Kartlegging og vurdering av stordata i offentlig sektor

RAPPORT TIL KOMMUNAL- OG MODERNISERINGSDEPARTEMENTET
UTARBEIDET AV VIVENTO AS OG AGENDA KAUPANG AS

AGENDA KAUPANG

vivento

V1.01

Forord

Stordata eller «Big Data» har lenge vært «hypet» opp til å være det som skal forløse data som den nye oljen. Mange private bedrifter har lenge benyttet avanserte analyser basert på stordata teknologi til å tilpasse markedsføring av produkter og tjenester til bedrifter og den enkelte forbruker. Men hva med offentlig sektor, i hvilken grad har offentlige virksomheter tatt i bruk mulighetene som ligger i denne teknologien? Og hvilke utfordringer står det offentlige overfor når store volum av data fra mange ulike datakilder skal utnyttes i sanntid?

Mulighetene er mange – *Teknologirådet* har pekt på at politiet kan bli flinkere å predikere når og hvor det kan skje kriminelle handlinger dersom slik teknologi tas i bruk. *Helsesektoren* kan bli flinkere å forutse arvelige sykdommer og å treffe bedre med behandling av pasientene. *Skatteetaten* og *NAV* kan bli flinkere å avdekke svindel. *Vegvesenet* kan styre trafikk og veivedlikehold, samtidig som mulighetene for å reagere raskt på miljøkatastrofer kan bedres.

De store byene kan ha oppdatert oversikt over sitt karbonfotavtrykk for å kunne sette inn målrettede tiltak for å bli klimanøytrale. Men ikke minst kan offentlige virksomheter benytte stordata til å skreddersy tjenester til den enkelte etter den enkeltes livssituasjon og behov. Hva er så konsekvensene av at det samles inn, sammenstilles og bearbeides store mengder data om innbyggerne? Har vi kompetanse i offentlig sektor til å utnytte mulighetene? Vil personvernet bli ivaretatt på en god måte når data som er samlet inn med ulike formål kobles sammen?

Denne rapporten er skrevet av Vivento as – i samarbeid med Agenda Kaupang – på oppdrag fra Kommunal- og moderniseringsdepartementet og gjengir en kartlegging og vurdering av bruk av stordata i offentlig sektor.

Vivento AS
Desember 2015

Innhold

Forord 3
1 Sammendrag 9
2 Innledning 11
2.1 Oppdrag 11
2.2 Disposisjon 12
2.3 Metode og avgrensninger 12
3 Stordata bakgrunn og begrep 14
3.1 Begrepet «Stordata» 14
3.2 Bakgrunn for stordata 15
3.3 Stordata-verdikjeden og -teknologier 16
4 Drivkrefter for bruk av stordata 21
4.1 Lavere kostnader og tilgang på skytjenester 21
4.2 Tingenes internett 22
4.3 Maskinlæring og kunstig intelligens 22
4.4 Hadoop 23
4.5 Åpne data 23
5 Utvalgte land 25
5.1 Australia 25
5.2 Finland 26
5.3 Nederland 26
5.4 Storbritannia 27
5.5 Sverige 27
5.6 Sør-Korea 28
5.7 USA 29
5.8 EU 29
6 Bruk av stordata i offentlig sektor i Norge 31
6.1 Samferdsel 31
6.2 Justis 36
6.3 Helse 39
6.4 Arbeid og velferd – NAV 43
6.5 Skatt, toll og avgift 45
6.6 Klima, miljø og beredskap 47

6.7 Smarte bygg og byer 49
6.8 Kartdata 53
6.9 Offisiell statistikk 54
6.10 Forskning 55
6.11 Oppsummering av status for bruk av stordata i offentlig sektor i Norge 56
7 Gevinstpotensialet 59
7.1 Effektive beslutningsprosesser 59
7.2 Bedre kvalitet og aktualitet i beslutningene 59
7.3 Raskere reaksjonstid 60
7.4 Presise prediksjoner 60
7.5 Økt brukerorientering 60
8 Utfordringer og hindringer 62
8.1 Stordata-økosystemet 62
8.2 Tilgang til data med god kvalitet 63
8.3 Infrastruktur og teknologi 67
8.4 Juridiske rammevilkår 68
8.5 Introduksjonskostnader 72
8.6 Kompetanse 73
8.7 Markedet 74
9 Anbefalinger 77
9.1 Læringspunkter for offentlige virksomheter 77
9.2 Tilgang til data og metadata med god kvalitet 78
9.3 Infrastruktur og teknologi 80
9.4 Juridiske rammevilkår 81
9.5 Kompetanse 83
9.6 Markedet 85
9.7 Koordinert gjennomføring 85
9.8 Oppsummering av anbefalte tiltak 86
Vedlegg 1: Tre utvalgte prosjekt 87
1 Helsedirektoratet: Prehospital strategisk satsing 87
1.1 Løsning for prehospital strategisk satsing 88
1.2 Datagrunnlag 89
1.3 Teknisk løsning 89
1.4 Gevinster 91
1.5 Utfordringer, løsninger og læring 92

1.6 Veien videre 94
1.7 Om Norsk pasientregister 94
2 Statens vegvesen: Datainn og intelligente trafikksystemer 95
2.1 Datagrunnlag i Datainn prosjektet 97
2.2 Løsning 97
2.3 Gevinster 101
2.4 Utfordringer, løsninger og læring 102
3 Statsbygg: Smarte bygninger 105
3.1 Datagrunnlag og teknologi 107
3.2 Gevinster 107
3.3 Utfordringer, løsninger og læring 108
Vedlegg 2: Relevante forskningsprogram 110
Vedlegg 3: Nærmere beskrivelse av bruk av stordata i utvalgte land 114
3.4 Storbritannia 114
3.5 Nederland 116
3.6 Finland 118
3.7 Sverige 120
3.8 USA 122
3.9 Sør-Korea 125
3.10 Australia 126
3.11 Utvalgte EU-initiativer 127
Vedlegg 4: Definisjoner 130
Litteratur 135

1 Sammendrag

De potensielle samfunnsmessige gevinster for Norge ved bruk av stordata er store. Dersom forholdene blir lagt til rette kan stordata lede til gevinster i mange sektorer, og gi muligheter for effektivisering av offentlig sektor generelt.

Stordata har potensiale til å effektivisere beslutningsprosesser, oppnå bedre kvalitet og aktualitet i beslutninger og gi aktører i samfunnet muligheten til å reagere raskere på avvik fra normalsituasjoner. Rapporten peker på eksempler hvor bruk av stordata kan ha stor effekt, slik som i forbindelse med miljøkatastrofer, ved å forutse utviklingsmønstre i helsesektoren, bedre planlegging av samferdsel, trygge samfunnet for potensielle farer og bidra til økt næringsutvikling og innovasjon.

Samtidig som stordata representerer store muligheter stiller det oss også ovenfor utfordringer, bl.a. hva informasjonssikkerhet og personvern angår. Det vil oppstå etiske dilemmaer hvor man må vurdere den samlede samfunnsnytten av å ta de nye mulighetene i bruk.

OECD hevder at stordata representerer et paradigmeskifte mot en mer datadrevet økonomi, der smart sammenstilling og analyse av store datamengder kan forbedre økonomisk konkurranseevne, drive innovasjon, og bidra til rettferdig fordeling og bærekraftig utvikling.

Stordata handler om utnyttelse av store mengder data, gjerne på tvers av virksomheter, datakilder og formater. Blant annet kan dette omfatte sanntids datastrømmer. Stordata er en ny teknologi som omfatter hele verdikjeden; datainnsamling, lagring, prosessering, analyse og visualisering av resultater. Det finnes en rekke definisjoner på begrepet stordata, men i denne rapporten har vi lagt følgende definisjon til grunn:

Stordata er analyse av massive samlinger av data («volume»), med stor variasjon i datakilder og formater («variety»), hvor datasettet oppdateres med høy frekvens («velocity») og hvor grunnlagsdataenes opprinnelse og kvalitet er avklart («veracity») og hvor analysen av datasamlingene gir økt verdi i forhold til hva datakildene enkeltvis ga («value»).

Stordata er særlig relevant nå fordi en rekke faktorer som driver mulighetene for stordata sammenfaller i tid. De viktigste faktorene er:

- I markedet har det vokst fram en mengde sensorteknologier som kan innhente og distribuere data («tingenes internett»).
- Maskinlæring og kunstig intelligens som teknologi har vi hatt noen tiår, men er nå i ferd med å bli moden.
- Vi ser en betydelig kostnadsnedgang for lagring og prosessering av data.
- Robust teknologi som gjør det mulig å prosessere store datamengder distribuert på mange datamaskiner er tilgjengelig (F.eks. Hadoop).
- Det er etter hvert gjort tilgjengelig flere åpne datasett fra private og offentlige aktører som er relevante for sammenstilling og analyser.

Denne rapporten beskriver bruk av stordata i offentlig sektor i Norge og i utvalgte land. Generelt synes bruken av stordata internasjonalt fortsatt å være i en tidlig fase. Det er

forskjeller mellom land når det gjelder i hvilken grad offentlig sektor har evnet å utnytte potensialet i stordata. Vi ser at utnyttelsen av stordata i stor grad henger sammen med i hvilken grad offentlige data er tilgjengeliggjort, og hvordan landene har tilrettelagt personvernlovgivningen og andre rammebetingelser. USA er ledende når det gjelder tilretteleggelse for utnyttelse av stordata, tett fulgt av Storbritannia og Nederland. Norge, samt andre EU-land og velutviklede nasjoner som Sør-Korea, Singapore, Japan, Kina, New Zealand og Canada, har til en viss grad strategier for utnyttelse av stordata. Samtlige har etablert stordata-prosjekter i offentlig sektor.

Flere norske forskningsmiljøer gjør verdifull forskning på stordata, samtidig som også stordata i seg selv utnyttes i ulike forskningsprosjekter. På områder som medisin, demografi og smarte byer finner vi interessante prosjekter og forsøk. Norske forskningsmiljøer har nøkkelroller i flere internasjonale forskningsprogrammer, slik at kunnskapsgrunnlaget og kompetansen er i ferd med å bygges opp i Norge.

Hovedinntrykket for Norges vedkommende er imidlertid at de fleste offentlige virksomheter har kommet kort i utforsking og bruk av stordata. Majoriteten av offentlige virksomheter følger med på utviklingen uten at de har konkrete satsinger på området, til tross for at de forvalter store datavolumer. En del offentlige virksomheter har interessante pilotprosjekter og forsøk, slik som Statsbygg, Skatteetaten og Nav.

De som har kommet lengst i å ta i bruk stordata kjennetegnes ved at de har etablert en strategisk satsing på området. Prosjektene er godt forankret i virksomhetsstrategier og rammevilkårene er avklart. I tillegg kjennetegnes virksomhetene av at de har lang erfaring med analysevirksomhet og forvaltning av data, i tillegg til at de har profesjonelle IT-miljøer.

Rapporten belyser noen utfordringer knyttet til utnyttelsen av stordata i offentlig sektor som

- å få tilgang til data med god kvalitet
- gjenbruk av data i andre sammenhenger enn opprinnelig formål
- spørsmålet om hvem som eier analysedata
- tilgang til kompetanse på området

Vi anbefaler at det utarbeides en helhetlig politikk på området som anerkjenner tilveksten av store datamengder som en samfunnsressurs. Politikken bør følges opp av en strategi som inneholder målrettede tiltak for å utnytte stordata, også for de områdene som er utfordrende. Gjennomføringen bør koordineres slik at det sikres at tiltakene virker positivt i samme retning og at den tradisjonelle silo-tankegangen unngås.

Vi anbefaler en rekke tiltak:

- Etablere et sentralt kompetansemiljø innen stordata og informasjonsforvaltning.
- Økt satsing på veiledning av offentlige virksomheter i informasjonsforvaltning.
- Økt satsing på tilgjengeliggjøring av åpne data for viderebruk.
- Utvikle bedre innsynsløsninger for stordataanvendelser.
- Vurdere å etablere felles testmiljø for utprøving av stordata teknologi.
- Utrede hvordan algoritmer som benyttes til analyser kan gjøres åpent tilgjengelig.
- Utrede og etablere policies for eiendomsrett til data og analyser.
- Vurdere offentlig-privat samarbeid om innovasjonsprosjekter.

2 Innledning

Denne rapporten er utarbeidet av Vivento AS på oppdrag fra Kommunal- og moderniseringsdepartementet i perioden 3.august til 1. desember 2015.

Prosjektgruppen har bestått av:

- Seniorkonsulent Bjørn Holstad, Vivento AS (prosjektleder)
- Seniorrådgiver Gjermund Lanestedt, Agenda Kaupang AS
- Konsulent Torbjørn Bartels, Vivento AS
- Konsulent Johan Nordin, Vivento AS

Rapportens datagrunnlag er dels en kartlegging av bruken av stordata i offentlig sektor og de utfordringer virksomhetene har møtt og de løsninger de har kommet opp med i den sammenheng – dels en kartlegging av bruken av stordata i andre land.

I oppdragsbeskrivelsen nedenfor er problemstillingene nærmere beskrevet.

2.1 Oppdrag

Følgende oppdrag er beskrevet i konkurransegrunnlaget fra Kommunal- og moderniseringsdepartementet:

1. Fremskaffe en oversikt og beskrivelse av hva Norge gjør på stordataområdet, innbefattet hvilke muligheter og utfordringer det er grepet fatt i og hvilke løsninger som evt. er foreslått eller implementert. Beskrivelsen bør omfatte hva som har vært nytten og lærdommen av prosjektene og hvordan man evt. har utnyttet kunnskapen.
2. Gi en kort oversikt med omtale av konkrete stordatasatsinger/prosjekter i inntil 6 andre presumptivt avanserte land, samt beskrive evt. overføringsverdi og grunnlag for disse prosjektene.
3. Beskrive minimum 3 brukercase på hvordan norsk offentlig sektor selv bruker stordata eller hvordan private aktører i Norge har tatt i bruk stordata med offentlige data som en del av datagrunnlaget. I dette ligger også hvordan prosessen fra ide til implementering har skjedd, hvilke resultater bruken av stordata har ført til eller forventes å føre til, hvilke hindringer en har støtt på i prosessen (eks. praktiske, juridiske etc.) og hvordan disse er løst, i tillegg til hvilke muligheter en ser fremover. Der det foreligger en overføringsverdi til andre områder enn det konkret omtalte så bør dette fremkomme. Omtalen skal inneholde beskrivelse av forventede gevinster og i hvilken grad disse ble realisert.
4. Gi anbefalinger om hvordan offentlig sektor i rollen som a) leverandør av rammevilkår kan legge bedre til rette for bruk av stordata på forskjellige samfunnsområder, og i rollen som b) offentlig tjenesteleverandør selv kan ta i bruk stordata, samt forutsetninger for dette.
5. Dersom det er andre forhold relatert til stordata i offentlig sektor som tilbyder mener bør omhandles, så kan dette foreslås og tas med i tilbudet.

2.2 Disposisjon

I kapittel 3 går vi nærmere inn på ulike definisjoner av begrepet stordata («Big Data») og bakgrunnen for fenomenet og teknologien.

I kapittel 4 beskriver vi drivkreftene som ligger til grunn for at bruk av stordata er særlig relevant nå.

I kapittel 5 gir vi en kort oversikt over bruken av stordata i noen utvalgte land; Australia, Finland, Nederland, Storbritannia, Sverige, Sør-Korea og USA. Disse er også nærmere redegjort for i vedlegg 3.

I kapittel 6 beskriver vi status for bruk av stordata i offentlig sektor i Norge. Beskrivelsen er delt inn etter utvalgte samfunnsområder. Ulike stordataprosjekt beskrives i tillegg til at vi beskriver utfordringer og læringspunkter. Vi analyserer hvorfor status er som den er på de ulike områdene, før vi gir en samlet vurdering av utfordringer, muligheter og læringspunkt. I vedlegg 1 beskrives prosjektene fra Statens vegvesen, Helsedirektoratet og Statsbygg mer i detalj. I kapittel 6 gir vi også en kort beskrivelse av ulike forskningsinitiativer i Norge. Tre av forskningsprogrammene er beskrevet ytterligere i vedlegg 2.

I kapittel 7 beskriver vi gevinstpotensialet som vi ser for offentlige virksomheter, basert på kartleggingen av bruk av stordata i utvalgte land og i offentlig sektor i Norge.

I kapittel 8 gir vi en analyse av de utfordringer og hindringer for bruk av stordata i offentlig sektor som vi har identifisert gjennom kartleggingen.

I kapittel 9 vurderer vi aktuelle tiltak for bruk av stordata i offentlig sektor og gir våre anbefalinger om tiltak som kan bidra til at gevinstpotensialet kan realiseres.

I vedlegg 4 redegjør vi for noen sentrale definisjoner på området.

2.3 Metode og avgrensninger

Kartleggingen er basert på intervjuer med representanter fra 20 virksomheter, hovedsakelig fra statlig sektor. I tillegg har vi hatt møter med deltakere i tre norske forskningsprosjekt og med markedsaktører.

Underveis i prosjektet har vi hatt tre møter med oppdragsgiver som har gitt tilbakemelding på arbeidet.

Metoden som er benyttet er i stor grad dokumentstudier av norske og internasjonale publikasjoner. Det eksisterer mye relevant litteratur på området, og mye av dette materialet er gjennomgått og gjengitt som referanser i litteraturlisten. I tillegg er det benyttet kvalitative intervjuer med representanter fra offentlige virksomheter. Intervjuobjektene har fått tema og spørsmålsliste tilsendt i forkant. Vi har også innhentet synspunkter, innspill og dokumentasjon på e-post fra noen virksomheter.

Oppdraget er avgrenset til å dreie seg om bruk av stordata i offentlig sektor, og bruk av stordata i privat regi – basert på offentlige data. Vi har imidlertid ikke lykkes med å finne gode eksempler på slik bruk av stordata i privat sektor og har derfor valgt å fokusere på rent offentlige tiltak. Mange private virksomheter benytter offentlige data, særlig grunndata, i sin

oppgaveutførelse, men det er foreløpig uklart om de benyttes i en kontekst som kan kalles stordata.

Vi har valgt ut tre caser som eksempel på prosjekter fra ulike sektorer som kan gi nyttig læring for andre som vurderer å ta i bruk stordata. Selv om Statsbygg ikke har kommet så langt i sine prosjekt som Statens vegvesen og Helsedirektoratet, har deltakelsen i EU-prosjektet «ProDatamarked» nyttig overføringsverdi til andre sektorer.

Det ble tidlig klart at det neppe finnes initiativer på området i kommunal sektor. Flere kommuner har etablert tradisjonelle datavarehus, men vi har ikke funnet kommunale initiativ eller initiativ fra KS som dreier seg om stordata. Da det faller utenfor mandatet for denne rapporten har vi ikke gått inn på å analysere hvorfor situasjonen er som den er i kommunal sektor.

Utvalget av land vi har analysert er gjort både ut fra tilgjengeligheten på informasjon og oppdragsgivers ønske. Andre land vi kunne tatt med i en slik sammenheng omfatter teknologisk utviklede land som Singapore, Japan, Kina, Tyskland, Belgia, Frankrike, Italia, Danmark, New Zealand og Canada. Det kunne også vært interessant å se hvordan stordata i økende grad utnyttes som teknologi for å håndtere problemstillinger i utviklingsland.

De tiltakene vi har foreslått er basert på de innspill vi har fått fra virksomhetene vi har vært i kontakt med, tiltak som er gjennomført i andre land og tiltak som er beskrevet i litteraturen vi har gjennomgått. Det har ikke vært mulig å undersøke nærmere hvor sterk påvirkning de ulike tiltakene vil ha på mulighetene for å realisere gevinstpotensialet.

3 Stordata bakgrunn og begrep

3.1 Begrepet «Stordata»

I litteraturen om stordata verserer det en rekke ulike definisjoner som til dels har utviklet seg over tid. Begrepet er i noen grad uklart og brukes ofte synonymt med relaterte konsept som «Business Intelligence» (BI) og «data mining». Stordata som konsept kan bare delvis sies å dekke disse andre konseptene. Alle de nevnte tre begrepene handler om analyse av data og i mange sammenhenger avanserte analyser. Men stordata-begrepet blir til forskjell fra de to andre først relevant når datamengdene, antall transaksjoner og antall datakilder blir så store og komplekse at de krever egne metoder og teknologier for å kunne trekke mening ut av dataene.

Dette danner også utgangspunktet for den mest utbredte definisjonen av stordata, de tre V'er[1]; «Volume», «Velocity» og «Variety»:

- Volume = volum: Store mengder av data, fra datasett med størrelser fra terrabytes til zettabytes
- Velocity = omløpshastighet: Store mengder data fra transaksjoner med høy oppdateringsfrekvens resulterer i datastrømmer som kommer i stor hastighet. Tiden til rådighet for å handle på bakgrunn av slike datastrømmer vil ofte være kort.
- Variety = variasjon: Data kommer fra ulike datakilder – både fra eksterne kilder og fra datakilder internt i virksomheten som skal bruke dem. Men dataene har også variabel form; transaksjons- og loggdata fra ulike applikasjoner, strukturerte data som rader av data, ustrukturerte data som tekst, bilder, videostrømmer, lyd m.m.

I tillegg til dimensjonene over har andre aktører tilført andre dimensjoner til V-definisjonen. SAS Institute legger til «variability» og «complexity» i betydningen varierende formater og kompleksitet.

IBM inkluderer også en fjerde «V» som er «veracity» som kan oversettes med sannferdighet eller kvalitet. Veracity speiler graden av usikkerhet knyttet til dataenes kvalitet eller opprinnelse, noe som er avgjørende for hva slags analyser og prediksjoner som kan gjøres på bakgrunn av datasettene.[2] Datakildenes sannferdighet og kvalitet er mellom annet avgjørende for i hvilken grad det kan benyttes algoritmer som tar hensyn til og retter opp kvalitetsbrister eller hull i datasettene. Poenget er særlig relevant i forbindelse med offentlige beslutningsprosesser som for eksempel forvaltningsvedtak, hvor dataenes sannferdighet er svært viktig for å ivareta hensynet til korrekt saksbehandling og innbyggernes rettssikkerhet.

Atter andre miljøer tilfører «Value» (verdi) som en dimensjon i definisjonen, altså at stordata må kunne gi verdi i form av ny innsikt og kunnskap ut over det de originale datakildene har gitt. Arkitekturen og oppsettet som benyttes må i en slik kontekst være kostnadseffektiv, og være designet for å kunne trekke verdi ut av data ved å legge til rette for datafangst med høy oppdateringsfrekvens, analyser og innsikt.[3] Tradisjonelle datavarehus er statiske og ikke

[1] (Løvoll, et al., 2015)
[2] (IBM, 2014)
[3] (O'Railly Radar Team, 2011)

designet for å legge til nye datasett. De er heller ikke tilrettelagt for å søke etter nye sammenhenger i dataene som ikke er definert på forhånd.

Gartners (2012) definisjon inneholder både knytningen til tradisjonell teknologi og det at bruk av slik teknologi må gi merverdi:

Stordata er høy-volum, høy-hastighet, og/eller høy-varierte informasjonsressurser som krever nye former for prosessering for å kunne understøtte bedre beslutninger, innsikt og prosess-optimalisering.

Det EU-finansierte forskningsprosjektet BYTE[4] definerer derimot stordata slik:

Stordata er bruk av stor-volum datasett, som oppdateres med stor hastighet, fra kilder med stor variasjon for å kunne trekke ut verdi (innsikt og kunnskap), og som er basert på kilder med sikker opprinnelse og kjent kvalitet. Som krever kostnadseffektive nye former for dataprosessering for å kunne forbedre innsikt, beslutninger og kontroll. I tillegg er disse kravene støttet opp under nye former for datamodeller, nye infrastrukturtjenester og verktøy som evner å produsere og prosessere data fra kilder med stor variasjon og som kan levere data i varierte former til ulike data og informasjonsforbrukere og utstyrsenheter.[5]

I denne rapporten har vi valgt å basere oss på flere av de eksisterende definisjonene og ender derfor opp med en definisjon bestående av *fem* V-er. Gjennom vår analyse av bruk av stordata i offentlig sektor i Norge og i utvalgte land, har vi lagt denne definisjonen til grunn:

Stordata er analyse av massive samlinger av data («volume»), med stor variasjon i datakilder og formater («variety»), hvor datasettet oppdateres med høy frekvens («velocity») og hvor grunnlagsdataenes opprinnelse og kvalitet er avklart («veracity») og hvor analysen av datasamlingene gir økt verdi i forhold til datakildene («value»).

3.2 Bakgrunn for stordata

Stordata som begrep kan spores tilbake til 1980-tallet og diskusjoner om forvaltning av store mengder data og datasett i både akademia og i industrien.[6] Den første formelle akademiske definisjonen kom først i 2000. Fenomenet har først og fremst hatt sin fremvekst på 2000-tallet med særlig relevans de siste årene.

Data har vært lagret og forvaltet i databaser helt siden 1960-tallet. Effektiv bruk av data har krevd høy grad av struktur på dataene slik at de raskt kan gjenfinnes og brukes. Det finnes i dag enorme mengder strukturerte data i ulike offentlige databaser og registre. Noen av disse er åpent tilgjengelige, mens andre inneholder informasjon som kun kan benyttes av personer med et klart definert tjenstlig behov som for eksempel sensitive personopplysninger eller forretningshemmeligheter.

Data, eller informasjonselement, har til alle tider blitt benyttet som grunnlag for å fatte beslutninger. Data gir ingen mening eller verdi i seg selv, om det ikke gjøres en tolkning eller analyse slik at dataene settes i en eller annen kontekst. Først da kan data bli til informasjon og kunnskap.

[4] (BYTE, Universitetet i Oslo, 2015)
[5] (Løvoll, et al., 2015)
[6] (Løvoll, et al., 2015)

Informasjon i form av tekster i ustrukturerte former, gjerne fremstilt i en tekstbehandler, er lagret digitalt i store mengder. Fremveksten av «world wide web» og internett-teknologi har medført en omfattende publisering og tilgjengeliggjøring av både strukturerte og ustrukturerte data.

Stordata representerer på mange måter et paradigmeskift i måten data utnyttes på, fra det tradisjonelle fokuset på uttrekk og analyser på strukturerte data, til gjenbruk av store mengder data fra alle tilgjengelige digitale kilder. Gjenbruk og analyse av data fra mange ulike kilder, ustrukturerte så vel som strukturerte, har krevd en helt ny generasjon arkitektur og teknologi.[7]

EU–prosjektet BYTE fremstiller paradigmeskiftet slik[8]:

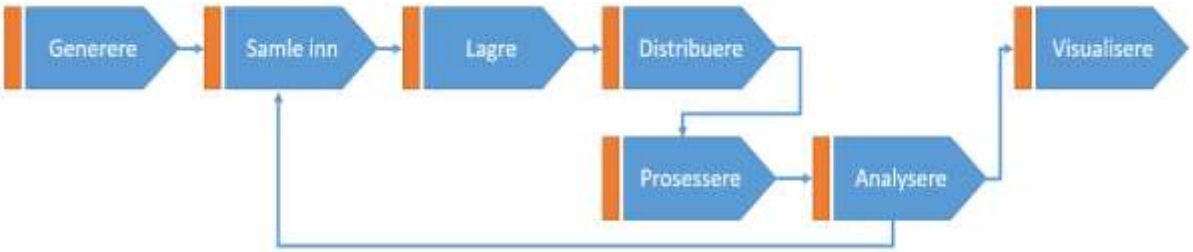
Tradisjonelt paradigme	Nytt paradigme
Noen av dataene: En online transaksjon blir dokumentert med noen hovedfelt slik som tidspunkt og IP-adresse	**Alle data**: Online transaksjoner blir dokumentert ved klikkestrømmer og søkeveier på web-sider, alle datafelt om transaksjon slik som tidspunkt, IP-adresse, geografisk plassering, kanalbruk osv.
Rene data: Datasettene er relasjonelle tabeller, definerte og avgrensede.	**Kaotiske data:** Datasettene er ikke alltid relasjonelle og strukturerte.
Deterministisk: I relasjonsdatabaser har dataene assosiasjoner, korrelasjoner, og avhengigheter som følger klassiske matematiske og statistiske prinsipper.	**Komplekse koblinger:** Data kan kobles, ha duplikater, være overlappende, ufullstendige og ha mange meninger noe som gjør at de ikke kan håndteres i klassiske relasjonelle statistiske læringsverktøy (som f.eks. SPSS) (Getoor & Taskar, 2009)
Utforsking av data for å teste hypoteser: Definerte datastrukturer gjør at utvikling og testing av hypoteser gjøres mot kjente datafelt og sammenhenger.	**Utforsking av innsikt:** Udefinerte datastrukturer krever utforsking for å oppnå kunnskap om sammenhenger som tidligere ikke har vært kjente.
Tids forsinkelse i data analyser: Data må defineres og struktureres før de kan brukes, så samles inn og sorteres. Tiden det tar å gjøre et datauttrekk vil variere, men representerer vanligvis en forsinkelse.	**Sanntids analyser av data: A**nalyser kan gjøres med en gang data er samlet inn.

Tabell 1 Stordata paradigmeskifte – BYTE

3.3 Stordata-verdikjeden og -teknologier

En stor mengde med data skaper ikke verdi i seg selv. Arbeidet med å skape nytte og verdi fra en rekke kilder med stort og variabelt datainnhold betegnes som «verdikjeden» for stordata. Denne verdikjeden starter med at data genereres og/eller identifiseres, deretter samles data inn og lagres før de blir redusert og distribuert til ulike datamaskiner som prosesserer dataene i klynger ved bruk av ulike algoritmer. I neste steg analyseres data ved hjelp av relevante analyseverktøy, og til slutt presenteres resultatet i ulike visuelle former og rapporter. Verdikjeden, som ikke er lineær, kan illustreres slik:

[7] (O'Railly Radar Team, 2011)
[8] (Løvoll, et al., 2015)



Figur 1: Vedikjede for stordata

Generere

Data genereres i økende grad av sensorer og måleenheter. «Tingenes internett» («Internet of Things» - IoT)[9] refererer i stor grad til slike sensorer som sender data i strømmer. Data fra satellitter og ubemannede fly får også en større og større betydning etter hvert som de gjøres tilgjengelig for viderebruk av andre. Havet og havbunnen kartlegges med SONAR[10] og landområder kartlegges med LIDAr[11]. Digitale kamera og videokameraer benyttes i økende grad og gir tilgang til store mengder data. Det genereres også store mengder data gjennom mobile enheter som mobiltelefoner og nettbrett, trafikkdata, GPS-data, posisjonering via mobildatanett og data fra ulike applikasjoner m.m. Fortsatt vil mye data også være menneskeskapt gjennom manuelle registreringer av data og tekster, men det store volumet av data vil i fremtiden komme fra maskiner.

Samle inn

Data samles inn mer og mer som oppslag i andre datakilder og lagres på lavkostenheter, ofte i skyen. I denne delen av kjeden dreier det seg også om å tilrettelegge data. Når data samles inn som mikrodata fra sensorer, web-services o.l. vil de lettere kunne konsumeres enn når de hentes fra ustrukturerte dokumenter fra tekstbehandlere, PDF-dokumenter og HTML-kode. Ulike typer skript benyttes ofte for å tilrettelegge dataene for analyse.[12] I dette steget må også data kvalitetssikres; er dataene riktige, mangler det data, har de riktig betydning i konteksten osv.

Lagre

Data har tradisjonelt blitt lagret i strukturerte relasjonsdatabaser med strenge krav til og regler for struktur. Spørringer mot databasene har vært basert på Structured Query Language (SQL)[13]. I stordatasammenheng er det behov for nye typer teknologier som kan lagre alle typer datasett. Det har derfor vokst fram en rekke produkter som kalles NoSql[14] (New Sql,

---

[9] (Wikipedia, Internet of Things, 2015)
[10] SOund NAvigation and Ranging er en teknikk som bruker lyd under vann til å navigere med eller detektere andre fartøyer. (Wikipedia, 2015)
[11] LIDAR (evt. LiDAR eller Lidar) er en optisk fjernmålingsteknikk som brukes til hurtig måling av fysiske objekters posisjon. Ved å måle tidsforskjellen eller endringer i bølgefasen mellom et emittert lasersignal og et reflektert lys kan avstanden til og andre egenskaper ved objekter beregnes. (Wikipedia, 2015)
[12] (O'Railly Radar Team, 2011)
[13] (ITpro, 2004)
[14] NoSQL omfatter et bredt spekter av forskjellige databaseteknologier som ble utviklet som svar på en økning i volumet av data lagret om brukere, gjenstander og produkter, frekvensen der denne informasjonen er tilgjengelig, og ytelses- og behandlingsbehov. NoSQL benyttes ofte til lagring og gjenfinning av data i XML-

Non-Relational elle Not Only Sql) databaser som kan lagre slike data og gi en viss grad av konsistens mellom dataelementene slik at ulike typer data kan lagres og gjenfinnes. Alle disse produktene har åpen kildekode, slik som for eksempel Google BigTable.

Distribuere og prosessere

Data prosesseres i verktøy som er programmert med algoritmer som sørger for å gjøre de rette uttrekkene, men det benyttes også maskinlæring og dataminimering. Store datasett får verdi først når man har tilgang til prosessorkraft for å behandle og gjøre beregninger på dem. Google introduserte i sin tid «MapReduce»-konseptet som går ut på å distribuere store datamengder på mange datamaskiner i såkalte klynger. Den for tiden mest populære implementasjonen av Google MapReduce[15] er Apache Hadoop[16]. Dette er en løsning som gjør det mulig å prosessere beregninger distribuert på tusenvis av datamaskiner. Opprinnelig var løsningen en batchbasert løsning, men med Hadoop 2.0 kan man kjøre spørringer og kalkulasjoner i sanntid og med god responstid. Andre tilsvarende produkt er Amazon sin Elastic map reducer[17] og Elastic sin Elasticsearch[18]. I følge Loukides[19] er Hadoop en fullstendig stordata-plattform som tilbyr sanntidsprosessering av store datamengder. MapReduce og Hadoop er i dag svært sentrale teknologier i mange stordata–løsninger. MapReduce kan derfor sies å være en de facto standard.

Analysere

Analysene gjøres i ulike typer verktøy. Til en viss grad kan tradisjonelle analyseverktøy benyttes på uttrekk av data. Det å kunne bygge statistiske modeller vil fortsatt være en viktig del av analysearbeidet. Men noe som virkelig gir muligheter for grundige analyser er verktøy for maskinlæring.[20] Til daglig forventer vi at nettsteder, applikasjoner og sosiale media skal gi oss tilbud basert på data om oss. Det finnes en rekke mobilapplikasjoner som baserer seg på slik maskinlæring. Et eksempel er at man tar et bilde av en person og applikasjonen finner dennes identitet ved å benytte ansiktsgjenkjenning og sammenligne med bilder som tidligere er publisert på nett. Det finnes en rekke maskinlærings-bibliotek og verktøy tilgjengelig, som for eksempel PyBrain for Pyton[21], Mahout[22] knyttet til Hadoop og Prediction API fra Google[23].

Resultatet av analyser vil ofte kunne inngå som datagrunnlag og samles inn til nye analyser i andre sammenhenger.

---

baserte dokumenter. Relasjonsdatabaser, derimot, ble ikke laget for å takle de skala- og smidighetsutfordringene som møter moderne applikasjoner, og de var heller ikke bygget for å dra nytte av billig lagringsplass og prosessorkraft tilgjengelig i dag. (MongoDB, u.d.)

[15] (Dean & Ghemawat, 2004)

[16] (Apache, 2015)

[17] (Amazon, 2015)

[18] (Elasticsearch, 2015)

[19] (Loukides, 2011)

[20] (O'Railly Radar Team, 2011)

[21] (PyBrain - The Pyton Machin Learning Library, 2015)

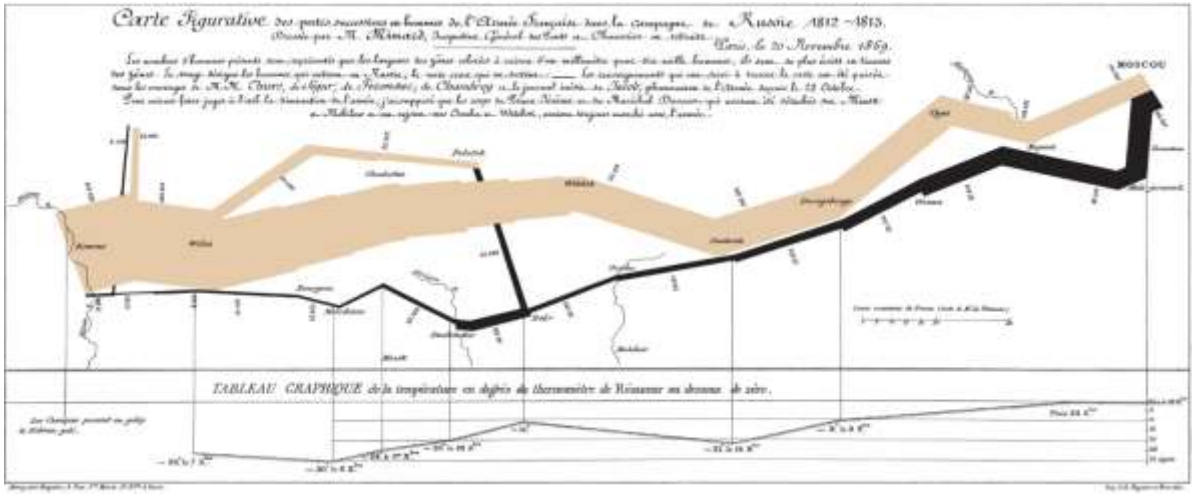
[22] (Apache, 2015)

[23] (Google, 2015)

Visualisere

For å få virkelig verdi av data er det nødvendig med en visualisering eller det som også benevnes som historiefortelling. Å presentere komplekse sammenhenger i forståelige grafer er en stor utfordring som krever spesialkompetanse. Etterhvert er dette blitt en egen profesjon. Bruk av åpne offentlige data og stordata har gjort at visualisering også er i ferd med å vokse fram som marked i seg selv. Norske Claylab[24] er et eksempel på dette, et firma som mellom annet har bidratt til utviklingen av innlandsutvalget.no[25] som betegnes som en interaktiv Norsk offentlig utredning (NOU).

Dersom man klarer å tilpasse det visuelle uttrykket til mottaker så kan en graf være «verdt tusen ord». Men samtidig kan en graf som er vanskelig å forstå for mottakere ha motsatt effekt. Presentasjon og rett visuelt uttrykk for informasjonen kan være like viktig som informasjonen selv.[26] Behovet for å visualisere data på en god måte for ulike interessenter må ikke undervurderes.



Figur 2 Charles Joseph Minard sin graf fra 1869 visualiserer tapene til Grand Arme i felttoget mot Russland 1812 – 1813

Charles Joseph Minards berømte graf fra 1869 viser hvordan størrelsen på Grande Armée ble mindre og mindre etter hvert som Napoleon marsjerte mot Moskva (brun linje fra venstre til høyre) og tilbake (svart strek fra høyre til venstre). Størrelsen på armeen tilsvarer bredden på streken. Denne grafen er eksempel på en grafisk framstilling som forteller en historie og teknikken benyttes den dag i dag til å visualisere strømmer.

De fleste analyseverktøy har også moduler som gjør det mulig å visualisere data på ulike måter. Det er imidlertid et raskt voksende marked for spesialiserte visualiseringsverktøy og tjenester knyttet til å visualisere komplekse sammenhenger. Eksempler på slike verktøy er

[24] (Clay Labs, 2015)
[25] (Innlandsutvalget, 2015)
[26] (Barlow, 2014)

Dygraphs[27], InstantAtlas[28] og –Fusion chart[29]. Datavisualisation[30] er et godt eksempel på en leverandør som leverer skreddersydde visualiseringer basert på ulike typer teknikker.

3.3.1 Løsninger tilgjengelig i markedet

Alle de tradisjonelle Business Intelligence-leverandørene innen analyse- og visualiseringsverktøy (SAS Institute, TABLEU, IBM, Datameter, Pentaho m.fl.) tilbyr i dag også stordataløsninger som støtter Hadoop[31]. Det betyr at deres analyseverktøy benyttes til å gjøre spørringer mot datasett som er tilrettelagt ved hjelp av en Hadoop-løsning.

Datasettene som analyseres er både strukturerte data og mer ustrukturerte dokumenter som for eksempel XML-dokumenter. NoSql har vokst fram som et supplement til tradisjonelle SQL-databaser og muliggjør spørringer mot mer ustrukturerte data. På dette området er de aller fleste løsningene basert på åpen kildekode og er som regel bygd videre på Google BigTable eller Amazon Dynamo. Eksempler er Cassandra[32], Cloudera[33], Hadoop/Hbase[34], Amazon simpledb[35], Elasticsearch[36], IBM Informix[37] m.fl.

Alle store programvarehus tilbyr i dag stordata-løsninger, Google, Microsoft, SAP, Pentaho osv. som dekker det meste av verdikjeden. Hadoop er også et merkenavn og er Apache sin løsning, denne dekker også hele spekteret fra lagring via distribusjon og prosessering til analyse og visualisering.

Tingenes internett og maskinlæring henger som nevnt tett sammen med stordata ved at dataene ofte er samlet inn fra sensorer i sanntid (tingenes internett) og at algoritmene som benyttes til analysene kan være tilrettelagt for å kunne lære på grunnlag av tidligere analyser (maskinlæring). Et eksempel på dette siste er når IBM kombinerer bruk av Watson og stordata i ulike sammenhenger.[38] I tillegg er cloud computing eller skybaserte[39] tjenester sentralt for realisering av stordata-løsninger. Både lagring, prosessering og analyser kan kjøres i skyen, enten disse skytjenestene er private, offentlige eller hybride (avhengig av dataenes sikkerhetsnivå).[40] Her finnes det en rekke leverandører, i Norge satser både Evry og IBM på stordata-tjenester i skyen.

---

[27] http://www.dygraphs.com
[28] http://www.instantatlas.com
[29] http://www.fusioncharts.com
[30] https://www.datavisualization.ch
[31] Hadoop er teknologi som muliggjør analyse av store datamengder ved bruk av mange datamaskiner samtidig
[32] (Apache Cassandra, 2015)
[33] (Cloudera, 2015)
[34] (Apache Hbase, 2015)
[35] (Amazon Simpledb, 2015)
[36] (Elastic, Elasticsearch, 2015)
[37] (IBM Informix, 2015)
[38] (IBM, 2015)
[39] Se definisjonslisten i vedlegg 4.
[40] (Marr, Big Data-As-A-Service Is Next Big Thing, 2015)

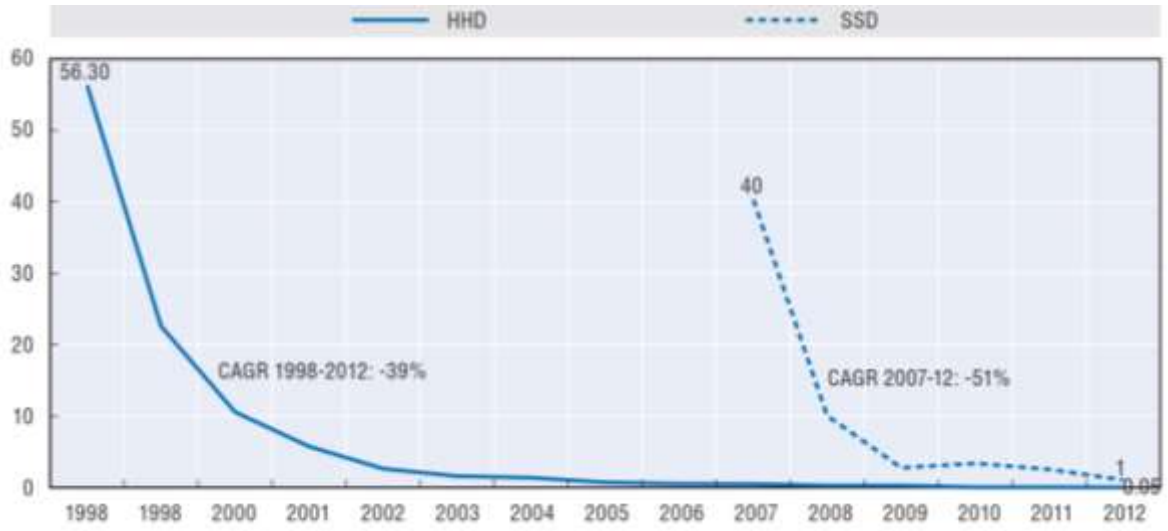
4 Drivkrefter for bruk av stordata

Stordata er en teknologi som er forholdsvis ny, men som vi har vært inne på har nyttepotensial også for offentlig forvaltning. En del forhold ved den teknologiske utviklingen gjør at dette er særlig relevant nå, og er drivkrefter for å begynne å utnytte mulighetene som stordatateknologien gir. Vi beskriver i det følgende disse mest sentrale driverne og muliggjørerne for økt bruk av stordata:

- Lavere kostnader på lagring og prosessering, og tilgang på skytjenester
- God båndbredde
- Tingenes internett
- Maskinlæring og kunstig intelligens
- Hadoop
- Åpne data

4.1 Lavere kostnader og tilgang på skytjenester

Det er blitt mer regningssvarende å generere og lagre store mengder data. Dette skyldes i hovedsak at prisen på lagring av data har gått drastisk ned de siste årene. Prisene har reelt sett gått ned på lagring av data på privatpersoners og virksomheters datamaskiner. I tillegg har framveksten av skybaserte tjenester (Cloud computing)[41] endret tilgjengligheten til og kostnaden for fjernlagring av data. Det har også blitt svært mye billigere å prosessere og analysere data gjennom nye teknologier som multikjerne-prosessorer (CPU) og ikke minst skybaserte tjenester som «data-as-a-service» og «muliti cloud applications». Stordata som en tjeneste i skyen vil framover bli mer og mer vanlig, ifølge Forbes magazine[42].



