lars Meinich Andersen

Senior Manager,
Performance & Information Management, PwC Consulting
+47 91 66 22 43, lars.meinich.andersen@no.pwc.com

Magne Bakkeli

Direktør,
Performance & Information Management, PwC Consulting
+47 91 66 22 69, magne.bakkeli@no.pwc.com

¹ SAS Institute, Intel (2015) – Do you perceive your company as data driven? The 2015 Nordic Survey on Big Data and Hadoop. Hentet fra: http://www.sas.com/content/dam/SAS/da_dk/doc/other1/SAS_Hadoop_PDF.pdf

² IBM (2012) - The Four V's of Big Data. Hentet fra ibmbigdatahub: http://www.ibmbigdatahub.com/sites/default/files/infographic_file/4-Vs-of-big-data.jpg

³ Simon, Phil. (2013). Too Big to Ignore - The Business Case for Big Data. John Wiley & Sons

⁴ Mayer-Schönberger, V., & Cukier, K. (2013). Big Data - A Revolution That Will Transform How We Live, Work and Think. Houghton Mifflin Harcourt

⁵ Lyse. (n.d.). Smartly. Hentet fra Lyse: <https://www.smartly.no/>

⁶ Halter, O., & Rao, A. (2013). Capitalizing on the promise of Big Data. PwC

⁷ IBM Institute for Business Value. (2014). The Speed Advantage - Why data-driven organizations are winning the race in today's market place. IBM

⁸ Van Rijmenjam, M. (n.d) Macy's Is Changing The Shopping Experience With Big Data Analytics. Datafloq

⁹ Bjelland, J. & Sundsøy, P. (2014). Big Data Analytics i Telenor Group. Hentet fra Norges Markedsanalyseforening: <http://www.analysen.no/latest-news/item/big-data-analytics-i-telenor-group>

¹⁰ INMA (2013), <http://www.inma.org/blogs/conference/post.cfm/schibsted-s-global-investment-in-advanced-data-analytics-is-paying-off>

¹¹ PwC, The Economist Intelligence Unit. (2014). Gut & gigabytes: Capitalising on the art & science in decision making. www.pwc.com/bigdecisions