Figur 3 Pris per gigabyte lagring på harddisker i USD 1998 til 2012[43]

Tidligere kunne lagringskostnader i seg selv være en grunn til at data som ikke lenger var nødvendige for sitt opprinnelige formål, ble slettet. I dag er lagringskostnadene på et slikt

[41] (Wikipedia, Nettskyen, 2015)
[42] (Marr, Big Data-As-A-Service Is Next Big Thing, 2015)
[43] (OECD, 2013)

nivå at data like gjerne kan lagres for evigheten. En illustrasjon er at kostnader per gigabyte (GB) for en tradisjonell PC-harddisk falt fra USD 56 i 1998 til USD 0,05 i 2012 (et gjennomsnittlig fall på 39 % per år). Dette er en trend som bare ytterligere forsterkes av teknologiutviklingen. Med nye typer disker faller kostnaden per GB enda mer.

4.2 Tingenes internett

Tilgangen på data har vokst enormt de senere årene. Dette skyldes bl.a. fremveksten av nye forbrukerdrevne teknologier som smarttelefoner og andre håndholdte enheter som nettbrett o.l. Som vi har pekt på har også utbredelsen av sensorer, videokameraer og andre datakilder i vår omgivelse også bidratt til nye, store datamengder. I følge Forbes magazine[44], SINTEF[45], IBM[46] m.fl. er 90 % av alle data i verden produsert de siste to årene. Innen 2020 vil det bli produsert 1,7 megabyte med nye data hvert sekund for hvert menneske på planeten!

Tingenes internett (IoT) er nettverket av alle fysiske «ting» som er sammenkoblet og som kan samle inn og utveksle data. Sensorer og målestasjoner som avgir data medfører en eksplosjonsartet vekst av data som kan bli gjenstand for analyse. Eksempler på slike dataproduserende sensorer i «ting» finnes nå praktisk talt overalt, for eksempel i biler, klær, elektronikk i hjemmet, hvitevarer, overvåkningskamera, bygninger, strømnett, internettrafikk, mobiltelefonbruk, sosiale media m.m.

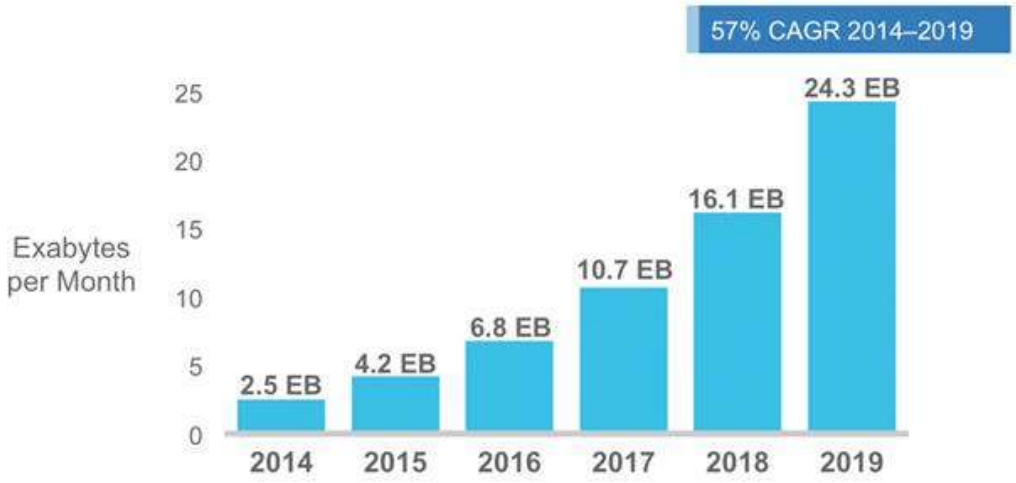
Fremveksten av smarttelefoner og deres applikasjoner (apper) viser tydelig hvordan data kan sammenstilles og danne grunnlag for ny informasjon og kunnskap. Mobiltelefoner kan kombinere trafikkdata, data om geografisk lokasjon, internettilknytning, bilder, lyd og video. Mange av appene baserer seg på innsamling og bruk av data om personer, og mobiltelefonene og deres apper er igjen en kilde for data til bruk for andre analyser og funksjoner.

Til sammen genereres enorme mengder data som transporteres over internett. Det er dette som er utgangspunktet for og gjør det mulig å foreta analyser i sanntid. Ciscos oversikt over mobil datatrafikk i verden er illustrerende. Den viser at mobil datatrafikk økte med 68 % på verdensbasis i 2014 og at det ble transportert 3,5 exabyte data per måned mot slutten av 2014. I 2014 ble det transportert 30 ganger mer data på mobilnettene enn for hele internett i 2000. Cisco spår en gjennomsnittlig årlig vekst i datatrafikken på 57 % i perioden fram til 2019.

4.3 Maskinlæring og kunstig intelligens

Som resultat av forskning og utvikling er såkalt lærende systemer i ferd med å bli en moden teknologi. Dette betegnes ofte som maskinlæring eller kunstig intelligens. Begrepet er ikke utelukkende brukt i forbindelse med stordata, da maskinell mønstergjenkjenning og prediktiv analyse også er begreper innen «data mining». Lærende systemer dreier seg i hovedsak om at datamaskiner er programmert til å gjenkjenne mønster og predikere et sannsynlig utfall basert på tilgjengelig data. Dess større og bedre datagrunnlag, dess mer presise prediksjoner.

[44] (Marr, Big Data: 20 Mind-Boggling Facts Everyone Must Read, 2015)
[45] (SINTEF, 2013)
[46] (IBM, 2015)



Figur 4 Cisco: Årlig vekst (CAGR) i datatrafikk på mobilnettet 2014-2019, exabyte per måned.[47]

4.4 Hadoop

Det synes å være en kjensgjerning at tradisjonell datavarehusarkitektur ikke skalerer til den fleksibilitet, de volumer og den kompleksitet som kjennetegner stordata-konteksten. Raskt økende volumer må kunne kvernes raskt og nye mønster må kunne avdekkes. Det tar for lang tid å laste og prosessere store datamengder i tradisjonelle datavarehus.

Stordata er også relevant fordi ny teknologi som Hadoop gjør det mulig å analysere store mengder data i form av tekst i ustrukturerte dokumenter, lyd, bilder, film m.m., men også strukturerte data - fordelt på mange datamaskiner som er lokalisert på ulike steder i såkalte klynger av datamaskiner. Dette gir tilgang til enorm lagringsplass og prosessorkraft som gjør at det er mulig å håndtere nesten ubegrensede samtidige oppgaver.

4.5 Åpne data

Offentlige data som er publisert på nettet og dermed tilgjengelig for viderebruk (fortrinnsvis maskinlesbare) kalles for åpne data[48]. McKinsey beregnet i 2013 at åpne data innen 7 samfunnsområder har et potensiale for en årlig global gevinst mellom USD 3,2 og 5,4 trillion.[49] I USA er det tilgjengeliggjort over 180 000 offentlige datasett og i Storbritannia over 20 000 datasett.

Mange land satser på økt tilgjengeliggjøring av åpne data som et bidrag til næringsutvikling, økt innsikt om offentlig sektors virksomhet og til effektivisering og innovasjon i offentlig sektor. Mange av dataene kan gjenbrukes i analyser basert på stordata, for eksempel er kart- og geodata ofte anvendt i denne sammenheng. EU ser flyt av data over landegrensene og sektorer som grunnleggende for fri flyt av mennesker, varer, tjenester og kapital i Europa. Et

[47] (CISCO, 2015)
[48] (Data Norge, 2015)
[49] (James Manyika, 2013)

av satsingsområdene i Connecting Europe Facility[50] er derfor etableringen av en europeisk katalog over åpne offentlige datasett i alle landene i EU og EØS.[51]

Økt tilgang til data gjør også at behovet for stordata-teknologi for å sammenstille datasettene og skape merverdi i form av analyser og presentasjoner av disse, øker. Stordata blir derfor mer og mer aktuelt i både offentlig og privat sektor.

[50] (European Commission, 2015)
[51] (European Commission, 2015)

5 Utvalgte land

Vi har gjennomgått status på stordata-området i noen utvalgte land, bl.a. for å ha noe å orientere oss etter når vi skal vurdere bruken av stordata i vår hjemlige kontekst. For hvert land har vi indikert noen sentrale anvendelsesområder for stordata, og hvordan myndighetene har lagt til rette for utnyttelsen av de nye mulighetene.

Generelt synes bruken av stordata internasjonalt fortsatt å være umoden. Det er imidlertid forskjeller mellom land når det gjelder hvor langt de har kommet med hensyn til å ta i bruk stordata som teknologi i forvaltningen. Vi ser at bruken av stordata bl.a. henger sammen med i hvilken utstrekning offentlige dataressurser er tilgjengeliggjort, og hvordan landene har tilrettelagt personvernlovgivningen og andre rammebetingelser for stordata. USA leder an i slik tilrettelegging. Land som Storbritannia og Nederland har også gjort mye – og flere EU-land, Norge og andre velutviklede økonomier som Sør-Korea, Singapore, Japan, Kina, New Zealand og Canada, har strategier eller i det minste stordata-prosjekter i offentlig sektor. Det er også interessant å se at stordata i økende grad omtales som en mulighet for utviklingsland, bl.a. knyttet til landbruk og folkehelse.

5.1 Australia

Australia var tidlig ute med å åpne sine offentlige data med lanseringen av Data.gov.au, etter mønster av USA og Storbritannia. Australia var også tidlig ute (2013) med en egen Big data-strategi[52] som bl.a. definerer «privacy by design», dataintegritet og prosesstransparens som prinsipper for utformingen av stordataløsninger i forvaltningen. Strategien omhandler bl.a. et ansvar for alle som er involvert til å ivareta personvernhensyn, sikre kvalitet i kildedata, opprettholde kontroll med personlige og sensitive data langs hele verdikjeden, og plikt til å gjennomføre risiko- og sårbarhetsvurderinger og høringer før nye stordata-anvendelser implementeres. Australia har for øvrig mange av de samme utfordringer knyttet til personvern og rammebetingelser for stordata som det Norge har.

Samfunnsområder som trekkes fram som aktuelle for stordata er landbruk og helsesektor. I tillegg representerer den australske gruveindustriens bruk av stordata til prosessoptimalisering en interessant case. På tross av at man var tidlig ute med hensyn til å utvikle policies, er det imidlertid vanskelig å finne mange andre konkrete anvendelser av stordata i Australia.

I 2015 utarbeidet regjeringen en «Better Practice Guide for Big Data»[53] som peker på en del forutsetninger for å håndtere stordata (kapasitet, kompetanse etc.), informasjonssikkerhet og personvern, samt viktigheten av ansvarlighet i dataanalysene. Et av tiltakene – etableringen av et «Data Analytics Centre of Excellence» organisert under australske skattemyndigheter – adresserer kompetanseutfordringene og tilbyr rammeverk og verktøy for analyser, og sørger for erfaringsutveksling på tvers av departementsområdene.

For Norge kan det være relevant å lære av de australske erfaringene med å utforme policies, prinsipper og kompetansetiltak for forvaltningen.

[52] (Australian Government, 2013)
[53] (Australian Government, 2015)

5.2 Finland

I Finland har flere sektorer planer om å ta i bruk stordata for mer brukerorienterte tjenester, større effektivitet og bedre underlag for beslutninger. Telekommunikasjoner, skole og utdanning, helse- og omsorgstjenester (vurdering av behandlingskvalitet), smarte infrastrukturer, industrielle prosesser, forskning, trafikkstyring og ressursforvaltning (bl.a. skogforvaltningen) trekkes fram som de mest aktuelle anvendelsesområder.

Åpne data er en viktig del av finsk IT-politikk. Som i andre land er bruken av stordata i stor grad koblet til satsingen på åpne data. Vi finner imidlertid få eksempler på bruk av stordata i offentlig sektor i Finland.

Aktører som Nokia synes å engasjere seg på området. Det nasjonale forskningsrådet Tekes[54] ser stordata som et viktig satsingsområde for finsk industri. Bl.a. gjennom «Industrial Internet»-programmet (2014-2019) allokeres store summer til utvikling av et industrielt, datadrevet økosystem[55]. Det er bl.a. etablert et offentlig-privat samarbeid om et «Data to Intelligence»-program, for å utvikle verktøy og metoder som grunnlag for nye offentlige tjenester og ny næringsvirksomhet.

Stordata ses også som en høyst aktuell eksportartikkel. Samtidig påpeker bransjen selv at man fortsatt er kommet kort[56]. En arbeidsgruppe oppnevnt av det finske samferdselsdepartementet la i 2014 fram et utkast til stordata-strategi[57] hvor de bl.a. pekte på behovet for tilgang til ekspertise og mer samarbeid på tvers av sektorer. De foreslo også at det etableres en stordata-hub og statlige støtteordninger for innovative stordata-anvendelser.

Norge kan lære av finnene når det gjelder samarbeid mellom privat og offentlig sektor i utviklingen av en datadrevet økonomi. Finnene har mer fokus på stordata-relatert næringsutvikling og verdiskapning i privat sektor enn det mange andre land har.

5.3 Nederland

Nederlandske myndigheter bruker stordata blant annet til strategiske prediksjoner, i intelligente trafikksystemer, til forutseende politiarbeid («predictive policing»), avdekking av økonomisk kriminalitet og naturressursforvaltning. Særlig interessant er samarbeidet med IBM og akademia om etableringen av «Digital Delta», et program for utnyttelse av stordata for forvaltning av kystområder og vannressurser og bedre kontroll med oversvømmelser. Andre anvendelsesområder er «Smart City»-konseptet[58] i Amsterdam, landbruk («smart matproduksjon») og helsesektoren.

Nederlenderne mangler en egen, uttrykt politikk for stordata, men mener at de har en unik posisjon på området, bl.a. på grunn av deres satsing på datavitenskap og økonometri ved landets universiteter.

---

[54] (TEKES, 2015)
[55] (TEKES, Industrial Internet, 2015)
[56] (Invest In Finland, u.d.)
[57] (Ministry of Transport and Communications, 2014)
[58] (Amsmarterdam City, 2015)

Nederland har også en offensiv politikk for åpne offentlige data i «Vision on Open Data»[59] fra 2013. Det er igangsatt flere åpne data-prosjekter som skal bidra til vekst og innovasjon, bl.a. på geodata-området. Næringslivet, akademia og myndigheter har nylig inngått et tettere samarbeid for å bringe Nederland helt i tet internasjonalt på stordata-området, gjennom etableringen av «Big Data Alliance»[60], som en arena for samarbeid og erfaringsutveksling for fagfolk og forskere på området.

Sannsynligvis har vi noe å lære av Nederland innenfor de aktuelle anvendelsesområdene (miljøområdet, geodata, kriseberedskap, «smart cities») hvor Nederland allerede har erfaringer over flere år.

5.4 Storbritannia

Storbritannia fikk i 2013 på plass sin strategi[61] for dataanalyse og stordata-anvendelser – *Seizing the data opportunity*, med forslag om betydelige investeringer i teknologiutvikling og forskning på området. Strategien har fokus på utvikling av kompetanse innen dataanalyse og rammebetingelser tilrettelagt for deling og linking av offentlige data. Britene var tidlig ute med å lansere en portal for åpne, offentlige data – Data.gov.uk. Portalen er en kilde bl.a. for stordata-relatert verdiskapning i privat sektor. Man regner med at stordata vil skape mange nye jobber innen et bredt spekter av bransjer i Storbritannia.

Viktige anvendelsesområder er strategisk «foresighting»[62], miljø- og ressursforvaltning, avdekking av økonomisk kriminalitet, samt «predictive policing». Storbritannia er et foregangsland med hensyn til å tilpasse og tilrettelegge lov- og regelverk til å understøtte stordataanvendelser[63]. De har bl.a. bestrebet seg på å tilpasse IPR-lovgivningen[64] til å gi enklere adgang til å analysere data på tvers av kilder med ulike eierskap. For områder som informasjonssikkerhet og personvern er det utarbeidet retningslinjer, prinsipper, verktøy og veiledningsmateriell knyttet til stordata. Akademia og myndigheter har gått sammen om å etablere et nettverk for anonymisering som gir ekspertråd om anonymiseringsteknikker.

Vi tror Norge med fordel kan lære noe av britene når det gjelder utvikling av verktøy, veiledninger og kompetanse på stordata-området, i tillegg til å tilrettelegge de juridiske rammebetingelsene slik at det blir enklere å implementere samfunnsnyttige stordata-anvendelser.

5.5 Sverige

Sverige har ikke vært blant de fremste til å satse på stordata i offentlig forvaltning, selv om Vinnova[65] har utformet en forsknings- og innovasjonsagenda for stordata[66]. Personvern-utfordringer, uklar lovgivning for skytjenester, for få åpne offentlige datasett og mangel på

[59] (Algemene Rekenkammer, 2014)
[60] (Big Data Alliance, 2015)
[61] (HM Government, 2013)
[62] (ScienceDirect, 2009)
[63] (Byte, 2014)
[64] Intellectual property rights (IPR)
[65] (Vinnova, 2015)
[66] (Vinnova, 2015)

arbeidskraft med dataanalyse-kompetanse har vært blant faktorene som trekkes fram som forklaring på den beskjedne utbredelsen av stordata i offentlig sektor i Sverige.

I BADA-prosjektet[67] samarbeider Volvo, Scania og det svenske Trafikverket om å utvikle et trafikk-økosystem av digitale enheter. Drivkraften for et slikt økosystem er ønsket om større effektivitet og transportkapasitet, bedre ressursutnyttelse, sikkerhet og positive miljøeffekter. I Stockholm har IBM lenge prøvd ut stordata-teknologier for trafikkoptimaliseirng –hvor datakilder som GPS fra biler, radarsensorer langs hovedveiene, registerinformasjon om veiarbeid og hendelser, målinger av luftkvalitet og værforhold, til sammen og i tilnærmet sanntid brukes til å beregne køer, tidsbruk og optimale transportruter mellom punkter i byen.

Enkelte private aktører i Sverige har vært tidlig ute med å utnytte stordata. Et selskap som Klarna bruker stordata til å vurdere risiko i forbindelse med finansiering. Svenske Spotify benytter stordata til å generere spillelister og foreslå musikk for brukerne[68]. Sveriges stordatasatsing synes å være mer innrettet mot tradisjonelle kjerneområder som telekom, nettverksbaserte tjenester, bilindustri/transport og industriautomatikk – enn mot offentlig sektor. På policyutvikklingssiden vektlegges FoU, infrastruktur, kompetanse og erfaringsutveksling. Det er bl.a. foreslått å etablere et åpent nettverk av industrielle aktører og representanter for akademia som kan virke som koblingspunkt mellom interessenter.

Det er verdt å merke seg det industrielle fokuset her som kan representere spennende læring for en eventuell tilsvarende norsk satsing.

5.6 Sør-Korea

Sør-Koreas arbeid med stordata skjøt fart i 2012, gjennom etableringen av et «Big Data Forum» og en «Big Data Master Plan», hvor kjernen er en trinnvis innføring av stordata på relevante samfunnsområder og etablering av en nasjonal infrastruktur for stordata, deling av erfaringer, teknisk support og utdanning av eksperter.

De tidlige pilotene inkluderer predikering og varsling av skogbranner, smart katastrofe- og beredskapskartlegging og analyser av medisinsk informasjon. Fra 2015 utvides satsingen til helsetjenesteområdet og ytterligere 10 nye tjenesteområder. I 2016-2017 vil fiskeri og maritim virksomhet og ytterligere nye samfunnssektorer inkluderes i satsingen. Et av de anvendelsesområder som vises til i flere sammenhenger er det å avdekke korrupsjon, noe som har fokus i mange asiatiske land. Andre anvendelsesområder er værmeldingstjenester (med teknologi fra IBM)[69], analyser av klager på offentlige transporttilbud, boligmarkedstiltak, arbeidsløshetstiltak, vurderinger av industriell kapasitet og råvareforsyningen, samt analyser av oppvekstsvilkår for å forutsi mulige sosiale problemer i byområder.

Som en del av sitt Government 3.0-initiativ i 2013 har koreanske myndigheter åpnet mye av sine data. Dessuten er de tidligere silogrensene mellom ulike deler av statsforvaltningen fjernet, for å muliggjøre mer samarbeid på tvers – bl.a. om bruk av stordata.

For Norge kan det være interessant å vurdere om den trinnvise innføringen av stordata i sektor etter sektor – slik Sør-Korea har hatt som strategi, kan ha noe for seg som

[67] (Vinnova, 2015)
[68] (Computer Weekly, 2014)
[69] (IBM, 2013)

fremgangsmåte for å målrette innsatsen inn mot de mest relevante tjenesteområdene og samfunnsutfordringene. Som for flere av de andre landene framstår det å gi tilgang til åpne offentlige data som en svært sentral komponent i Sør-Koreas politikk på området.

5.7 USA

USA har gjennom et tett samarbeid mellom myndigheter og teknologileverandørene (IBM, EMC, SAS, Amazon m.fl.) utviklet teknologier og infrastruktur for dataanalyse. Fokuset har vært på prediktiv analyse og forebyggende innsats på områder som kriminalitet, skatt, sikkerhet, forskning, naturressursforvaltning og helsetjenester. Mange enkeltmyndigheter har lansert sine egne stordata-program. F.eks. har skattemyndighetene et program for å avdekke skatteunndragelser og bedrageri[70]. En utredning fra mai 2014 – «*Big data: Seizing opportunities, preserving values*» peker på helsesektoren (prediktiv medisinering, diagnoser etc.), skole og utdanning (learning analytics; optimaliserte utdanningstilbud), samfunnssikkerhet (trusselbilde-analyse) og kriminalitetsforebygging («predictive policing») som sentrale samfunnsområder med stort gevinstpotensial for bruk av stordata.

Myndighetene har publisert hundretusener av åpne datasett som omhandler mange samfunnssektorer og er råvare for innovative anvendelser, også stordata-anvendelser. Sammen med den tidlige utprøvingen av stordata gjør tilgangen på offentlige data at USA i dag kan vise til anvendelser på mange områder – både på føderalt og lokalt nivå – og i frivillighetssektoren. F.eks. har myndighetene i flere byer sammen med IBM lansert «Smart City»-satsinger med stordata som en viktig del.

Uten generell personvernlovgivning synes terskelen for dataanalyse på alle typer offentlige data å være lavere i USA. Dette er sannsynligvis en av grunnene til at USA har kommet langt på området. Vi ser ellers at rammebetingelsene er tema også i den amerikanske debatten. Det hvite hus har bl.a. lagt fram policydokumenter på områder som prising av offentlig informasjon og personvern. Nylig er muligheter og utfordringer knyttet til bruk av stordata i høyere utdanning til å avdekke svak framdrift blant studentene (og sette inn tiltak i tide) diskutert. Policyutviklingen er ellers i hovedsak rettet mot tilgang til data. Det hører til bildet at det opphavsrettslige rammeverket i USA – i mindre grad enn i Europa – utfordres av nye teknologier som stordata, bl.a. gjennom unntak i loven for «fair use».

Vi bør kunne lære av USA når det gjelder tilrettelegging av de juridiske rammebetingelsene for stordata. På «smart city»-området kan norske myndigheter lære noe av amerikanske storbyer som var tidlig ute med å planlegge en bærekraftig og datadrevet, urban økologi.

5.8 EU

EU-kommisjonens strategi «Veien mot en blomstrende, datadrevet økonomi»[71] påpeker at EU må gripe de mulighetene stordata gir, til nytte for innbyggere, økt økonomisk vekst, flere og bedre jobber, skreddersydde personlige produkter og tjenester av høy kvalitet. Samtidig vektlegges hensynet til det europeiske verdisynet på personvernområdet. I EUs tilnærming til stordata pekes på at det er nødvendig med tiltak på følgende områder:

[70] (Kim & Trimi, Big-Data Applications in the Government Sector. Communications of the ACM. Vol. 57, no. 3, 2014)
[71] (EU, 2014)

- Tilgjengelighet til data – gjennom tekniske standarder og relevante lovverk
- Kompetanse – gjennom bl.a. kompetansesentre og nettverk
- Samspill – fasilitere partnerskap og samarbeid på tvers av EU-landene
- Teknologi – besvare utfordringer som måtte oppstå på hvert nivå i verdikjeden
- Prosessering og datalagring – støtte distribuerte e-infrastrukturer for Europa
- Personvern og datasikkerhet – gjennom et balansert juridisk rammeverk som bidrar til høy tillit til den digitale økonomien.

I 2014 var det ca. 150 pågående EU-finansierte prosjekter (50 mill. euro i 2014 og 89 mill. euro i 2015) innenfor sektorene helse, transport, miljø, jordbruk, tjenesteyting og salg[72].

[72] (EU, 2015)

6 Bruk av stordata i offentlig sektor i Norge

I dette kapittelet gjengir vi eksempler på samfunnsområder og enkeltvirksomheter som behandler og analyserer store datamengder fra ulike kilder og som har tatt i bruk eller vurderer å ta i bruk stordata. Vi har lagt vekt på å trekke frem utfordringer de har møtt, løsninger de har benyttet og viktig læring de har gjort.

Vi gjennomgår eksemplene område for område. Gjennomgangen er ikke ment å være uttømmende, men peker på prosjekter som har gitt nyttig læring.

Til sist i kapitlet gir vi en samlet analyse av status for bruk av stordata i norsk offentlig sektor. Tre sentrale prosjekter fra Helsedirektoratet, Statens vegvesen og Statsbygg beskrives mer i detalj i vedlegg 1.

6.1 Samferdsel

På samferdselsområdet både produseres og utnyttes det svært store mengder data. Her er også graden av sanntidsdata høy, særlig når det gjelder trafikkovervåking på vei, i lufta og til sjøs. Sektoren vil derfor presumptivt også ha god nytte av dataanalyser. De fleste samferdselsdata er stedbundne, det vil si at de kan berikes med geografiske lokasjonsdata. Derfor blir tilgangen til gode kartdata svært viktig for analyser og bruk i denne sammenheng.

6.1.1 Statens vegvesen

Statens vegvesen har lenge samlet inn og benyttet store mengder trafikkdata fra ulike kilder og har klare strategier for utnyttelse av data på ulike områder. Det er også de som har kommet lengst i å planlegge og innføre bruk av stordata på sitt område. Satsingene er forankret i ulike strategier og handlingsplaner. Vegvesenets strategiske posisjon er å være sentral leverandør av kvalitetssikrede veidata og -analyser. Et eksempel på dette er tilgjengeliggjøring av Norsk veidatabank (NVDB) som er åpent tilgjengelig gjennom eget API og tilgjengelig på norsk katalog over åpne data på data.norge.no[73]. Et eksempel på anvendelse av denne datakilden er www.vegkart.no.

Intelligente transportsystemer (ITS) som konsept omfatter løsninger som benytter informasjons- og kommunikasjonsteknologi (IKT) i et trafikk- eller transportsystem, med den hensikt å bedre informasjonstilgang, påvirke atferd og forbedre transportløsninger og trafikkavvikling slik at man oppnår bedre fremkommelighet, trafikksikkerhet, miljø, tilgjengelighet og brukertilfredshet.[74] Sentralt i ITS står bedre utnyttelse av trafikkdata.

Vegvesenets strategi på utnyttelse av trafikkdata er nedfelt i Handlingsplanen for trafikkdata 2013-2018[75]. Første steg i strategiens handlingsplan er å etablere et nasjonalt nett av trafikkregistreringsstasjoner på riks-, fylkesveger og noen sentrale kommunale veger.

[73] (Direktoratet for forvaltning og IKT, 2015)
[74] (Statens vegvesen, 2014)
[75] (Vegdirektoratet, 2012)

Datagrunnlaget til Statens vegvesen dreier seg mer og mer om sanntidsdata enten det gjelder trafikkdata, reisetid, godstransport og veikvalitet. Dataene vil være basert på målinger fra sensorer, kamera osv., men også på manuell registrering av data.

Case: Datainn
Datainn dreier seg om innsamling og behandling av trafikkdata. Prosjektet etablerer i overkant av 5000 målepunkt som skal utgjøre et nasjonalt nett av trafikkregistreringsstasjoner for sykkel og motorkjøretøy.

Formål:
De overordnede målene er å sikre bedre beslutningsgrunnlag for alle typer veiprosjekt, bedre trafikkavvikling og ikke minst bedre trafikksikkerhet. I tillegg kan mye av dataene frigis til relevant forskning.

Nytte:
Vegvesenet ser allerede gevinster i form av langt bedre informasjon til trafikantene og dermed bedre informerte trafikanter. I tillegg er det tatt ut gevinster i form av at styringen av drift og vedlikehold er bedre. Bedre data gir mye bedre grunnlag for å lage prognoser og å planlegge vedlikehold. Dette gjør at det spares vesentlig på vedlikeholdskostnadene.

Læring:
Bruk av stordata-teknologi har gitt mye verdifull læring. Virksomheten har mellom annet endret måten IT og fag samhandler på. Dette dreier seg både om tilgang på kompetanse og kapasitet, men også om fleksible arbeidsprosesser og måter å levere IT-tjenester på. En av de store utfordringene for Vegvesenet framover er tilgang på tilstrekkelig prosessorkraft. Spørsmål om bruk av privat og offentlig sky, etablering av egne datasenter m.m. står sentralt i de vurderingene som gjøres.

Stordata i Statens vegvesen er nærmere beskrevet i vedlegg 1 – tre utvalgte case.

Det arbeides også med å utvikle en løsning for å kunne benytte rutedata og avviksdata som vil gi bedre avvikling av kollektivtrafikk. Regjeringen har besluttet å opprette et et eget selskap[76] som skal stå for innsamling av rute- og avviksdata og utvikle ulike tjenester basert på disse dataene, som for eksempel en nasjonal ruteplanlegger.

6.1.2 Kystverket

Kystverket har utviklet en rekke tjenester for informasjon om kysttrafikken. Noen av disse er skreddersydde løsninger for andre offentlige myndigheter slik at de kan hente ut relevante skipsopplysninger uten å be om enkeltvis rapportering fra skip og rederier. Tidligere måtte fartøy som seilte på norskekysten ofte rapportere samme informasjon flere ganger til ulike myndigheter. Det nasjonale meldingssystemet SafeseaNet forenkler det administrative

[76] (Nasjonal Reiseplanlegger, 2015)

arbeidet hos sjøfarende og myndigheter samtidig som det bidrar til sjøsikkerhet, havnesikring og effektiv sjøtransport ved å lagre, hente og utveksle fartøysopplysninger.[77]

En annen viktig datakilde er AIS[78]. AIS (Automatisk Identifikasjons System) er et antikollisjonshjelpemiddel som er innført av FNs sjøfartsorganisasjon IMO for å øke sikkerheten for skip og miljø, samt forbedre trafikkovervåking og sjøtrafikktjenester. En AIS-transponder[79] ombord på et skip skal automatisk og med nødvendig nøyaktighet og oppdateringsrate forsyne andre skip og kyststaters myndigheter med informasjon fra skipet. Gjennom å sende sin egen informasjon og å motta andre skips informasjon, kan alle skip med AIS om bord danne et bilde av trafikksituasjonen i sitt nærområde. Det finnes også AIS-transpondere på land (i f.eks. fyr), ombord i lystfartøy (AIS klasse B) samt i redningshelikopter og -fly.

Case: Forprosjekt AIS-data

Formål:
Utrede og teste ut konsepter for lagring og tilgjengeliggjøring av AIS-data.
Målet er å etablere en løsning hvor det er gis tilgang til bl.a. hele arkivet av historiske AIS-data (rådata med høy oppløsning), samt tilrettelagt for kopling mot andre registerdata.
Nytte:
AIS-data har et svært stort potensial, og forvaltning og anvendelse av dem vil få et betydelig fokus i Kystverket i årene fremover. Anvendelsesområdene for slike tjenester er mange, det kan for eksempel legge til rette for at det kan gjøre sannsynlighetsanalyser for uønskede hendelser i framtid, men også predikere trafikkvolum og miljøutslipp.
Læring
Dette er nå i en planleggingsfase. Teknologivalg er en del av dette. Valg av stordata-teknologi og skytjenester er deler av arkitekturvalgene man står overfor. Dette er en betydelig kompetanseutfordring for Kystverket.

Kilde: E-post med informasjon fra Kystverket

Tjenesten Havbase[80] inneholder historiske posisjonsdata fra AIS. Disse posisjonsangivelsene berikes så med tall om utslipp av ulike stoffer på distansene som seiles. Utslippstallene er estimert på grunnlag av skipets fart og informasjon om skipets motorstyrke, drivstoff, størrelse, type m.m. hentet fra skipsregisteret IHS Fairplay. Mulighet for kobling mot ulike skipsregistre gjør det mulig å presentere detaljert stedfestet informasjon om seilt distanse.

[77] (Kystverket, 2015)
[78] (Kystverket, 2015)
[79] (Kystverket AIS, 2012)
[80] (Kystverket "Havbase", 2011)

Tjenesten Kystinfo[81] er basert på data fra mellom annet AIS, men også en rekke andre kilder. Det er en web-basert tjeneste som gjør det mulig for hvem som helst å gjennomføre analyser på et kartgrensesnitt på problemstillinger knyttet til kystfart.

Barentswatch[82] har etablert tjenester om hav- og kystområder lettere tilgjengelig for myndigheter, beslutningstakere og allmenne brukere. Dette skal både forenkle tilgang til og sikre utveksling av offentlig informasjon. Det er etablert tjenester som viser data om bølgehøyde, fiskeri, havner og værforhold i polare strøk.

6.1.3 Luftfartstilsynet

Luftfartsmyndighetene i Norge ser sikkerhetsdata som fundamentalt for den entydig positive sikkerhetsutviklingen som luftfarten har hatt de siste årene – og som enda mer sentralt i fremtiden. Særlig sentralt står informasjon om risiko og forståelse og styring av risiko, uansett om det er snakk om små aktører eller store internasjonale organisasjoner. Reguleringer og læring fra enkeltsaker må suppleres med metodisk utvikling av risikostyring av flysikkerheten.

Tilsynet sammenstiller informasjon fra egne inspeksjoner med informasjon om hendelser og avvik som rapporteres inn til tilsynet gjennom Altinn (om lag 8000 hendelser i året). Dette gir grunnlag for kommunikasjon og tilsyn med aktuell organisasjon. I tillegg rapporteres informasjonen videre til en europeisk database som skal legge grunnlaget for bedre analyser. Flight Data Monitoring (FDM) -systemer er etablert for flyselskapene, og kan gi verdifullt supplement til slike analyser av flysikkerheten.

6.1.4 Utfordringer og læring innen samferdselssektoren

Luftfartstilsynet produserer en rekke statistikker og analyser av luftfartssikkerheten på ulike indikatorer[83]. Men foreløpig har ikke Luftfartstilsynet egne initiativ for å ta i bruk stordata. Dels skyldes dette at de internasjonale initiativene på området er noe umodne, dels skyldes det begrenset datavolum og datakvalitet i alle fall når det gjelder internasjonale datakilder. Disse forholdene arbeides det med og ulike IT-verktøy vil være helt sentralt i analysearbeidet knyttet til det å avdekke og forstå risiko i luftfarten.

En av utfordringene innen samferdsel er personvernet, knyttet til det å skulle overvåke den enkeltes posisjon og bevegelser. Men personvernproblemstillingene blir først aktuelle når man kommer inn på opplysninger om hver enkelt bilfører, båtfører eller passasjer. Disse dataene er bare unntaksvis sensitive og i mange sammenhenger uinteressante.

Vegvesenets data er i stor grad åpent tilgjengelige og registrert i datakatalogen på data.norge.no[84]. De har også tatt et strategisk valg om å være leverandør av kvalitetsdata om vei og samferdsel samtidig som de utnytter dataene selv. Kystverket bør vurdere det samme slik at også deres data kan benyttes i nye tjenester. Vi kaller data som er «publisert» og dermed tilgjengelig for viderebruk (fortrinnsvis maskinlesbart) for åpne data. Disse åpne

[81] http://www.kart.kystverket.no/
[82] https://www.barentswatch.no/
[83] (Luftfartstilsynet, 2015)
[84] Norges offisielle oversikt over åpne offentlige data finnes på http://www.data.norge.no, deling av åpne data og bruk av data.norge.no er forankret i digitaliseringsrundskrivet hvor regjeringen gir føringer for digitalisering av offentlig sektor https://www.regjeringen.no/no/dokumenter/digitaliseringsrudskrivet/id2462793/

dataene kan være gratis, eller de kan være underlagt ulike betalingsregimer. Lukkede data er derimot data som ikke er direkte tilgjengelig for brukere utenfor virksomhetene. Tilgang skal her forstås vidt, dvs. ikke bare som fysisk tilgang i den forstand at de er mulig å finne fram til og lese på nettet, eller å laste ned. Åpne data må i praksis også være logisk og semantisk tilgjengelig gjennom data om datasettet (metadata), dvs. at de er lagret og merket på en slik måte at de er mulig å forstå og sette inn i riktig kontekst.

Ser vi på andre land så er det klart at de landene som har en offensiv politikk på å tilgjengeliggjøre offentlige data for viderebruk også er kommet lengst på stordataområdet. Dette skyldes nok oppmerksomheten om dataenes verdi som ressurs og dermed fokus på forvaltning av ressursen. Vegvesenet er et norsk eksempel på et slikt syn på verdien av dataressursene, og er med dette på mange måter en foregangsvirksomhet både med hensyn til å tilgjengeliggjøre og til å utnytte data.

Statens vegvesen finner at den tekniske kompetansen er tilgjengelig i markedet, men peker også på at det først og fremst er viktig å sette veiingeniørene i stand til å forstå og analysere data. I tillegg synes det å være en mangel på data- eller informasjonsvitere som både har kompetanse på å forstå og tilrettelegge datasett og behersker statistisk analyse. Statens vegvesen arbeider selv med å bygge kapasitet og kompetanse på disse områdene.

Kystverket har lang erfaring med forvaltning av data om kystfarten, men står nå overfor valg som krever ny kompetanse i virksomheten. I lys av at det er mange som står overfor de samme valgene knyttet til å få merverdi ut av tilgjengelige data uten å ha kompetanse nok om de teknologiske mulighetene som finnes, ser derfor Kystverket behov for at det offentlige etablerer arenaer for kunnskapsdeling og innovasjon.

Utfordringene som nevnes knyttet til tilgang på kompetanse er reelle både når det gjelder å kunne gjøre analyser på aktuelle data, men også når det gjelder å gjøre teknologiske valg. Kystverket står midt opp i disse vurderingene nå og savner samarbeidsparter å henvende seg til i offentlig sektor. Behovet for kompetanseutveksling og nettverksbygging burde kunne imøtekommes med forholdsvis beskjedne ressurser.

Spørsmålet om tilgang på kompetanse i markedet besvares likevel ikke av dette alene. Selv om den kanskje viktigste hindringen for å kunne ta i bruk stordata-teknologi er kompetanse, kan det synes som om utdanningsinstitusjonene og markedet henger etter med å utdanne og rekruttere teknologer og datavitere.

En av grunnene til at noen virksomheter, slik som Statens vegvesen, ligger godt an i sitt arbeid med å ta i bruk mulighetene som stordata gir, er etter vår vurdering at de har lang erfaring med datafangst i sanntid og det å forvalte data. Dette setter dem i stand til å utnytte data og å tilrettelegge data i tjenester rettet mot ulike typer brukere.

Et annet relevant forhold er det strategiske perspektivet. Virksomheten må se verdien av dataanalyse som strategisk satsingsområde. Statens vegvesen har klare strategier for å hente inn og å utnytte data, og de har posisjonert virksomheten i et helhetlig data-økosystem som tilbyder av kvalitetsdata på sitt område.

Et interessant perspektiv som er reist av Statens vegvesen dreier seg om at stordata-konseptets fleksible egenskaper skaper behov for nye leveransemodeller av IT-tjenester til og i virksomhetene. Fleksibiliteten og dynamikken i stordata gjør at fagsiden krever større

fleksibilitet i IT-leveransene. Endringer må gjøres raskt og behovene vil variere over tid. Vi snakker her ikke bare om smidig utvikling, men også om smidig drift og forvaltning. Dette minner om DevOps konseptet som nettopp dreier seg om å få til smidig og fleksibel utvikling, kvalitetssikring (test) og drift, for å kunne sette IT-løsninger enda raskere i produksjon.

Det er for tidlig å si noe om dette blir en trend etter hvert som sensorteknologi og stordata tas mer i bruk i offentlig sektor. Mye tyder i hvert fall på at DevOps som konsept vokser i utbredelse og at dette ikke bare er knyttet til smidig utvikling av programvare, men også til økende bruk av skytjenester og stordata. Dette er et område som Vegvesenet har erfart som en utfordring og læringsområde – og som andre virksomheter som skal inn i samme prosesser som Vegvesenet må være oppmerksomme på og håndtere i sine virksomheter.

Et siste viktig punkt i erfaringene fra denne sektoren er spørsmålet om tilgang på prosessorkraft. Offentlig sektor har tradisjon for å drifte IT-løsningene selv eller sette driften ut til norske driftsleverandører. Dette er en modell som ikke skalerer i særlig grad og som er langt dyrere enn å ta i bruk skybaserte tjenester for både lagring og prosessering. Statens vegvesen har behov for å vurdere nye skybaserte driftskonsepter, siden behovet for prosessorkraft vil øke etter hvert som datavolumene øker og analysene blir mer og mer avanserte. Problemet er ikke løst og vil være et viktig strategisk spørsmål for Vegvesenet.

6.2 Justis

I Overordnet nasjonal strategi for bekjempelse av datakriminalitet fra mai 2015 (Datakrimstrategien) peker politiet på bruk av stordata som en metode for å kunne analysere veldig store mengder digital informasjon. Det pekes på at økte mengder digital informasjon vil kreve nye metoder for å trekke ut relevant kunnskap. Strategien foreslår å etablere et datakrimsenter som mellom annet utvikler metoder for å trekke kunnskap ut av store datamengder tilgjengelig på ulike datamaskiner. Kripos har allerede satt i gang utviklingsprosjekter for å ta i bruk stordata-teknologi.[85]

Teknologirådets ferske rapport «Forutseende politi» vurderer i hvilken grad teknologien setter politiet i stand til å være mer forutseende og å kunne predikere kriminelle handlinger framfor å reagere i etterkant.[86]

I rapporten foreslås det at politiet tester ut mulighetene som de nye metodene gir, basert på positive erfaringer i andre land. Men samtidig peker rådet på en del forhold som vil være utfordrende ved å ta i bruk slik teknologi:

- Algoritmene som legges til grunn for analysene må være nøytrale og transparente
- Analysene må baseres på kvalitetsdata som er fullstendige og korrekte
- Det må vurderes nøye på hvilke områder prediktive analyser er gyldige

Som vi redegjorde for i kapittel 5 er forutseende politiarbeid («predictive policing») tatt i bruk i utstrakt grad i flere andre land som USA, Storbritannia og Nederland. I Norge er justissektoren kommet relativt kort på dette området selv om potensialet er stort. Analyse av mønster fra politiets egne databaser kombinert med kartdata, matrikkeldata, digitale spor fra

[85] (Politidirektoratet, 2014)
[86] (Teknologirådet, 2015)

ulike elektroniske enheter osv. kan for eksempel bidra til at politiet settes i stand til å predikere hvor det mest sannsynlig vil skje kriminelle handlinger. Eksempler fra andre land viser at sikkerhetsmyndighetene også benytter slik teknologi til å predikere terrorhandlinger og andre sikkerhetstrusler.

6.2.1 Oslo sentrum politistasjon

Sentrum politistasjon i Oslo har gode erfaringer med bruk av dataanalyse på data om lommetyveri kombinert med geografisk informasjon. Gjennom slike analyser har de avdekket hvor tyveriene skjer og hvem som er typiske offer. Selv om dette neppe kan kalles analyse på stordata så viste metoden seg å være effektiv i forhold til å treffe forebyggende tiltak.

6.2.2 Politidirektoratet

Politidirektoratet har etablert et datavarehus som benyttes til analyser og etterforskning basert på data fra saksbehandlingssystemene i politietaten. Løsningen driftes og forvaltes av Politiets IT-tjeneste. Det finnes derfor kompetanse i etaten på dataforvaltning og analyse.

Muligheter til å håndtere stordata på en god måte vil spare tid i etterforskningssammenheng, men det antas også at antallet saker som fanges opp vil øke drastisk. Særorganet Økokrim er blant de som har erfart både mer effektiv etterforskning og økt tilgang på saker.

6.2.3 Utfordringer og læring justis

Vi har ikke hatt tilgang til informasjon om Kripos og PST sitt arbeid på området, men i justissektoren for øvrig er ikke stordatateknologi tatt i bruk i særlig grad. Det finnes dog mange eksempler på god erfaring med analyser basert på datavarehus. Det er vår vurdering at justissektoren kan ha stor nytte av å videreutvikle sine arbeidsformer i retning av prediktive analyser slik Teknologirådet foreslår[87]. Erfaringene med stordatabaserte analyser i Økokrim er i stor grad positive, selv om potensialet for å videreutvikle metodene også er store her.

Vi tror det også er behov for å se på måten data behandles og benyttes mer grunnleggende i sektoren. Behovet for økt deling av data og samordning av metadata og formater er sentralt. Det er også en vesentlig begrensning at modernisering av IT-løsningene går sakte. Politiet sliter med gamle distribuerte løsninger som ikke kan kommunisere ved hjelp av standardiserte grensesnitt og formater på en effektiv måte. Dette er ikke bare et hinder for tilgang til bedre analysegrunnlag, men en digitaliseringsutfordring for hele sektoren. Både domstolene og kriminalomsorgen ville hatt verdi av å kunne berike sine data med data fra politiet, og vice versa.

Selv om Økokrim har benyttet stordatateknologi i mange år allerede finner de det foreløpig vanskelig å selge inn mulighetene som ligger i stordata i sin egen sektor.

[87] (Teknologirådet, 2015)

Det synes for øvrig å være et trekk at de ulike spesialenhetene i politiet som Kripos, Økokrim og PST har egne initiativer innen dataanalyse - uten at det er tatt initiativ til et slikt noe samarbeid. Her kan det være synergier å hente gjennom økt samarbeid om analysemetoder og tilknytning til data, og en annen og bedre organisering av den samlede kompetansen i politietaten.

Case: Analyse av økonomiske transaksjoner i Økokrim
Økonomiske transaksjoner over en viss størrelse skal rapporteres til Økokrim via Altinn (fra finansinstitusjoner, forretninger osv.). Disse dataene kobles med data fra valutaregisteret, enhetsregisteret, folkeregisteret, mva-registeret, kjøretøyregisteret osv. I tillegg kan de kobles til data som inngår i beslaglagt materiale i enkeltsaker. Løsningen har vært i drift siden 2008.
Formål:
Analyser av data om økonomiske transaksjoner for å avdekke mulige straffbare handlinger.
Nytte:
Til dette benytter Økokrim en løsning som finner potensielle saker automatisk, som så blir gjennomgått og ofte etterforsket. Det avdekkes slik også mange saker som politiet per i dag ikke har ressurser til å etterforske.
Læring:
Det er avgjørende med tilgang på god kompetanse. Det må derfor utdannes flere og forskes mer på området. Foreløpig er det også for dyrt for mange å kjøpe slik kompetanse. Økokrim har valgt å rekruttere internt, men det kommer også av at straffeprosessloven gjør at man ikke kan oversende data til kommersielle aktører for analyse.

I følge Økokrim er det mange utfordringer knyttet til bruk av stordatateknologi i politiet. Bl.a. er politiets saksbehandlingsløsninger utdaterte og gir få muligheter til analyser på tvers i en distribuert arkitektur. En annen utfordring som heller ikke er enkel å håndtere er behovet for bruk av data på tvers i justissektoren og offentlig sektor for øvrig. Dette er til dels et kompetansespørsmål, men også et spørsmål om tilgjengelighet (mangel på metadata, ulik semantikk og teknologihindringer i gamle IT-løsninger).

Løsningen til Økokrim er designet for stordata og er strukturert for å håndtere inndata og oppslag i flere kilder. Regelmotoren i løsningen sørger for at de mest relevante sakene genereres. Løsningen er også designet for å kunne ta i bruk maskinlæring – dvs. at mønster og sammenhenger danner grunnlag for mer presise treff i neste omgang. Per i dag mangler det ressurser til videreutvikling i denne retningen. Dette gjør at løsningen ikke er tatt i bruk i sin fulle bredde. Det finnes heller ikke ressurser hos Politiets IT-tjeneste til å bidra til å videreutvikle løsningen.
Kilde: Intervju med Økokrim

Kompetanseproblemet og svak tilgang på støttetjenester fra Politiets IT-tjeneste er nevnt som en utfordring. Dette dreier seg dels om kapasitet, men også om porteføljen av tjenester som IT-enheten skal levere. Erfaringene fra Statens vegvesen viser at behovet for raske og

smidige endringer aktualiseres gjennom bruk av stordata, noe som utfordrer gjeldende paradigme for IT-forvaltning – hvor endringstakten ofte er underlagt et kø-regime, f.eks. ved at endringer prioriteres i et endringsråd.

Det er likevel stort potensiale i sektoren, og det Økokrim har fått til er en god demonstrator i så måte. Datakrimstrategien peker i retning av bedre utnyttelse av stordata som virkemiddel, og indikerer en vilje i ledelsen til å prioritere bruk av stordata på dette området.

Utfordringene når det gjelder sanntidsanalyse av datatrafikk i datakrim-sammenheng og bruk av avanserte analyser for et mer forutseende politi er bare delvis av teknisk karakter. I mange sammenhenger vil disse nye metodene forutsette endringer i måten politiet planlegger og utfører sitt arbeid på, slik også Teknologirådet påpeker i sin rapport. Bl.a. må innsamling og analyse knyttes nært til det operative politiarbeidet, slik at prediksjonene fra analysene kan følges opp av strategiske og taktiske vurderinger av hva som er riktig respons, og hvordan responsen virket.[88]

For å kunne nyttiggjøre seg stordata som teknologi vil justissektoren generelt og politiet spesielt ha behov for en omstilling – til nye arbeidsprosesser, støttet av nye IKT-løsninger og med bedre tilgang til relevante data. En slik omstilling vil være krevende både strategisk og operativt, og den vil være ressurskrevende.

6.3 Helse

Innen helsesektoren finnes det mye data som kan benyttes i analyser – både til planlegging/ressursstyring, medisinsk forskning og i forebyggende behandling. Vi har identifisert en rekke eksempler hvor stordatateknologi er i ferd med å bli tatt i bruk.

I tillegg finner vi også flere eksempler hvor mer tradisjonelle datavarehusløsninger er benyttet. Tilgangen til kvalitetsdata i helsesektoren har aldri vært bedre enn i dag, samtidig innføres det i flere og flere sammenhenger sensorer og digital apparatur som leverer data og gjør det mulig å gjøre sanntidsanalyser.

6.3.1 Helsedirektoratet

Helsedirektoratet har lenge arbeidet med å standardisere sine metadata slik at data skal være lett tilgjengelige i kvalitetsregistre. Mulighetene for å utnytte stordata basert på direktoratets data er derfor gode.

Helsesektoren har flere enkeltstående prosjekter som er av interesse – bl.a. innen medisinsk forskning og innen ressursplanlegging. Eksemplet fra prehospital strategisk satsing er strategisk forankret og gjennomføres i en virksomhet som er «moden» i sitt arbeid med kvalitetssikring, forvaltning og formidling av informasjon. Innen medisinsk forskning benyttes kliniske data i kombinasjon med genom data og data fra helseregistrene.

Dette er et lovende område, for eksempel kan man innen kreftbehandling bli langt mer forebyggende dersom man kan forutsi hvilke grupper og individer som har størst sannsynlighet for å bli syke. Helsesektorens aktører kan da i større grad arbeide forebyggende enn det de gjør i dag.

[88] (Teknologirådet, 2015)

Case: Prehospital strategiske satsing i Helsedirektoratet
Direktoratet benytter stordata for å koordinere ressurser (transport, medisinsk personale og akuttmottakene) for å styre de prehospitale aktivitetene bedre (AMK-sentralene, ambulansetjenesten m.m.).
Det er utviklet et nasjonalt system for innsamling og bearbeidelse av data fra AMK-sentraler og ambulansetjenesten.

Formål:
Å bedre akuttmedisinske prosesser.

Nytte:
Raskere responstid på ambulansetjenesten, bedre ressursplanlegging og beslutningsgrunnlag.

Læring:
Gjennom de ulike datakildene som benyttes i de prehospitale løsningene, håndteres over 250 millioner rader med data fra en rekke kilder. Det har derfor vært nødvendig å satse på stordata-teknologi som skalerer bedre enn mer tradisjonell BI-teknologi.

At dataene ikke er definert som helseopplysninger har vært avgjørende for å kunne berike dem med datasett fra kilder utenfor helsesektoren, noe som hadde vært vanskelig innenfor Helsedirektoratets forvaltningssone. I tillegg er det laget policies og rutiner for at ikke personopplysninger skal kunne tolkes ut av de aggregerte opplysningene som behandles.

Prosjektet har også hatt som oppgave å utvikle en kodestandard på prehospital-feltet. Kontroll med datamodell, kodestandard og infrastruktur og å se disse i sammenheng har gitt viktig læring og på mange måter vært en forutsetning for å lykkes.

Prehospital strategisk satsing i Helsedirektoratet er beskrevet nærmere i vedlegg 1 om utvalgte case.

Stordata gir også nye muligheter innen medisinsk forskning. Innen hjerneforskning utvikles det for eksempel statistiske metoder og andre teknikker som kan bidra til at kostbare prøver og analyser kan gjenbrukes av flere forskergrupper. Hittil har det vært vanlig at de enkelte forskergruppene har samlet inn og analysert egne prøver uten at disse er blitt delt og gjenbrukt av andre grupper.[89]

[89] (Forskningsrådet, 2014)

I desember 2013 konkluderte «G8 Global Dementia Summit» med at bedre bruk av tilgjengelige data, deling av ressurser og forskersamarbeid var nøkkelen til å finne svaret på demensgåten. Ambisjonen var å finne en kur eller passende terapi innen 2025.[90]

Case: Kreftforskning basert på stordata
Oslo universitetssykehus og Kreftregisteret er deltakere i forskningsprosjektet Big Insight i regi av Norsk regnesentral (jamfør vedlegg 2).

Formål:
Utvikle nye og mer presise metoder for kreftforskning basert på kunnskap.

Nytte:
Stordata kan bidra til å bedre kliniske metoder og behandlingsformer gjennom bruk av ulike datakilder. Og dermed økt pasienttilfredshet.

Nye statistiske analyser av data i pasientjournalene i kombinasjon med Kreftregisteret kan gi effekt i denne sammenheng.

Kreftregisteret har samlet inn kvalitetssikrede data om kreft systematisk fra ulike kilder siden 1953. Dataene benyttes til vitenskapelige studier med mål om å forhindre kreft. Mengden av data fra 60 års innsamling er et perfekt datasett for bruk i stordata-sammenheng. For eksempel finnes det over 10 million registreringer av prøver av livmorhalser. Bruk av stordata kan for eksempel endre prosedyren fra at det tas prøve hvert tredje år til personlig testing, basert på den enkelte kvinnes risikoprofil.

En annen gevinst er bruk av data til å tilpasse et oppfølgingsregime til personer som nettopp har avsluttet kreftbehandling.

*Kilde: B*rev fra Oslo universitetssykehus og Kreftregisteret til Norsk Regnesentral om deltakelse i Big Insight og intervju med Norsk Regnesentral.

Med innføring av velferdsteknologi[91] i hjemmene til pleietrengende og på pleiehjem vil mulighetene til å følge opp den enkelte pasient bli bedre. Godt over halvparten av norske kommuner har allerede tatt i bruk slik teknologi og mange planlegger å ta den i bruk.

Helsedirektoratet peker på at en vellykket implementering av velferdsteknologi i stor grad dreier seg om hvordan teknologien innarbeides som verktøy i tjenestenes oppgaveløsning, og at organisasjonen bygges rundt velferdsteknologiske verktøy.[92]

[90] (Ulrike Deetjen, 2015)
[91] NOU 2011:11 gir en oversikt over hva velferdsteknologi er: Trygghets- og sikkerhetsteknologi, kompensasjons- og velværeteknologi, teknologi for sosial kontakt og teknologi for behandling og pleie. (Helse og omsorgs departementet, 2011)
[92] (Helsedirektoratet, 2012)

Bruk av stordatateknologi på data fra velferdsteknologi kan bidra til å gjøre tjenestene mer kunnskapsbaserte. For eksempel kan data fra ulike typer sensorer bidra til at langtidsplanlegging av omsorgskapasitet kan forbedres og at tilbud i større grad kan skreddersys den enkelte. Stordatateknologien kan også brukes til å forske på effektene av velferdsteknologi på ulike grupper av personer med behov for omsorgstjenester. I denne kartleggingen har vi ikke funnet eksempler på muligheten for stordata-analyser basert på data fra velferdsteknologi er utnyttet i norske kommuner.

Case: Hjerne- og hjerteforskning basert på stordata
Ved Universitetet i Stavanger er det etablert en forskergruppe som arbeider med en rekke prosjekter som anvender og utvikler metoder for håndtering, modellering og analyse av biomedisinske data.

De pågående prosjektene innenfor systembiologi og bioinformatikk dreier seg om analyse av gjenopplivingsdata fra akuttmedisin, magnetresonansavbildninger av hjerteinfarkt og av hjernen hos demente.

Formål:
Bedre behandlingsformer og metoder for demens og ulike hjertelidelser.

Nytte:
I løpet av programområdets tre første år har gruppen utviklet automatiske metoder for tolkning av gjenopplivingsdata og identifisering av teksturer som diagnostiske informasjonsbærere i magnetresonansavbildningene fra både hjerte og hjerne.

Kilder:
(Universitetet i Stavanger, 2013)

6.3.2 Folkehelseinstituttet

Folkehelseinstituttets kjerneoppgaver er å generere kunnskap om helsestatus i samfunnet og om beredskap for smittevern. Selv om instituttet ser mulighetene som stordata gir med hensyn til å skape muligheter for vesentlig raskere utnyttelse av data, så har de per i dag ingen konkrete prosjekter for å ta teknologien i bruk. En årsak er de tekniske utfordringene det er å håndtere datalagring og -prosessering på en god måte. I dette ligger det at genom-datasettene er veldig store. Dersom man skal kunne benytte seg mer også av sensordata fra helseundersøkelser –vil det innebære en vesentlig økning i datamengder og lagrings og prosesseringsbehov.

6.3.3 Utfordringer og læring innen helsesektoren

Så langt er vurderingen at Folkehelseinstituttet selv må bygge opp en egnet infrastruktur for stordata-analyser. Grunnen til dette er at de per dags dato ikke kan benytte seg av skytjenester, av hensyn til informasjonssikkerheten. Her må det gjøres ytterligere analyser for å kunne vurdere om og hvordan stordata kan være en del av instituttets strategi på sikt.

Helsedirektoratets konklusjon er at kompetanseutfordringene må løses gjennom å bygge egen kompetanse på området. I Folkehelseinstituttet som har kommet kortere i bruk av stordata, er

det i hovedsak statistikere og noen analytikere som bruker statistikkverktøy for å håndtere dagens datamengder, men instituttet har satset på å styrke kompetansen innen bioinformatikk til det doble det siste året.

Bruk av stordata-teknologi gjør at personvernet og sikkerhetsarkitekturen i sektoren utfordres. Poenget med stordata er å koble store volum av data fra ulike kilder og å gjøre verdiøkende analyser på tvers av datasettene. Det kan være formålstjenlig å berike data fra kilder utenfor helsesektoren slik som for prehospital strategisk satsing. Så lenge det ikke behandles personidentifiserbare helseopplysninger er dette uproblematisk. Problemet oppstår når det er behov for å behandle personopplysninger, eller når helseopplysninger kan tilbakeføres til enkeltpersoner slik dette er definert i helseregisterloven[93].

For eksempel sorteres genom-data i dag i henhold til ulike sikkerhetssoner når de samles inn og lagres. Data avidentifiseres alltid før de frigjøres til forskningsformål. Samtidig kan dataene gi muligheter for forebyggende medisin hos enkeltindivider, så fremt dataene ikke er anonymisert. Denne utfordringen er ikke løst, men det kan være en mulig vei at den som er behandlingsansvarlig kan tillates å kunne reidentifisere personen dersom stordataanalyser gir interessante treff for den enkelte pasienten.

Helsesektoren samler inn og behandler store volum av data. Lagringskapasitet, prosesseringskraft og tid det tar for å utvikle nye rapporter og analyser vil kunne bli en utfordring på sikt. Det er behov for å vurdere alternative måter å lagre og prosessere data på. Bl.a. blir spørsmålet om bruk av skytjenester svært aktuelt.

På mange måter er helsesektoren et eget økosystem, der Norsk helsenett representerer den nødvendige sikre kommunikasjonsinfrastrukturen, og hvor aktørene utveksler data. I tillegg finnes kompetanse tilgjengelig i sektoren. Det er når data skal prosesseres og berikes av data utenfor økosystemet at mange av utfordringene oppstår.

6.4 Arbeid og velferd – NAV

NAV vurderer bruk av stordata for å kunne forebygge og avdekke trygdemisbruk bedre. Men de ser også muligheter for mer målrettede arbeidsmarkedstiltak og mer skreddersydde servicetilbud til innbyggerne, men dette er foreløpig på idéstadiet.

Det er vår vurdering at NAV er i en tidlig fase mht. å ta i bruk stordata-teknologi, men samtidig gjør nyttebetraktningene at potensialet er stort. Bare en liten økning i antall avsløringer av trygdesvindel kan gi en vesentlig samfunnsgevinst. Dersom arbeidet i Big Insight-prosjektet viser seg å gi gode resultater kan det sannsynliggjøres at en videre satsing kan være selvfinansierende.

[93] (Lovdata, 2002)

NAV har erfaring med bruk av datavarehus til analyser, men da først og fremst for å framskaffe styringsinformasjon. Bruk av stordata til kontrollvirksomhet vil medføre endringer i arbeidsprosessene og i måten kontrollvirksomheten drives på. Behovet for bedre tilgang til kompetanse innen datavitenskap blir også aktualisert.

Case: Avdekke trygdemisbruk
NAV deltar i forskningsprogrammet Big Insight i regi av Norsk regnesentral.

Formål:
Utvikle bedre kontrollmetoder for å avdekke trygdesvindel. I tillegg vil nye metoder kunne gi bedre muligheter til også å predikere mulig trygdesvindel og dermed arbeide forebyggende. Med økt mulighet for å kunne bruke informasjonen i de datasettene som etaten besitter vil dette gjøre det mulig å etablere nye måter å arbeide med kontroll på.

Nytte:
Proba samfunnsanalyse har estimert at sosial svindel kan dreie seg om rundt 8 milliarder kroner per år eller om lag 5 % av utbetalingene av foreldrepenger, dagpenger, uførepensjon, arbeidsavklaringspenger og overgangsordninger. I dag er NAV bare i stand til å avsløre en begrenset del av dette. I 2013 ble det avslørt svindel for om lag 230 millioner kroner. Å kunne forebygge og i vesentlig større grad avdekke svindel, er hovedgevinsten med bruk av stordata og drivkraften bak NAVs deltakelse i Big Insight.

Læring:
Mange ulike og til dels gamle IT-systemer gjør at selve datainnsamlingen fra egne løsninger er utfordrende og tar tid. IT forvaltningsmodellen er krevende – hver enkelt endring i algoritmer eller datauttrekk kan være ressurskrevende og må derfor inn i prioriteringskøen i IT-avdelingen. Dersom det besluttes at tjenesten skal over i produksjon i linjen så er det behov for å etablere tjenester som sørger for at endringer kan gjøres raskt og effektivt. En annen læring er knyttet til at folketrygdloven med dagens ordlyd er til hinder for kobling av datasett på tvers av NAV. Her er departementet i gang med å endre loven slik at det er mulig å gjennomføre analyser på tvers av datasett for å forebygge svindel uten at dataene er anonymiserte. Men per i dag gir loven vesentlige begrensinger mht. masseinnhenting og mulighetene for samkjøring av data.

Kilder: Intervju med NAV, Brev fra NAV til Norsk regnesentral om deltakelse i Big Insight, (PROBA samfunnsanalyse, 2013)

6.4.1 Utfordringer og læring i NAV

Selv om de utfordringene NAV har støtt på i arbeidet med å oppdage og forebygge trygdesvindel nok i stor grad kan løses i regi av forskningsprogrammet, peker de allerede nå på sentrale policyområder som må håndteres nasjonalt. Tydeligere føringer for stordataanalyser på personopplysninger og behov for opprydding i lovverket er en del av dette.

NAV understreker at det er behov for klare felles føringer knyttet til bruk av stordatateknologi i Norge. De nevner mellom annet personvern, standardisering av dataformater og enkel tilgang til metadata om andre datasett (både åpne og lukkede).

Et annet viktig forhold er spørsmålet om finansiering. Når nye metoder og teknologier er tilgjengelige er det viktig at det finnes midler og reelle muligheter til å teste ut de ulike mulighetene dette gir.

NAV påpeker videre at det er behov for åpenhet om hvordan informasjon om den enkelte behandles. Innbyggere og næringsdrivende må enkelt kunne få informasjon om hvordan data om dem selv behandles på tvers av forvaltningen. Dette er et rimelig krav og et behov som er forsøkt løst helt siden norge.no ble etablert i 2004 og gjelder digitaliseringen av offentlig sektor generelt. Det viser seg svært utfordrende å samkjøre informasjon om hvilke data som benyttes til hva, og redegjøre for dette på en enkel måte for innbyggerne når de er i kontakt med det offentlige. For å kunne løse dette er det behov for standardisering av metadata, sannsynligvis standardisering av formatene som data utveksles på og ikke minst at metadata er tilgjengelig for hele forvaltningen – slik at informasjon om hva de ulike datasettene brukes til kan presenteres på en enkel måte for brukerne.

Kvalitet på metadata nevnes også som en sentral forutsetning for å kunne analysere datasett på tvers av forvaltningen. Dette omtales ofte som «semantisk interoperabilitet», og dreier seg om at de begrepene som dataelementene representerer er definerte, beskrevet og lett tilgjengelig for den som vil gjenbruke dataene. Dette gjelder uavhengig av om datasettene er åpent tilgjengelige (åpne data) eller om de er lukket og krever hjemmel for å kunne gjenbrukes. NAV har arbeidet mye med dette de siste årene og har kommet et langt stykke på vei – ikke minst gjennom etablering av A-ordningen[94].

6.5 Skatt, toll og avgift

Både Toll- og avgiftsetaten og Skatteetaten har initiativer for å benytte stordata. Toll- og avgiftsetaten har planer om økt bruk av stordata for å avdekke mønster i smugling m.m. Skatteetaten vurderer bruk av stordata-teknologi for å beregne risiko for feil i skatte- og avgiftsberegningene og risiko for skatteunndragelser.

Skatteetaten har benyttet datavarehus til å gjøre analyser på egne data de siste 10 årene. Formålet har vært å trekke ut aktuelle tilfeller hvor det bør gjennomføres kontroll i form av for eksempel bokettersyn. Skatteetaten har startet med å se på stordata-teknologi da det har vist seg at den tradisjonelle datavarehus-arkitekturen er for lite fleksibel. Det tar for lang til å få gjort endringer. Dette har blitt aktualisert i forbindelse med at Skatteetaten tok over forvaltningen av Toll- og avgiftsdirektoratets system TVIST og behovet for å kunne gjennomføre sanntidsanalyser basert på både Skatteetatens og Toll- og avgiftsetatens datakilder. Det er også behov for å kjøre analyser på andre offentlige datakilder.

Skatteetaten og Toll- og avgiftsetaten har begge forutsetninger for å lykkes med å få god effekt av stordata. De har nå et stort og slagkraftig felles IT-miljø, hvor det også finnes god kompetanse på dataarkitektur og informasjonsforvaltning. De har også lang praktisk erfaring med datavarehus og avanserte analyser.

[94] (Altinn, A-ordningen, 2015)

6.5.1 Utfordringer og læring i skatt, toll og avgift

Økte datamengder og behov for fleksibilitet i både tilknytning av datakilder og analysemetoder gjør at det fremtvinger seg bruk av andre teknologier enn tradisjonelle datavarehus-løsninger. Både data- og textmining er aktuelt og vil sannsynligvis medføre at det blir behov for ny kompetanse.

Det er også en kjensgjerning at datavarehusarkitekturen ikke skalerer, raskt økende datavolumer må kunne kvernes raskt igjennom, og nye mønster må kunne avdekkes effektivt. Det tar for lang tid å laste og prosessere store datamengder (som f.eks. ligningsdata) i dagens løsninger. I tillegg er det et etterslep på videreutvikling av modellene i datavarehuset. Stordata kan i større grad gi en skalerbar arkitektur som kan håndtere store volum.

Case: Avdekke skatteunndragelser

Skattedirektoratet er del av Big Insight prosjektet som gjennomføres av Norsk regnesentral med fire områder hvor de ønsker å utvikle prediktive modeller:

- *Prediktive modeller for å drive inn underbetalt skatt*
- *Velge ut næringsdrivende og privatpersoner for kartlegging*
- *Overvåke registrering av næringsvirksomhet og andre økonomiske aktiviteter for å avdekke konkurser*
- *Predikere næringsdrivende og privatpersoner som ikke leverer skatteoppgjør i tide og finne mulige metoder for å påvirke adferd*

Formål:

Planen er å etablere en prosess som identifiserer risikoer for skatteunndragelser og svart økonomi. Ved å benytte informasjon direkte fra regnskapssystemene til næringsdrivende kan man eksempelvis gjøre sanntidsanalyser på ulike risikofaktorer og dermed avdekke risiko tidlig. Eksempelvis har Portugal gått over til månedlig rapportering av regnskapsdata. Bruk av regnskapsdata og data fra A-ordningen (EDAG) vil gi et godt datagrunnlag.

Nytte:

Samfunnsgevinster i å avdekke flere skatteunndragelser. Besparelser for etaten ved å bruke mindre ressurser på skjønnsmessige vurderinger. Kan benytte mer risikobaserte strategier og sette ressursene inn på de tilfellene som har høyest risiko.

Læring:

Stordata gir flere muligheter for å gjøre ulike typer analyser. Det er mer effektivt å analysere sammenhenger framfor å teste forhåndsdefinerte problemstillinger. Skatt ønsker å komme over i bruk av mer utforskende analyse framfor tradisjonell datavarehusanalyse. Datavarehusene skalerer heller ikke godt nok. Det er en rekke etiske dilemmaer knyttet til mulighetene for behandling av personopplysninger.

Kilder: Intervju med Skattedirektoratet; Brev fra Skattedirektoratet til Norsk Regnesentral av 20.1.2014 som bekrefter intensjon og vilje til deltagelse i Big Insight. (Wikipedia, Text Mining, 2015)

Bruk av stordata-teknologi vil imidlertid medføre behov for smidige IT-prosesser og fleksibilitet når det er behov for endringer, noe som kan være en utfordring for en strømlinjeformet IT-avdeling. Sanntidsanalyser gjennom bruk av stordatateknologi innebærer ofte behov for raske endringer i løsningene.

Gjennom både MAG[95]-prosjektet og innføring av A-ordningen[96] har Skatteetaten fått god erfaring med samordning av metadata både internt og med NAV og SSB. Dette gjør at data som benyttes som grunnlag i Big Insight vil være av god kvalitet med avklart kontekst, noe som øker mulighetene for å lykkes med analysene.

Økt risikobasert analyse implementert i IT-systemene kan sannsynligvis gi bedre effektivitet ved at man kan respondere raskere på saker med høy risiko. En utvikling i retning av et mer og mer internasjonalt næringsliv krever at Skatteetaten må kunne handle raskt. Dagens metoder med etterkontroll tar for lang tid.

Mindre bruk av skjønn og økt bruk av maskinell analyse av data kan gi høyere grad av likebehandling av like tilfeller. I tillegg kan det gjøre at Skatteetaten kan utvikle bedre tjenester mot publikum.

Bruk av teknologi og data om virksomheter og personer til å avdekke skatte- og avgiftsunndragelser, og svindel generelt for den saks skyld, medfører naturligvis noen etiske dilemmaer. Spørsmålet om hvor langt det offentlige skal gå i å samle inn og analysere data om innbyggernes privatliv står her sentralt.

Det er i dag fullt mulig rent teknisk å hente data fra sosiale media og slik kartlegge enkeltpersoners disposisjoner. Dersom Skatteetaten og Toll- og avgiftsetaten i tillegg kombinerer disse dataene med egne data som de deler seg imellom – og også får tilgang til data fra NAV og Økokrim, vil det sannsynligvis gi mye større muligheter for å finne svindlere. Dersom de i tillegg samarbeider med banker og forsikringsselskap øker mulighetene ytterligere. Disse virksomhetene vil kunne ha et fullstendig bilde over personers disposisjoner i sanntid, noe som gjør at spørsmålet om personvern blir svært aktuelt.

6.6 Klima, miljø og beredskap

Offentlige myndigheter og institusjoner sitter med en stor mengde data om vær, vannforhold, fjell, jorderosjon m.m. som sammenstilt og bearbeidet på en god måte kan predikere naturkatastrofer. For varsling av flom og skred har NVE, i nært samarbeid med Meteorologisk institutt, Statens vegvesen og Jernbaneverket, i operativ bruk systemer som sammenstiller observasjoner og prognoser fra MET og NVE, modellerte griddete datasett for mer enn hundre tema som ajourføres daglig, satellittdata, vegmeldinger, webkamera, farekart, etc – og har gjort disse åpent tilgjengelig gjennom www.xgeo.no. Ytterligere sammenstilling og videreutvikling av lignende verktøy kan gi muligheter for å håndtere naturkatastrofer på en bedre måte, og ikke minst i større grad gi grunnlag for å forutse slike katastrofer mer presist enn i dag.[97]

[95] (Skatteetaten, MAG, 2013)
[96] (Skatteetaten, A-Ordningen, 2015)
[97] Det EU-finansierte forskningsprosjektet IQMULUS ser også på muligheter for å utnytte stordata i denne sammenheng, jamfør http://www.iqmulus.eu

Selv om dataene finnes i virksomhetenes systemer kan det være et problem at dataene ikke alltid er tilgjengelig for sammenstilling og bruk når katastrofen først er der. De offentlige virksomhetene på dette området samler inn og behandler til dels store mengder data for sine formål og alle ser potensialet i å hente inn og analysere sanntidsdata i enda større grad. Det er likevel bare NVE som har kommet så langt at de vurderer bruk av stordatateknologi i forbindelse med flomovervåking. Her kan det hentes inspirasjon fra Nederland hvor vannmyndighetene (se vedlegg 3) har tatt i bruk stordata for å forutse og å håndtere oversvømmelser.

6.6.1 Norges vassdrags- og energidirektorat (NVE)

NVE har et nettverk av målestasjoner som samler inn anslagsvis 50-60 millioner enkeltmålinger pr. år. Dette håndterer NVE i tradisjonelle databaser og videreforedler til flom- og skredvarsling m.m. De klarer å prosessere dataene raskt nok i dag til å kunne gi flom- og skredvarsling tidsnok.

Case: Sanntidsanalyse av datastrømmer i NVE

NVE har startet et prosjekt hvor det settes det opp et regelsett for automatisk sjekk av unormale verdier i datastrømmene. Sanntidsanalyser gjøres maskinelt basert på at alle data fra NVE sine målestasjoner sjekkes i en regelmotor. Dette dreier seg om å ta tak i den ganske massive datastrømmen og prosesserer løpende i stedet for å analysere i ettertid.

Formål:
Raskere respons på unormale verdier i datastrømmer om høy vannføring og skredfare.

Nytte:
Bedre og raskere informasjon til publikum og eventuelt generere varsler til rette personer og systemer kan hindre tap av liv og helse og materielle verdier.

Kilde: Informasjon på e-post fra NVE

6.6.2 Direktoratet for samfunnssikkerhet og beredskap (DSB)

DSB har en rekke egne datakilder om blant annet transport av farlig gods, infrastruktur og om sårbare objekter som fredet natur o.l. I tillegg kobles disse dataene til NVEs data om skred- og flomfare, data om fredete kulturminner, sårbare bygninger og befolkningsdata fra SSB. I kartinnsynsløsningen til DSB vises disse dataene i en kartflate som kan benyttes til ulike formål. I tillegg er det utviklet en påloggingsløsning for brannvesen og brannalarmsentraler med informasjon om lager med farlige stoffer. Denne løsningen planlegges utbygd med mer data.

6.6.3 Miljødirektoratet

Statistisk sentralbyrå (SSB) og Miljødirektoratet beregner årlig totale nasjonale utslipp til luft av 20 utslippskomponenter. Dette datasettet publiseres halvannet år etter at utslippene faktisk

fant sted, mens utslippene som vises i sektorene landbasert industri og petroleumsindustri til havs publiseres et halvt år etter at de fant sted[98].

I regnskapet inngår blant annet utslippstall fra bedriftenes årsrapporter til myndighetene. Det er da utslippene fra de største punktkildene som benyttes i regnskapet. For utslippskilder hvor det ikke er rapporteringsplikt beregnes utslippene på bakgrunn av aktivitetsdata og utslippsfaktorer. For andre utslippskomponenter, og for utslipp til vann generelt, er det ikke gjort en slik beregning av totale utslipp på tvers av sektorer.

Miljødirektoratet har også en rekke datakilder med informasjon om miljøet i Norge. Disse dataene er i all hovedsak stedbundne og gir derfor størst mening når de visualiseres i kart. Data baseres på registreringer fra for eksempel fylkesmenn, kommunene og Statens naturoppsyn. Informasjonen har stor interesse for både næringsdrivende, media og publikum. Noen sentrale datasett er Artsdatabanken, Naturbase, Rovdyrdatabasen og utslippsdata. Dataene er i stor grad tilgjengeliggjort og visualisert, slik som for eksempel Naturbase[99].

6.6.4 Utfordringer og læring i klima, miljø og beredskap

Virksomhetene på dette området har tradisjon for å samle inn og håndtere store datamengder. Alle har nå også mulighet til å gjøre seg nytte av data fra det europeiske satellittprosjektet Copernicus[100] som etter hvert kan levere satellittdata som kan avdekke miljøutslipp som for eksempel oljesøl, overvåke snøsmelting og flom, CO_2-utslipp og annen luftforurensning, miljøkatastrofer m.m. Dataene vil være fritt tilgjengelige og kan utnyttes av alle disse offentlige virksomhetene. Dette gir nye muligheter for virksomhetene, men vil også medføre behov for nye og endrede arbeidsprosesser og analysemetoder. Dette representerer i sin tur også kompetanseutfordringer.

Både NVE og Miljødirektoratet har registrert åpne datasett på data.norge.no og har dermed lagt til rette for at andre kan gjøre analyser på deres datasett. Her er imidlertid potensialet åpenbart større, begge har flere datasett som kunne vært delt for viderebruk og dermed brukt til å berike andre datasett i andre virksomheter og sektorer.

Det er vår vurdering at det er store muligheter for å utnytte data bedre på dette området, både gjennom at data i større utstrekning deles og analyseres på tvers av virksomhetene, men også ved at data i økt grad stilles til rådighet for viderebruk av private aktører. Det er også stort potensiale for å utvikle prediktive analyser av både flom, forurensning og naturkatastrofer basert på sanntidsdata, noe som aktualiseres gjennom at sanntidsdata for mange relevante miljøparametere nå gjøres tilgjengelig av Copernicus-prosjektet.

6.7 Smarte bygg og byer

Byggsektoren er Norges største landbaserte «sektor» økonomisk sett – med over 600 milliarder i årlig omsetning. Her finner vi også en rekke spennende initiativer, også i en digital kontekst. Alle byggeprosesser er dataintensive og alle nye bygg er konstruert i digitale modeller framstilt i digitale verktøy og basert på BIM (bygningsinformasjonsmodellering). I tillegg har mange moderne bygg elektroniske drifts- og vedlikeholdssystemer som kan styre alt fra ventilasjon og varme til planlagt vedlikehold.

---

[98] http://www.norskeutslipp.no
[99] http://kart.naturbase.no/
[100] (Copernicus, 2015)

Både Direktoratet for byggkvalitet (DiBK) og Statsbygg utforsker nå mulighetene som stordata kan gi i denne sammenheng. Begge etatenes vurderinger så lang er at teknologien virker lovende. For disse er satsingen på stordata en naturlig konsekvens av øvrig digitalisering i sektoren. Flere og flere digitale løsninger og tjenester gir også flere og flere data og datakilder som skal behandles – og som igjen kan utnyttes til nye tjenester.

«Smarte byer» (Smart Cities) har vært et begrep lenge. Smarte byer som konsept innebærer bl.a. energieffektivisering og optimalisert ressursbruk generelt i byer. Konseptet berører bygninger, infrastruktur, byplanlegging, mobilitet, helse, livskvalitet og mye mer. Det foregår en rask utvikling i feltet internasjonalt og i Norge –både i regi av lokale myndigheter og innen industri, forskning og utdanning.

Et eksempel er TRIANGULUM-prosjektet hvor Stavanger samarbeider med Manchester og Eindhoven. De tre byene gjennomfører pilotprosjekter med løsninger som integrerer energi, mobilitet og IKT på nye måter. Smarte byer som konsept henger nøye sammen med stordata, siden det å kunne styre energiutnyttelse og mobilitet avhenger av at data er tilgjengelig i sanntid og kan fremstilles slik at mønstre kan gjenkjennes og utnyttes for styringsformål. I denne sammenheng forsker Universitetet i Stavanger på hvordan stordata kan bidra til energieffektivitet i byene[101].

6.7.1 Statsbygg

Statsbygg har tilrettelagt for datafangst fra en rekke systemer og sensorer i bygningene ved å etablere et datanettverk som binder sammen om lag 300 eiendommer. Strategien har vært å legge til rette infrastrukturen før en starter med storskala datafangst.

6.7.2 Utfordringer og læring i smarte bygg og byer

Både Statsbygg og DiBK har erfart at det er vanskelig å få tilgang til datasett som andre virksomheter forvalter. Selv om offentleglova og digitaliseringsrundskrivet pålegger offentlige virksomheter å tilgjengeliggjøre data som ikke er taushetsbelagt, sitter dette likevel langt inne.

Erfaringene som de to etatene har gjort seg når det gjelder å få tilgang til data, ser vi som et tegn på en sektor med manglende tradisjon for å dele data. En medvirkende årsak er nok at sektoren er umoden i bruk av digitale verktøy og dermed i liten grad ser nytten utveksling av data kan ha. Men dette er i ferd med å snu, en digitaliseringstrend og høy grad av modernisering i det siste medfører mer tilgjengelige data. Men har også den konsekvens at tidsserier kan være vanskelig å sammenligne og at løsninger som benyttes må være fleksible og kunne utnytte stadig nye datakilder.[102]

[101] (Stavangerregionen, 2014)
[102] (KPMG, 2015)

Det er en sentral utfordring i denne sektoren at datasettene ofte er krevende å ta i bruk fordi de mangler gode metadata. Behovet for en felles ontologi på området er stort. Sintef har tidligere tatt initiativ på dette området, men i skrivende stund er det ingen som har noe pågående koordineringsinitiativ på området. DiBK ser derfor dette som en viktig oppgave fremover – ikke bare for å få økt nytte av stordata, men for å få fart på digitaliseringen generelt og for å kunne utveksle og analysere data på tvers av sektoren.

Case: Drift og vedlikehold basert på stordata i Statsbygg
Statsbygg har samlet inn og analysert data fra egne bygg for å kunne styre energiforbruket bedre og etter hvert også ha oversikt over bygningenes tilstand og planlegge vedlikehold bedre.
Hovedkilden av data er eiendommenes forvaltning, drift og vedlikeholdssystemer (FDV), men også andre typer systemer kan avgi data som for eksempel løsninger for energimåling, temperaturmålere og sensorer som måler luftkvalitet.
I tillegg deltar Statsbygg i et internasjonalt pilotprosjekt i regi av ProDataMarket-prosjektet som er finansiert over EUs Horizon 2020-satsing. Prosjektkonsortiet består av 8 partnere, med Statsbygg og SINTEF fra Norge. Løsningene er foreløpig på pilotstadiet. I piloten hentes data fra et begrenset antall bygninger, men tanken er at alle bygningene som staten eier skal kunne levere tilstandsdata digitalt.
Formål:
Det overordnede målet er å finne muligheter for innovasjon basert på eiendomsdata på tvers av relevante aktører.
Nytte:
Energi kan spares ved å treffe tiltak på avvik i energiforbruket og måle effektene av energiøkonomiseringstiltak.
Bedre vedlikeholdsplaner for de byggene som leverer tilstandsdata.
Statsbygg har allerede hatt gevinst på de byggene som har vært med i piloten.
Læring:
Tidkrevende og ofte vanskelig å få data fra andre virksomheter (også offentlige), selv om dataene i utgangspunktet er åpne.
Dårlige metadata gjør det krevende å ta i bruk andres data.
Det er viktig med standardiserte grensesnitt mot bygningenes driftssystemer.
Stordata i Statsbygg beskrives nærmere i vedlegg 1.
Kilder: Intervju med Statsbygg, Innlegg på Sintef seminar om stordata

Nedslagsfeltet for stordata-analyser på bygg- og anleggsområdet er omfattet av hele 15 ulike statlige og kommunale sektormyndigheter som har et eller annet ansvar for tilsyn med ulike aspekter ved bygninger. Digital samhandling og gjenbruk av data fordrer en semantisk interoperabilitet mellom virksomhetene og datasystemene. Det å harmonisere begrepsbruk og å utvikle en ontologi for sektoren er derfor et naturlig sted å starte. Vår vurdering er at det her er et åpenbart grunnlag for samarbeid mellom aktørene i denne sektoren.

Mangelen på metadata med begrepsbeskrivelser og de øvrige inntrykk av en «lukket» sektor når det gjelder å dele og å utveksle data tyder mer generelt på svak informasjonsforvaltning i sektoren. En god forvaltning av en virksomhets informasjonsressurser er preget av at

datasettene (både strukturerte og ustrukturerte) er godt beskrevet og at disse beskrivelsene (metadataene) er publisert og tilgjengelige for andre. God informasjonsforvaltning vil også dreie seg om at dataenes kvalitet er kjent og at sikkerhetshensynene er vurdert. Vår vurdering er at tiltak for å bedre informasjonsforvaltningen i sektoren vil kunne bidra til å muliggjøre bruken av stordata på dette spennende området.

Case: Forprosjekt bedre beslutningsgrunnlag i DiBK

Forprosjekt som vurderer bruk av stordata for å analysere data om byggkvalitet i norsk bygningsmasse som kunnskapsgrunnlag for regulatoriske tiltak.

DiBK har gjennomført et forprosjekt/prøveprosjekt for å få inn mer data om bygningsskader, fra ulike kilder – ved hjelp av stordata-tilnærming. DiBK ser om det er mulig å hente inn loggdata fra forsikringsselskap, rapporter fra takstmenn, rapporter fra ettårsbefaringer, overtakingsdata fra klagesaker, data fra forvaltning/drift/vedlikehold systemer, eiendomsmeglere, tilsynsmyndigheter, Byggforsk/Sintef sitt skadearkiv m.fl.

Formål:

Avklare om metodikk som faller inn under begrepet stordata kan benyttes til å kartlegge byggkvalitet og omfanget av byggefeil. DiBKs overordnete formål er å skaffe bedre grunnlag for regelverksutforming.

Nytte:

Finne ut om det finnes datagrunnlag slik at DiBK kan måle om vedtatte reguleringer har forventet effekt på byggekvaliteten. Få nødvendig erfaringer som kan brukes til å planlegge et eventuelt hovedprosjekt der denne metodikken legges til grunn.

Læring:

Generelt har DiBK erfart at datagrunnlaget fra de ulike kildene ofte har dårlig struktur og kvalitet.

Datakildene er alt for ofte lukket. Det har vært vanskelig å få data fra en del organisasjoner – særlig i privat sektor, men også fra offentlige virksomheter. Det er tid- og kompetansekrevende å skaffe tilstrekkelige data. Når dataene først er tilgjengeliggjort kan det også være vanskelig å benytte dem på grunn av ulik begrepsbruk og at datasett innsamlet for et annet formål ofte mangler viktige dataelementer som er avgjørende for bruk i nye kontekster. Det mangler også en gjennomgående nøkkel som gjør det mulig å koble datasett, fortrinnsvis gjennom matrikkelen. Det å identifisere bygningstypen kan være komplisert.

Løsningen for DiBK er minste motstands vei, dvs. å benytte datasett som kan benyttes uten større problemer, ved at det er enkelt å finne dem og ved at de kommuniserer greit. En sentral del av arbeidet videre vil være å rydde i begrepsbruken basert på tidligere arbeid utført av blant andre Byggforsk. Det blir viktig å ta tak i de begrepsdefinisjonene som finnes og utvikle omforente ontologier for sektoren.

Kilder: Intervju med DiBK og (KPMG, 2015)

6.8 Kartdata

Svært mange sanntidsdata fra de eksemplene som vi har sett på kan knyttes til et geografisk punkt eller areal, dvs. de er stedsrelaterte. Dette gjør at verdien av dataene øker betraktelig når de kan berikes med kartdata. Dette gjelder både for analyser av data knyttet direkte til forhold ved geografien og for visualisering av andre typer data. Statens kartverk har som hovedmål å gjøre offentlige kart- og geografiske data lett tilgjengelig. En stedsrelatert, digital virkelighet er et viktig bidrag både til effektivitet i offentlig sektor, verdiskapning i næringslivet, til bedre beredskap og et tryggere samfunn. I denne sammenheng er etableringen og forvaltningen av en ny nasjonal høydemodell basert på laserdata og fotomatching av stor betydning – ikke minst som grunnlag for analyser knyttet til naturfarer, hydrologi, geologi og terrengforhold.

Derfor har Kartverket også prioritert etableringen av en sentral, nasjonal node for tilgang til stedsrelaterte data som offentlig sektor forvalter. Målet er at noden ikke bare skal ha tilgjengelig kartdata, men alle typer egenskaper som er relevante for å kunne gjengi virkeligheten. Tilfanget av data vil være stort og omfatte eiendomsdata, marine og maritime data, topografiske data, geologiske data, hydrologiske data, ulike typer miljødata og en rekke andre tematiske data. I Noden vil det være funksjonalitet for å få tilgang til geografisk stedfestet informasjon på en rekke ulike måter: Som metadata om stedsrelaterte datasett inkludert ontologier, katalog over ulike datasett, APIer og tjenestegrensnitt som gjør datasettene tilgjengelige for offentlig sektor, næringslivet, akademia og publikum generelt.

Begrepet «Noden» illustrer at denne infrastrukturen skal kunne kobles med datasett og datakilder på andre områder og i andre land. Noden er en del av Kartverkets og Norge digitalt-partenes felles strategiske satsing. Planen er å basere utviklingen av Noden på Geonorge-løsningene. Fram mot 2018 planlegger Kartverket å videreutvikle løsningen med mer avanserte måter å tilgjengeliggjøre data på.[103]

6.8.1 Utfordringer og læring på kartdataområdet

Som flere av informantene har pekt på er tilgangen til kartdata sentralt i mange av de datadrevne «økosystemene». De fleste som benytter slike data i dag benytter Geonorge-portalen[104] og mange benytter programmeringsgrensesnitt (API) til Matrikkelen, eller henter data fra Hoydedata.no for detaljerte høydedata. Behovet for geografisk stedfestet informasjon er i mange sammenhenger stort og helt sentralt for å få merverdi av andre datasett. På en rekke områder er det behov for nye eller langt bedre geografiske data enn i dag for å utløse disse gevinstene. For eksempel mangler det gode dybdedata for innsjøer og elver.

Det er vår vurdering at Kartverkets satsing på å utvikle en nasjonal node for geografisk informasjon vil kunne bidra til at mange kan få enda mer verdi ut av sine datasett, i tillegg til å forenkle tilknytningen til slike data – forutsatt at vilkårene for bruk av data gjennom Noden vil tilrettelegge for slik bruk.

[103] Beskrivelsen av Noden er basert på skriftlig innspill fra Statens kartverk
[104] http://kartverket.no/Geonorge

6.9 Offisiell statistikk

Statistisk sentralbyrå (SSB) har ingen egne prosjekter eller initiativ som ser på bruken av stordata i arbeidet med offisiell statistikk i Norge, og har heller ingen konkrete planer. Byrået følger imidlertid med på internasjonalt arbeid på området, blant annet i regi av FN gjennom The United Nations Economic Commission for Europe (UNECE).[105]

Offisiell statistikk må være basert på data som er kvalitetssikret på basis av definerte objektive kvalitetskriterier og beste praksis, siden slik statistikk i mange sammenhenger benyttes som grunnlag for planlegging og beslutninger (jamfør for eksempel FNs prinsipper for offisiell statistikk[106]). All innsamlet statistikk gjennomgås derfor grundig med tanke på kvalitet før de publiseres.

For eksempel er svært mye statistikk basert på forhåndsdefinerte populasjoner (f.eks. personer i aldersgruppen 16-74 år bosatt i Norge) over tid og måler ulike egenskaper ved disse gjennom ulike variabler (arbeidssted, inntekt osv.). En utfordring i vår kontekst er at mye av datatilfanget en får fra stordata ikke vil gi samme muligheter til å kontrollere populasjonen. Det kan også være et problem at en i store datamengder får veldig mange variabler (og målinger), noe som gjør det vanskelig å finne variablene som er viktigst for å forklare det fenomenet en ønsker å undersøke.

UN Economic and Social Council[107] peker på at dataenes (tidsmessige) aktualitet blir stadig viktigere. Årsaken til at stordata ofte er mer aktuelle er at data hentes inn i sanntid uten forsinkende kvalitetssikringsprosesser. Samtidig fastslår de at stordata neppe kan oppnå samme kvalitet som data til statistikkformål som på tradisjonelt vis er gjennomgått og kvalitetssikret før publisering. Dermed blir det en reell avveining mellom nøyaktighet og aktualitet.

6.9.1 Utfordringer og læring i utarbeidelse av offisiell statistikk

I følge FN[108] har flere land og Eurostat startet arbeidet med å ta i bruk stordata-teknikker som supplement til mer tradisjonelle statistiske metoder når de utarbeider offisiell statistikk.

FN peker på at det sannsynligvis også må utvikles nye metoder for analyse av mange heterogene datakilder for eksempel gjennom «data mining» i kombinasjon med maskinlærings-algoritmer. Stordata-teknologi i kombinasjon med algoritmer som prosesserer regler og teknologier for maskinlæring er en reell mulighet også i sammenhenger hvor det produseres offisiell statistikk. Det synes klart at tradisjonelle statistiske metoder ikke skalerer nok med hensyn til å kunne gjøre sanntidsanalyser på store datamengder.

I sin studie av mulighetsrommet for bruk av stordata til offisiell statistikk finner OECD at en rekke kilder på Internett kan supplere mikrodata som er samlet inn for statistikkformål og berike disse. Slik får en nye muligheter for predikere i sanntid, ut fra tidsaktuelle data. Områder som nevnes som særlig interessante er sysselsettingsstatistikk, statistikk om bruk av IT og telefoni og prisindekser. Som et eksempel publiseres 95 % av alle ledige stillinger på Internett. En analyse av søk på ledige stillinger via søkemotorer og stillingsannonser på

[105] (UNECE Statistics, 2015)
[106] (SSB, u.d.)
[107] (UNECE Statistics, 2015)
[108] (UNECE Statistics, 2015)

sentrale kilder som Finn, Linkedin osv. kan supplere statistikken som utarbeides av SSB og NAV og slik bidra til bedre prediksjoner.[109]

6.10 Forskning

Som vi har sett så benyttes stordata som teknikk innen forskning på helserelaterte problemstillinger. Men i tillegg har flere norske forskningsmiljøer programmer på området hvor det forskes på ulike anvendelser av stordata i ulike sammenhenger. Ved Universitetet i Oslo, Universitetet i Bergen, NTNU og SINTEF forskes det både på modeller og algoritmer som kan gi større grad av nytte av tilgjengelige data, på modeller for å nedskalere datasett samtidig som verdi beholdes, teknologier for å sammenstille datasett, muligheter for anonymisering og reidentifisering av personer, visualisering m.m. Ved Oslo universitetssykehus benyttes stordata mellom annet i forbindelse med forskning på arvelige sykdommer.

SINTEF har også hatt en sentral rolle i det EU-finansierte IQMULUS-programmet[110] som har som mål å utvikle metoder og teknikker for å utnytte informasjon som ligger skjult i store heterogene geografisk datasett og gjøre dem praktisk tilgjengelig som pålitelig støtte i beslutningsstøtteprosesser for eksempel knyttet til kriser.

Universitetet i Oslo er prosjektkoordinator for «Optique»[111] programmet som skal ta fram en portefølje av verktøy for å kunne utnytte stordata bedre. Metoder for å benytte skytjenester, sluttbrukerverktøy for å kjøre spørringer mot datasett osv.

Norsk regnesentral har etablert programmet «Big Insight»[112] i samarbeid med Universitetet i Oslo og Universitetet i Bergen. Programmet skal utvikle nye statistiske metoder og modeller for bedre å kunne predikere fenomener som er ustabile, forstå uforklarte mekanismer og sammenhenger, avdekke skjulte mønstre, m.m. Målet er at modellene som utvikles skal være generiske basert på nye metoder innen statistikk og maskinlæring som håndterer store datamengder.

NTNU har en rekke forskningsprosjekter innen stordata på tvers av fagområdene datateknikk og statistikk/matematikk. De konsentrerer forskningen på to områder:[113]

- Industri-relevant innovasjon («web-mining» for å lage trendanalyser, anbefale produkter og tjenester til kunder)
- Bedre helsevesen (pasientjournal informasjon/bioinformatikk, genom forskning på aldersrelaterte sykdommer m.m.)

NTNU forsker også på såkalte åpne og autonome digitale økosystem (OADE[114]), miljøovervåking av aktiviteter i petroleumsindustrien (VERDIKT[115][116]) og personalisert medieinnhold. De forsker også på muligheter for å identifisere enkeltpersoner ved å kombinere åpne offentlige datasett.

---

[109] (Reimsbach-Kounatze, 2015)
[110] (IQmulus, 2015)
[111] (Optique, 2015)
[112] http://biginsight.no/biginsight/drupal/
113 http://www.ntnu.edu/ime/bigdata
114 http://www.ntnu.edu/ime/oade
115 http://research.idi.ntnu.no/doil/doku.php
116 http://lovedev.azurewebsites.net/Documentation

I tillegg er forskere fra NTNU sentrale i en rekke internasjonale forskningsprogrammer som for eksempel EU-prosjektene AMIDST («Analysis of MassIve Data Streams»)[117], PRESIOUS («predictive geometric augmentation technologies for cultural heritage»)[118] og Climate KIC CTT («Carbon Track and Trace»).

Institutt for Informatikk ved Universitet i Oslo deltar i EU prosjektet BYTE som skal bistå europeisk forskning og industri med å få fram de positive sidene ved stordata og å minimere de negative sidene slik at de kan få en større del av stordata markedet innen 2020. Prosjektet er basert på casestudier hvor man tar utgangspunkt i, og undersøker ulike eksterne forhold og rammevilkår slik som økonomi, lovverk, sosiale forhold, etikk og politikk.

Prosjektene Optique, Big Insight og BYTE er beskrevet nærmere i vedlegg 2.

Universitetet i Stavanger forsker på bruk av stordata og energieffektivisering i regi av Triangilum prosjektet (finansiert av Horizon 2020). Ved dette universitetet forskes det også på systembiologi og bioinformatikk. Senter for IP-basert serviceinnovasjon CIPSI er etablert ved UiS og tar mål av seg til å innta en ledende posisjon innen behandling av stordata.[119]

Ved universitetet i Tromsø forskes det må maskinlæring, mellom annet for å håndtere manglende data i store datasett.[120]

6.11 Oppsummering av status for bruk av stordata i offentlig sektor i Norge

Flere forskningsmiljøer gjør verdifull forskning på stordata og flere bruker også stordata i sin forskning. Norske forskningsmiljøer har nøkkelroller i flere internasjonale program, og kunnskapsgrunnlaget og kompetansen er derfor i ferd med å bygges opp i Norge.

Vårt hovedinntrykk er likevel at de fleste offentlige virksomheter har kommet kort i utforskning og bruk av stordata, men at flere virksomheter nå utforsker mulighetene gjennom pilotprosjekt og forsøk. Målet for flere virksomheter er å utforske mulighetsrommet som ny teknologi og nye analysemetoder gir, slik at de i neste omgang kan sikre en mer strategisk forankring i organisasjonen. Dette gjelder for eksempel Statsbygg, Skatteetaten, Nav m.fl.

6.11.1 Strategisk forankring

De som har kommet lengst i å ta i bruk stordata, slik som Statens vegvesen og Helsedirektoratet, kjennetegnes ved sine strategiske satsinger på området. Prosjektene er godt forankret i virksomhetsstrategier og rammevilkårene for prosjektene er derfor avklart. Selv om de også er i en utviklingsfase har de kommet langt i å tilgjengeliggjøre sanntidsdata for analyser i ordinær drift og forvaltning. De har allerede kunnet ta ut gevinster for eksempel i form av bedre informasjon til beslutningstakere og publikum.

6.11.2 Kunnskap om tilgjengelig data

Flere av de virksomhetene vi har vært i kontakt med oppgir at det er utfordrende å sette seg inn i og å forstå både interne og eksterne datasett. Ofte er metadataene om datasettene begrensende i og med at begrepene som dataene representerer ikke er definert, kodeverk er

[117] http://amidst.eu/
[118] **http://www.presious.eu/**
[119] (Universitetet i Stavanger, 2014)
[120] (Universitetet i Tromsø, 2014)

ikke forklart eller at konteksten i form av regelverk og populasjon ikke er beskrevet. De virksomhetene som har kommet lengst har lang erfaring med forvaltning av data og metadata, og flere har også erfaring med bruk av tradisjonelle datavarehus og analyser.

6.11.3 Avventende holdning

For hoveddelen av offentlige virksomheter virker det som at de følger med på utviklingen uten at de har konkrete satsinger på området. Selv om mange har store volum av data som de forvalter fra ulike kilder, og til og med har sanntidsdata tilgjengelig, har de ikke kommet så langt at de har kunnet vurdere hva slags verdi deres virksomhet kan ha av teknologien. Dette skyldes delvis manglende kompetanse, men også manglende evne til å drive forsøks- og utviklingsarbeid. Det er et fellestrekk at det er de store tunge dataforvalterne med store ressurser og med store og profesjonelle fag- og IT-miljø som er kommet lengst.

6.11.4 Tilgang til data og delingskultur

Et annet område som trekkes fram er at det er vanskelig å få tilgang til og utnytte data hos andre offentlige virksomheter. Dette er data som i utgangspunktet bør være åpne, og altså ikke behøver å være lukket (f.eks. av hensyn til taushetsbestemmelser). Muligheten til å kunne benytte åpne data (og grunndata for den saks skyld) fra andre virksomheter til berikelse av egne datasett er med på å øke verdien av analysene. Vi ser også at flere etterspør åpne data samtidig som de har mange datasett selv som kunne vært åpnet.

Holdninger til åpne data og delingskultur henger nøye sammen med interessen for stordata. Det er en tendens til at de som har et strategisk standpunkt med hensyn til å dele data også har en uttrykt strategi for bruk av stordata.

Det kan dessverre synes som om forvaltningen av data i virksomhetene generelt preges av tradisjonell silotenking. Innsamling, behandling og forvaltning av data settes ikke inn i en større sammenheng, eller knyttes til tverrgående prosesser i offentlig forvaltning og andre virksomheters og sektorers behov.

6.11.5 Personvern

Gjenbruk av informasjon reiser prinsipielle spørsmål knyttet til personvern; i hvilken grad bør det offentlige kunne benytte personrelaterte data i andre virksomheter for å kunne gjøre mer presise analyser?

Deling av data mellom Skattedirektoratet, Toll- og avgiftsetaten, Nav og Økokrim gir for eksempel mye høyere grad av muligheter til å avsløre økonomisk svindel. Det kan også være aktuelt å samarbeide med banker og forsikringsselskap for å øke denne sannsynligheten enda mer. Slike anvendelser reiser ikke bare spørsmål om personvern og hva sekundærbruken kan innbefatte, men også om eierskap til dataene og om resultatene av analyser basert på eksterne datakilder uten videre kan deles videre.

6.11.6 Kompetanse

Alle virksomhetene vi har snakket med nevner at tilgang til kompetanse på området er begrenset. Særlig gjelder dette kompetanse innen datavitenskap, det å forstå datakildene, datamodellene og hvordan datasettene bør beskrives (metadata) slik at de kan tas i bruk i kombinasjon med andre datakilder. Vi kan fastslå at dette er en viktig og knapp kompetanse. I kombinasjon med kompetanse innen statistisk metode er denne kompetansen avgjørende for å kunne utnytte de mulighetene som stordata gir. Dette er heller ikke noe man kan skaffe til

veie i markedet. Leverandørene kan først og fremst tilby teknisk kompetanse og kunnskap i verktøybruk. I tillegg nevner flere at det også er behov for at tradisjonelle studier suppleres med analysekompetanse. Særlig er dette relevant for de ulike ingeniørstudiene.

6.11.7 Operativt verktøy

Eksemplet fra prehospital strategisk satsing i Helsedirektoratet viser at stordata ikke kan begrenses til å dreie seg om å utarbeide beslutningsgrunnlag, ressursplanlegging og forskning. Teknologien kan også ha stor betydning i selve oppgaveutførelsen. Bruk av stordata-teknologi i ambulanser med tilgang til god informasjon om den til enhver tid raskeste og beste kjøreruten vil kunne ha livsviktig betydning. Både Statens vegvesen, Skatteetaten og NAV vil kunne effektivisere arbeidsprosessene sine og oppnå store gevinster gjennom bruk av stordata i sin oppgaveløsning.

Vi har ikke funnet eksempler i Norge på at stordata tenkes brukt i forbindelse med å effektivisere selve saksbehandlingen knyttet til myndighetsutøvelse, selv om dette nevnes som en mulighet både av våre informanter og i litteraturen.

6.11.8 Tjenesteutvikling og innovasjon

Mange i privat sektor benytter stordata-teknologien til å skreddersy tjenester til sine kunder ut fra historiske data. Det finnes datagrunnlag i ulike offentlige virksomheter som vil kunne benyttes til å forutsi hva slags type offentlige tjenester en person trenger eller for å automatisere rettighetsbasert saksbehandling.

Muligheten til å skreddersy tjenester etter livssituasjon har lenge vært diskutert[121]. Individuelt tilpassede tjenester blir først mulig når man har tilgang til detaljerte personopplysninger om den enkelte. Er f.eks. en person arbeidsledig vil dette kunne utløse rettigheter hos både kommunen og statlige virksomheter, og som tilbys vedkommende automatisk. Slik gjenbruk av data utfordrer de tradisjonelle skillelinjene mellom virksomheter, og vil kreve at virksomhetene samhandler tettere seg imellom, koordinerer sine systemer, rutiner, implementering av brukerrettigheter mv.

6.11.9 Smidige leveransemodeller

De som har kommet lengst på stordata-området nevner behovet for nye IT-leveransemodeller. Når det gjøres analyser på sanntidsdata er det avgjørende at tiden er kort fra endring av algoritmene som ligger til grunn for analysene til de er satt i produksjon og gir muligheter for nye analyser. Det vanlige i dag er at virksomhetenes leveransemodeller er basert på ITIL[122], med bruk av standardiserte prosesser for endringsstyring og driftssetting. Nye mer smidige modeller hvor utvikling, test og driftssetting gjøres hurtigere, som for eksempel DevOps, synes å være et behov som aktualiseres når stordatabaserte løsninger tas i bruk.

For samtlige virksomheter vil kanskje den største utfordringen være i det hele tatt å gjøre de riktige analysene. Det å komme fram til hvilke analyser virksomheten trenger er et eksplorativt arbeid. Det er av stor betydning i denne sammenheng å enkelt kunne teste ut ulike problemstillinger på ulike datasett. Derfor er det viktig også å ha en teknisk arkitektur og et IT-regime som muliggjør slik uttesting over tid.

---

[121] (Difi, LivsIT, 2007)
[122] IT Infrastructure Library jamfør eksempelvis http://www.itsmf.no/

7 Gevinstpotensialet

Som vi har sett innebærer stordata interessante nye måter å produsere verdi og nytte på. Eksemplene er også mange på bruk av stordata til å løse forvaltningsoppgaver. Teknologien representerer et potensial for kvalitativ forbedring og effektivisering av planleggings- og beslutningsprosesser og ikke minst til å målrette og brukerorientere offentlig tjenesteproduksjon[123].

7.1 Effektive beslutningsprosesser

Mange av de initiativene vi har sett på dreier seg om å framskaffe bedre beslutningsgrunnlag. DiBK ønsker bedre grunnlag for å beslutte reguleringer på bygningsområdet, Statens vegvesen ønsker bedre grunnlag for beslutninger om veiutbygging gjennom Nasjonal transportplan. Statsbygg ønsker å kunne ta beslutninger om vedlikehold av eiendommer og dermed spare utgifter og forbedre tjenestene.

Både Skatteetaten og NAV har sett gevinstpotensialet som stordata gir med hensyn til å avsløre og forebygge skatteunndragelser og svindel av velferdsgoder. Bare innen 5 av velferdsordningene som NAV forvalter er det estimert at svindel kan dreie seg om så mye som 8 milliard kr per år.

I tillegg er det klare effekter ved at selve saksbehandlingen i større grad baseres på stordata framfor manuelle og mer skjønnsbaserte arbeidsprosesser. McKinsey beregnet i 2012[124] at europeisk offentlig forvaltning kan redusere sine utgifter med 15-20 % samtidig som tjenestenivået opprettholdes eller forbedres, bare ved å basere seg på funksjonalitet fra stordata. Dette tilsvarer en gevinst på mellom €150 og €300 milliard per år, en gevinst som McKinsey beregner vil øke med 0.5 prosentpoeng per år.

Et eksempel er Bundesagentur für Arbeit (arbeidsdirektoratet i Tyskland) som har et budsjett på €54 milliard og 120,000 fulltidsansatte. Ved hjelp av bl.a. en stordata-strategi har de forbedret kundeservicen og kuttet kostnadene til arbeidsmarkedstiltak med €10 milliard per år uten at arbeidsledigheten har økt.[125]

7.2 Bedre kvalitet og aktualitet i beslutningene

Ved å bruke stordatateknologi og å tilrettelegge for å kombinere relevante datasett er også potensialet til stede for at beslutningene blir basert på bedre datagrunnlag. Å berike egne datasett med grunndata, stedsbestemmelser i form av kartdata, supplerende data fra andre og egne datasett bidrar. Høy oppdateringsfrekvens av sanntidsdata gjør også dataene mer aktuelle. Totalt gir dette et bedre beslutningsgrunnlag.

OECD peker på at offisiell statistikk kan bli mer dagsaktuell gjennom bruk av stordata. Vi har gjennom eksempler fra klima-, miljø- og bygningsområdet sett at det tar lang tid å utarbeide offisielle statistikker. Her er det et vesentlig potensiale for både å hente inn aktuelle data og å produsere statistikker mye raskere.

[123] (Chris Yiu, 2012)
[124] (McKinsey, 2011)
[125] (McKinsey, 2011) s 60

Det samme gjelder utarbeidelse av plandokumenter til vegutbygging og byplanlegging. Her vil mangelen på kvalitative og aktuelle grunnlagsdata ofte gjøre at planene utarbeides på feil premisser.

7.3 Raskere reaksjonstid

Gode analyser og presentasjoner av sanntidsdata gir også mulighetene for å analysere og reagere raskere på avvik fra normalsituasjoner. For eksempel viser tall fra Richmond i USA at responstiden på hasteoppdrag for ambulanser er langt raskere enn i Oslo (ca. 9 min. i Richmond mot ca. 17 min. i Oslo), selv om de to byene er sammenlignbare i utstrekning og befolkningstall.

Med gode datasett vil beredskapsaktørene kunne reagere raskere og dermed forhindre uheldige konsekvenser av kriser – ved hjelp av bedre tilgang til sanntidsdata. I Nederland benyttes om vi har sett stordata til å overvåke dikene og aksjonere raskt på avvik som kan medføre oversvømmelser.

Algoritmer som finner avvik fra normalsituasjoner basert på stordata benyttes i det hele tatt i mange sammenhenger som for eksempel drift av veier, bygninger og IT-infrastruktur. Ved hjelp av stordatateknologi har Statsbygg allerede spart energikostnader ved å avdekke avvik på energiforbruket i sanntid.

7.4 Presise prediksjoner

Stordata gir muligheter for å gjøre analyser i forbindelse med policyutvikling. Et område som i Storbritannia har framstått som særlig aktuelt med hensyn til utnyttelse av stordata-teknologier, er strategisk «foresighting» og støtte for policyutvikling[126].

Erfaringer fra USA, Nederland og Storbritannia viser at politiet kan bli mer effektive ved å ta i bruk forutseende politiarbeid, prediksjoner basert på stordata, som metode.

Innenfor transport- og byplanlegging vil det være viktig å kunne forutse befolkningsutviklingen, utviklingen i transportmønstre, forurensing m.m. for å kunne planlegge både boligutbygging, infrastruktur, offentlige tjenester og kulturtilbud og fritidsfasiliteter.

7.5 Økt brukerorientering

Å skreddersy tilbudet av produkter og tjenester til kunder basert på tilgjengelige data har lenge vært vanlig i privat sektor. I offentlig sektor har det tilsvarendevært snakk om å tilpasse tjenester bedre til brukernes livssituasjon gjennom utnyttelse av data, men det finnes så langt få eksempler på at dette har lykkes.

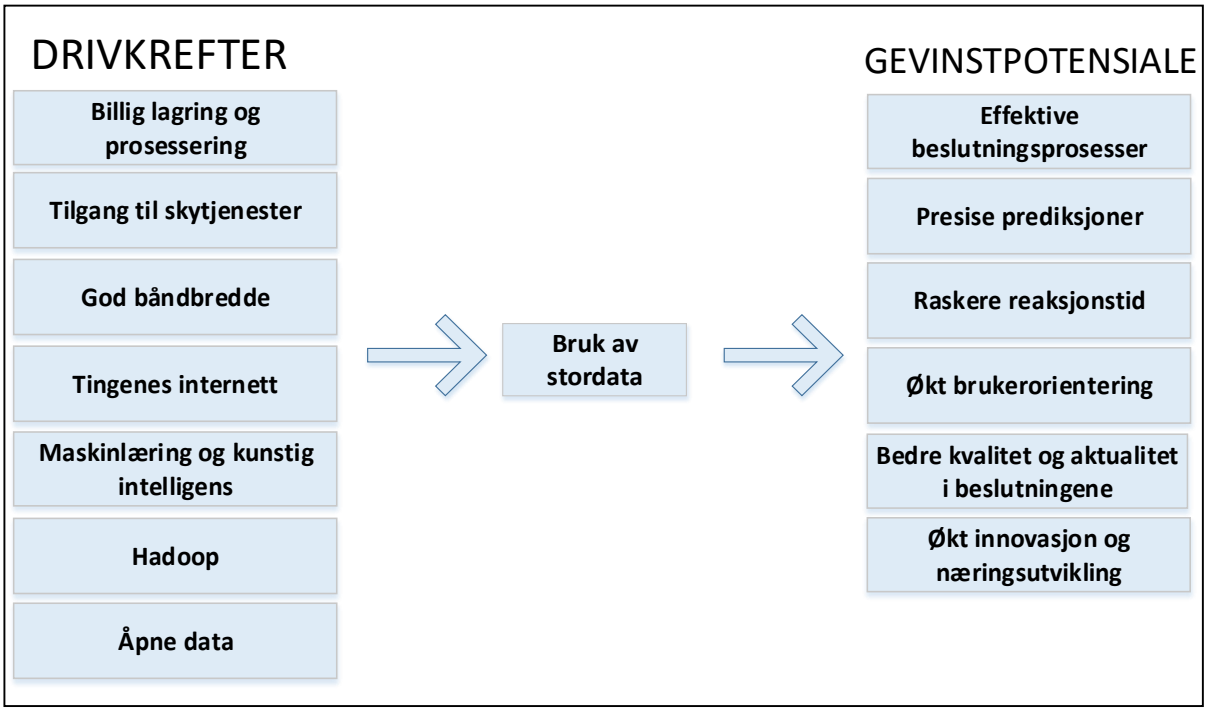
Eksemplet over fra Bundesagentur für Arbeit i Tyskland dreier seg imidlertid også om å tilrettelegge tjenester for ulike brukersegmenter.[127]

[126] (Elsevier, 2010)
[127] (McKinsey, 2011)

Statens vegvesen peker på at deres strategiske satsinger på å utnytte data allerede har gitt gevinster i form av mer informert publikum, noe som har medført at den enkelte kan planlegge hvordan de kan reise og dermed spare både tid og ulemper.

Analyser av brukernes livssituasjon og mulighetene for å kombinere datasett fra ulike deler av offentlig sektor legger til rette for helt nye måter å tenke brukerorientert tjenesteutvikling og tjenesteproduksjon på.

I figuren under oppsummerer vi forholdet mellom drivkrefter for bruk av stordata (kapittel 4) og gevinstpotensialet vi har beskrevet i dette kapittelet.



Figur 5 Forholdet mellom drivkrefter for bruk av stordata og gevinstpotensialet

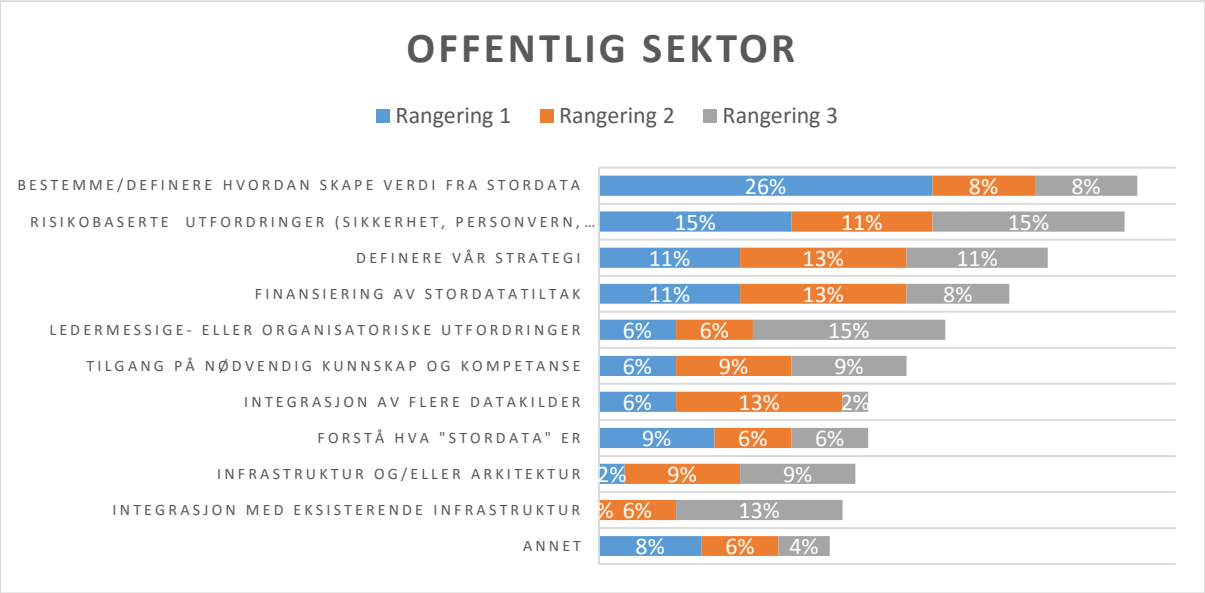
8 Utfordringer og hindringer

Stordata som teknologi og arbeidsform representerer også utfordringer. En politikk for stordata må handle om å utnytte potensialet samtidig som man håndterer utfordringer og hindre for å unngå eventuelle negative virkninger. I dette kapittelet ser vi på aktuelle utfordringer og hindre for å få effekter av stordata-teknologi i offentlig sektor.

Gartners[128] årlige stordata-undersøkelse viser at respondentene i offentlig sektor mener de største utfordringene er det å finne ut hvordan de i det hele tatt skal få verdi av stordata, ved siden av risikoen knyttet til blant annet informasjonssikkerhet og personvern. Dette overensstemmer med de vurderinger som figurerer i øvrig litteratur.

8.1 Stordata-økosystemet

EU-kommisjonen har pekt på at stordata-teknologien må settes inn i en større kontekst for at teknologien skal ha verdi. De ser for seg denne konteksten som et «data-økosystem»[129]. Også OECD setter i en rapport fra 2015[130] om datadrevet innovasjon stordata inn i en slik «økosystem»-kontekst.



Figur 6 Gartner Research Circle, Big Data Survey June 2015

Essensen i økosystem-tilnærmingen er at flere dimensjoner – og egenskaper hos aktørene i systemet – må adresseres for at en over tid skal kunne sikre bred nytte av stordata-teknologien. I følge EU-kommisjonen[131] bør en særlig fokusere på:

- <u>Tilgjengelige data med god kvalitet fra det offentlige:</u> Offentlig sektor står for en betydelig del av produksjonen av data i samfunnet, gjennom den myndighetsutøvelse

[128] (Gartner, u.d.)
[129] (Marta Nagy-Rothengass, EU, 2014)
[130] (OECD, 2015)
[131] (Marta Nagy-Rothengass, EU, 2014)

og de kontrollsystemer som er etablert. Offentlige datasett må gjøres tilgjengelig for viderebruk og det må være kjent for forvaltningen hvilke lukkede datasett ulike forvaltningsorgan behandler. Det kreves god oversikt over hvilke data som finnes, et regelverk som er tilpasset viderebruk – herunder til stordataformål – og tekniske standarder som gjør det mulig å gjenbruke data på en enkel måte.

- Offentlige rammevilkår som stimulerer til bruk av stordata:
    - Det trengs lovgivning, regelverk og policy-avklaringer som er tilpasset bruk av stordata.
    - Det bør finnes støtteordninger til oppstart av ny virksomhet basert på stordata: Siden området og markedet er nytt vil stimuleringsmidler til oppstart og nyetableringer være et viktig bidrag til det å utvikle et marked og å øke modenheten og kunnskapen på området.
    - Den nødvendige infrastruktur og teknologi må være på plass: En viktig rammebetingelse er tilgangen til nettverk med tilstrekkelig båndbredde, fasiliteter for datalagring, tilstrekkelig datakraft til prosessering – og alt med den nødvendig sikkerhet og robusthet.
- Tilgjengelig kompetanse: Det trengs kompetanse innen datavitenskap, informatikk, matematikk, statistikk og dataanalyse for å utnytte mulighetene i teknologien. Utdanningssektoren må forsyne både næringsliv og offentlig sektor med datavitere og annen nødvendig kompetanse for at virksomhetene til fulle skal kunne utnytte potensialet. I tillegg bør det finnes miljøer med dybdekompetanse, som også kan gi råd til virksomhetene i bruk av stordata-teknologi.
- Forskning og innovasjon: Det er behov for forskning på ulike områder, både for å øke kunnskapen på området og for å legge grunnlaget for nye teknologier og konsepter som kan avstedkomme ny produkt- og tjenesteinnovasjon.
- Et velfungerende marked: Markedet må kunne tilby ulike produkter og tjenester for at både private og offentlige virksomheter bredt skal kunne etterspørre og ta i bruk teknologien.

Denne oppfatningen av at et datadrevet samfunn avhenger av flere faktorer overensstemmer i stor grad med de utfordringsbilder som beskrives i casene både internasjonalt og i Norge og i den litteratur vi har gjennomgått. Den synes også å være reflektert i de policies som trekkes opp i land vi har sett nærmere på. Vi har her også i stor grad strukturert våre vurderinger av utfordringsbildet i Norge og vår videre analyse etter en slik «økosystem» -tilnærming.

Det synes riktig å oppsummere at økosystemet rundt stordata fortsatt er ungt og umodent, selv om flere elementer etter hvert kommer på plass. Når vi konkluderer med at økosystemet er umodent, tenker vi ikke primært på teknologien men på det litt større bildet – med aktører, kompetanse, samspill mellom offentlig og privat sektor, tilgang til infrastruktur, lagringstjenester, offentlige og private datakilder og analyse- og visualiseringsteknologi

8.2 Tilgang til data med god kvalitet

Sammen med tilgangen til infrastruktur (bredbånd, prosesseringskapasitet, skytjenester) og verktøy for å utføre analyser er tilgang til *data* naturlig nok den viktigste grunnforutsetningen for å kunne utnytte stordata som teknologi. I et samfunn som vårt norske, hvor det offentliges engasjement er stort på veldig mange samfunnsområder, er en meget stor andel av de data som naturlig kan inngå i analyser i offentlig eie. Dette handler om at det offentlige gjennom

sin myndighetsutøvelse forvalter data om et bredt spekter av temaer med høy aktualitet både for den enkelte og for næringslivet, og at det offentlige er delaktig i utformingen av tjenester på mange samfunnsområder. En rekke offentlige organer (slik som Statens vegvesen, NGU, NVE, Kystverket, Statens kartverk, kommunene) har som en uttalt del av sitt samfunnsoppdrag å samle inn og forvalte data – som bidrag til samfunnets felles kunnskapsgrunnlag.

8.2.1 Åpne data og kultur for datadeling

Offentlig sektor er på denne måten en svært sentral og direkte bidragsyter til et velfungerende data-økosystem. Det offentliges rolle i økosystemet er samtidig avhengig av at dataene som forvaltes, også er åpent tilgjengelige for viderebruk og bruk. Dette at det offentliges disposisjoner på dataområdet er så sentralt påpekes av flere av de virksomhetene vi har vært i kontakt med, men også i mange av de internasjonale studiene vi har gjennomgått. I de fleste land er derfor også politikkutviklingen knyttet til stordata nært knyttet til arbeidet med åpne data.

At det offentliges data har god kvalitet er også et viktig aspekt i denne sammenheng. Det hjelper lite at offentlige data er åpent tilgjengelige, så lenge de er av så lav kvalitet at de ikke kan brukes i de anvendelser som er aktuelle. Et eksempel her er Kartverkets høydedatamodell som det på 2016-budsjettet er bevilget midler til å revidere og forbedre. I mange nyttesammenhenger i samfunnet er det høyst relevant å bruke høydedata og annen data som gir geografisk lokalisering til å berike data i analysene. Nesten viktigere enn at data er av høy kvalitet, er ofte at de er av *kjent* kvalitet – slik at denne kan tas hensyn til i analysene som gjøres.

De siste årene har man i den utviklede delen av verden sett et økende krav om deling og gjenbruk av offentlig sektors data. EU har gjennom flere direktiver satt tilgangen til og viderebruk av offentlig informasjon på dagsorden. Stadig flere lands myndigheter får øynene opp for at deres data kan være av stor forretningsmessig verdi for næringslivet og for tjenesteinnovasjon i offentlig sektor selv (dvs. andre deler av forvaltningen). Dessuten er dataene også etterspurt av innbyggerne. I tillegg til at de kan ha direkte nytteverdi for sluttbruker (f.eks. innen transport, helsetjenester etc.) handler tilgangen til data også om innsyn, sporbarhet, transparens og åpenhet. Åpenhet om hva forvaltningen sitter med av kunnskapsgrunnlag har en åpenbar demokratidimensjon, i det at det gir innbyggerne lettere tilgang til beslutningsgrunnlag og prioriteringer som gjøres. Åpne data gir bredere innsikt i offentlige prosesser og styrker tilliten til det offentlige og det politiske systemet. I tillegg er de en viktig forutsetning for verdiskaping og innovasjon, også i forbindelse med bruk av stordata.

I Norge satte fellesføringene for statlig sektor i 2011 det formelle startskuddet for en mer systematisk satsing på åpne data. Fellesføringene forutsatte at etatene, så langt det lar seg gjøre, skal tilgjengeliggjøre sine rådata i maskinlesbare formater. Noen typer data skal ikke offentliggjøres eller være tilgjengelig for viderebruk. Dette gjelder blant annet personsensitive opplysninger, taushetsbelagt og gradert informasjon, og data som tredjepart har opphavsrett til.

Difi[132] har som nevnt ansvar for data.norge.no - Norges offisielle nettsted for åpne offentlige data. Her kan offentlige virksomheter registrere datasett som er tilgjengelige for viderebruk, og finne veiledningsmateriell og informasjon om bl.a. Norsk lisens for offentlige data (NLOD).

Den tette koblingen og avhengigheten til åpne offentlige data for berikning og analyser i stordatasammenheng i et samfunn som Norge gjør at de viktigste hindrene for deling av offentlige data også blir viktige hindre for utbredelsen av stordata. I et slikt lys kan f.eks. begrensende pris- og betalingsarrangementer være uheldig. Dette handler dels om økonomi for den som skal bruke dem i dataanalyser – og som ofte ikke er sikker på verdien før analysene er gjort, dels handler det også om at analysene kompliseres rent praktisk. Det kan også være juridiske begrensninger gjennom tredjeparts opphavsrett, slik det ikke er uvanlig å se f.eks. innen kultursektoren, eller komplekse eller suboptimale lisensregler. Det forekommer at offentlige virksomheter publiserer data med en viderebrukslisens som forhindrer kommersiell viderebruk, noe som naturligvis er til hinder for næringslivets stordataanvendelser.

I forlengelsen av dette er også offentlig sektors kultur – eller mangel på slik – for deling noe en bør rette oppmerksomheten mot. Flere av våre informanter ser dette som sentrale hindre mange steder noe som gir grunnlag for å hevde at det eksisterer en kultur for silotenking framfor en kultur for deling. Det kan synes å være behov for et tverrsektorielt påtrykk og mer koordinering for å få til en delingskultur som også er tuftet på prinsipper for personvern og tilfredsstillende informasjonssikkerhet.

Viktige elementer i en kultur for deling er at virksomhetene har:[133]

- Kunnskap om verdien av data (både nå og i framtiden);
- Forståelsen for viktigheten av riktig datainnsamlingsdesign (krav til datainnsamler);
- Forståelsen av dekoblingen mellom innsamling og bruk (at bruk ikke kan forutses av datainnsamler) – og dermed viktigheten av gode metadata;
- Tilgangen til delingsarenaer og enkelt tilgjengelige tekniske fasiliteter for deling.

Det er også verdt å fremheve her at alle data som samles inn skal kunne deles – enten de er åpent tilgjengelige eller lukket (tilgjengelige for kun visse typer anvendelser). Tilgjengelighetsprinsippet[134] knyttet til tilfredsstillende informasjonssikkerhet dreier seg nettopp om dette[135]. Uten gode metadata er det vanskelig for andre enn den som konstruerte eller samlet inn dataene å dra nytte av dem i nye kontekster eller på nye tidspunkt.

I stordatasammenheng må det påpekes at tilgang til data for analyse ikke bare handler om at det offentlige gjør data tilgjengelig for viderebruk under åpen lisensiering, men også om metadata som kan beskrive datafeltenes betydning i form av begrepsdefinisjoner og kodeverk som benyttes (semantikk), hvilken kontekst data er samlet inn i og dataenes kvalitet og relevans. Logiske eller semantiske forhold kan ellers gjøre det svært krevende å bruke dataene uten videre fortolkning og bearbeiding. Flere av virksomhetene vi har snakket med fremhever dette som et problem. Både Statsbygg, DiBK, Statens vegvesen og

[132] (Difi, Åpne data, 2015)
[133] (Regjeringen, 2005)
[134] Konfidensialitet, integritet og tilgjengelighet jamfør f.eks. https://no.wikipedia.org/wiki/Datasikkerhet
[135] (Difi, Åpne data, 2015)

Helsedirektoratet nevner at det er både kompetanse- og ressurskrevende å sette seg inn i og forstå ulike datasett.

Det kan f.eks. være at begreper er brukt ulikt i de ulike kilderegistre som er tenkt å inngå i analysen, og dermed gjør en sammenstilling umulig. I slike sammenhenger vil det være nødvendig å ha tilgang til de domenespesifikke ontologiene. Begrepsmessige siloer i offentlig sektor trekkes av flere fram som en åpenbar stopper for innovative stordata-anvendelser på tvers av forvaltningen. Dette vil ikke bare være et hinder for nytte i offentlig sektor selv, men også for verdiskapning i næringslivet og for nytteanvendelser for oss alle.

8.2.2 Sannferdighet, kvalitet og relevans

Innen noen domener og kontekster framheves *sannferdighet* (eller på engelsk «veracity» - jamfør definisjonen av stordata som benyttes i denne rapporten) som en sentral utfordring for bruken av stordata. Dette handler først og fremst om at det kan finnes grunner til å stille spørsmål ved kildedataenes kvalitet eller korrekthet. I litteraturen vises det bl.a. til case innen krise- og katastrofehåndtering, hvor kildedataenes korrekthet naturligvis er helt avgjørende for at de riktige og livsviktige beslutninger blir fattet. Man kan f.eks. tenke seg at rykter som spres i sosiale medier feilaktig kan bli tolket som fakta, og derfor må verifiseres gjennom innlemmelse av andre typer datakilder i analysen.

Viktigere enn *høy* kvalitet er som sagt ofte *kjent* kvalitet, dvs. at de data som inngår i analyser og sammenstillinger har beskrivelser som tilkjennegir oppløsning, feilmarginer, tidsstempel eller andre kvalitetsindikatorer. Slik kan algoritmene og analyseverktøyene ta hensyn til datakvaliteten.

Graden av detaljering og datasettets granularitet[136] og kvalitet på data er også en åpenbar utfordring i mange sammenhenger. I mange caser og i litteraturen på området påpekes at varierende datakvalitet og heterogenitet i datakildene (f.eks. kobling av sensordata, åpne registerdata og innhold f.eks. fra Twitter og andre sosiale medier) gjør det krevende å få til en meningsfull analyse og presentasjon av resultater. Dette handler om interoperabilitet og håndtering og sammenstilling av ulike dataformat, domenespesifikk begrepsbruk og ulikeartede kvaliteter ved selve dataene. Både Luftfartstilsynet, DiBK og Statsbygg nevner at det kan være utfordrende å gjenbruke data med uklar eller dårlig kvalitet. Innen miljø- og naturovervåkingen ser man eksempelvis at data er samlet inn for et bestemt formål, med de krav som har vært relevante for dette formålet. Over tid blir det aktuelt å benytte dataene til helt andre formål, men feil granularitet eller lav kvalitet kan være til hinder for de nye anvendelsesområdene.

Et annet aspekt ved dette er at metadata beskriver et avgrenset og spesifikt bruksområde for datasettene, og derved gjør det vanskelig i det hele tatt å identifisere datasettene for de nye analyseformålene.

I forlengelsen av dette – og som vi pekte på under avsnittet om åpne data – kan utfordringene også handle om usikkerhet knyttet til om dataene i det hele tatt er *relevante* for den analysen som skal gjøres eller den beslutningen som skal tas. Beskriver datasettene forhold som har

[136] Høyeste grad av granularitet betyr at dataene ikke kan deles opp i flere detaljerte data, navn kan for eksempel deles opp i fornavn og etternavn eller fornavn, mellomnavn og etternavn osv. med stadig høyere granularitet. Dette referer også til 1. normal form i relasjons databaser hvor alle dataelement er atomiske og kan ikke deles videre opp i nye dataelement (E.F.Codd, 1970).

noe med saken å gjøre, og vil analysen være basert på en sammenkobling av datasett som faktisk har noe med hverandre å gjøre?

8.3 Infrastruktur og teknologi

Utnyttelsen av stordata kan for mange virksomheters og sektorers vedkommende fordre *teknologiinvesteringer* langs «verdikjeden», i form av sensorer for datafangst, fasiliteter for datalagring, prosesseringskapasitet, verktøy for analyse og kompetanse for bruk.

Norge har god tilgang til både fast og mobilt bredbåndsnettverk, ifølge OECD er vi helt i teten i OECD. Tilgang til nettverk med god båndbredde er en forutsetning for bruk av stordata både når det gjelder utveksling av data og ikke minst innhenting av data fra sensorer. Økte datamengder krever at nettverkene bygges ut videre med enda bedre kapasitet slik at dette ikke blir en flaskehals i framtiden.

Mens den tekniske infrastrukturen er i ferd med å bli handelsvare, har mange fortsatt til gode å investere i den nødvendige kompetansen og relevante verktøy. Særlig for SMB-segmentet og for deler av offentlig sektor vil tilgangen til bestillerkompetanse i egen virksomhet være en vesentlig utfordring. Dette er i og for seg ikke en utfordring som er spesielt for dette området, men et spørsmål om modenhet for en ny teknologi.

Andre forhold knyttet til teknologien og infrastrukturen for stordata er av mer reell karakter ved at en del av de aktuelle teknologiene og infrastrukturelementene fortsatt er uprøvde, eksklusive, lukkede og proprietære[137]. Særlig gjelder dette analyseverktøyene som benyttes av de store mer tradisjonelle BI-leverandørene.

Tilgang til *datalagringsfasiliteter* er et grunnleggende behov i «økosystemet». Skyløsninger er ofte svaret på de behov som oppstår når lagringsbehovene for en virksomhet kommer opp i terabyte (TB) og petabyte(PB). Internasjonale markedsaktører tilbyr løsninger som kan dekke disse behovene. Avhengigheten til utenlandske tjenesteytere reiser imidlertid nye sikkerhets- og personvernrelaterte problemstillinger. Den pågående diskusjonen om EU-domstolens kjennelse[138] om ugyldigheten til Safe Harbour-avtalen med USA minner oss om dette.

Både Statens Vegvesen og Folkehelseinstituttet ser tilgang til nødvendig prosessorkraft som en utfordring. Dette kan til dels skyldes at dagens driftsmodeller er basert på en velprøvd modell med intern drift eller hostede tjenester og at det vil være krevende å skulle drifte nok prosessorkapasitet på denne måten til å kjøre tunge algoritmer på svært store datasett. Skytjenester er på mange måter løsningen på dette problemet, men for helsesektoren medfører dette vesentlige utfordringer i forhold til informasjonssikkerhet. Mens det for de fleste offentlige virksomheter vil medføre utfordringer å ha tilstrekkelig bestillerkompetanse.

Som redegjort for tidligere i denne rapporten har nye databaseteknologier og løsninger som Hadoop muliggjort prosesseringen og forvaltningen av enorme datamengder. En sentral del av teknologiaspektet er behovet for enkel tilgang til dataene – med andre ord *tilgjengelighet*. Det hjelper lite i en krisesituasjon at råvaren for beslutningssystemene ligger lagret i skytjenester som for øyeblikket er utilgjengelige. På et samfunnsnivå er dette et høyst relevant innsatsområde; det å sørge for robust tilgang til sikre lagringstjenester. Datasettene

[137] (Network Computing, 2015)
[138] (The Praliament Magazine, 2015)

trengs i stadig nye applikasjoner, som støtte for beslutninger og i sammenstillinger med andre data. Både lagrings- og andre tekniske fasiliteter blir i en framtidig, stordata-drevet økonomi viktig samfunnskritisk infrastruktur – og må håndteres deretter.

De tidlige forsøkene med stordata i våre norske caser, så vel som i litteraturen, viser at det ofte er problematisk å innarbeide bruken av stordata i virksomhetenes eksisterende arbeidsprosesser og beslutningssystemer. Stordata forblir noe eksperimentelt og «utenfor» de ordinære arbeidsmetodene. Analysene er kanskje også noe som «settes bort» til eksterne analysebyråer. Denne integrasjonsutfordringen har en datateknisk og arkitekturmessig side, men etter vår vurdering er det først og fremst kompetansemessige – og kanskje også kulturelle, administrative eller praktiske grunner til dette. Flere av de brukercase vi har fått innblikk i indikerer at det ikke uten videre er enkelt å endre på arbeidsflyt, arbeidsdeling og funksjonell spesialisering i virksomhetene. Det å ta i bruk stordata er disruptivt. Noen er for eksempel av den oppfatning at kvantitativ analyse skal og bør gjennomføres på «klassisk» vis, av folk utdannet innen statistisk analyse – og med de verktøy og metoder som alltid har vært brukt.

8.4 Juridiske rammevilkår

8.4.1 Personvern

Personvernhensyn er som vi har sett i tidligere kapitler en sentral problematikk ved bruk av stordata. Personvern er ett av de temaer som er mye framme i arbeidet med policyutvikling og diskusjonen om stordata som teknologi, også i andre land. Selv om en stor del av de case og anvendelser vi ser konturene av – også i offentlig sektor – ikke handler om personopplysninger i det hele tatt, men om f.eks. naturfenomener, trafikkvolumer og varelogistikk, så er det likevel grunnlag for å si at dette er et sentralt hinder for stordata.

I utgangspunktet skal data som samles inn om en person, kun benyttes for det formål de ble samlet inn – og for øvrig i samsvar med de forutsetninger som er gitt i henhold til lovverket. Grunnlaget for å samle inn data om person er gitt av lov eller forskrift, gjennom den aktuelle aktørens mandat som myndighet, eller gjennom den enkeltes samtykke.

Det å bruke personrelaterte opplysninger i stordatasammenheng kan bryte med disse forutsetningene. Stordata handler om gjenbruk av data til nye formål, og utfordrer derfor prinsippet om formålsbegrensning – så fremt ikke anvendelsen kan knyttes direkte til formålet med datainnsamlingen. Stordata representerer også en verdsetting av data i seg selv – hvor verdipotensialet nettopp ligger i framtidige og til dels ukjente bruksmuligheter. Det er derfor vanskelig å se for seg hvordan formålet skal kunne beskrives på forhånd, slik at det eksempelvis kan gis samtykke fra dem det gjelder.

Skatteetaten peker i denne sammenheng på spørsmålet om hvor langt det offentlige bør gå i å innhente informasjon om innbyggerne. Er det greit at det offentlige bruker «webcrawling»[139] teknologi og tekstanalyse for systematisk analyse av offentlig tilgjengelige facebookprofiler? Er det greit med enkeltoppslag? Hvor lang kan det rettferdiggjøres at man går? Dette er spørsmål som er generelle, men som aktualiseres av nye teknologier og muligheter.

På dette området er det klart at dagens regelverk kan være til hinder for økt gjenbruk. Det kan for eksempel være en mulighet å kjøre analyser på anonymiserte data og likevel identifisere

[139] (Computer, how stuff works, 2015)

personer og selskaper dersom man får treff. Dette har særlig relevans innen helserelatert forskning, men også innen politiet og i jakten på svindel med offentlige midler. I kriminalitetsforebyggende arbeid bør det vurderes om informasjon i større grad kan deles på tvers, men samtidig se på hvilke mekanismer som likevel kan ivareta det nødvendige personvernet.

En sentral bekymring i denne sammenheng er mulighetene for sammenstilling av personopplysninger til nye og kanskje uønskede formål. Fordi data om person i en stordatakontekst ikke vil bli slettet, men tvert imot tilrettelegges og gjøres tilgjengelig for framtidige anvendelser, mister den enkelte kontrollen med hvordan dataene faktisk blir benyttet. Dette har i de fleste land en klar juridisk side; uten tilstrekkelig hjemmel for behandling av dataene for det nye formålet vil muligheten til å gjøre analyser være begrenset. Noen land har som vi har sett innført unntak for forsknings- og analyseformål, og med en mer eller mindre romslig fortolkning av hva som kan dekkes under dette.

I Norge og mange andre land vil behandling av dataene med et nytt formål kreve en ny hjemmel i form av lov eller forskrift, konsesjon, eller samtykke fra de personer opplysningene dreier seg om. I en stordatasammenheng er det et spørsmål om dette er realistisk; uansett hvilket hjemmelsgrunnlag som velges vil det være tidkrevende å skaffe lovhjemmel eller konsesjon. Å innhente informert samtykke vil i mange sammenhenger være praktisk vanskelig, i alle fall dersom det dreier seg om bruk av personopplysninger samlet inn i en annen virksomhet med et helt annet formål – og for lenge siden.

Et potensielt viktig anvendelsesområde for stordata er data om geografisk lokalisering (hvor personer befinner seg, og på hvilket tidspunkt). Slik lokaliseringsinformasjon sammenstilt med andre data om personer er et ikke uvanlig «problemområde» i personvernsammenheng. Dataene om personens disposisjoner trenger ikke i seg selv å være direkte intimiderende, men kan når de «berikes» med sted- og tidfesting gi en samlet framstilling av vedkommendes livsførsel eller adferd – og som bryter med den avgrensning av bruksområdet som opprinnelig ble gitt samtykke til. Det er også nevnt andre eksempler på at data som i sin originale form ikke omfatter data som identifiserer person (ved at de er anonymisert e.l.) – like fullt kan utgjøre en trussel mot personvernet ved at de kan berikes og sammenstilles med andre data og derved gjøre det mulig å finne fram til den aktuelle personen på uønskede måter.

Mange datasett som beskriver personer handler ikke direkte om sensitive forhold, f.eks. er data om en persons eiendommer ikke å betrakte som sensitiv informasjon. Det er imidlertid en utbredt oppfatning at data om personer som ikke er sensitive, sammenstilt med andre data kan oppfattes som sensitive selv om de ikke nødvendigvis er det etter personopplysningsloven eller helseregisterloven. For eksempel kan enkelte finne at data om økonomiske disposisjoner sammenstilt med lokasjonsdata e.l. ønskes hemmeligholdt – bl.a. av kriminalitetsforebyggende grunner.

Datatilsynet peker på re-identifisering som en av de virkelig store utfordringene ved stordata-analysere i sin rapport fra 2013[140]. Selv om data som benyttes i analysene er anonymiserte data, vil de sammenstilt med andre data likevel kunne vise til person og dermed være personopplysninger. Vi har ikke fått klarlagt noen dokumenterte tilfeller av dette i Norge, og

[140] (Datatilsynet, 2013)

vi har heller ikke i litteraturen sett noen eksempler på at dette har vært noe faktisk problem i praktiske sammenhenger. Vi er imidlertid kjent med at NTNU har forskningsprosjekt på området. Vi mener likevel at det er riktig å fastslå at den som behandler personopplysninger må være oppmerksom på denne risikoen for re-identifisering. Anonymisering av data vil i de fleste, men ikke alle, tilfeller hindre at personene dataene beskriver blir identifisert.

Vi pekte over på betydningen av metadata og kunnskap om dataenes kvalitet. Dette er potensielt også en problemstilling i personvernsammenheng. Dersom et datasett som omfatter personopplysninger har lav kvalitet eller er samlet inn i en spesifikk kontekst og med et bestemt formål, kan bruk av disse dataene i en helt annen sammenheng gi feilaktige slutninger, f.eks. i en tenkt stordata-støttet, automatisert saksbehandling. Mangelfulle metadata om formål og begrepsbruk kan på den annen side også medføre at datasettet feiltolkes og forkastes uten at man har fått avklart på en god måte om det kunne vært anvendt – gitt visse forutsetninger.

Dersom bruk av stordata skal inngå i saksbehandlingsprosess rettet mot enkeltvedtak om en ytelse eller tjeneste til en bestemt person, kan datagrunnlaget naturligvis ikke anonymiseres. Spørsmålet om hjemmelsgrunnlag og formål blir da aktuelt slik det er for alle spørsmål om gjenbruk av personopplysninger på tvers av virksomhetsgrenser i offentlig forvaltning.

Datatilsynet har i sin rapport om stordata påpekt at personvernlovgivningen er dekkende også for stordata-området og at det ikke er behov for tilpasning av lovgivningen i en ny virkelighet. Det er Datatilsynets oppfatning at gode rutiner for anonymisering vil bidra til å redusere faren for at personer identifiseres. Dersom personopplysninger likevel skal behandles bør hjemmelsgrunnlaget være basert på samtykke fra den som personopplysningene omhandler.

I Datatilsynets veileder til anonymisering av personopplysninger[141] gjøres det klart at personopplysningsloven ikke stiller krav til behandling av anonyme eller anonymiserte opplysninger, og at en korrekt identifisering av hvor grensen går mellom personopplysninger og anonyme data kan være av stor betydning for vurdering av i hvilken grad datasettet kan benyttes i stordata-sammenheng. Datatilsynet definerer anonyme opplysninger slik:

Opplysninger som det ikke er mulig å knytte til et identifiserbart individ når man tar i betraktning alle de hjelpemidlene som med rimelighet kan tenkes brukt for å identifisere vedkommende, enten av den behandlingsansvarlige eller av en hvilken som helst annen person.

Anonymisering av personopplysninger kan altså medføre at andre kan gjenbruke de anonyme dataene. Anonymisering og pseudonymisering er vanlige tiltak innen helsesektoren bl.a. for å muliggjøre forskning på og nye anvendelser av helserelatert informasjon på et aggregert nivå (f.eks. kartlegging av mønstre i sykdomsutbredelser).

Vi ser både fra litteraturen og de virksomhetene vi har vært i kontakt med i vår kartlegging (f.eks. Statens vegvesen) at hensyn til personvernet representerer en reell hindring for flere mulige bruksområder for stordata. Det er bl.a. en utfordring å skulle behandle sammensatte datasett som potensielt inneholder personopplysninger fra ulike eksterne kilder i sanntid – så lenge man ikke har kontroll på informasjonssikkerheten i kildene og ved overføringen av

[141] (Datatilsynet, 2015)

dataene. Spørsmålet om hva som er tilfredsstillende informasjonssikkerhet må besvares gjennom grundige risiko- og sårbarhetsanalyser. Likevel kan dette være utfordrende i og med at virksomhetenes kompetanse på området ofte er begrenset. Behovet for å ivareta personvernet gjennom sikringstiltak gjør sannsynligvis også at sekundærbruken av helserelatert informasjon foreløpig er begrenset i Norge[142]. Særlig gjelder dette dersom helseopplysninger skal kombineres med datakilder utenfor helsesektorens sikkerhetsinfrastruktur.

8.4.2 Eiendomsrett

Immaterielle rettigheter (IPR) og lisensbestemmelser knyttet til de data som er aktuelle for stordata-anvendelser, er noe som går igjen som en reell barriere i mange av de pilotprosjekter vi finner omtalt i litteraturen. Opphavsretten vanskeliggjør deling av data og stordata-anvendelser innen kreative sektorer, ikke bare i privat sektor – men også innen den offentlige kultursektoren, og i anvendelser som kombinerer kulturdata med andre data. Opphavsretten trekkes fram av flere som et «utdatert» rammeverk, som spenner ben for mange mulige anvendelser av stordata. På sosiale medier gir brukerne opp sine rettigheter til dataene, men i mange sammenhenger er data – også slike som er åpent tilgjengelige – beheftet med opphavsrettslige klausuler og bestemmelser som gjør stordata-anvendelser vanskelige eller utenkelige.

For offentlig sektor gjelder ellers i utgangspunktet at data skal publiseres under en åpen lisens. Det offentlige bør for sin egen del ikke hevde opphavsrett eller ta betalt for data som er innsamlet som en del av sin myndighetsutøvelse og for skattebetalernes midler, så fremt det ikke foreligger særskilte behov for dette – og dette er hjemlet i lov- og regelverk. Dersom det må tas betalt for data, bør dette i utgangspunktet begrense seg til å dekke kostnadene. En spesiell utfordring er at en del etater benytter lisenser som tillater viderebruk, men ikke kommersielt.

Selv om det er teknisk realiserbart å sammenstille data som er åpent tilgjengelige, kan opphavsretten og lisensbestemmelsene reise spørsmål om hvorvidt det sammenstilte eller berikede datasettet og produktene (analysen og de visuelle framstillingene) skal betraktes som verk avledet av de underliggende datakildene, eller som noe eget – og hvordan dette nye i så fall skal lisensieres. Skal man i en framstilling av et beslutningsgrunnlag framstilt på basis av stordata måtte navngi alle rettighetshaverne knyttet til datakildene? Hvordan skal en analyse basert på kildedata og berikende data publisert under en åpen lisens, i seg selv lisensieres?

Under denne overskriften må en også vurdere hvor godt eierskapet til data er definert – og hvordan det bør være i et framtidig økosystem. Gartner har pekt på at innen 2019 vil 90 % av alle virksomheter ha data i sine systemer som de ikke eier eller kontrollerer.[143] Dette handler dels om ansvarliggjøring av rett aktør for datakvalitet m.v., men også om eierskap og råderett. Hvem eier dataene om et kjøretøy som beveger seg i et veinett – bileieren, bilprodusenten eller veimyndighetene? Hvem eier egentlig kassakvitteringene (POS-data) som forteller på varenivå hva forbrukeren har handlet – forbrukeren eller butikken? Kan eierskapet knyttes til objektet (varen – som jo skifter eier ved handelen) eller ved transaksjonen (som jo initieres av

[142] (Helse SørØst, 2014)
[143] (Gartner, 2013)

kunden, men som muliggjøres av butikken og baseres på butikkens infrastruktur og datasystemer). Alle vil være interessenter når det gjelder å utnytte dataene til sine formål, den dag de blir bevisst at dataene har en verdi for dem. Det er grunn til å anta at uavklarte rettighetsforhold på slike områder bidrar til usikkerhet og risiko, og i hvert fall at de bremser innovasjonen på området.

8.4.3 Andre juridiske utfordringer

Vi ser også en rekke andre egenskaper ved stordata som vil utfordre lovverket – uten at vi i denne omgang har tatt mål på oss å behandle dette uttømmende. Det kan f.eks. pekes på at vår norske offentlighetslovgivning opererer med et innsynsbegrep som er knyttet til dokumentet som format. Det er tvilsomt om algoritmer som ligger til grunn for maskinbaserte beslutninger i sanntid vil følges av en dokumentasjon som gir mulighet for en slik type innsyn som det som er tenkt i offentlighetsloven. Tilsvarende stiller bestemmelsene i forvaltningslovens §§17 og 18 krav om innsyn i forvaltningsorganets «saksdokumenter» og de fakta som ligger til grunn for en beslutning. Hvordan stordata-anvendelser skal gjøres koherent med forvaltningsloven er i utgangspunktet vanskelig å se. Vi er imidlertid kjent med at det nå er nedsatt et utvalg som skal se på behovet for å modernisere forvaltningsloven[144].

I sin rapport *Forutseende politi* problematiserer Teknologirådet[145] mulighetene for tilsyn og kontroll med de algoritmer som benyttes i forbindelse med politiets bruk av stordata og prediktive analyser. Selv om det er åpenbare taktiske grunner til at politiets (og tilsvarende Skatteetatens, NAVs og mange andre etaters) arbeidsmetoder må være delvis hemmelige, er det samtidig slik at publikum må kunne ha innsikt i de antagelser og vurderinger som ligger til grunn for beslutninger som treffer dem direkte. Offentlig sektors beslutningssystemer skal i hovedprinsipp være transparente. En utvikling som ikke sikrer dette, vil kunne bidra til fremmedgjøring, stigmatisering av enkeltindivider og virke frihetsberøvende. Det er imidlertid en betydelig pedagogisk utfordring å sikre at hele befolkningen forstår hva prediktiv analyse er for noe, og hvordan de prediktive algoritmene fungerer. Det er her også et spørsmål om hvordan man skal sikre seg tilgang til algoritmer som kanskje er del av forretningshemmelighetene til leverandørene av analyseverktøy og stordata-relatert programvare.

Stordata kan også berøre spørsmålet om inkludering. For befolkningsgrupper som er mindre digitalt kompetente vil stordata som fenomen og verktøy i forvaltningen være fremmedgjørende og vanskelig å forstå eller sette seg inn i. I tillegg til at dette reiser etiske spørsmål om deltakelse og demokratiets virkemåte, kan det være aktuelt å se på om stordata-anvendelser i visse konkrete sammenhenger vil kunne utfordre diskriminerings- og tilgjengelighetslovens formålsparagraf og § 4.

8.5 Introduksjonskostnader

Vårt inntrykk fra arbeidet med kartleggingen er at mange virksomheter er nysgjerrige på stordata-teknologien og hva den kan gi av nytte for dem. For mange vil nok likevel introduksjonskostnadene for å ta i bruk stordata framstå som høye, sett i forhold til den usikkerhet som fortsatt hersker om nytteaspektene. På overordnet nivå vil en offentlig

[144] (Regjeringen.no)
[145] (Teknologirådet, 2015)

politikk på området også måtte handle om å adressere akkurat dette og gjøre aktørene mer sikre på nytten; ved å tilrettelegge for virksomhetenes utprøving av mulighetene (test-bed for løsninger), og ved å synliggjøre case som illustrerer nytteeffektene av stordata til prediktiv analyse og mer effektiv saksbehandling og beslutningsprosesser.

Den andre siden av denne utfordringen er å påvirke markedet, slik at priser og kostnader synker. Dette vil handle om anskaffelsespolitiske grep som stimulerer til konkurranse i markedet for løsningskonsepter og tjenester, herunder motvirke innelukking og vertikal «bundling» av tjenester for stordata. Dette bør for øvrig kunne være et viktig innovasjonsområde, eksempelvis for Difi og KS. Videre vil det påvirke markedet å sikre en god bestillerkompetanse i forvaltningen. Dette vil handle om å understøtte vurderingene i virksomhetene med kunnskap om god praksis, viktige premisser og krav som bør stilles, og (nøytrale) råd om hvilke løsningsvalg som egner seg for ulike formål.

Introduksjonskostnader dreier seg ikke bare om budsjetter til å dekke behovet for anskaffelser av ny teknologi eller innleie av kompetanse. Introduksjonskostnader er også det å overvinne intern motstand og skepsis i etablerte fagmiljøer og i møtet med innøvde arbeidsformer. Vi merker oss i ulike sammenhenger at ikke alle vurderer stordata som noe relevant fenomen for dem. Det er også grunn til å utgå fra at mange på leder- og beslutningsnivåene i virksomhetene vil være konservative, og forsiktige med å ta inn over seg de nye mulighetene som ligger i stordata. Slik vi også observerer i våre caser vil bruken av stordata derfor lett bli noe som skjer litt ved siden av og i tillegg til de etablerte strukturer.

8.6 Kompetanse

Slik vi har sett vil tilgangen på kompetanse for mange virksomheter være en betydelig utfordring. Selv om mange er nysgjerrige, er det antagelig få i offentlig forvaltning som har tatt stilling til de muligheter stordata gir for dem. En god del vil sannsynligvis ikke være i stand til å uttrykke krav om leveranser på dette området i en anbudsprosess, eller vurdere kvaliteten og innholdet i de tilbud de mottar. For at de som sitter sentralt i beslutningssystemene, eller som bemanner analyseenheter, informasjonsavdelinger og strategiske stabsfunksjoner, skal være gode bestillere må dette bildet endres. Det må naturligvis også være tilsvarende kompetanse og kapasitet (til å redegjøre for mulighetene og begrensninger i teknologien) på tilbydersiden – enten dette er i lokal IT-avdeling, hos løsningsleverandører eller blant tjenesteytere i markedet.

Det ligger i sammenhengen at «kompetanse» her handler om flere ting. Dels må virksomhetene kjenne til hva som finnes av løsninger og produkter på området, og hvilke spesielle styrker og svakheter de eventuelt er beheftet med. Dels bør de også kjenne til aktørbildet, og relevante kunde-case, slik at de kan orientere seg bredt om de ulike tilbyderne og deres teknologier og løsningskonsepter.

Det som sannsynligvis er aller viktigst, og som vil kunne sette fart og retning i stordata-satsingen til virksomhetene, er de relevante fagenhetenes kunnskaper om det som kan kalles «data science» eller datavitenskap. Dette er i dag en ettertraktet kompetanse på verdensbasis. Forbes kårer «data scientists» til den aller «hotteste» jobbkategorien for tiden[146]. Virksomhetene trenger å jobbe mer systematisk med å kapitalisere på de store datamengdene

[146] (Forbes, 2014)

i egen virksomhet og i omgivelsene. Men til dette trenger de disse nye ekspertene. Gjennom metoder som «data mining», maskinlæring og oppbygging av prediksjonsmodeller kan de bistå med å predikere utfall og etablere kvalitativt bedre grunnlag for anbefalinger, beslutninger, markedsføringer og utvikling av nye tjenester.

Men hvor får virksomhetene tak i disse nye ekspertene? I dag mangler datavitenskap som fagfelt innen de fleste profesjonsutdanninger, og feltet er smalt selv innen akademia, i hvert fall her i Norge. I flere internasjonale rapporter blir denne mangelen på kompetanse i samfunnet trukket fram som en betydelig utfordring på IT-området. Gjennomgående pekes det på at de teknologiske mulighetene – bl.a. på stordataområdet – ligger i forkant av virksomhetens evne til å utnytte mulighetene. Så lenge ikke utdanningsinstitusjonene ikke oppgraderer og oppdimensjonerer dette området, vil det med andre ord være mangel på adekvat arbeidskraft i virksomhetene. I tillegg til at dette vil være en bremse i utnyttelsen av stordata på bred front, er det grunn til å tro at det også vil bidra til å låse inne virksomheter som ønsker å eksperimentere med stordata i arrangementer i regi av proprietære løsnings- og tjenesteleverandører.

Kompetanse handler heller ikke bare om spesialisert kompetanse innenfor ingeniør-, matematikk- og informatikk-miljøene. På basis av det vi finner i litteraturen og i omtalen av temaet fra de virksomheter vi har intervjuet i kartleggingen, blir det i økende grad viktig med kombinasjonskompetanser, hvor nettopp datavitenskap appliseres på et definert fagområde – f.eks. natur- og miljøforvaltning, meteorologi, revisjon, trafikkøkonomi, logistikk eller politifaglig metode. Det er grunnlag for å hevde at det i hovedsak ikke finnes slike tilbud ved studiestedene i dag.

Kompetanse handler til slutt også om brukerkompetanse, dvs. både evnen til å vurdere når og hvordan stordata skal benyttes – og til hva, og om å kunne vurdere kritisk både datagrunnlag og analysemodeller som benyttes (noe som forutsetter transparens), og den resultatframstilling og -visualisering som er output av analysene. Dette er en form for digital kompetanse som i framtiden blir viktig så vel hos offentlige myndigheter og i andre beslutningssystemer som baserer seg på stordata – som blant oss alle som sluttbrukere og forbrukere. Samlet sett kan en si at vi må bli flere som evner å utlede innsikt fra de data vi omgir oss med i samfunnet – og ikke minst å forstå hvordan andre gjør det, om oss.

8.7 Markedet

Markedet for stordata-tjenester domineres av noen store aktører, som IBM, HP, EMC og Amazon. Selv om mange aktører innen det tradisjonelle Business Intelligence-segmentet (slik som SAS Institute) nå også opererer innen det en kan kalle «stordatamarkedet», og det dukker opp nye stordata-baserte forretningsmodeller, må markedet totalt sett fortsatt sies å være umodent. Konkurransen og tilfanget av alternative løsningskonsepter, programvarer og tjenester er forholdsvis begrenset – selv om dette er i rask endring på verdensbasis.

Markedets umodenhet har antagelig vært ett av hindrene for en omfattende bruk av stordata. Særlig for offentlig sektor vil det være vanskelig å basere seg på spesialiserte, kommersielle kunnskapsmiljøer med proprietære, ofte lukkede teknologier – som gjør analyse- og visualiseringsjobben for kundene. Offentlig sektor ønsker tradisjonelt ikke å være tidlig ute i å ta i bruk nye teknologier, og ta risiko for feilinvesteringer. I flere land ser vi at det offentlige går sammen med f.eks. IBM i utprøvingen av teknologi for lagring og dataanalyse,

i ulike former for offentlig-privat samarbeid. For andre er det mer naturlig med et løsere forhold til leverandøren, og at analysene bestilles fra gang til gang. Kundenes bestillinger besvares da med en rapport eller et visuelt produkt tilbake, riktignok ikke helt i sanntid.

I tillegg til at offentlige innkjøpere er redde for innlåsing i en enkelt leverandørs rammeverk og løsningskonsepter, er det også viktig å ha kontroll med og transparens i analysemodeller og algoritmer. Krav om åpne data og åpen kildekode vil være et mulig virkemiddel for å motvirke den type innlåsing som vil være uheldig på lenger sikt. Dette handler også om prisen på programvare og tilgangen på kompetanse i virksomheten; for mange vil det ikke finnes noe valg – det vil falle for dyrt å investere i en «in-house» tilnærming, og en må forholde seg til de leverandørene og tilbudene som er til stede i markedet.

Gjennom offentlige anskaffelser kan det offentlige bidra til å utvikle et mer dynamisk og velfungerende marked for stordata-baserte tjenester (lagring og forvaltning, «data mining», berikning, analyse, visualisering). Som tidligere påpekt er det viktig at det offentlige også er med på å stille krav ved slike anskaffelser, blant annet til åpne modeller, standarder og grensesnitt.

Samtidig er det antagelig en betydelig kompetanseutfordring å spesifisere krav og rigge til anbudskonkurranser på dette området. Det å utvikle sjekklister og standardiserte mal-dokumenter for kravstilling for anskaffelser av produkter og tjenester på området er derfor et mulig tiltak som kan gjøre terskelen lavere for mange aktører.

Et annet område som vil påvirke markedet i riktig retning er standardiseringsarbeid, og myndighetenes bruk av de standarder som utvikles. Mellom annet nevnes behovet for en standardisert måte å beskrive offentlige data på og standardiserte formater for overføring av data. Sammen med bransjeorganer bør kanskje staten utrede mulighetene for et standardiseringsprogram på området.

Det offentlige kan også inngå samarbeid med private aktører (OPS[147]) om å utvikle demonstratorer og piloter som kan bane vei og validere tekniske konsepter, lagringsfasiliteter, standarder, forretningsmodeller, transparenskrav og mulige reguleringsmekanismer. I denne sammenheng er det også høyst relevant for offentlige aktører – gjerne ledsaget av innovative norske markedsaktører – å søke inn i internasjonale samarbeidsprosjekter knyttet til faktiske forvaltningsanvendelser og hvor vi kan lære av andre land (slik som f.eks. Nederland og Storbritannia innen områder som intelligente trafikksystemer og avdekking av økonomisk kriminalitet).

Et av de mer urovekkende forholdene er den avhengighet som skapes mellom offentlige virksomheter og enkeltaktører som opererer i tjenestemarkedet for stordata. I noen av de case vi har studert har aktørene gitt uttrykk for bekymring for en økende avhengighet til USA-baserte tjenester og tjenesteytere på stordata-området, i det at analyser utføres av spesialiserte amerikanske spesialistmiljøer – med proprietær programvare og «lukkede» analysemodeller. EU-kommisjonens Neelie Kroes ga i 2013[148] uttrykk for sin bekymring over at hele 17 av de 20 største aktørene i stordatamarkedet er amerikanske selskap; bare to er europeiske. I tillegg til at den geopolitiske avhengigheten kan være urovekkende, vil slike konstellasjoner også

[147] Offentlig privat samarbeid se http://www.anskaffelser.no/temaer-bae/offentlig-privat-samarbeid-ops/hva-er-offentlig-privat-samarbeid-ops
[148] (Big Data for Europe, EU, 2013)

kunne true den transparens og muligheter for sluttbrukers innsyn i algoritmer og analysemodeller som vi ønsker.

Det offentlige kan til slutt også bidra til et fremvoksende marked gjennom sine disposisjoner på området åpne data. Tilgangen til åpne datasett er som vi har resonnert rundt tidligere i dette kapitlet en fundamental forutsetning for at det skal finnes råvare for analyser og foredling til kunnskap. I en del andre land er myndighetene langt mer offensive på å publisere offentlige datasett åpent enn det vi er i Norge[149]. Selv om vi ligger på en god 7. plass (Danmark ligger på 1. plass) på Global Open Data Index i 2014 er mye ugjort på dette området som mest handler om åpenhet i forvaltningen. I USA hadde man i 2014 hele 112.300 åpne datasett i forvaltningen, noe som selvsagt bidrar til å spore an til datadrevet tjenesteinnovasjon i både offentlig og privat sektor – og som muliggjør en større satsing på stordata.

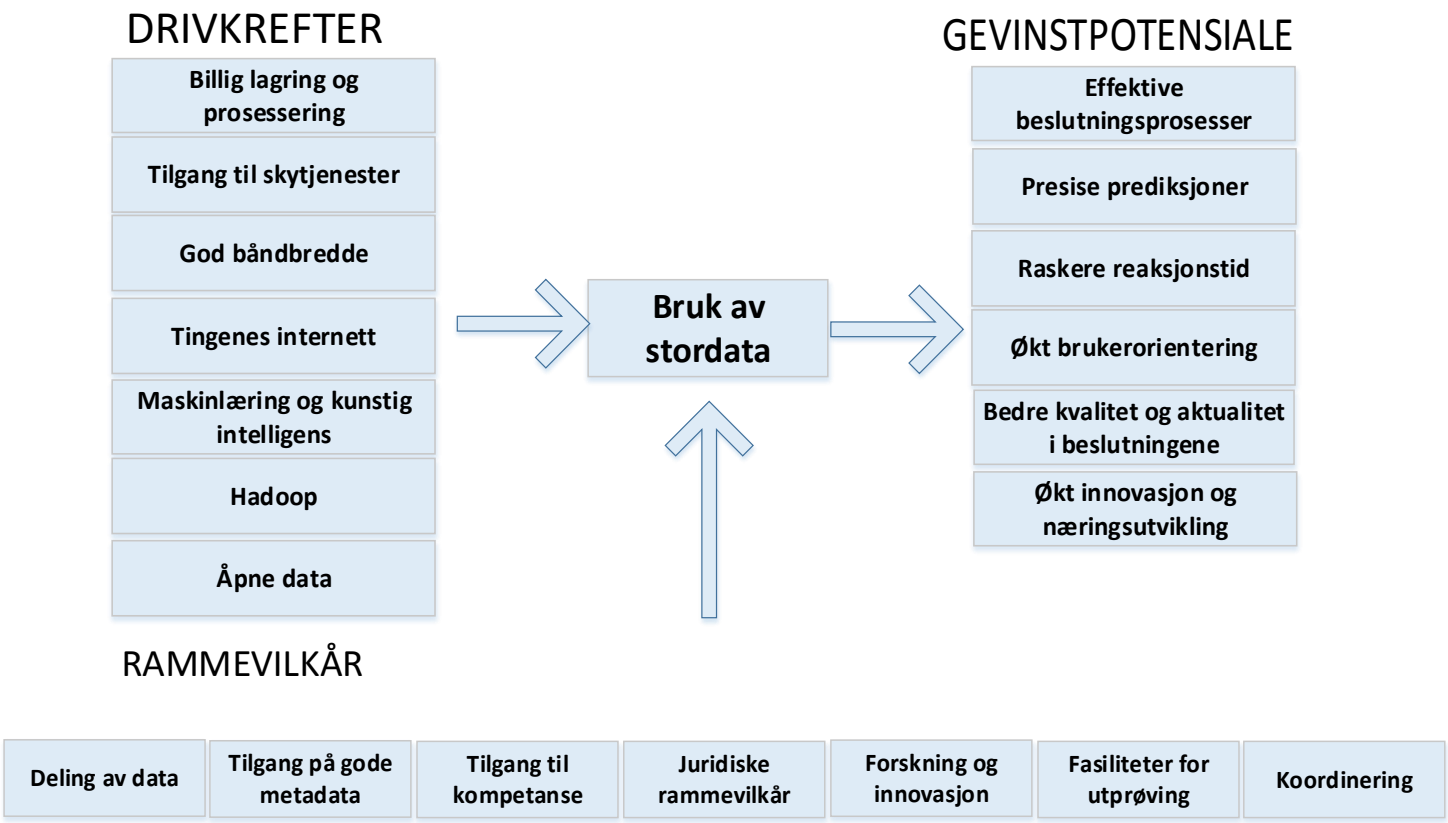
For at markedsaktørene skal være med og utnytte potensialet knyttet til stordata i kontekst av offentlig sektor, må de etter vår vurdering se over og remodellere de eksisterende forretningsmodellene. Det forretningsmessige potensialet kan være svært stort for de aktører som har skaffet seg tilgang til de største datamengdene og som har de mest effektive, skalerbare og robuste algoritmene. Utviklingen på området forutsetter foruten nye forretningsmodeller også en videreutvikling av ulike typer infrastruktur, programvare, analysemetoder og matematiske algoritmer som kan håndtere stordata. Vi har en del utvikling foran oss på disse områdene. Aktørbildet er sammensatt og mange vil være involvert i de løsninger som introduseres. Som samfunn vil vi i marsjen komme til å bli utfordret på de spørsmålene vi har omtalt ovenfor; personvern, datalagring, eierskap til data og transparente analyser.

[149] I Norge er det gjort tilgjengelig om lag 1600 datasett.

9 Anbefalinger

Vi tror det er viktig å ha det større utfordringsbildet som er beskrevet i kapittel 8 for øyet når det utformes nasjonal politikk. Skal man ta økosystem-tanken på alvor så er det viktig å se alle disse forholdene i sammenheng. En politikk på området bør følges opp av en strategi som har som mål å gjennomføre en poltikk som både har perspektiv på data som felles ressurs og inneholder målrettede tiltak for å utnytte ressursen og tiltak for de områdene som er problematisk, slik de for eksempel har gjort i Storbritannia og USA. En felles strategi og handlingsplan som gjennomføres med god koordinering og styring kan også sikre at tiltakene som treffes virker positivt i samme retning.

Figur 7 viser sammenhengen mellom drivkreftene for bruk av stordata (kapittel 4), gevinstpotensialet (kapittel 7) og hvilke rammevilkår som kan påvirke bruken av stordata slik at det blir større muligheter for å ta ut gevinstene i offentlig sektor (kapittel 8 og 9).



Figur 7 Forholdet mellom drivkrefter for bruk av stordata, gevinstpotensialet og rammevilkår.

I tillegg til at rammevilkårene tilrettelegges for bruk av stordata så kan virksomhetene selv treffe en del tiltak, de viktigste læringspunktene som vi kan trekke ut av andre virksomheters arbeid med stordata gjengir vi i punkt 9.1.

9.1 Læringspunkter for offentlige virksomheter

Basert på denne kartleggingen har vi identifisert læringspunkter hos ulike offentlige virksomheter som har tatt i bruk stordata og som andre kan dra nytte av. De to første momentene, god informasjonsforvaltning og deling av data, kan sees som forutsetninger for å

kunne få god nytte av å ta i bruk stordata. Mens utprøving og pilotering og strategisk forankring bør sees på som steg på veien mot innføring.

God informasjonsforvaltning

De virksomhetene som har tatt i bruk eller tester ut bruk av stordata-teknologi på en effektiv måte arbeider generelt godt med forvaltning av egne data. Dette betyr at dataene er beskrevet med gode metadata og er enkelt tilgjengelig. God forvaltning av data krever at metadata er underlagt standardiserte begrepsdefinisjoner, kodeverk og regelverk, og at populasjon og ikke minst datamodeller er kjent. Egen og andres bruk av data i nye sammenhenger forenkles betraktelig gjennom gode metadata.

Del egne datasett

Flere peker på behovene for tilgang til datasett som andre virksomheter forvalter. Dette gjelder særlig datasett som ikke er åpent tilgjengelig, men uten at det foreligger noen åpenbare grunner til dette. Samtidig sitter de samme virksomhetene på egne datasett som burde vært åpnet for maskinell viderebruk. Her er det visse nettverkseffekter. Det å tilby egne datasett gjennom data.norge.no kan også medføre at andre finner at de kan dele egne data.

Vi anbefaler alle offentlige virksomheter å vurdere hvilke data som kan åpnes for viderebruk, i sammenheng med arbeidet med bedre forvaltning av egne data. I tillegg bør virksomhetene ha oversikt over hvilken grad av sikring av de ulike dataene som er tilfredsstillende. I den sammenheng vil de også kunne ta stilling til hvem data kan deles med samtidig som personvern og andre hensyn ivaretas. Det er viktig at virksomhetene i større grad er i stand til å forstå verdien av egne data – også for omgivelsene – og gjør risikovurderinger og analyser for å vurdere hva som er tilfredsstillende informasjonssikkerhet.

Utprøving og pilotering

Bruk av stordata-teknologi bør prøves ut i begrenset omfang før videre satsing besluttes. Bedre forståelse av dataenes kvalitet og kompleksitet, identifisering av tekniske og ressursmessige hindringer, og det å avdekke mulighetene til å få verdi av egne og andres datasett er viktige læringseffekter av slik innledende forsøksvirksomhet.

Strategisk forankring

De som har kommet lengst på området har også god strategisk forankring av bruken av stordata i virksomheten. En datadrevet virksomhet utarbeider strategier og handlingsplaner for satsing på data, slik det er gjort i f.eks. Statens vegvesen. I slike virksomheter sees kjerneoppgavene og utnyttelsen av data i nær sammenheng. Etter hvert som utprøving viser resultater er det viktig å treffe og dokumentere strategiske valg for den videre satsingen på stordata. Strategiene må adressere hele prosessen fra innsamling til utnyttelse og deling av data.

9.2 Tilgang til data og metadata med god kvalitet

Som redegjort for over er tilgang til data med god kvalitet en forutsetning for å ta i bruk stordatateknologi. Tilgang til data fra andres og egen virksomhet dreier seg i stor grad om å ha oversikt over hvilke datasett som finnes, hvilke dataelement som inngår og hvilke begrep

disse representerer, i tillegg dreier det seg om informasjon om formål, hjemmel i regelverk, sammenheng mellom de ulike dataelementene og avklaringer om datakvalitet.

De som har tatt i bruk stordata gir tydelige signal om at det er behov for tilgang til flere datasett og at det synes å mangle en kultur for deling av data i offentlig forvaltning. Flere tar til orde for at det tas i bruk ulike virkemidler for å få flere virksomheter til å legge sine data til rette for viderebruk.

Fordi åpne data er så vidt viktig også for stordata-anvendelser, er det også nærliggende å peke på behovet for en fortsatt og forsterket satsing på arbeidet med åpne data i offentlig sektor. Det å arbeide for å utvikle en kultur for deling i det offentlige er en del av dette. Stordata medfører en tilleggsdimensjon i vurderingen av nytte i arbeidet med publisering av åpne data fra virksomheter og myndighetsorganer.

I tillegg er det behov for at det offentlige bedrer egen informasjonsforvaltning, ved at de blant annet beskriver sine metadata bedre. Det er særlig viktig å ta hensyn til den dekoblingen av ansvaret for datainnsamling og ansvaret for bruk som kjennetegner den datadrevne økonomien, ved å sørge for gode beskrivelsesdata og kvalitets- og risikovurderinger i forbindelse med publiseringen av nye datasett. Det er blant annet viktig å sikre at data som det hefter sikkerhetsbegrensninger eller forretningshemmeligheter ved, ikke gjøres tilgjengelig for analyser og sammenstillinger som på en eller annen måte kan kompromittere den skjermingen som må være der.

Bedre tilgjengeliggjøring av ontologier (oversikt over begrepsapparat) innen ulike domener vil også være et viktig tiltak for å øke mulighetene for å utnytte data på tvers av samfunnssektorer og fagdomener. Når det gjelder metadata generelt er det relevant å vurdere om disse dataene bør beskrives på en standardisert måte slik at gjenbruksverdien øker.

Flere tiltak har en klart pedagogisk karakter; dvs. mye handler om å gi den riktige veiledning til virksomhetene, slik at de på ulike måter legger til rette for at deres data (og metadata) blir tilgjengelige i maskinlesbare formater. Det er naturlig i denne sammenheng å vise til anbefalingene i Difi-rapporten *Informasjonsforvaltning i offentlig sektor*[150] fra 2013.

Videre synes det å være generelt viktig for tilliten til analysene at *usikkerheter* om datakvalitet, validitet og tilmed relevans framkommer i de aktuelle stordata-baserte beslutningsgrunnlagene – dvs. at det legges opp til en gjennomført transparens i og publisering av de forbehold og usikkerheter som vil ligge innebygd i presentasjonene. Til dels handler dette også om digital kompetanse og evne til medie- og kildekritikk hos sluttbrukerne. Forskning og ny teknologi kan bidra til en løsning på dette med manglende kvalitet i datasett. Bl.a. ved Universitetet i Tromsø[151] forskes det på såkalt informasjonsteoretisk maskinlæring og tetting av hull i store datasett.

Anbefalinger:

- **Utarbeide krav til, og anbefale standarder for beskrivelse av metadata om offentlige datasett.**

[150] (Difi, Informasjonsforvaltning i offentlig sektor, 2013)
[151] (Universitetet i Tromsø, 2014)

Sterkere vektlegging av kravet til metadata om datasett i offentlig sektor, inkludert forskning og andre aktører som utfører oppdrag for det offentlige. I tillegg er det behov for å standardisere måten metadata beskrives på og at alle metadata om offentlige datasett, åpne og lukkede, i størst mulig grad er publisert og åpent tilgjengelig.

For å øke tilgangen til informasjon om dataene i ulike datasett kan det være på sin plass å standardisere beskrivelsene av metadata.

- **Økt satsing på veiledning av offentlige virksomheter i informasjonsforvaltning**

Mye av mangelen på delingskultur synes å dreie seg om kompetanse og modenhet. Vi anbefaler derfor at det etableres et veiledningsopplegg for arbeidet med data i offentlige virksomheter.

- **Økt satsing på tilgjengeliggjøring av åpne data for viderebruk i og utenfor forvaltningen**

Satsingen på tilgjengeliggjøring av åpne offentlige data for viderebruk bør styrkes. I tillegg til veiledning, krav og pålegg så bør økonomiske virkemidler vurderes. Det bør vurdere om kostnadene som virksomheter har ved å tilgjengeliggjøre data ikke lenger legges på den som ønsker tilgang til dataene eller på dataforvalter selv, men at dette for eksempel sentralfinansieres.

Det er her viktig å understreke at det også er vår anbefaling at arbeidet med informasjonsforvaltning generelt og metadata og åpne data spesielt sees i sammenheng med arbeidet med personvern og tilstrekkelig informasjonssikkerhet.

9.3 Infrastruktur og teknologi

En type tiltak som er relevant her er å sikre allmenn tilgang til de infrastrukturer som er nødvendige for stordata-anvendelser i både offentlig sektor og næringslivet. Det er fortsatt behov for å stimulere utbyggingen av høyhastighets bredbånd, både fast og mobilt, særlig i grisgrendte strøk hvor tilgangen fortsatt er ujevn[152].

Fellesføringene i digitaliseringsrundskrivet for 2016[153] stiller krav til statlige virksomheter om å vurdere skytjenester. Dette er en start og det blir viktig å evaluere i hvilken grad dette følges opp. Det å etablere skytjenester for offentlig sektor kan være et annet relevant tiltak. Flere samfunnssektorer vurderer å etablere skytjenester for brukere i egen sektor. En felles, sikker og robust offentlig sky med datakataloger, ontologier, data og datanalyse-tjenester ville naturligvis gjort terskelen for stordataanvendelser i offentlig sektor lavere.

De skytjenester virksomhetene gjør seg avhengig av, må ha en tilgjengelighet som tillater å utvikle rutiner og beslutningssystemer med basis i dem – i tillegg til å tilfredsstille personvern- og andre viktige hensyn. Tilsvarende bør nye og robuste forretningsmodeller i privat sektor kunne baseres på sky-baserte tjenester og dataanalyse som en integrert del av virksomhetens forretningslag.

Selv om mye er basert på åpen kildekode er en del av stordata-teknologiene som vi har påpekt fortsatt til dels eksklusive, lukkede og proprietære. Her er det betimelig å utvikle en

[152] (Nasjonal Kommunikasjonsmyndighet, 2015)
[153] (Regjeringen, 2015)

offentlig innkjøpspolitikk på området. Når de større virksomhetene anskaffer stordata-løsninger er det viktig at det stilles krav som utvikler markedet i retning av åpne og felles standarder, og åpne og dokumenterte modeller og teknologier.

Det vil dessuten være behov for leverandørnøytral veiledning på teknologiområdet, om produkter, tjenester, avhengigheter, avtaleformer, teknologitrender etc. – slik at terskelen for virksomhetene som vurderer mulighetene for å ta i bruk stordata blir noe lavere enn i dag. Målgruppen for slik veiledning vil primært være offentlige virksomheter, men privat sektor særlig i SMB-segmentet vil ha mye av de samme behovene. Dette bør etter vårt skjønn kunne løses av konsulentmarkedet eller ved at det etableres offentlige nettverk og samarbeid.

Anbefalinger:

- **Utrede bruk av skytjenester i offentlig sektor og etablere policy**

Det bør vurderes om det er grunnlag for å etablere felles offentlige skytjenester eller om slike tjenester skal anskaffes i markedet. Det bør også utredes i hvilken grad skytjenester i markedet som kan benyttes av det offentlige i stordatasammenheng, avhengig av om dataene er å betrakte som åpne eller lukkede.

- **Videreutvikle norsk geoinfrastruktur**

Som beskrevet i kapittel 6.8 er sanntidsdata ofte stedbundne. Mange virksomheter påpeker viktigheten av tilgang til geografisk informasjon for å berike andre data og til visualisering.

9.4 Juridiske rammevilkår

EU-kommisjonen har pekt på at stordata-teknologien må settes inn i en større kontekst i form av et data-økosystem for å få verdi[154]. Offentlig sektor må tilgjengeliggjøre datasett og legge rammevilkårene til rette for at data skal kunne viderebrukes, og også selv utnytte data på tvers av forvaltningen. Det offentlige må også legge til rette for innovasjon og gode muligheter for å starte opp virksomhet på området.

9.4.1 Personvern

Utfordringene vi ser på personvernområdet kan delvis håndteres med pedagogiske virkemidler, men må antagelig også håndteres regulatorisk. Dette kan også inkludere utviklingen av teknologier som understøtter en ønsket praksis. Det vil bl.a. kunne være aktuelt med ulike metoder for automatisert kontroll med tilgangen til og bruken av sensitive dataelementer.

Det kan være aktuelt å vurdere regulatoriske bestemmelser om distribusjon og bruk av analyseresultater og sammenstillinger som kan identifisere person. En kan også vurdere om det bør stilles strengere krav om teknologier og algoritmer (såkalt «privacy-by-design»[155]) som spesifikt ivaretar personvernhensynene i stordata-anvendelsene.

Et generelt krav om transparens i analysemodellene vil også være med på å bygge tillit til og selvkorrigerende mekanismer inn i anvendelsene. Det er også aktuelt å etablere egnede klageinstitutter for eksempel i form av en uavhengig nemnd som vurderer etiske spørsmål

[154] (Marta Nagy-Rothengass, EU, 2014)
[155] (Wikipedia, Privacy by design, 2015)

ved bruk av stordata-teknologi på personopplysninger[156]. En slik nemnd bør i så fall kunne vurdere saker før de er kommet så langt at Datatilsynet eller Personvernnemnda skal vurdere saken. For eksempel kan man tenke seg at nemnda uttaler seg til virksomheter som ønsker råd allerede på idéstadiet.

På den annen side mener vi det også kan være grunn til å vurdere om dagens utbredte bruk av nokså spesifikke og generelt restriktive samtykkeerklæringer ved innsamling av data bør endres. I mange sammenhenger kan man se for seg at personopplysninger av mindre sensitiv karakter (slik som alder, yrke, bostedskommune etc.) kan vurderes som ufarlig å viderebruke av den det gjelder. Det bør utredes om samtykkeerklæringer i større grad kan gjøres åpne og hjemle framtidige, ikke-spesifiserte analyse- og forskningsbehov, i det minste når de aktuelle data som samles inn i utgangspunktet ikke omfatter sensitive opplysninger. Ved innsamling av nye personopplysninger bør det være en fast vurdering om datasettet kan ha framtidig verdi i stordatasammenheng, og at samtykkeformuleringene utformes deretter.

Dette igjen vil måtte følges opp med gode innsynsordninger der enkeltpersoner, enten i rollen som innbygger eller næringsdrivende, kan finne informasjon om hvor hans personopplysninger behandles av det offentlige. Dette er et rimelig krav for å ivareta rettssikkerheten til den enkelte og en transparent offentlig sektor.

Anbefalinger:

- **Vurdere om samtykkeinstituttet bør endres**

Vi anbefaler at det utredes hvilke følger et forenklet samtykkeinstitutt får, særlig med hensyn til samtykke for bruk av personopplysninger som ikke bør betraktes som særlig sensitive.

- **Utvikle bedre innsynsløsninger for stordataanvendelser**

For å styrke personvernet anbefaler vi at det etableres løsninger som gir innbyggere og næringsdrivende innsyn i hvilke data om dem som viderebrukes, av hvem og til hva. Vi anbefaler at offentlige virksomheter etablerer oversikt over hvilke datasett som viderebrukes av andre og hvilke datasett de viderebruker fra andre kilder.

9.4.2 Eiendomsrett

Selv om man ikke kommer utenom at eiendomsforhold må vurderes i hvert enkelt tilfelle er det et klart identifisert behov for at problemstillingen avklares tydeligere på policynivå og gjennom reguleringer.

Dette er også et område hvor det kan gis tydeligere føringer i veiledningsmateriell og gjennom andre pedagogiske tiltak.

Anbefaling:

- **Etablere policy for eiendomsrett til data og analyser**

Eierskap til analyser (inkludert visuelle framstillinger) basert på ulike datasett fra ulike kilder og virksomheter må avklares. Forholdet til innsynsretten nedfelt i offentlighets- og forvaltningslovgivningen må også avklares i denne sammenheng.

[156] (Gartner, 2013)

9.4.3 Andre juridiske utfordringer

Behovet for regelverkstilpasninger for å legge bedre til rette for stordataanvendelser og dataanalyse som verktøy i forvaltningen, må vurderes nærmere. På et generelt nivå er det krevende å se for seg at retten til innsyn og kravene til transparens uten videre kan ivaretas på noen enkel måte når stordata tas i bruk i sammenhenger knyttet til forvaltningsvedtak som treffer enkeltindivider.

I veiledningsmateriell som presenterer rettigheter og plikter innenfor områder hvor stordata er tatt i bruk, må det redegjøres på en enkel måte hvordan transparens og innsyn for den enkelte er sikret.

Anbefaling:

- **Utrede hvordan algoritmer som benyttes til analyser kan gjøres åpent tilgjengelig for de som har behov**

Selv om det ofte vil være krevende for lekfolk å forstå algoritmene og hvordan de benyttes anbefaler vi likevel at det utredes om det kan stilles et generelt krav om at disse alltid skal være tilgjengelige for innsyn for dem som ønsker det.

9.5 Kompetanse

Sett fra virksomhetene kommer ikke stordata av seg selv. En bred implementering av stordata krever at de som skal ta i bruk teknologien har tilstrekkelig datavitenskaplig kompetanse og kjennskap til aktuelle analyse- og framstillingsverktøy – enten konteksten er beslutningssystemer i offentlig sektor, offentlig og privat tjenesteyting, eller innen salg og markedsføring. Kompetanse i virksomhetene er med andre ord en vesentlig forutsetning for å lykkes med å realisere potensialet. Dette understrekes i mange av de utredninger og strategidokumenter vi har gjennomgått i forbindelse med denne kartleggingen.

Det bør vurderes å styrke studietilbudet i dataanalyse innen informatikk eller informasjonsvitenskap, og å innarbeide dataanalyse som fag i etablerte studier hvor dette kan være relevant (jf. det som er sagt om kombinasjonskompetanser over).

Vi ser at det vil være krevende å bygge opp tilstrekkelig kompetanse i forvaltningen, til at man skal kunne ta i bruk stordata på en effektiv måte. Det er derfor aktuelt å bygge opp et sentralt statlig kompetansemiljø som virksomhetene kan trekke på i sine bestrebelser – både i vurderinger av teknologier, utforming av kravspesifikasjoner, vurdering av personvernhensyn (f.eks. teknikker for anonymisering), risiko- og sårbarhetsvurderinger når det gjelder design av datasett, metadata og ontologier, planlegging av kompetansebehov, bruk av ulike analyseverktøy, god praksis i visualisering, innarbeiding i offentlig sektors rutiner og gevinstrealisering. Etablering av et sentralt kompetansesenter vil sannsynligvis være meget kostnadsbesparende i forhold til om hver sektor eller forvaltningsenhet bygger opp egen kompetanse. Et slikt tiltak vil være i tråd med det mange andre land gjør på området.

Offentlige virksomheter må styrke sin kompetanse om mulighetene stordata gir. Det er behov for kompetanse om teknologien, de etiske og juridiske vurderingene som må gjøres, om krav til forvaltning av egen og andres informasjon og kompetanse om selve analysemetodene og om visualisering. Virksomhetene må settes i stand til å bestille godt og til å følge opp sine

leverandører på en god måte. Det er behov for et kompetansemiljø som kan bidra med kunnskapsdeling, kurs og veiledninger for hele offentlig sektor.

Spørsmålet på policynivå er her hvordan man kan påvirke at virksomhetene tar inn stordata som en muliggjører og omstillingsagent for en mer datadrevet økonomi og en kvalitativt bedre offentlig tjenesteyting. Et åpenbart tiltak er å innarbeide stordata som et av de teknologiområder som topp- og mellomledere i offentlig sektor skoleres i, gjennom de kompetansetilbud som finnes for denne gruppen.

Et annet tiltak er å demonstrere hvordan stordata kan være en muliggjørende teknologi for en radikalt mer effektiv offentlig sektor. Dersom en kunne lansert en helt ny måte å håndtere en utfordring på et «ufarlig» og uproblematisk område i forvaltningen, men som innebar dataanalyse i sanntid – som erstatning for en større «tidstyv», ville dette ha en god pedagogisk effekt.

Det å kunne teste ut sammenstilling og bruk av ulike datasett fra ulike virksomheter i et trygt miljø er noe som bør kunne tilbys gjennom at det etableres et offentlig testmiljø eller et offentlig-privat samarbeid. Det er da viktig at rammevilkårene for dette settes slik at lovgivningen ivaretas på en god måte.

Det er for få uteksaminerte studenter som har kompetanse innen dataanalyse (herunder stordata). Dette gjelder både innen IKT-utdanningene og innen andre fag hvor analysekompetanse vil være en viktig kompetanse i arbeidslivet. Utdanningspolitikken bør derfor i større grad bidra til å sikre kompetanse om innsamling, forvaltning og utnyttelse av data på ulike nivå.

Det satses allerede en del på forskning innenfor dette området, men det er fortsatt behov for slik forskning. Her må både nasjonale og internasjonale forskningsmidler og -miljøer utnyttes. Siden dette allerede er dagens politikk setter vi ikke dette opp som eget forslag til tiltak.

Anbefalinger:

- **Vurdere å etablere testmiljø for utprøving av stordata-teknologi**

Flere virksomheter har tatt til orde for at det er behov for miljø der de kan teste ut bruk av stordata på egne og andres datasett. Mange virksomheter har ikke muligheter til å etablere slike miljø på egenhånd og mener at dette bør etableres som samarbeid mellom flere virksomheter og gjerne i samarbeid med private virksomheter og leverandører. Et alternativ er å etablere et slikt miljø i forbindelse med at det etableres en felles offentlig sky-tjeneste.

- **Etablere sentralt kompetansesenter innen stordata og informasjonsforvaltning**

Et sentralt kompetansemiljø vil ha som oppgave å veilede de offentlige virksomhetene slik at de settes i stand til å vurdere stordata i sin virksomhet. Dessuten kan dette kompetansesenteret bidra med veiledning på de områdene som vi nevner tidligere.

I tillegg bør stordata inngå som del av opplæring i digitalisering på strategisk nivå. Vi anbefaler også at et eventuelt kompetansesenter ser på mulighetene for å etablere demonstratorer i form av eksempelprosjekt som kan være til inspirasjon for andre virksomheter.

- **Økt satsing på stordata i ulike studietilbud**

Med et så tydelig signal om mangelen på tilgjengelig kompetanse er det åpenbart for oss at studietilbudene innen stordata må styrkes. Særlig gjelder dette innen datavitenskap og dataanalyse.

9.6 Markedet

I tillegg til de mulige tiltak som allerede er pekt på, er det viktig at en vurderer hvordan det offentlige i større grad kan påvirke helheten. I dette ligger å vurdere aktuelle former for offentlig-privat samarbeid, hvordan det kan skapes forutsigbarhet og et velfungerende marked, med tilstrekkelig kunnskap både på tilbuds- og etterspørselssiden, med muligheter også for private innovatører, samtidig som behovet for transparens, åpenhet, viderebruk og andre viktige hensyn er ivaretatt. Det handler til syvende og sist om hvordan det offentlige kan bidra til framveksten av et velfungerende data-økosystem.

Siden stordata er ny teknologi og fordi det kreves ulike typer kompetanse langs de ulike delene av verdikjeden er det også potensiale for nyetableringer og vekst i dette markedet. Det bør vurderes å etablere støtteordninger som er målrettet mot markedet for datautnyttelse. Oppstartstøtte og støtte til forskning og utviklingsarbeid vil kunne gi effekt.

Anbefaling:

- **Vurdere offentlig-privat samarbeid om innovasjonsprosjekter**

Det bør vurderes nærmere hvordan man kan etablere ordninger for samarbeid mellom privat og offentlig sektor, for å skape mer innovasjon innen stordata.

9.7 Koordinert gjennomføring

Det er et behov for å se politikken og tiltakene på stordataområdet og andre relevante områder som åpne data, informasjonsforvaltning (metadata) og informasjonssikkerhet og personvern i sammenheng. Dette kan etter vårt skjønn best gjøres ved at det etableres en helhetlig strategi på området. De fleste av tiltakene vi har nevnt bør inngå i en slik strategi.

Flere virksomheter peker på behovet for tverrsektorielt påtrykk og koordinering av arbeidet med å utnytte data som ressurs i offentlig forvaltning, dette gjelder også for bruk av stordata. Påtrykket bør skje på vanlig måte i linjen, noe som medfører behov for koordinering og påtrykk på departementsnivå.

Uansett om det utarbeides strategier eller ei, anbefaler vi også at ansvaret for å gjennomføre de ulike tiltakene koordineres for å sikre en mest mulig helhetlig og smidig gjennomføring med lav risiko. Oppgavene kan gjennomføres av ulike virksomheter, men det er viktig at det overordnede ansvaret og styringen av et eventuelt gjennomføringsprosjekt tildeles en virksomhet. Her er det naturlig å peke på Difi, men i så fall med bred forankring i forvaltningen for eksempel gjennom Skate og egen referansegruppe på lavere nivå.

Flere av de virksomhetene vi har snakket med mener at det er behov for sentral koordinering på området, og at Difi bør få en tydeligere rolle i dette.

9.8 Oppsummering av anbefalte tiltak

Oppsummert anbefaler vi at følgende tiltak gjennomføres som koordinerte aktiviteter – fortrinnsvis i et eget prosjekt eller program:

Organisatoriske tiltak:

- **Etablere sentralt kompetansesenter innen stordata og informasjonsforvaltning**

Pedagogiske tiltak:

- **Økt satsing på veiledning av offentlige virksomheter i informasjonsforvaltning**
- **Økt satsing på tilgjengeliggjøring av åpne data for viderebruk i og utenfor forvaltningen**
- **Økt satsing på stordata i ulike studietilbud**

Tekniske tiltak:

- **Utarbeide krav til, og anbefale standarder for beskrivelse av metadata**
- **Utrede bruk av skytjenester i offentlig sektor og etablere policy**
- **Utvikle bedre innsynsløsninger for stordataanvendelser**
- **Vurdere å etablere testmiljø for utprøving av stordata-teknologi**

Regulatoriske tiltak:

- **Utrede hvordan algoritmer som benyttes til analyser kan gjøres åpent tilgjengelig**
- **Utrede og etablere policy for eiendomsrett til data og analyser**
- **Vurdere om samtykkeinstituttet bør endres**

Økonomiske tiltak:

- **Vurdere offentlig-privat samarbeid om innovasjonsprosjekter**

Vedlegg 1: Tre utvalgte prosjekt

1 Helsedirektoratet: Prehospital strategisk satsing

I 2014 håndterte de akuttmedisinske kommunikasjonssentralene (AMK sentralene) 880 214 hendelser, noe som resulterte i 605 300 ambulanseoppdrag. Totalt ble det håndtert 1 270 513 akuttmedisinske tiltak i Norge. Det å koordinere ressurser (transport, medisinsk personale og mottaksapparatet) på en god måte er avgjørende for både liv, helse og for å sikre effektiv utnyttelse av tilgjengelige ressurser.

Den prehospitale strategiske satsingen er Helsedirektoratets satsing på å benytte data for å bedre akuttmedisinske prosesser. For å få til dette skal det utvikles et nasjonalt system for innsamling og bearbeidelse av data fra helsetjenestens AMK-sentraler og ambulansetjenesten. Disse settes sammen med store datasett fra andre offentlige kilder (SSB, Folkeregisteret, Vegdirektoratet og Politidirektoratet) og benyttes til å modellere logistikk mellom boligområder og ulike institusjoner/tjenester (ambulansestasjoner, politistasjoner, brannstasjoner, sykehus, legevakt, fastleger, Røde Kors m.fl.).

Helsedirektoratet har gjennom arbeidet tatt initiativ til samarbeid med Sverige, Finland, Danmark, Island og USA om bruk og videreutvikling av løsningene. Det finnes allerede nasjonale installasjoner i Danmark og USA. Samlet har Norden og USA opp mot 35 millioner ambulanseoppdrag per år, noe som gir data nok til å gjøre prediktive analyser.

Et av målene er å kunne få til dynamisk flåtestyring av ambulanser. Ved å basere seg på prediksjoner vil flåten kunne plasseres og være tilgjengelig der det sannsynligvis er størst behov og dermed kunne redusere responstiden ved uttrykning. Dynamisk flåtestyring av ambulanser er etablert i en rekke land som for eksempel USA, Canada, Thailand, Israel, Australia, England og Nederland. I disse landene benyttes historiske data til å predikere sannsynlighet for akuttmedisinske situasjoner (og dermed utrykninger) i geografisk begrensede områder.

Gjennom de ulike datakildene som benyttes i de prehospitale løsningene håndteres over 250 millioner rader («records») med data fra en rekke kilder. Det har derfor vært nødvendig å satse på stordata teknologi i stedet for mer tradisjonell BI-teknologi.

Hovedelementene i arbeidet er:

- Utvikling av en nasjonal informasjonsmodell for ambulansetjenesten
- Standardisert og harmonisert kodeverk og begrepsbruk
- Nasjonalt system for innsamling og publisering av data
- Internasjonalt samarbeid med de nordiske land og USA

Arbeidet er delt inn fem delprosjekt. Delprosjektene «Oppfølging AMK», «IKT løsning» og «Web publisering» eies av Avdeling norsk pasientregister. Delprosjektene «Nasjonalt kodeverk» og «Nordisk samarbeid» eies av Avdeling medisinske nødmeldetjeneste.

Data brukes til virksomhetsutvikling og forskning, og vil være et viktig grunnlag i arbeidet med kvalitetsforbedring, organisering og pasientsikkerhet. Løsningene brukes til beslutningsstøtte, planlegging, beredskap, forskning og analyse – og blir tilgjengelige for publikum.

Bakgrunn

Befolkningsøkning, høyere levealder og større andel eldre øker behovet for helsetjenester, også akuttmedisinske tjenester utenfor sykehus. Veksten i antall henvendelser til medisinsk nødnummer har økt de siste årene, og har vært større enn befolkningsutviklingen. Sentralisering og spesialisering både i spesialisthelsetjenesten og kommunehelsetjenesten, inkludert legevakt, øker også behovet for prehospitale tjenester. Behandling flyttes i økende grad ut av sykehus og det er i dag bredere mulighet for diagnostisering utenfor sykehus. Dette gir også økt kompetansebehov i de prehospitale tjenestene.

Samtidig viser tall fra Norsk pasientregister stor vekst i akuttmedisinske tjenester, og at akuttmedisin målt i antall pasientkontakter er det største medisinske fagområdet i spesialisthelsetjenesten. Samtidig er det lite forskning på fagområdet og små fagmiljøer. En helhetlig elektronisk infrastruktur for kommunikasjon og informasjonsdeling vurderes som nødvendig for en effektiv, sikker og kvalitativt bedre pasientbehandling, økt samfunnssikkerhet og beredskap. Dette er bakgrunnen for at Helsedirektoratets ledelse i mai 2014 behandlet sak om etablering av Prehospital strategisk satsing.

1.1 Løsning for prehospital strategisk satsing

Løsningen er flyttet fra utviklingsmiljøet hos NTNU til ordinær drift og produksjon hos Helsedirektoratet i november 2015.

Figur 8: Tjenester i prosjektet - hold inne CTRL tasten og klikk på de ulike boksene for å få opp løsningen.

Området for prediksjoner er under utvikling og testing og de første løsningene forventes i drift i løpet av første halvår i 2016.

1.2 Datagrunnlag

Løsningen er basert på data fra Norsk pasientregister som har data fra sykehusenes Akuttmedisinske informasjonssystem (AMIS) og samtaledata til nødtelefon 113 fra Telenor og Nødnett. Data fra AMIS oppdateres hvert tertial mens samtaledata oppdateres daglig.

Av helseeksterne kilder benyttes det åpne datasettet fra Nasjonal vegdatabank (vegvesenet)[157], matrikkelen fra Folkeregisteret, kartdata fra Statens kartverk, ulike demografiske datasett fra SSB, fastlegedata fra Min Fastlege (NAV), Politidirektoratets «Politianalysen 2016», med mer. Helsedirektoratets egne data av god kvalitet om ulykker, mens politiet har svært gode data om alvorlige ulykker og dødsfall.

For å forbedre løsningen ytterligere ønsker NPR å utvikle løsninger for daglig oppdatering av flere datakilder. Dette gir grunnlag for å benytte løsningen som et «ferskvare» planleggingsverktøy, overvåkningsverktøy og til å øke presisjon i prediktive løsninger.

Det er et savn at det ikke finnes EPJ i Ambulanser. Dette er til utprøving på et par steder, men svært begrenset i forhold til det nasjonale behovet. Prinsippet om at rapporteringsbyrden må oppveie nytteverdien står sterkt i helsesektoren, det jobbes derfor med å etablere en løsning som gir nytte både for ambulansene og for planleggingsformål. Målet er å utvikle kodeverk for ambulanse EPJ før det utvikles løsning. For andre pasientjournalløsninger så har situasjonen vært motsatt og medført kostnader når kodeverket standardiseres og implementeres i eksisterende løsninger.

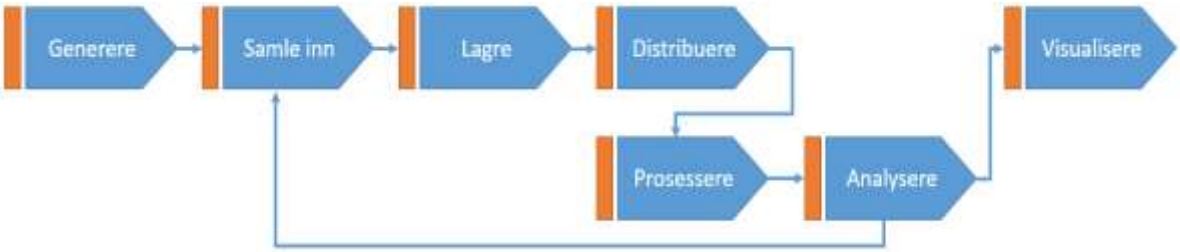
Dette prosjektet omfatter derfor å utarbeide både en kodestandard og en felles datamodell for «elektronisk ambulansejournal».

På sikt vil løsningen også hente data fra nytt kommunalt pasientregister som er under planlegging. Det er ønskelig at løsningen inneholder data fra hele den prehospitale kjeden.

1.3 Teknisk løsning

Stordata-løsningen til *prehospital strategisk satsing* er basert på verktøy utviklet som åpen kildekode, dette har vært et bevisst valg for å kunne dele på kildekoden med de andre samarbeidslandene og dermed dra gjensidig nytte av videreutviklingen. I tillegg har det vært viktig for prosjektet å satse på teknologi som er godt utbredt, per nå er det ifølge Google over 200 000 installasjoner av samme teknologi. I det følgende forklares de ulike teknologiene som benyttes i de forskjellige stegene i verdikjeden[158]:

[157] http://data.norge.no/data/statens-vegvesen/nasjonal-vegdatabank-api
[158] Beskrivelsen er basert på tekstlig innspill fra Prehospital strategisk satsing i Avdelig for norsk pasientregister



Figur 9: Verdikjede stordata

1. **Samle inn:** Data samles inn fra AMK-sentralene, pasientjournalene osv. og sendes som XML-filer til Norsk pasientregister. Et eget skript leser inn dataene fra XML filene som kommer inn til Norsk pasient register. Dataene blir så lest inn i SOLR[159] og tilrettelagt for prosessering og etter hvert søk.
2. **Lagre:** For å oppnå optimalisering med hensyn til løsningens responstid og fleksibilitet, blir enhetene i XML filen «flatpakket». For AMK data medfører dette at hierarkiske element (hendelse, oppdrag, henvendelse, osv) organiseres som en rad per element i en flat struktur.
3. **Distribuere:** SOLR klarer å håndtere datasett opp til 300 GB per server, mens ElasticSearch er designet for ubegrenset størrelse på datasettet. Begge tjenestene er bygget på samme grunnelement, så det er lett å oppskalere løsningene ved å skrive om funksjonene som i dag benytter SOLR til ElasticSearch. Beregning av percentiler kan være en tung prosess på store datasett, derfor benytter løsningen et enkelt system ved å fordele spørringer ut til flere ulike SOLR servere. At man enkelt kan legge til flere maskiner, eller bytte til ElasticSearch gjør dette til en både skalerbar og fleksibel løsning.
4. **Prosessere:** Det er laget en mellomtjeneste basert på tornado[160] servere (laget av Facebook) som gjør spørring opp mot SOLR og Elasticsearch og viser data i brukerens visualiserings dashbord (i JASON format[161]). Mellomtjenesten har også en modul som lager flate pivot tabeller som kan benyttes direkte i Excel, SPSS, Open Office, osv. I denne mellomtjenesten er det bygget inn en mekanisme som forhindrer at spørringer med mindre enn 5 sendes til dashboardet/bruker for å sikre sensitivitet og personvern i løsningen.
5. **Analysere/visualisere:** Analysene gjøres av sluttbruker direkte i dashbordene. Her kan man gjøre brukerstyrte analyser av de ulike datasettene. Dashbordene er skrevet i javascript og de ulike dashbordene bruker ulike åpne bibliotek for plotting av data.

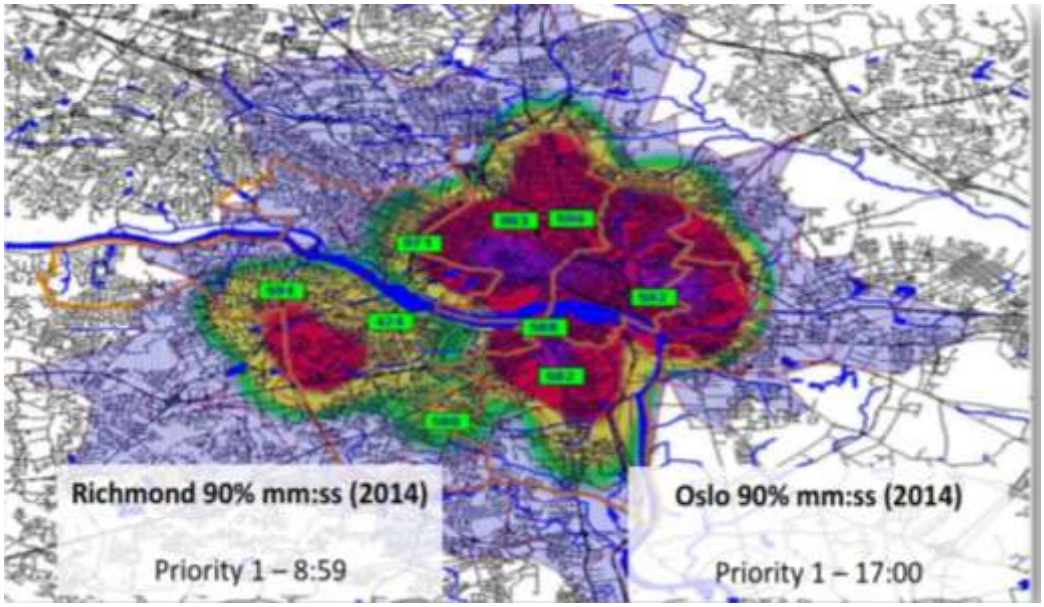
[159] Løsning for å klargjøre data for prosessering (Solr, 2015)
[160] https://github.com/tornadoweb/tornado
[161] JSON: JavaScript Object Notation, uttales er en enkel tekstbasert standard for datautveksling. Den er opprinnelig avledet fra JavaScript for å representere enkle datastrukturer. Standarden er imidlertid uavhengig av JavaScript eller andre programmeringsspråk. (Wikipedia, 2015)

1.4 Gevinster

For helsesektoren har løsningen allerede gitt gevinst gjennom at det er levert datagrunnlag til utarbeidelse av ny akuttmedisinforskrift (Helsedirektoratet, 2015). Utvalget undersøkte blant annet hva det vil kreve av legeressurser dersom det skulle gjennomføres et mål om legedeltakelse i 75 % av utrykninger med rød respons (førsteprioritet ambulanseutrykninger). Ved hjelp av den prehospitale løsningen fant man raskt ut at i 2014 ville det dreie seg om 143 367 oppdrag som ville kreve minst 70 legestillinger.

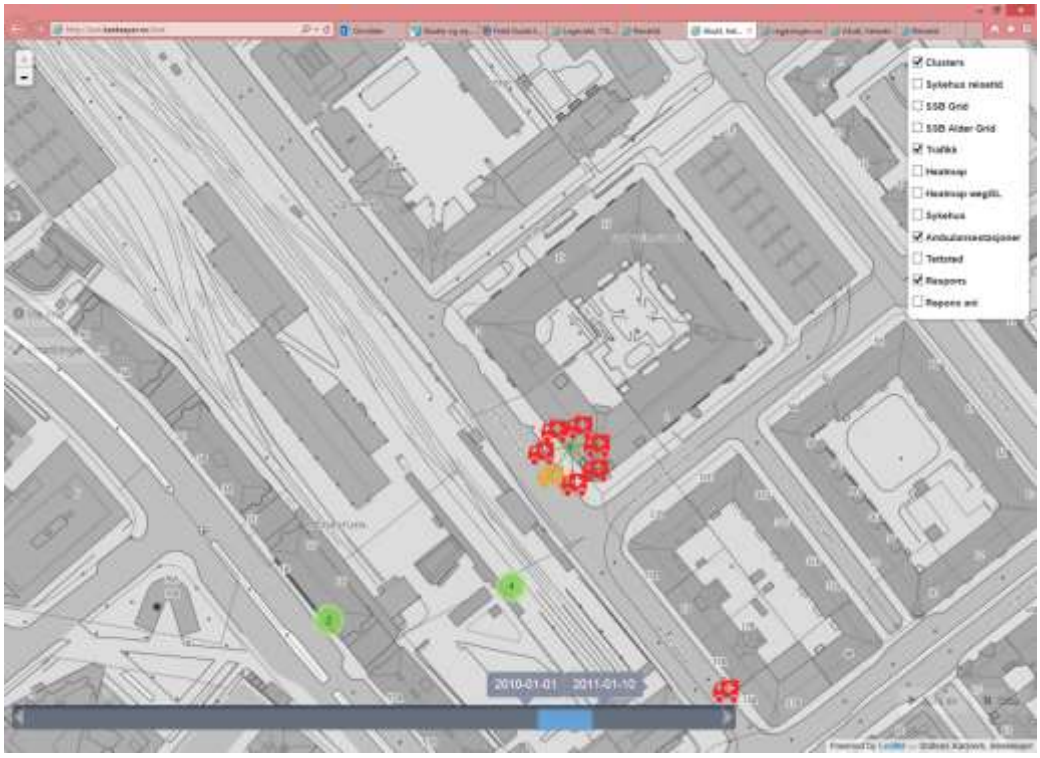
Løsningen gir et godt grunnlag for å planlegge og forbedre de prehospitale aktivitetene. For eksempel er det mulig å følge med på responstiden til alle akutte ambulanseutrykninger i hele landet. Man kan nå også sammenligne reisetid til sykehus i sentrale og grisgrendte strøk. De som bor i Trysil kommune må belage seg på gjennomsnittlig reisetid til sykehus på 1t og 45 minutter (95 prosentil), mens de som bor i Lørenskog har litt over 5 minutter reise med bil til sykehus (95 prosentil).

Gjennom det internasjonale samarbeidet legges det til rette for at Norge kan sammenligne seg med de andre landene. For eksempel viser dataene at utrykningstiden for ambulanser er nesten dobbelt så høy i Oslo som i Richmond, USA. Dette til tross for at byene har om lag samme befolkningstall, men langt flere utrykninger i Richmond enn i Oslo. Dette har Richmond oppnådd blant annet fordi de kan predikere hvor det vil bli hendelser og hvordan ressursene dermed kan prioriteres både i tid og på sted.



Figur 8 Utsnitt fra Richmond sin akuttmedisin løsning. Kilde: Helsedirektoratet.

Videre er det etablert et samarbeid med veimyndighetene og politiet, selv om trafikkulykker bare utgjør 1% av de prehospitale hendelsene. Statens vegvesen kan for eksempel se nærmere på trafikkavviklingen på steder hvor det ofte skjer ulykker. Dette er det lett å finne i denne løsningen. Illustrasjonen under viser trafikkulykker på hjørnet av Sørkedalsveien og Harald Hårfagres gate, i 2010 var det 7 utrykninger dit – hvorav seks var førsteprioritet.



Figur 9 Løsning for oversikt over ambulanseutrykninger, http://kart.beekeeper.no/kart

Med dette som grunnlag kan så veimyndighetene gå inn å vurdere om det bør treffes trafikksikringstiltak.

1.5 Utfordringer, løsninger og læring

Teknologi:

Det var tidlig klart at tradisjonelle «Business Intelligence» verktøy ikke kunne benyttes. For det første fordi det var behov for at sluttbrukere skulle kunne velge datakilder og å sette sammen data selv og få dette presentert i et meningsfylt grensesnitt. Brukerne kjenner selv sine problemstillinger best og disse vil variere fra brukergruppe til brukergruppe. Prosjektet har vurdert at tradisjonelle datavarehus er for statiske til å kunne oppnå den nødvendige variasjonen. En annen utfordring med tradisjonelle BI løsninger har vært responstid. Disse forholdene har medført at det har vært nødvendig å se på stordata teknologier.

Mellom annet på bakgrunn av dette valgte man SOLR og Elasticsearch som er bygget for stordata og som gir mulighet for sluttbrukerdefinerte søk med god responstid.

Informasjonssikkerhet og personvern:

For å lykkes er prosjektet utført utenfor forvaltningssonen til Helsedirektoratet. Sikkerhetsløsningene vurderes som for strenge til at man kunne gjennomføre utviklingsprosjektet innenfor forvaltningssonen, blant annet fordi det ble problematisk å hente data fra eksterne kilder.

For å omgå dette så er datagrunnlaget vurdert i forhold til om de er å betrakte som helseopplysninger. Det er vurdert at løsningen ikke behandler data som kan defineres som helseopplysninger. Prosjektet har dermed kunnet gjennomføre utviklingen utenfor forvaltningssonen til Helsedirektoratet.

Dagens sikkerhetsløsning er etablert for å beskytte informasjon mot uautorisert innsyn i, og endringer av data. Mens hele poenget med den prehospitale satsingen er å utnytte eksterne datakilder og gjøre data tilgjengelig for analyse i ulike sammenhenger så raskt som mulig. Anonymisering av data slik at de ikke inneholder personopplysninger har derfor vært eneste alternativ.

En annen viktig læring har vært at prosjektet har funnet det nødvendig å ha full kontroll på infrastrukturen som benyttes. Dette først og fremst siden det er et utviklingsprosjekt og dermed nødvendig å kunne avdekke feil i alle ledd.

For å unngå revers/indirekte identifisering basert på andre data enn personopplysninger så returnerer ikke løsningen datasett som har mindre enn fem treff. Prosjektet har erfart at det er viktig å vurdere aggregeringsnivå på dataene som for eksempel årstall, geografisk enhet som kommune, legevakt osv. Detaljerte data kan gjøre det mulig å identifisere enkeltpersoner, mens høyaggregerte data reduserer brukerstyrt visning. Dette må vurderes nøye i alle datakilder slik at en opprettholder både personvern og fleksible brukerløsninger.

For løsningene som er utviklet er personopplysninger i liten grad relevant, det er anonyme styringsdata som er interessant i denne sammenheng. Likevel stammer mange av dataene fra hendelser hvor enkeltpersoner har vært involvert.

Informasjonsstruktur

Siden informasjonsstruktur har vært en av hovedleveransene i prosjektet så har det vært viktig å ha kontroll på informasjonsmodellene (felles datamodeller og harmonisert kodeverk). Det oppleves som viktig at prosjektet eies av de samme avdelinger som har ansvaret for utvikling av informasjonsstrukturen.

Det er også viktig å ha god kjennskap til brukernes behov slik at utviklingen kan skje i tråd med dette. Dette kan være utfordrende, men løses gjennom ulike fora og dialog.

Kompetanse

Det kan være vanskelig å skaffe datavitenskap («data science») kompetanse. I tillegg er det viktig å ha kombinasjonskompetanse på både fag og datavitenskap.

Prehospital strategisk satsing prosjektet har løst dette ved at prosjektet har bygd egen kompetanse på datavitenskap. Dette er selvsagt sårbart på grunn av at fagmiljøet er lite, samtidig er det en erkjennelse at det generelt finnes få eksperter på området. Kombinasjonskompetansen mellom helsefag og datavitenskap er spesiell og uansett vanskelig å anskaffe i markedet fordi det dreier seg om ulike typer helsefag.

Datakvalitet

Datasettene som kommer inn til løsningen er i all hovedsak XML-filer og den eneste direkte kvalitetskontrollen som gjøres er på duplikater. Men den viktigste kontrollen gjøres av brukerne selv, når de får visuell tilgang til data gjennom ulike dashboard – da kan de kvalitetssjekke data selv og si fra om feil de oppdager.

1.6 Veien videre

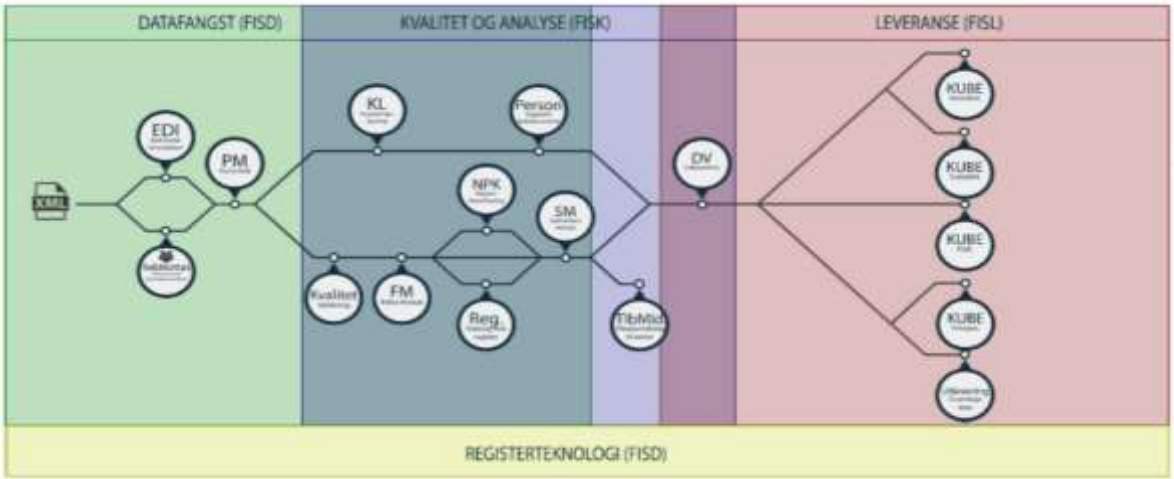
Løsningen ble satt ble satt i ordinær drift i Helsedirektoratet i november i år og er med det kommet over i en forvaltningsfase. Prosjektet vil fortsette med å utvikle løsningen og koble på nye og forbedrede datasett etter hvert som de er tilgjengelige. Dette gjør at teknologien gjøres tilgjengelig for bruk i andre sammenhenger.

Løsningen er allerede installert og tatt i bruk i Nordjylland (Danmark) og USA. I løpet av desember 2015 blir den også installert i København. Videre er det planlagt installasjon i Finland første kvartal 2016. Løsningene vil derfor bli forvaltet og videreutviklet under en åpen kildekodelisens, noe som vil medføre at løsningen kan forbedres i fellesskap med de ulike landene som deltar i samarbeidet.

1.7 Om Norsk pasientregister

Norsk pasientregister inneholder data fra spesialisthelsetjenesten. Registeret ble etablert av SINTEF i 1997, men ble overført til Helsedirektoratet i 2007. Fra og med 2008 inneholdt registeret personidentifiserbare data innen somatikk, psykisk helsevern, rehabilitering, personskader, spesialist behandling av rusmiddelavhengige, sykestuene i Finnmark og data fra AMK-sentralene. Etter hvert legges også data fra radiologiske enheter til.

Norsk pasientregister er i dag egen avdeling underlagt Divisjon for Helseøkonomi og finansiering i Helsedirektoratet og består av om lag 50 ansatte. Avdelingen har som hovedoppgave å forvalte dataene i registeret på en god måte, gjenbruk av dataene i registeret er et av formålene til registeret, hvor det tilbys både anonyme og personidentifiserbare data til sektoren. Den er organisert i tråd med arbeidsprosessene i verdikjeden:



Figur 10 Norsk pasientregister arbeidsprosess og organisering.

Seksjonene i avdelingen består av Datafangst som har ansvaret for mottak av data fra nesten 2000 virksomheter, Kvalitet og analyse seksjonen har ansvaret for kvalitetssikring av data inkl. dekningsgradsanalyser før data legges over i datavarehuset -som Leveranse seksjonen benytter til å utarbeide leveranser av data på de ulike områdene som registeret dekker. I tillegg har Registerteknologi seksjonen ansvaret for helhetlig forvaltning av teknologien som benyttes.

Det er slik vi forstår det en viktig forutsetning for Prehospital strategisk satsing at prosjektet gjennomføres i regi av Norsk pasientregister og at innsamling av data skjer i tråd med registerets etablerte arbeidsprosesser, men med valg av åpen kildekode verktøy til både database og publiseringsløsninger. Et tungt fagmiljø med erfaring fra både innsamling, kvalitetssikring, analyse og distribusjon har vært en viktig suksessfaktor.

2 Statens vegvesen: Datainn og intelligente trafikksystemer

Statens vegvesen er pålagt til enhver tid å ha oversikt over utviklingen av trafikken på riks- og fylkesvegnettet både i sanntid (dvs. dagens trafikk) og historisk utvikling. Men også å kunne predikere utviklingen i trafikken på vegnettet.

Intelligente transportsystemer (ITS) omfatter løsninger som benytter informasjons- og kommunikasjonsteknologi (IKT) i et trafikk- eller transportsystem, med den hensikt å bedre informasjonstilgang, påvirke atferd og forbedre transportløsninger og trafikkavvikling slik at man oppnår bedre fremkommelighet, trafikksikkerhet, miljø, tilgjengelighet og brukertilfredshet.[162]

Handlingsplan for ITS (2014 -2017) skisserer 21 prosjekter på området som skal bidra til bedre informasjonssystemer om trafikk. Et av områdene er bedre utnyttelse av trafikkdata.

Sammen med EUs ITS-direktiv, ITS Action Plan og Statens vegvesens ITS-strategi, er Samferdselsdepartementets «Report on national ITS action envisaged over the following five year period» til EU-kommisjonen en viktig vegviser for det videre arbeidet med ITS i Statens vegvesen. [163] [164][165]Handlingsplan for ITS er en oppfølging og konkretisering av ITS-strategien og skal være et styringsdokument for framtidsrettet, helhetlig og målrettet bruk av ITS i tråd med Nasjonal transportplan. Tiltakene i handlingsplanen dreier seg først og fremst om infrastrukturtiltak.

Vegvesenets strategi for utnyttelse av trafikkdata er nedfelt i Handlingsplan for trafikkdata 2013-2018[166]. Første steg i strategiens handlingsplan er å etablere et nasjonalt nett av trafikkregistreringsstasjoner på riks-, fylkesveger og noen sentrale kommunale veier. Statens vegvesen skal i samarbeid med kommunene vurdere behov for registreringer på det kommunale vegnettet. Kommunen har ansvar for etablering og innsamling av data fra disse punktene.

Vegdirektoratet har ansvaret for utarbeidelse av retningslinjer og godkjenning av det nasjonale nettet. I tillegg har Vegdirektoratet ansvar for utviklingen av nasjonalt trafikk- og transportdatasystem hvor dagens NorTraf (Norsk trafikkdatabase)[167] inngår.

[162] (Statens vegvesen, 2014)
[163] (Statens Vegvesen, u.d.)
[164] (EU, ITS Directive, 2010)
[165] (EU ITS Action Plan, 2015)
[166] (Vegdirektoratet, 2012)
[167] (Statens vegvesen, 2015)

Statens vegvesen har satt opp følgende mål for sin satsing på innhenting av trafikkdata:

- *Trafikkdata skal ha kjent kvalitet*
- *Trafikkdata skal samles fra riktig vegnett*
- *Trafikkdata skal ha riktig detaljeringsgrad*
- *Trafikkdata som fremskaffes skal etterspørres*

Effektmålene for satsingen er:

- *Gir god beslutningsstøtte*
- *Gir mulighet for hendelsesovervåkning*
- *Effektiviserer innsamlingen av trafikkdata – mindre manuelt arbeid*
- *Tilbyr trafikkdata i sann tid med bedre kvalitet*
- *Automatisere feilmeldinger fra kontinuerlige trafikkregistreringspunkt*

Generelt sett er datagrunnlaget til Statens vegvesen i dag tredelt:

1. Data om veginfrastrukturen – målebiler som registrerer ujevnheter, tar bilder hver 20 meter osv. Data i nasjonal vegdatabank.
2. Tilstandsdata om vær, nedbør, snø, is, eller annen type tilstand som beslutningsstøttedata. Egne data kombinert med data fra Metereologisk institutt.
3. Trafikkbeskrivelser. Mer og mer data og mer og mer sanntid. Bruker online registrering i sanntid (sensorer m.m.) i tillegg til bilder, veimeldinger osv.

I tillegg forskes det på om man i fremtiden kan buke data fra bilparken, fra mobiltelefoner, posisjoneringsdata o.l. For eksempel har Statens vegvesen et prøveprosjekt i samarbeid med Volvo hvor det hentes inn sanntidsdata fra 300 Volvoer i Oslo området.

Vegvesenet har også etablert samarbeid med eksterne dataleverandører. Et eksempel er en pilot på TOMTOM data i samarbeid med globale og lokale leverandører hvor det utforskes om det er mulig å koble data fra navigasjonssystemer som TOMTOM, Google, Garmin, busselskaper osv. med offentlig data.

Det er flere anvendelsesområder av vegvesenets trafikkdata, noen eksempler er:

- Utarbeidelse av Nasjonal transportplan,
- Veiovervåking i veitrafikksentralene,
- Planlegging/dimensjoner av veiutbygging, drift og vedlikehold,
- Analyse av trafikksikkerhet,
- Planlegging av beredskap,
- Grunnlag for informasjon til publikum og media,
- Datagrunnlag for politiet,
- Statistikkformål til Statistisk sentralbyrå og
- Datagrunnlag for ulike interesseorganisasjoner.

2.1 Datagrunnlag i Datainn prosjektet

Datainn-prosjektet er forankret i Handlingsplan for trafikkdata og dreier seg om innsamling og behandling av trafikkdata. Gjennom prosjektet etableres det i overkant av 5000 målepunkt som skal utgjøre et nasjonalt nett av trafikkregistreringsstasjoner for sykkel og motorkjøretøy.

Figur 10 Nasjonalt nett av trafikkregistreringsstrasjoner for sykkel og motorkjøretøy.

Datagrunnlaget til Statens vegvesen dreier seg mer og mer om sanntidsdata enten det gjelder trafikkdata, reisetid, godstransport og veikvalitet. Dataene vil være basert på målinger fra sensorer, kamera osv., men også på manuell registrering av data.

Utbyggingen av nettet har følgende plan:

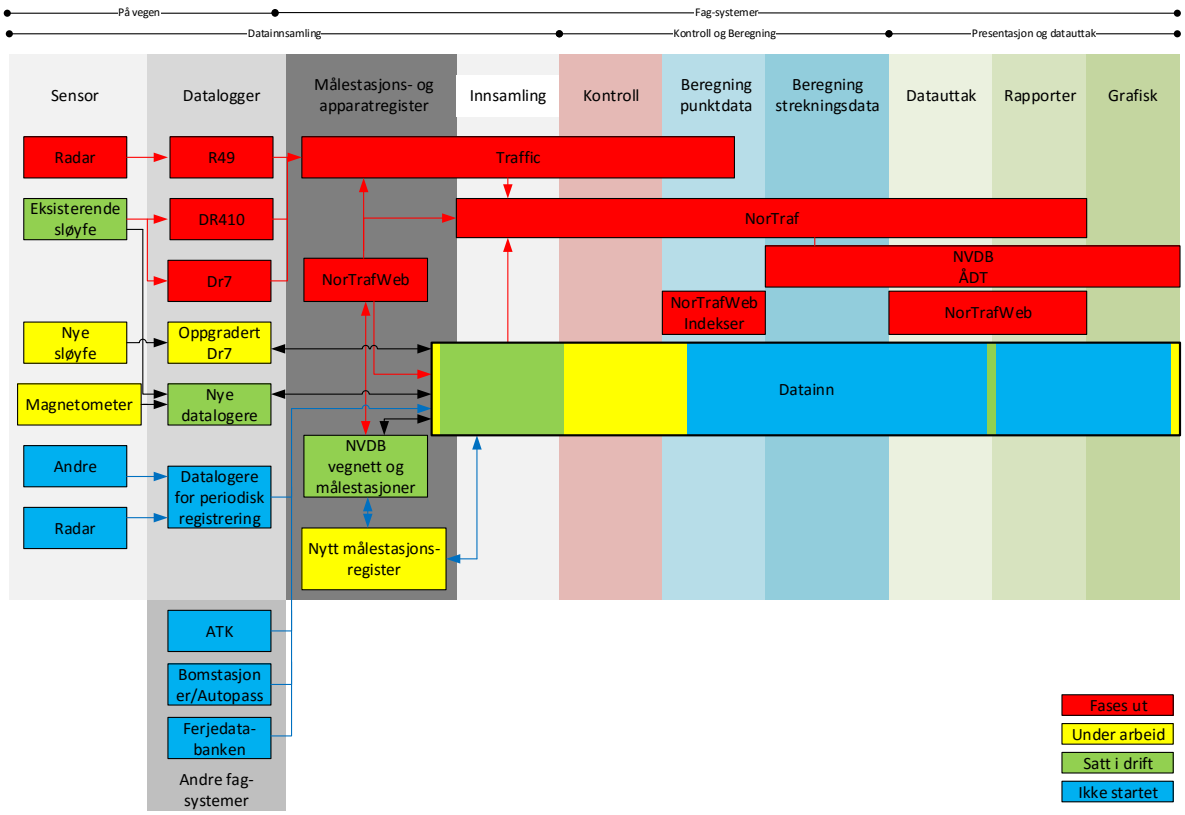
Utbyggingen skal gjennomføres på en mest mulig kostnadseffektiv måte ut fra følgende prioritering:

1. Punkt som benyttes til beregning av byindekser (2014)
2. Gammelt trafikkregistreringsutstyr skiftes ut på eksisterende punkter (2015)
3. Punkt som inngår i støykartlegging (2015)
4. Punkt som benyttes i beregning av vegtrafikkindeksen (2016)
5. Punkt som benyttes i beregning av fartsindeks (2017)
6. Punkt som benyttes av drift og vedlikehold (2017)
7. Punkt som benyttes i luftberegninger (2018)
8. Resterende punkt som benyttes til ÅDT-beregning (ÅDT: ÅrsDøgnTrafikk = trafikkmengde) (2018)

2.2 Løsning

Statens vegvesen konsoliderer trafikkdata satsingen i Datainn prosjektet slik at dette dekker både innsamling av data, kontroll, beregning av punktdata og strekningsdata, datauttak, rapporter og grafisk fremstilling/visualisering.

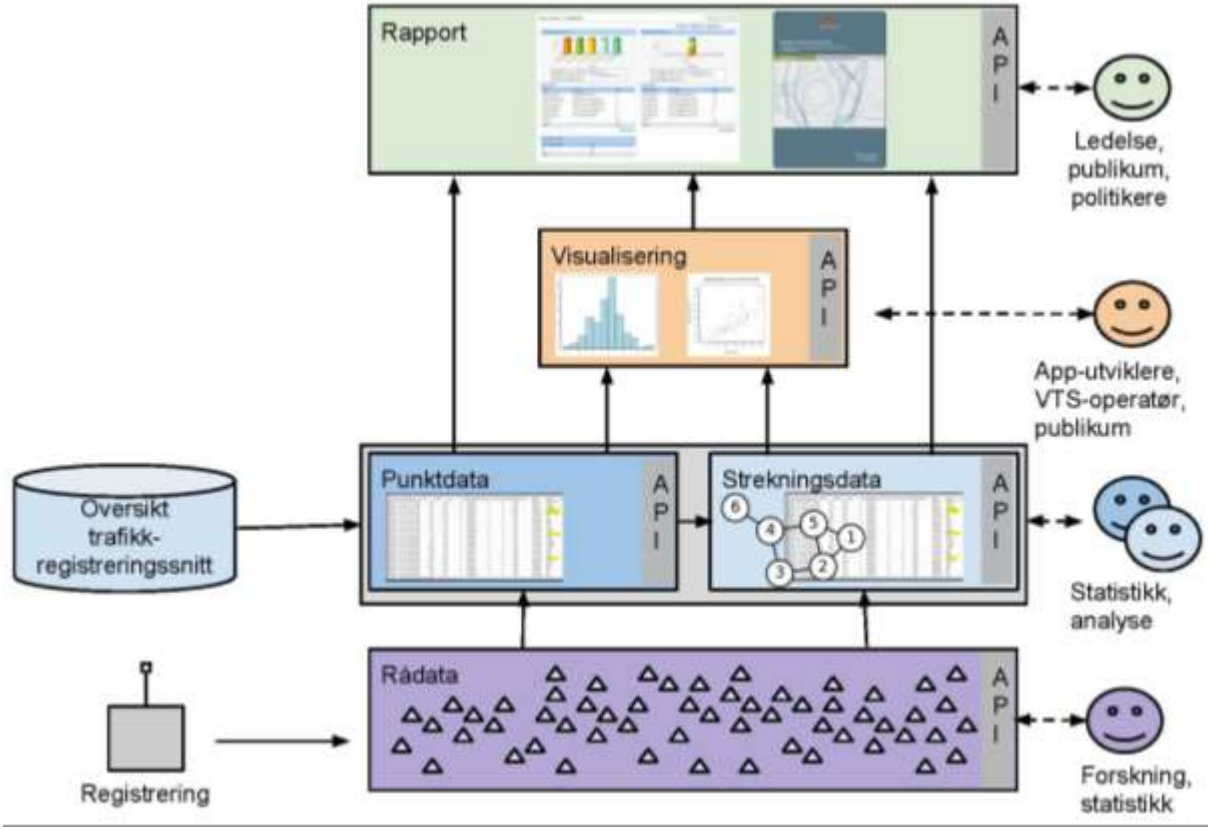
Data hentes fra sensorer på veiene i sanntid, mens data om plassering av målestasjoner med tilhørende sensorer og datalogger hentes fra Norsk veidatabank (NVDB).



Figur 11 Datainn prosjektet og tilhørende datakilder og portefølje. Kilde: Presentasjon av Kristin Gryteselv, Statens Vegvesen

Overordnet er systemarkitekturen firedelt, hvor rådata fra ulike kilder både i sanntid og historiske data er tilgjengelig. Disse dataene er tilgjengelige for gjenbruk blant annet for forskningsformål og statistikk.

Punktdata og strekningsdata er tilgjengelig for analyse for ulike formål.



Figur 12 Overordnet arkitekturskisse trafikkdata i Statens vegvesen

Norsk veidatabank (NVDB)

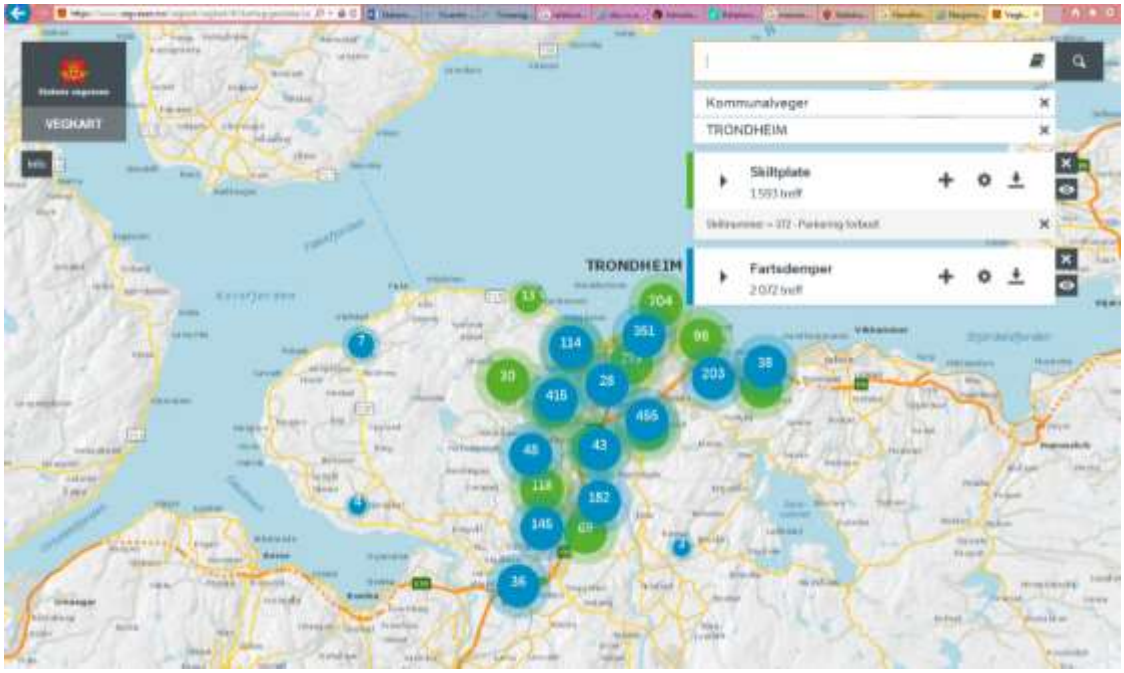
Dataene fra Norsk veidatabank (NVDB) er et viktig grunnlag for Datainn prosjektet. Dette datasettet er åpent tilgjengelig gjennom eget API slik at private viderebrukere kan benytte dette til å utvikle tjenester på web og i applikasjoner/apper (jamfør også data.norge.no)[168]. Statens vegvesen har registrert flere datasett på data.norge.no og har gjort åpne data til et strategisk valg.

I tillegg er det laget grensesnitt rettet mot publikum og media. Disse kan benytte visuelle framstillinger av data. Disse brukerne har da blant annet mulighet til å se på ulike typer data visualisert på kart. Tjenesten vegkart er eksempel på dette[169].

Tjenestene Vegkart og NVDB API gir tilgang til de fleste data langs vegnettet, som fartsgrenser, skiltplater og trafikkmengde. Tjenestene er foreløpig i testversjon (beta), men er allerede i utstrakt bruk.

Vegkart-tjenesten gjør det mulig å søke på en rekke data og visualisere dette på nett. Vegvesenets datakatalog definerer begreper som det er mulig å søke på. For eksempel kan en bruker søke på skiltplate, fotocelle eller fartsdemper. I eksemplet i figuren under er det søkt fram data om skiltplater og fartsdempere i Trondheims- området.

[168] (Direktoratet for forvaltning og IKT, 2015)
[169] (Statens vegvesen, 2015)



Figur 13 Sluttbrukertjenesten vegkart, Statens vegvesen

Teknologi

Innsamling skjer i fra egne målestasjoner og registreringer, men det er et klart ønske om å benytte trådløse nett for overføring av dataene (trådløse mobilnett eller lignende). På sikt ønsker prosjektet å benytte «connected devices» som for eksempel sender posisjon hvert 20 sek fra TOMTOM selv om dette ikke er del av dagens handlingsplan.

Statens vegvesen er i ferd med å utvikle en portefølje av verktøy og løsninger som dekker hele verdikjeden av stordata. Noen av teknologiene som benyttes er kort beskrevet under.

Løsningene for Datainn prosjektet er basert på åpne kildekodeverktøy. Basis i de nye tjenestene er Elasticsearch som er et stordataverktøy som støtter søk og sammenstillinger over multiple datakilder, fulltekstsøk og sanntidsanalyser av strukturerte data.[170]

For hver Målestasjon instansieres det en applikasjon som er ansvarlig for å lytte på datastrømmen fra målestasjonen (trafikkpasseringer og systemhendelser), og å lagre data for disse i Elasticsearch.

Det instansieres også en «ryddegutt» (pr målestasjon) i form av en applikasjon som jevnlig leter etter hull i meldingssekvensen som er lagret i Elasticsearch for en gitt målestasjon, og for å hente eventuelle manglende meldinger fra målestasjonen slik at indeksen i Elasticsearch blir komplett.

Det er også utviklet grensesnitt slik at data kan eksporteres til andre systemer og importere data fra målestasjonsregisteret (Nortraf databasen). Løsningen er satt opp med overvåking av de ulike applikasjonene som sørger for å starte applikasjonene ved behov eller restarte ved feil under kjøring.[171]

---

[170] (Elasticsearch, 2015)
[171] Basert på systemdokumentasjon motatt fra Statens vegvesen

For eksempel testes Visim og Aimsum som simuleringsverktøy i forbindelse med spesialanalyser og simuleringer på dataene.[172] Simuleringene gjøres både i mikrosammenhenger ved for eksempel bygging av veikryss og på mer makroorienterte simuleringer slik som utarbeidelse av nasjonal transportplan.

Vegvesenet satser på hyllevare så langt som mulig, men også på bruk av åpen kildekode og åpne standarder for å slippe leverandørinnlåsing.

2.3 Gevinster

I følge Handlingsplan for trafikkdata var vegtrafikkdatasystemene fragmenterte og dataene ikke tilgjengeliggjort på en tilfredsstillende måte, verken for interne eller eksterne brukere når strategien ble skrevet i 2012.

Arbeidet i Statens vegvesen skal håndtere denne situasjonen og følgende effektmål er satt opp for satsingen[173]:

- *Sikre tilgjengelighet til vegtrafikkdata i ulike fagsystemer og fagmiljøer.*
- *Etablere et robust forvaltnings- og formidlingssystem for vegtrafikkdata.*
- *Effektivisere innsamlingen og bearbeiding av vegtrafikkdata.*
- *Etablere en god beregnings- og analysemetodikk.*
- *Sikre at alle relevante vegkantsystemer blir utnyttet i innsamlingen av vegtrafikkdata (AutoPASS, ATK, klimastasjoner etc.).*
- *Sikre at ny infrastruktur langs vegnettet tilrettelegges for å samle inn vegtrafikkdata (for eksempel trekkrør og koblingskummer).*
- *Etablere kvalitetsmål for vegtrafikkdata fra ulike datakilder.*
- *Sikre en enkel tilgang til vegtrafikkdata for alle brukere.*
- *Etablere gode grunnlagsdata for effektiv trafikkstyring.*

De overordnede målene dreier seg om å sikre bedre beslutningsgrunnlag for alle typer veiprosjekt, sikre bedre trafikkavvikling og ikke minst bedre trafikksikkerheten. I tillegg kan mye av dataene frigis til relevant forskning.

Vegvesenet ser allerede gevinster i form av langt bedre informerte trafikanter. I tillegg er det tatt ut gevinster i forhold til at styringen av drift og vedlikehold er bedre. Bedre data gir mye bedre grunnlag for å lage prognoser og å planlegge vedlikehold. Dette gjør at det spares vesentlig på vedlikeholdskostnadene.

En annen effekt er store forbedringer i flere av arbeidsprosessene, man trenger for eksempel ikke lenger å ha folk ute på veiene, men kan omdisponere personell til andre oppgaver. Omleggingen skjer gradvis siden en del personell har kompetanse som ikke er så lett å gjenbruke på andre områder.

Statens vegvesen peker også på at de aller fleste data kan gjenbrukes til andre formål og dermed gi gevinster ut over eget ansvarsområde.

[172] (MathWorks, 2015)
[173] (Vegdirektoratet, 2012)

2.4 Utfordringer, løsninger og læring

Informasjonssikkerhet og personvern

Strategi for ITS slår fast at ITS kan medføre personvernutfordringer for både privat og offentlig sektor. Målet er at det skal legges vekt på dette i utformingen av ITS-løsninger, slik at mulighetene for misbruk av personlig informasjon blir så små som mulig eller eliminert. ITS-systemer bør utvikles etter prinsippet «privacy by design».[174]

«Privacy by design» har medført at vegvesenet har brukt mye ressurser på å innarbeide personvern i de nye stordata-løsningene. Gjennom dialog med Datatilsynet om hvordan dataene fra Autopass skal anonymiseres for bruk til analyser har man kommet fram til en metode som sikrer slik anonymisering. Det er derfor utviklet en egen løsning for å anonymisere data før de tas i bruk i stordatasammenheng og eventuelt gjøres tilgjengelig for andre.

Det er også etablert rutiner for sletting av overskuddsinformasjon. For eksempel kan Autopass-systemene lese registreringsnummer på biler, slike data slettes før de benyttes i stordata sammenhenger.

Vegvesenet har egne hjemler for å behandle personopplysninger, slike data behandles kun etter formålet internt i etaten. Et eksempel på dette er ulykkesdata i nasjonal veidatabank som ikke er tilgjengeliggjort som åpne data.

I et samfunnssikkerhetsperspektiv så er det en vurdering om enda mer informasjon (anonymisert) også burde vært tilgjengelig for analyseformål.

Det er likevel et spørsmål knyttet til å innhente hjemler når data viderebrukes fra eksterne kilder som for eksempel biler. Dersom det skal innhentes samtykke fra eiere av f.eks. 300 000 biler så sier det seg selv at dette blir upraktisk. Spørsmålet er da om dette da må forskriftsfestes.

Organisatoriske utfordringer

Å arbeide med stordata som forskningsbasert utvikling ligger i stor grad i grenselandet mellom IKT og fagområdene til Vegvesenet. Dette utfordrer de tradisjonelle bestiller (fag) og leverandør (IKT) rollene. Ikke minst møter IKT-driftsorganisasjonen utfordringer i forhold til ITIL (IT Infrastructure Library)[175] prosesser og roller. Det har vist seg tidkrevende å få på plass nye arbeidsmetoder. Problemet er å få til en effektiv flyt fra data genereres til de kan presenteres og brukes, dette er både en organisatorisk og en kompetansemessig utfordring.

[174] (Statens vegvesen, 2014)
[175] (QPR International, 2015)

I utgangspunktet var ikke IKT-organisasjonen rigget for å ta imot slike store datamengder i sanntid, verken teknologisk eller organisatorisk. Det har så langt manglet et stordata miljø på IKT-området, men også et miljø med tilstrekkelig analysekompetanse og kapasitet. Her har virksomheten kommet godt i gang med å bygge egen kapasitet, noe som har tatt tid. I tillegg utvikles det helt nye arbeidsprosesser for dette formålet.

Vegvesenets policy er at de skal være den offisielle dataleverandøren av kvalitetssikret veg data som er mest mulig åpne for viderebruk. Å sikre kvalitet og legge til rette for innovasjon er en viktig oppgave i denne sammenheng, dette er også krevende organisatorisk.

Teknologiske utfordringer

På stordataområdet går den teknologiske utviklingen raskt, det kommer stadig nye og forbedrede teknologier som det er krevende å holde tritt med. Samtidig skaper dette en dynamikk og stadig nye muligheter. Statens vegvesen bruker tid og ressurser på å forske på mulighetene.

Det er et valg at vegvesenet ikke ønsker å investere i egen infrastruktur for å overføre data, men basere seg på eksisterende trådløse nettverk som finnes. For å kunne hente data fra tilknyttede enheter som f.eks. navigasjonssystemer i biler så vil det kreve nettverkskapasitet, dette er delvis et budsjettspørsmål.

Konkret benyttes mobilnettet til datainnsamling, men det er deler av vegnettet som ikke har dekning i mobilnettet. Vegvesenet ønsker seg bedre dekning langs vegnettet slik at vi kan få overført data i sanntid.

De store datamengdene krever også stor grad av prosessorkraft. Det er behov for å etablere et prosjekt som ser nærmere på hvordan prosessorkraften kan spres på flere maskiner for å håndtere de enorme datamengdene. For øyeblikket er det ikke etablert prosjekt for å løse dette.

Eierskap til data

Statens vegvesen har funnet at det er en problemstilling å avklare eierskap og viderebruk av data fra eksterne kilder. Særlig gjelder dette data fra kollektivtrafikk og private biler. Som tidligere beskrevet har vegvesenet i dag et forskningsprosjekt som henter data fra 300 Volvoer i Oslo området. Viderebruk av disse er sikret gjennom avtale med Volvo og basert på samtykke fra bileier. Men hva om det på sikt dreier seg om 300 000 biler? Det er da behov for at eierskap og vilkår for viderebruk avklares nærmere, er det bileier, leasingselskap, bilselger, importør eller produsent som eier dataene? Og i neste omgang vil det bli spørsmål om hvem som skal lagre dataene og hvem som kan analysere dem.

Det er også et spørsmål om framtidige forretningsmodeller skal innebære at bilprodusentene skal sørge for at det leveres data til veimyndighetene.

Kompetanse

Den knappeste faktoren når det gjelder kompetanse i vegvesenet har vært tilstrekkelig analyse kompetanse. Det er behov for flere statistikere eller datavitere som kan utnytte dataene bedre. Det krever også en god del kompetanse å skulle bestille slik kompetanse da det kan være vanskelig å kravspesifisere dette i et konkurransegrunnlag. Dessuten er det tvilsomt om det finnes særlig kompetanse i markedet.

Hovedproblemet er knyttet til kombinasjonen mellom klassiske trafikkfag, statistikk og data kunnskap. Dette er per i dag en utfordring.

Når det gjelder IKT området så er man bedre stilt, det synes å være en drøm for utviklere å jobbe med løsninger knyttet til stordata. Her henter Vegvesenet kompetanse via rammeavtaler.

For eksempel utdanner NTNU klassiske trafikkingeniører og IT-ingeniører. Men ingen av studiene inneholder så langt ITS. Det synes å være behov for både egen ITS-utdanning og ITS som del av trafikkstudiene. Men det synes å ta lang tid å få med utdanningsinstitusjonene og å bygge opp studier.

Datakvalitet og viderebruk av åpne data

Som nevnt er det vegvesenets rolle å levere kvalitetssikrede veidata for viderebruk i andre sammenhenger. Datagrunnlag som beskriver transportinfrastrukturen med tilhørende egenskaper og dynamiske data som beskriver aktuell situasjon for trafikken er en forutsetning for ITS.

Statens vegvesen har også besluttet at etatens data normalt skal være fritt tilgjengelig. Etaten ønsker å stimulere private aktører til utvikling av produkter og tjenester til nytte for trafikantene. Det er også slik at økt gjenbruk av data gir mulighet for tilbakemeldinger og bedre datakvalitet.

Det er en forutsetning for anvendelse av dataene at det inngås en NLOD-avtale (Norsk lisens for offentlige data). Statens vegvesen har sammen med Trafikvärket i Sverige og andre internasjonale veiadministrasjoner, etablert en felles organisasjon (TNITS) for å støtte dataflyt mellom ITS-tjenester og nasjonale vegdatabanker i Europa.[176]

Interoperabilitet

Det er mange ulike aspekter som må tas i betraktning for å oppnå interoperabilitet. Etterlevelse av tekniske standarder er ikke tilstrekkelig, og interoperabilitet bør derfor identifiseres som en utfordring for ITS-utvikling.

Terminologi, lovgivning, ITS arkitektur, standarder, forretningsmodeller, kontrakter og testregimer må være på plass for å oppnå tilstrekkelig interoperabilitet.

I ITS-strategien slås det fast at etaten skal være aktivt med i arbeidet for åpne spesifikasjoner og standarder for ITS. Det norske rammeverket ARKTRANS[177] er et godt verktøy for å sikre interoperabilitet og gjenbruk av informasjon.

Det er et mål å skaffe åpne data som genereres fra GPS-baserte systemer i privat sektor.[178]

---

[176] (Statens vegvesen, 2014)

[177] «ARKTRANS brukes i nasjonale og europeiske prosjekter som adresserer co-modale så vel som multimodale løsninger. Basert på disse prosjektene blir ARKTRANS kontinuerlig oppdatert og videreutviklet. ARKTRANS fremmer mer effektive co-modale løsninger gjennom å være et multimodalt rammeverk. Systemsamhandling, effektiv informasjonsflyt, koordinering over ulike transportformer, etc. er støttet av felles konsepter og spesifikasjoner som dekker nåværende gap i semantikk og legger til rette for interoperabilitet og effektivitet».(http://www.arktrans.no/)

[178] (Statens vegvesen, 2014)

3 Statsbygg: Smarte bygninger

Statsbygg har startet arbeidet med å utvikle en ny løsning for digital datafangst fra statens ca. 2350 bygninger på 600 eiendomskompleks i inn- og utland, noe som utgjør om lag 2,7 millioner m^2. Tanken er på sikt å kunne hente data fra ulike kilder og gjøre det mulig å analysere og å presentere bedre beslutnings- og planleggingsgrunnlag for energieffektivisering, vedlikehold og utbygginger.

I inneværende IT-strategiperiode har Statsbygg prioritert utbygging av et felles eiendomsnettverk som gjør at data kan utveksles mellom eiendommene digitalt. Per i dag er ca. 300 eiendommer tilknyttet nettverket. Hovedkilden av data er eiendommenes forvaltning, drift og vedlikeholdssystemer (FDV), men også andre typer systemer kan avgi data over dette nettverket. Eksempler på dette er løsninger for måling av energiforbruk, temperaturmålere og sensorer som måler luftkvalitet.

Når det gjelder stordata så er det særlig to initiativer som er relevante i denne sammenheng, et pilotprosjekt for sanntidsanalyse av energiforbruk og deltakelse i det EU-finansierte ProDataMarket prosjektet.

Pilotprosjekt for sanntidsanalyse av energiforbruk dreier seg om å kunne agere på endringer i energiforbruk i enkeltbygninger, eiendommer eller i porteføljen. Dermed kan man agere raskt dersom det skjer endringer i forbruket. I tillegg kan man analysere virkningene av tiltak for å spare energi. Datagrunnlaget er ulike typer sensorer og målere som leverer data om energiforbruket, og eventuelt energiproduksjonen, til en sentral database hver time. Dataene kobles så til ulike datasett som værdata, geografisk lokasjon basert på grunndata fra matrikkelen og eiendomsregisteret, enhetsregisteret, energimerking, veidata m.fl.

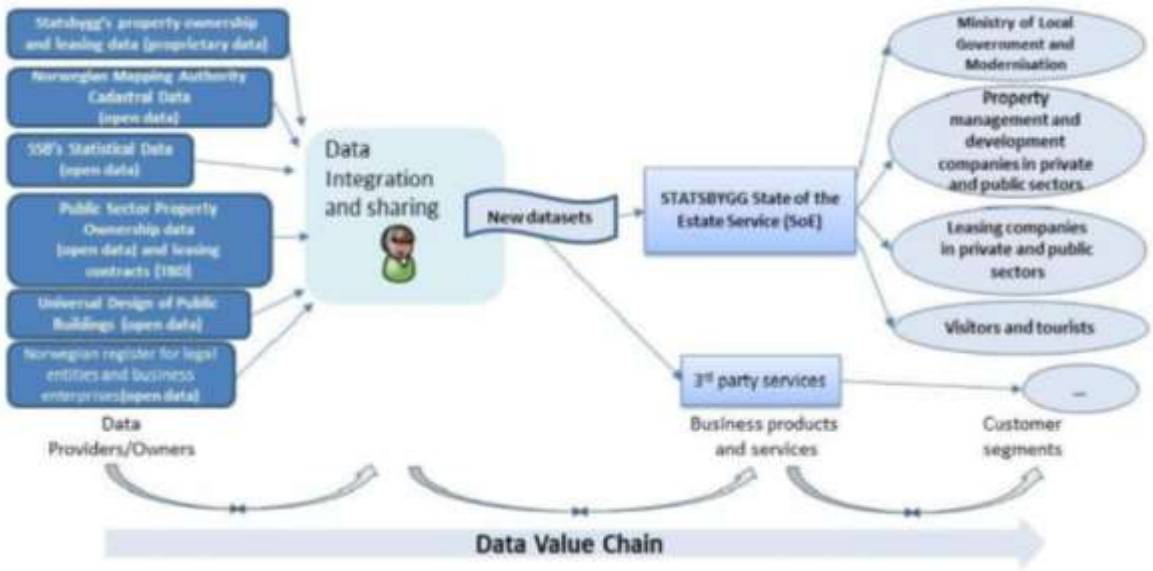
Selv om piloten bare omfatter 3 eiendommer så langt, så er resultatene lovende og har gitt verdifull læring. Dagens rapportering av energiforbruk til Statsbygg skjer manuelt, noe som er arbeidskrevende og medfører vesentlig forsinkelse og dermed også forsinkelser i iverksetting av tiltak for å hindre vedvarende forhøyet energiforbruk. Bare automatisert datafangst kan dermed gi vesentlige kostnadsbesparelser, men i tillegg kommer store besparelser effekter av redusert energiforbruk.

I tillegg deltar Statsbygg i et internasjonalt pilotprosjekt i regi av ProDataMarket prosjektet som er finansiert over EUs Horizon 2020 satsing. Prosjektet består av 8 partnere i konsortiet, med Statsbygg og SINTEF fra Norge. Det overordnede målet er å finne muligheter for innovasjon basert på eiendomsdata på tvers av relevante aktører.

ProDataMarket vil skape en «digital markedsplass» for både åpen og lukket eiendomsdata slik at eiere av disse dataene skal kunne gjøre dem tilgjengelige for de aktører som kan nyttiggjøre seg datasettene. [179]

Løsningen er foreløpig på pilotstadiet hvor det hentes data fra et begrenset antall bygninger. Men tanken er at alle bygningene som staten eier skal kunne levere tilstandsdata digitalt.

[179] (Prodatamarket, 2015)



Figur 14 Data value chain

Flere analysescenarioer kan støttes gjennom løsningen:

1. Rapportere kontinuerlig om tilstanden på statens eiendommer
2. Energieffektivisering
3. Analyser av tilgjengelighet til kontorbygg
4. Analyse av leiepriser
5. Analyser av grunnforhold
6. Tilgjengeliggjøring av data for andre tredjepartstjenester
7. Risiko og sårbarhetsanalyser, f.eks. risiko for at vernede bygg skal bli oversvømt.

Figur 11 Eksempel på simulering av oversvømmelsesfare for vernede bygninger, Kilde: Statsbygg SoE

3.1 Datagrunnlag og teknologi

Statsbygg har en rekke egne datakilder fra bygningenes styringssystemer og ulike drifts og vedlikeholdssystemer. Disse kan avgi mye forskjellig informasjon om bygningenes tilstand. I tillegg har Statsbygg også andre datakilder som for eksempel data om leiekontrakter og – priser, disse er ikke gjenstand for tekstanalyse men registreres manuelt. I tillegg benyttes en egen database med oversikt over bygninger og universell utforming.

Data fra matrikkelen og data fra norsk eiendomsinformasjon kobles med kartdata og eiendomsregisteret.

I EU pilotprosjektet er det også gjort forsøk på å få koblet på værdata slik at det kan gjøres analyser på vedlikeholdsbehov, energiforbruk o.l. i forhold til hvor værutsatt bygningen er eller for eksempel styre oppvarming i henhold til værforhold.

I ProDataMarket prosjektet benyttes teknologi fra Ontotext, som er et Bulgarsk selskap som leverer produkter og løsninger for semantisk interoperabilitet og dataintegrasjon. Ontotexts teknologi for data forvaltning er basert på at det etableres ontologier for å modellere datamodeller, som så kan benyttes til analyse av ustrukturert tekst for å hente ut strukturert informasjon fra teksten.

Ontotext skal levere en skybasert løsning for data og informasjonsforvaltning til ProDataMarket. Løsningen består av en metadatakatalog som viser datakildenes datamodeller (RDF-grafer) og andre metadata, samt kobling til begrepsmodeller (ontologier). Dette vil være en av nøkkelkomponentene i infrastrukturen som tas fram i ProDataMarket.[180]

3.2 Gevinster

Bedre datagrunnlag om offentlige bygninger vil kunne gi effekter i form av bedre grunnlag for å planlegge vedlikehold og oppgradering av bygg, men også bidra til å få ned energiforbruk og vurdere verdi på bygget på en bedre måte. Dessuten ville «oversikt over statens eigedomar» kunne være tilgjengelig med riktig tilstandsbilde nesten til enhver tid. Det samme kan sies om andre rapporteringer som for eksempel energiforbruk og arealutnyttelse.

Moderne bygg inneholder en rekke sensorer og mekanismer som håndteres av bygningens sentrale styringssystem, og som integreres med byggets FDV-system som igjen avgir og mottar data fra Statsbyggs sentrale FDV-systemer. Gode analyser lokalt og sentralt gjør at man kan oppnå mer kostnadseffektive bygg ved for eksempel at vedlikehold kan planlegges bedre og dermed spare penger både på selve vedlikeholdet, men også på gjenoppretting etter skade og eventuelt unngå verditap.

Moderne bygninger er også smartere ved at teknologien i større grad styrer forhold som energiforbruk og luftkvalitet. Data analyser i kombinasjon med maskinlæring kan automatisere driften ytterligere.

Når det gjelder energiforbruk så har Statsbygg klare resultatindikatorer fra eierdepartementet (KMD), per i dag er målet 200 kwh per m^2 i gjennomsnitt for hele porteføljen. Men den overordnede visjonen er at bygningene skal være selvforsynte med energi, noe som selvsagt ikke kan gjelde eldre bygg.

[180] (Prodatamarket, 2015)

Men i tillegg til redusert energiforbruk, så har teknologien et langt større potensial, for eksempel kan energiforbruk predikeres og tiltak kan planlegges. Man kan tenke seg at det på sikt tilbys grønne leieavtaler hvor det for eksempel garanteres null utslipp av klimagasser.

3.3 Utfordringer, løsninger og læring

Teknologiske utfordringer

Det er en utfordring å hente data fra eldre bygg hvor det må gjøres vesentlige investeringer i sensorer og IT-infrastruktur for å kunne hente data. I slike tilfeller vil det være en kost nytte vurdering om det skal etableres infrastruktur i bygget og tilknytning til datanettverket.

Tilgang til data og metadata

De teknologiske utfordringene dreier seg også om å få orden på datakildene og oversikt over datastrukturer. For eksterne datakilder så er grunndataregistrene tilgjengelige og nyttige. Kartverkets data er beskrevet på en god måte i deres metadata katalog. Men samtidig er datastrukturene i matrikkelen krevende å sette seg inn i på detaljnivå og krever til dels spesialkompetanse.

For øvrig er det en generell utfordring at det finnes mange interessante datasett som kan gi merverdi, men det er alltid en utfordring å sette seg inn i og forstå datasettene. Dette dreier seg ofte om at metadataene er begrensede eller beskrevet i en særskilt kontekst. Men det dreier seg også om at Statsbygg selv har mange og spredte datakilder med data som er tidsbestemte og med mange avarter av informasjon. Det er krevende informasjonsforvaltningsarbeid å få oversikt og et riktig bilde av informasjonen.

Slik vi oppfatter det har det vært en viktig erkjennelse at arbeidet med å få nytte ut av datasett er til dels krevende, noe som igjen til dels skyldes kvaliteten på metadata og arbeidet som har vært gjort med å forvalte datakildene. Dette i seg selv har gitt god læring. For å bøte på dette har Statsbygg standardisert grensesnittet mot de ulike løsningene som utveksler data med de sentrale systemene.

En annen problemstilling er tilgang til datasett som andre virksomheter forvalter. Selv om offentleglova og digitaliseringsrundskrivet pålegger offentlige virksomheter å tilgjengeliggjøre data som ikke er taushetsbelagt så sitter dette langt inne. Et godt eksempel er tilgang til datasett med energimerkingssystemet fra NVE, Statsbygg ønsker tilgang til hele datasettet, men må spørre for hver enkelt bygning. Et annet eksempel er tilgang til reguleringsplaner og kartutsnitt fra kommunene, selv om noen kommuner nå er digitale og gir tilgang til digitale kart -så er det fortsatt vanlig at all slik informasjon kommer på papir og for hver enkelt bygning eller byggeprosjekt.

For private aktører dreier skepsisen til å dele data seg om at dataene representerer en verdi i form av et konkurransefortrinn. For offentlige virksomheter synes dette å dreie seg om holdninger i form av en iboende skepsis til å dele datasett både hos enkeltpersoner og virksomheter. Ofte har enkeltpersoner et sterkt eierskap til dataene de har ansvaret for. Statsbygg erkjenner selv at de har flere datasett som kunne vært åpnet opp og registrert på data.norge.no, for eksempel data registrert i forbindelse med grunnprøver.

Læringen her har vært at det er krevende å få til innovasjon dersom man er avhengige av eksterne parter, selv om disse er offentlige og at arbeidet med å få tilgang til data ikke må undervurderes.

En annen problemstilling er knyttet til data som ikke finnes fordi de ikke er digitalisert. Det mest åpenbare er oversikt over anlegg som finnes i grunnen (kabler, rør etc.) hvor det per i dag finnes lite digitale data. Som oftest er dette dokumentert på papir og oppbevares i kommunene. Dette medfører ofte en fordyrende prosess med kartlegging i hvert enkelt byggeprosjekt.

Det synes å mangle insentiver for å digitalisere data som kunne ha stor nytte for samfunnet totalt, i stedet brukes det ressurser i hver enkelt virksomhet når de har behov.

Informasjonssikkerhet og personvern

Noe informasjon som Statsbygg forvalter er unntatt offentlighet. Dette gjelder særlig informasjon om bygg som er sikkerhetsklarerte og fengsler. I tillegg behandler de en god del informasjon som er definert som forretningshemmeligheter. Men de aller fleste data er i utgangspunktet offentlige.

Informasjonssikkerhet og personvern har derfor ikke vært problematisk så langt i pilotene, men må vurderes når de kommer over i en fase der permanente løsninger skal etableres.

Kompetanse

Pilotprosjektene baserer seg på både intern og ekstern kompetanse. Pilotprosjektene har gitt mye læring både av teknisk, mer faglig og analytisk karakter. I tillegg har de gitt mye erfaring og læring som kan gi grunnlag for strategisk tenkning framover.

Stordata krever ulik type kompetanse. Man trenger en blanding av eksisterende fagmiljø på bygg og anlegg m.m., men behovet for kompetanse dreier seg også om ekspertise på data analyse som kan støtte eksisterende fagmiljø. Det er viktig at datasettene tilrettelegges for at byggetekniske fagfolk kan nyttiggjøre seg dem i sine analyser, uten at det krever særskilt IT-kompetanse.

Datakvalitet og eierskap til data

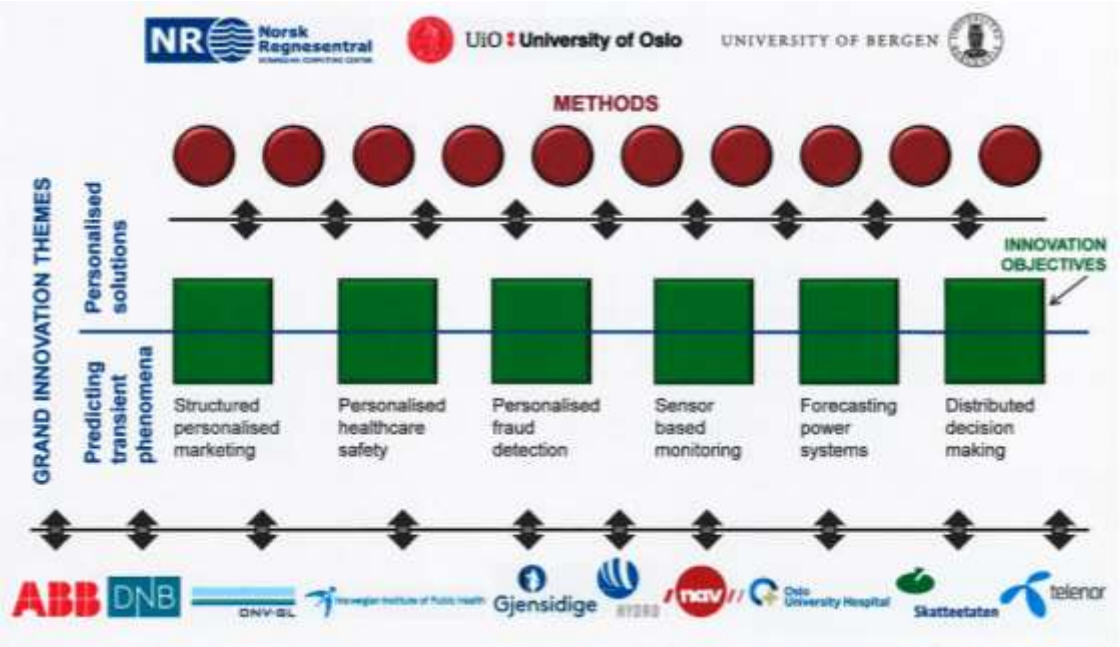
Det er en utfordring å skulle dele data og analyser som er basert på eksterne datakilder. Kan man eksempelvis dele matrikkeldata og kartdata videre? Hva om man vet at noen kilder har variabel kvalitet, bør man da dele dataene? Statsbygg finner det også uklart hvem som eier analyser som er basert på eksterne kilder. Foreløpig er ikke dette løst, men må vurderes i hvert enkelt tilfelle. Per i dag deles ikke data og analyser videre dersom det er tvil om eierskap og rettigheter.

Vedlegg 2: Relevante forskningsprogram

«Big Insight» ble startet opp som senter for statistisk innovasjon i 2015 og har en varighet på 8 år. Programmet ledes av Norsk regnesentral med universitetene i Oslo og Bergen som samarbeidsparter. Fra offentlig sektor er NAV, Skatteetaten og Oslo Universitetssykehus sammen med Kreftregistrene partnere, programmet er delvis finansiert av forskningsrådet, resten dekkes av medgått arbeidstid blant partnerne

Programmet skal utvikle nye statistiske metoder og modeller for bedre å kunne predikere fenomener som er ustabile, forstå uforklarte mekanismer og sammenhenger, avdekke skjulte mønstre, m.m. Målet er at modellene som utvikles skal være generiske, og basert på nye metoder innen statistikk og maskinlæring som håndterer store datamengder.

Tanken er at beslutninger som i dag er basert på typisk og gjennomsnittlig oppførsel utvikles mot individualiserte løsninger for eksempel innenfor medisinsk behandling, servicetilbud eller markedsføring.



Figur 12 Big Insight - samarbeidsparter, innovasjonsområder og metodeutvikling

Programmet håndterer faktiske case fra brukerstedene og skal utvikle metoder basert på disse. Det er i skjæringspunktet mellom utvikling av den generiske metodikken og de spesifikke utfordringene hos hver enkelt partner at innovasjonen skjer. Gjennom partnerne skal også innovative løsninger iverksettes i praksis.

Både NAV og Skatteetaten har et stort potensiale for å kunne utnytte stordata bedre. I Big Insight er de inne med case hvor man ser på mulighetene for å redusere trygdesvindel og skatteunndragelser. Data fra a-ordningen (inntektsdata) gir helt nye muligheter i den sammenheng.

Big Insight programmet nevner at utfordringer knyttet til personvern er løst gjennom anonymisering, men når det skal utvikles metoder som kan bidra til å utvikle individuelt tilpassede tjenester så kan dette bli en utfordring. Regnesentralen har gått flere runder med Norsk samfunnsvitenskapelig datatjeneste når det gjelder gode rutiner for å anonymisere data.

De ser at det mangler først og fremst kompetanse på analyse av data i virksomhetene. Stordata medfører at man trenger enda mer kompetanse på modellering og statistisk analyse enn tidligere. Det er utopisk å tro at maskinene kan lage analyse modellene.

Når det gjelder datakvalitet så er denne generelt sett variabel, data må uansett alltid kvalitetssikres. Norsk Regnesentral fremhever den gode datakvalitet i Skatteetaten.

BYTE - «The Big data roadmap and cross-disciplinarY community for addressing socieTal Externalities»[181]

BYTE prosjektet skal bistå europeisk forskning og industri med å få fram de positive sidene ved stordata og å minimere de negative sidene slik at de kan sikre en større del av stordata markedet innen 2020. Prosjektet er finansiert av EU gjennom Horizon 2020.

Målet er å etablere et «policy roadmap» slik at det skal være lettere å iverksette tiltak og lage føringer for en effektiv tilnærming til stordata i EU. Norge er også representert i prosjektet via Institutt for Informatikk ved Universitet i Oslo. Prosjektet er basert på casestudier hvor man tar utgangspunkt i, og undersøker ulike eksterne forhold og rammevilkår slik som økonomi, lovverk, sosiale forhold, etikk og politikk.

Videre vil det bli gjort en horisontal analyse hvor man ønsker å avdekke hvordan de positive rammevilkårene kan forsterkes, og hvordan de negative kan reduseres.

Prosjektet undersøker 7 case studier innen kultur, olje og gass, shipping, transport, observasjon av miljø fra bakken og fra satellitter, helse, kriseinformasjon og smarte byer. Foreløpige funn fra case studiene kan oppsummeres slik:

Kultur: På dette området utvikles det ontologier som i neste omgang kan gjøre det mulig å få lettere tilgang til kulturdata. Men området er umodent, forretningsmodellene er uklare slik at foreløpig er det forskere og akademia som har størst nytte av data på området. Opphavsrettslige problemstillinger gjør det vanskelig å utvikle nye anvendelser på dette området.

Olje og gass: Olje og gassindustrien er en teknologiintensiv sektor med enorme mengder data. Stordata teknologi er tatt i bruk for å overvåke oljereservoar, overvåke behov for vedlikehold av installasjoner og til miljøovervåking. Her har Norge gått foran og påpekt behovet for åpne data på tvers av ulike lands myndigheter, selskaper og operatører.

Men foreløpig er det utfordrende å få til deling av data på grunn av arbeidsdelingen mellom aktørene og mangel på incentiver til å dele data. I denne bransjen er det behov for å etablere nye samarbeidskonstellasjoner mellom drilling, analyse og operatører. Problemstillingen om hvem som eier data er sentral her, er det de som eier installasjonene, eller de som bistår dem med analyser som eier dem?

Shipping transport: Byte har funnet at det er et veldig tidlig stadium på dette området og at det ser ut til at skipsrederne ikke har fokus på bruk av data for kommersiell nytte. Dette står i kontrast til det norske Kystverkets som er langt fremme i både å samle inn, utnytte og dele data for navigasjon og skipstrafikk.

---

[181] (BYTE, Universitetet i Oslo, 2015)

Miljø – jordobservasjon/satellittobservasjon: På dette området er det tilgjengelig enorme mengder data, mange applikasjoner slik som for eksempel værmelding, klima, etc. Er godt eksempel er utnyttelse av data fra Copernicus satellitten som kontinuerlig leverer sanntidsdata om jordoverflaten. I tillegg gjøres det LAIR scanninger av jordoverflaten fra fly og SONAR scanninger av havbunnen fra båt. Det gjenstår likevel at dataene blir utnyttet bedre og tatt i kommersiell bruk.

Helse: BYTE har funnet at stordata primært utnyttes for forskningsformål særlig knyttet til genetikk og for eksempel forskning på arvelige sykdommer. Store datasett i form av forskningsapplikasjoner. Ikke overraskende er det store personvernutfordringer i denne sektoren, og en rekke etiske spørsmål knyttet til i hvilken grad data kan deles. Behandling av personopplysninger utenfor landegrensene er også en utfordring, siden kombinasjoner av data på tvers av skytjenester gjør det krevende å definere grensene mellom kildene.

Kriseinformasjon: På dette området er det fokus på dataintegrasjon mellom mange datasett som data om redningsaksjoner, sosiale media, sensor data (f.eks. oversvømmelser), geodata, værdata m.m. Forretningsmodellene er utfordrende siden både teknologiene og utstyret kan brukes til flere formål, utstyr som er installert i bygninger kan også brukes til varsling, overvåking o.l. For eksempel er krisevarsling og energieffektivisering to anvendelser på basis av samme utstyr. Som for all overvåking blir personvern og informasjonssikkerhet en utfordring.

Smarte byer: På dette området er det i ferd med å vokse fram anvendelser i forbindelse med å styre bilkøer, kartlegge forsamlinger, ha oversikt over ventetider, styre transport-logistikk m.m. Mobilitetsdata fra ulike kilder kan drive utviklingen av et levende stordata-økosystem i byene. I tillegg til mobilitetsdata fra mobiltelefoner, trafikksensorer m.m. Så er det mer og mer data tilgjengelig fra bygninger, biler og arrangementer. På dette området er åpne data driveren for innovasjon noe som også gjør personvern og informasjonssikkerhet til en utfordring.

Optique[182]

Optique programmet skal skape en portefølje av verktøy for å kunne utnytte stordata bedre. Metoder for å benytte skytjenester, sluttbrukerverktøy for å kjøre spørringer mot datasett osv. Universitetet i Oslo er prosjektkoordinator for Optique. Arild Waaler, professor ved Institutt for informatikk står som ansvarlig. Samtidig som de er prosjektkoordinator, er de selvsagt en av partnerne – sammen med 9 andre partnere (heriblant DNV GL (Veritas) og Statoil).

Optique ble startet opp i 2012. Prosjektet er finansiert over forgjengerprogrammet – Det 7. rammeprogram for forskning (FP7), under ordningen «Integrated projects». Varighet på prosjektet er fram til november 2016.

I dette prosjektet arbeides det med generiske metoder for å tilgang på data på tvers av kilder og ulike typer datasett. Løsningen består av et lag med metadata i form av ontologier og søkelogikk og et lag med kobling («mapping») til data kilder. Det er et mål at løsningen skal

[182] (Optique, 2015)

være mest mulig generisk, men at den også kan tilpasses ulike bruksområder. Løsningen har så langt blitt prøvd ut på Statoils mange ulike kildedata, og hos Siemens.

Optique skal også prøves ut i helsesektoren neste år, blant annet på Akershus universitetssykehus. I første omgang vil dette bli gjort i form av en forenklet versjon. Tanken er at Optique også skal kunne skape integrasjon mellom flere systemer i helsevesenet slik at data kan utnyttes på tvers av kildene.

Ontologien som benyttes gjør det mulig å enkelt formulere en spørring for å hente dataene som trengs. Optique teknologien kan legges som et lag på toppen av et hvilket som helst datasett (også datavarehus) med et grensesnitt som gjør det mulig for flere å hente ut data, uten at de nødvendigvis har spisskompetanse rundt datauthenting.

Vedlegg 3: Nærmere beskrivelse av bruk av stordata i utvalgte land

3.4 Storbritannia

Storbritannia var et av de landene som var først ute med å implementere stordata-satsinger i offentlig regi. I 2013 fikk de på plass en egen strategi[183] med fokus på å bygge opp evnen til dataanalyse og stordata-anvendelser. Strategien innebærer betydelige investeringer i teknologiutvikling og forskning på bruk av offentlige datasett. Blant de sentrale områdene som adresseres i strategien er behovet for å utvikle kompetanse innen dataanalyse, og hvordan juridiske og øvrige rammebetingelser for deling og linking av offentlige data kan tilrettelegges.

Etableringen av Alan Turing-instituttet[184] i 2014 skal bidra til å gjøre Storbritannia internasjonalt ledende på stordata-området. Instituttets kjerneområde er forskning på dataanalyse, matematikk, statistikk og algoritmer.

Anvendelser

Troen på stordata som ny vekstnæring synes å være spesielt stor i Storbritannia. I privat sektor regner man med at stordata vil skape mange nye jobber (ett anslag sier 58.000) innen et bredt spekter av bransjer og samfunnssektorer[185].

Storbritannia var også tidlig ute med å lansere sin portal for åpne, offentlige data –, som en kilde bl.a. for stordata-anvendelser, herunder verdiskapning i privat sektor[186].

Et område som i Storbritannia har framstått som særlig aktuelt med hensyn til utnyttelse av stordata-teknologier, er strategisk «foresighting» og støtte for policyutvikling[187]. Utgangspunktet var at man så at mange problemstillinger i samfunnet var multidisiplinære og tverrsektorielle, og krevde håndtering av store mengder variert informasjon fra mange kilder – slik som innen miljøområdet. Gjennom det såkalte HSC-programmet (Horizon Scanning Centre) som ble etablert i 2004, skaffet britene seg et kraftfullt verktøy for å analysere klimaforandringer og effektene på matproduksjon, vannkvalitet og andre viktige forhold på både nasjonalt, regionalt og internasjonalt nivå[188].

Et område som har vært mye diskutert både i Storbritannia og i mange andre land – som i Norge, er bruk av stordata til å avdekke skatteunndragelser, svindel og annen økonomisk kriminalitet. I noen analyser opereres det med gevinstpotensialer i Storbritannia bare på dette området på 20-30 milliarder pund pr år.

Når det gjelder kriminalitetsbekjempelse har man introdusert mulighetene for å «scanne» relevante miljøparametere for organisert kriminalitet, som verktøy for å predikere hendelser. Det er en lang rekke eksempler på slikt «forutseende» politiarbeid, særlig i land som USA, Storbritannina og Tyskland. Teknologirådet gir i en fersk rapport[189] en oversikt over

[183] (HM Government, 2013)
[184] (Wikipedia, 2015)
[185] (SAS Institute, Tech Partnership, 2014)
[186] (DATA.GOV.UK, 2015)
[187] (Elsevier, 2010)
[188] (HM Government, 2015)
[189] (Teknologirådet, 2015)

eksempler på slikt forutseende politiarbeid i flere land. I et EU-støttet prosjekt (ePOLICE[190]) under 7. rammeprogram har et konsortium bestående av bl.a. politiet i Yorkshire, Europol, akademia og verktøyleverandøren SAS, testet ut verktøy og metoder for dataanalyse og predikering av ulike typer kriminalitet som for eksempel datakriminalitet.

Politiet i Kent har siden 2013 benyttet programvaren PredPol (som også bl.a. Los Angeles- og Atlanta-politiet i USA har tatt i bruk) for prediktiv analyse. To ganger daglig analyseres automatisk informasjon om leilighetsinnbrudd, gatebråk og tyverier fra biler – og sammenstilles med historiske data for de siste fem år. Ut fra dette genereres kartframstillinger over såkalte «hot-spots» hvor det er mer sannsynlig at kriminalitet vil foregå de neste 12 timer. Politiet innarbeider så disse kartene i valget av patruljeruter. Politiet har gode erfaringer med PredPol så langt, men understreker samtidig at kvalitet på analysene bestemmes av den informasjon som mates inn i systemet: jo mer relevant informasjon, jo bedre blir resultatet. Slik sett har PredPol-implementeringen et begrenset «scope» – bestående av tre typer informasjon: hvilken type kriminalitet det er, hvor det skjer, og når det skjer. Siden systemet ikke omfatter informasjon om lovbryterne (f.eks. hvor de bor og hvordan de ser ut) eller om ofrene (f.eks. hvor de bor og hva som kjennetegner befolkningen eller omgivelsene for hendelsene), er analysene begrenset. Man har altså unngått å bake inn sosioøkonomiske faktorer i modellene, bl.a. fordi dette lett kunne ført til (feilaktig) profilering av mennesker som kriminelle eller utsatte.[191]

Politiet i West-Yorkshire har utviklet sin egen programvare for å analysere sammensatte data om hus- og bilinnbrudd. De har hatt gode erfaringer med å bruke verktøyet til å identifisere «hot-spots» slik at de kan varsle beboere med hensyn til å forebygge innbrudd. Lignende erfaringer har også politiet i Manchester hatt med en tilsvarende tilnærming.

Innenfor mer generell teknologiutvikling har britene i samarbeid med Nederland og Sveits – og IBM – etablert et stort data- og forskningssenter kalt DOME i Manchester – hvor det i forbindelse med innsamling av enorme mengder astronomidata fra radioteleskop forskes på såkalt «exascale computing»[192], datatransport, lagring og «streaming» analyse[193].

Rammevilkår

Storbritannia har en klar plan for sin satsing på stordata, nedfelt i nevnte strategi fra 2013 – *Seizing the data opportunity*. Britene har også klare ambisjoner på EUs vegne, bl.a. vedrørende bedre tilgang til offentlige data – for innovasjon og viderebruk[194].

Stordata har ellers vært gjenstand for politisk oppmerksomhet, bl.a. på grunn av de forannevnte anvendelsene av teknologien i politiet. The Parliamentary Office of Science and Technology oppsummerte i 2014 mulighetene og utfordringene på områdene kriminalitetsbekjempelse og nasjonal sikkerhet[195]. De konkluderte da med at de verktøy som er tatt i bruk kan være et godt supplement i arbeidet med å beskytte samfunnets interesser på

[190] (ePOLICE, 2013)
[191] (abc, 2015)
[192] «Exascale computing» som begrep - se: (Wikipedia, Exascale Computing, 2015)
[193] «Streaming» analyse er det samme som sanntids dataanalyse; analyse av data som typisk «strømmer» fra sensorer, videokameraer og andre måleinstrumenter.
[194] (HM Government)
[195] (Houses of Parliament, 2014)

disse områdene, men oppsummerer samtidig med at det fortsatt er en noe begrenset evidens for at de aktuelle applikasjonene har bidratt til større effektivitet enn tradisjonelt etterforsknings- og forebyggingsarbeid.

Storbritannia er et foregangsland med hensyn til å utfordre og tilpasse lov- og regelverk til å understøtte en utvikling i retning av flere stordataanvendelser[196]. Når det gjelder IPR har man arbeidet aktivt for å revidere lovgivningen og åpne for flere unntak fra de opphavsrettslige reglene, for å understøtte mulighetene til å analysere data på tvers av kilder med ulike eierskap. Fra 2014 har loven en omfattende liste med unntak, bl.a. når det gjelder adgangen til tekstanalyse («text mining») og dataanalyse. Storbritannia har også hatt til formål å påvirke EU til å revidere sine direktiver i «Storbritannias retning».

På informasjonssikkerhets- og personvernområdet har myndighetene utarbeidet klare retningslinjer for datasikkerhet ved stordata-anvendelser. Det er også utviklet verktøy for å vurdere personvernfølgene av ulike typer anvendelser, veiledning om hvordan personvern kan bygges inn i IKT-løsningene («privacy by design»), veileder i god anonymiseringspraksis, og prinsipper for transparens. Akademia og myndigheter har gått sammen om å etablere et nettverk for anonymisering (UKAN – UK Anonymisation Network) som gir ekspertråd om anonymiseringsteknikker.

Når det gjelder åpne data – som er en grunnforutsetning for stordata i offentlig sektor – skriver Cap Gemini i 2013[197] at Storbritannia er den desiderte trend-setteren internasjonalt, med den høyeste tilgjengeligheten av åpne datasett innen EU. På Open Data Barometer (2013) er de rangert som nummer 1 i verden og de er øverst på europeiske oversikter om åpne data[198]. Hensikten med å gjøre dataene tilgjengelige for allmennheten er å stimulere til innovasjon. Etableringen av Open Data Institute[199] i 2012 har vært et av tiltakene for å bistå gründere med å skape verdi av offentlige data. Åpne data skal i tillegg bidra til en transparent offentlig forvaltning, og derved styrke demokratiet.[200]

Det er etablert prinsipper for åpne data, som i stor grad befester en brukerorientert tilnærming. De siste årene har fokus i økende grad vært å legge til rette for kobling av data fra åpne kilder, og at enkelte datasett utpekes til å være essensielle deler av en «nasjonal informasjonsinfrastruktur» - både for allmenn kommersiell og sosial utnyttelse, og for å understøtte en bedre effektivitet og etterprøvbarhet i offentlig sektor. Man er samtidig klar over de personvernrelaterte risikoer som ligger i åpne data, bl.a. gjennom muligheten for å re-identifisere personer ved å kombinere datasett som hver for seg ikke kan knyttes til person. I det siste har derfor myndigheten varslet at man vil se strengere på personvernproblematikken, noe som bl.a. innbefatter obligatorisk bruk av personvernfaglig ekspertise i vurderingen av om datasett skal publiseres, og testing for å avdekke den aktuelle risiko for re-identifisering.

3.5 Nederland

Stordata er et viktig satsingsområde for Nederland. Blant de områder som Nederland har tatt i bruk stordata på er forvaltning av vannressurser, intelligente trafikksystemer, såkalt

[196] (BYTE, Universitetet i Oslo, 2015)
[197] (Capgemini Consulting)
[198] (Open Data Barometer, 2013)
[199] (ODI Open Data Institute, 2015)
[200] (Open Data Barometer, 2013)

«predictive policing» og avdekking av økonomisk kriminalitet. Nederland var også (sammen med Storbritannia) tidlig ute med å eksperimentere med «foresights» og analyser av data på tvers av tradisjonelle sektorgrenser.

Erfaringene fra de nederlandske «foresight»-studiene har vært at brede strategiske scan gir verdifulle innspill til politikkutforming – i det de identifiserer, vurderer, og kløstrer trender, problemer og utviklingstrekk[201].

Anvendelser

Basert på et samarbeid med IBM og akademia etablerte nederlandske vannmyndigheter i 2013 «Digital Delta», et banebrytende innovasjonsprogram med utnyttelse av stordata for å få bedre kontroll med oversvømmelser og forvaltning av vannressursene. Digital Delta integrerer og analyserer vanndata (tidevannsnivå, vannkvalitet, demningers forfatning, avrenning, fordamping, radardata, strømforhold etc.) fra mange eksisterende datakilder, inkludert modellberegninger og ferske så vel som historiske vedlikeholdsdata fra sluser, pumper og dammer[202]. Digital Delta er bygget opp som et skybasert system rundt IBMs programvareløsninger (og deres konsulenttjenester) – med et «intelligent dashbord» som myndighetene kan bruke for å overvåke vannressursene og predikere oversvømmelser på en effektiv måte. Siden oversvømmelser er et problem også for mange andre land, setter Digital Delta-programmet en tydelig standard for hvordan stordata kan adressere og bidra til å håndtere alvorlige, globale utfordringer.

Amsterdam-området benytter stordata som en del av sitt «Smart City»-konsept[203]. Transport- og kommunikasjonsinfrastrukturen er sammen med stordata og åpne data en plattform for en rekke applikasjonsområder («Smart Living», «Smart Economy», «Smart Mobility», «Smart Areas» etc.). Stordata benyttes på mange områder i «økosystemet», f.eks. for å finne ledige parkeringsplasser i byen. Amsterdam Smart City er et unikt partnerskap mellom bedrifter, myndigheter, kunnskapsmiljøer og innbyggerne i Amsterdam.

Inspirert av bl.a. politiet i USA og Storbritannina som tidlig tok i bruk datamodeller for å analysere kriminaliteten og finne fram til mønster og sannsynligheter på ulike lokasjoner, har også politiet i Amsterdam tatt i bruk såkalt «predictive policing» som arbeidsmetode. Dette handler om å analysere historiske data om hendelser, og koble dem til andre inputdata – slik som avstand til vei, kommunikasjonsmidler, hvorvidt det finnes bedrifter i nærheten som tidligere har vært involvert i kriminell virksomhet, og ulike typer sosiodemografiske data[204]. Ved å sammenstille data på denne måten kan en identifisere «hot spots» og intensivere innsatsen og politikapasiteten i disse områdene.

Andre anvendelsesområder for stordata i Nederland er landbruk («smart matproduksjon»), e-helse (bl.a. innen kreftmedisinske områder) og miljøforvaltning. Innen privat sektor er bl.a. TIE Kinetix[205] (med kontorer både i Nederland og flere andre land) en fremstående leverandør av dataanalyse-løsninger for markedsføring og «e-business/e-commerce», med

[201] (Elsevier, 2010)
[202] (People 4 Smarter Cities, 2015)
[203] (Amsterdam Smart City, 2015)
[204] (issuu; Predictive Policing, 2015)
[205] (Tie Kinetix, 2015)

kunder innen mange samfunnsområder – som telekom., næringsmiddelindustri, varehandel, energimarked og helsesektor.

Netherlands eScience Center (NleSC) ble etablert i 2011 og er landets ekspertisesenter for utnyttelse av stordata i forskningskontekst. Utgangspunktet for senteret er at de store samfunnsmessige utfordringene i årene som kommer, knyttet til bl.a. klimaendringer, befolkningsvekst, en aldrende befolkning og urbanisering, vil gjøre forskerne mer og mer avhengige av data-drevne og beregningsintensive tilnærminger – altså stordata. Nylig ble senteret sikret en ny 5-årig finansiering.

Rammevilkår

Selv om det mangler en egen, uttrykt generell politikk for stordata, har den nederlandske regjeringen en offensiv politikk for åpne offentlige data[206]. I den strategisk orienterte «Vision on Open Data» fra 2013 legges det til grunn prinsipper om at offentlige data må tilgjengeliggjøres for viderebruk så fremt det ikke foreligger juridisk grunnlag for ikke å gjøre det. Det er igangsatt flere åpne data-prosjekter som skal bidra til vekst og innovasjon, bl.a. på geodata-området – jf. «veikart for åpne geodata»[207] Selv om det er stor variasjon mellom sektorene ligger Nederland generelt ganske godt an på de internasjonale rankingene over tilgjengeliggjøring av offentlige data.

Nederlenderne mener selv at de har en unik posisjon på stordata-området, bl.a. med utgangspunkt i den langvarige satsingen på datavitenskap og økonometri ved landets store universiteter.

Næringslivet, akademia og myndigheter har nylig inngått et tettere samarbeid for å bringe Nederland i tet internasjonalt på stordata-området. Sommeren 2015 ble således Big Data Alliance[208] etablert, som en samarbeids- og erfaringsutvekslingsarena for fagfolk og forskere på området. Alliansen har ambisjoner om å være et utgangspunkt og støtte for prosjekter innen en lang rekke områder – slik som smarte byer, havneforvaltning, energi, mobilitet, vannressursforvaltning, e-helse og e-handel.

3.6 Finland

Myndighetene i Finland synes å ha en forholdsvis offensiv tilnærming til stordata, selv om det er begrenset med eksempler på anvendelser i dag. I et oppslag[209] fra sommeren 2015 hevdes det at 3 av 5 store finske departementer og myndighetsorganer planlegger å benytte stordata for å understøtte sin løpende virksomhet og som beslutningsstøtte. Det synes likevel fortsatt å handle mest om satsinger i fremtiden, og det er vanskelig å finne noen vesentlige caser fra forvaltningen –med unntak fra helsesektoren (se nedenfor). Mangel på ekspertise og brukerkompetanse trekkes fram som en forklaring på at en ikke har kommet godt i gang ennå.

Åpne data er en viktig del av finsk IT-politikk, som en plattform og en forutsetning for en digital økonomi, offentlig tjenesteyting og effektive beslutninger[210]. Inntreden av stordata som teknologi kobles i stor grad til satsingen på åpne data – slik også flere andre land gjør.

[206] (Algemene Rekenkamer, 2014)
[207] (Rijksoverheid, 2014)
[208] (Big Data Alliance bda, 2015)
[209] (Management Events, 2015)
[210] (Finnish Ministry of Finance, 2015)

Anvendelser

De anvendelsesområder som finske myndigheter særlig trekker fram som mest aktuelle for stordata er områder de også for øvrig er sterke på:

- Telekommunikasjon
- Skole og utdanning
- Helsetjenester
- Smarte infrastrukturer
- Industrielle prosesser
- Forskning
- Trafikkstyring
- Ressursforvaltning

Det finnes allerede i dag løsninger hvor finnene kan vurdere kvaliteten på helsetjenestene på basis av åpne data fra et stort antall indikatorer og data som sier noe om ulike behandlingsformer og prosedyrer ved institusjonene. Selv om det kanskje ennå ikke kan defineres som stordata, er valg av helsetilbud et åpenbart anvendelsesområde for nye, stordata-baserte analyseverktøy og beslutningsstøtte rettet mot innbyggerne.

Et eksempel fra ressursforvaltningsområdet, men som foreløpig også bare er «på tegnebrettet», er neste generasjons konsept for forvaltning av de finske skogressursene. Visjonen for prosjektet *Forest Big Data* (FBD)[211] er et system som kan benyttes gjennom hele verdikjeden og som kan håndtere mer presis, helhetlig og oppdatert informasjon om skogressursene og relaterte miljøforhold, enn det som er tilgjengelig i dag. Visjonen forutsetter samarbeid mellom mange offentlige og private aktører, om tilgjengelighet til og kvalitet på data. Den forutsetter også at det utvikles verktøy og teknologier for raffinering av de enorme datamengdene om skogressursene, for utnyttelse i den daglige forretningsdriften. Dataene omfatter både åpne offentlige data, delte data og enkeltaktørers kommersielle data. Målet for et nytt forvaltningskonsept er en bedre logistikk og styring langs hele forsyningskjeden – både skogproduksjon, trevirkekjøp og –salg – noe som kan bidra til større konkurransekraft internasjonalt for hele den finske skogsindustrien.

Industriaktører som Nokia er generelt tungt inne i stordata-satsingene i privat sektor i Finland. Det nasjonale forskningsrådet Tekes[212] ser stordata som et svært viktig satsingsområde. Bl.a. gjennom «Industrial Internet» -programmet (2014-2019) allokeres store summer for å stimulere fremveksten av et industrielt datadrevet økosystem i Finland[213]. Det er også et stort antall private aktører i markedet som jobber med tjenester relatert til stordata og dataanalyse.

Det er etablert et offentlig-privat samarbeid knyttet til strategiske sentre for vitenskap, teknologi og innovasjon (såkalte «SHOK» -sentre på finsk[214]). Et av disse sentrene er Digile[215] – som er et selskap eid av et stort konsortium av aktører. Deres «Data to

[211] (Metsäteho, 2015)
[212] (Tekes, 2015)
[213] (Tekes, 2015)
[214] (Tekes, 2015)
[215] (DIGILE, 2015)

Intelligence» -program er direkte relatert til stordata og utvikling av nyttige verktøy og metoder som kan danne grunnlag for nye tjenester og ny virksomhet.

Man ser ikke bare for seg at Finland skal lykkes med nasjonale stordata-satsinger gjennom investeringer og nasjonalt samarbeid mellom industriaktører og andre interessenter. Stordata-applikasjoner og –tjenester vil også kunne være en høyst aktuell eksportartikkel, mener bl.a. Transport- og kommunikasjonsdepartementet. Samtidig påpeker bransjen selv at man fortsatt er kommet kort, og at det er et stort uutløst potensial[216].

Innen offentlig administrasjon vil stordata kunne gi betydelige fordeler, blant annet i form av mer brukerorienterte tjenester, effektivitet og bedre underlag for beslutninger.

Rammevilkår

Som for flere andre europeiske lands tilsvarende policydokumenter sier den digitale agendaen for Finland[217] utarbeidet i 2011, lite eksplisitt om stordata. Som påpekt ovenfor står stordata likevel høyt på den IT-politiske agendaen, om enn ikke så mye for offentlig sektors egen del som nå det gjelder næringsutvikling og verdiskapning i privat sektor.

Det finske samferdselsdepartementet nedsatte i 2013 en arbeidsgruppe for å skissere en strategi for stordata, og denne la fram sitt utkast til strategi høsten 2014[218]. Arbeidsgruppen pekte bl.a. på behovet for ekspertise og samarbeid. En forberedende undersøkelse gjennomført av departementet i 2013 viste at nettopp mangel på ekspertise – sammen med manglende datakvalitet – er den mest krevende enkeltfaktoren i en nasjonal stordata-tilnærming.

Arbeidsgruppen foreslo en rekke tiltak som bedre kan understøtte selskapers, myndigheters og enkeltpersoners utnyttelse av data. For å adressere tilgjengeligheten og markedsføringen av data foreslår arbeidsgruppen etableringen av en stordata-hub. Arbeidsgruppen mener også at det bør gis støtte til å initiere innovative stordata-anvendelser i samfunnet, både til SMB-virksomheter og til utviklingsmiljøer.

3.7 Sverige

Offentlig sektor i Sverige har ikke vært blant de fremste til å satse på stordata, uten at det er kjent hvorfor det er slik. Heller ikke i Sveriges digitale agenda[219] fra 2011 er stordata viet noe plass. I samtale med forsker Daniel Gillblad ved SICS[220] fremkommer det at fraværet av satsing i offentlig sektor er en indikasjon på generelt manglende modenhet på området i vårt naboland.

Vinnova[221] har riktignok fastlagt en såkalt forsknings- og innovasjonsagenda for stordata[222]. Fra denne agendaen vil de rette oppmerksomheten mot en strategisk satsing innen ITS-området (se nedenfor). Hvis man ser bort fra slike forskningsprosjekter har det ikke vært

[216] (Invest in Finland, 2015)
[217] (Productive and inventive Finland, 2011)
[218] (Ministry of Transport and Communications, Finland, 2014)
[219] (Regjeringskansliet, 2015)
[220] (Swedish ICT SICS, 2015)
[221] (Vinnova, 2015)
[222] (Vinnova, 2015)

mulig å identifisere «ordinære» caser innen offentlig sektor –som kan betegnes som stordata[223].

Anvendelser

Vinnova hevder at ”Big Data Analytics” er på sterk frammarsj også i Sverige, og at en rik flora av skybaserte medie- og datatjenester gir sluttbrukerne enkel tilgang til relevant og strukturert informasjon fra et enormt, sammensatt og i stor del ustrukturert kildemateriale. Et selskap som Klarna (finansiering og e-business) benytter seg av stordata til å vurdere risiko (ifb med finansiering). Svenske Spotify benytter stordata til å generere spillelister og foreslå musikk for brukerne[224].

Vinnovas perspektiver plasserer stordatasatsingen i Sverige langt i retning av tradisjonelle industrielle utfordringer. Som de påpeker, vil mange av framtidens tjenester bygge på tilgang til store datamengder. Tradisjonelle svenske kjerneområder som telekom, nettverksbaserte tjenester og industriautomatikk kan plassere Sverige i fremste rekke ved at selskapene evner å nyttegjøre seg av de enorme datamengdene som skapes.

Det caset som ellers trekkes fram av Vinnova er den pågående satsingen på intelligente trafikksystemer som ligger i BADA-prosjektet[225], og hvor samarbeidspartnerne omfatter Volvo, Scania og det svenske Trafikverket. I prosjektet ser man for seg et framtidig økosystem av sameksisterende transportenheter, som styrer seg selv og som prosesserer de store datamangder som produseres som en del av transporten. Stadig flere kjøretøy, førere og infrastrukturelementer er koblet opp i nettet, og stadig mer data om kjøretøyene og trafikksituasjonen gjøres tilgjengelig. Drivkraften bak framveksten av et slikt økosystem er effektivitet med hensyn til transportkapasitet, ressursutnyttelse, sikkerhet og miljøeffekter.

I den svenske hovedstaden Stockholm prøvde for øvrig IBM allerede i 2010 ut sine stordata-teknologier i en banebrytende anvendelse på trafikkdata sammen med forskere ved Kungliga Tekniska Högskolan, KTH[226]. Sanntidsdata fra en rekke kilder slik som GPS fra biler, radarsensorer langs hovedveiene, informasjon om veiarbeid og hendelser, luftkvalitet og værforhold, ble brukt til å beregne køer, tidsbruk og optimale transportruter mellom punkter i byen. I IBMs egen omtale av dette caset betegnes det som en ubetinget suksess, med betydelig samfunnsnytte som resultat –i form av reduserte transporttider og karbonutslipp[227].

Rammevilkår

Blant årsaker på den tilsynelatende mangelen på anvendelser innen svensk offentlig sektor pekes det bl.a. på de «vanlige» utfordringene; personvern, uklar lovgivning for skytjenester, for få åpne offentlige datasett og mangel på arbeidskraft med kompetanse på dataanalyse.

På policyutvikklingssiden mener Vinnova at det er viktig framover å:

- *gi økonomisk støtte til FoU-aktiviteter på stordata-området*

[223] Epost fra Daniel Gillblad, seniorforsker på SICS Swedish ICT (www.sics.se)
[224] (Computer Weekly, 2014)
[225] (Vinnova, 2015)
[226] (IBM, 2012)
[227] (IBM, 2010)

- *iverksette tiltak for å utvikle ulike typer stordata-anvendelser innen industrien, og etablere en infrastruktur og en forretnings- og serviceøkologi rundt stordata*
- *iverksette tiltak for å sikre tilgang på nøkkelkompetanse på området*
- *iverksette tiltak for å sikre effektiv kompetansedeling og erfaringsutveksling mellom aktører*

Som en modell for samvirke på området foreslås det at det etableres et sterkt, åpent nettverk av industrielle aktører og representanter for akademia som kan virke som koblingspunkt mellom interessenter. Det bør oppfordres til samarbeid både på teknologi-, anvendelses- og forretningsutviklingssiden, og på utdannings- og forskningssiden.

Svenske myndigheter synes å ha en noe mindre aktiv holdning til mulighetene med stordata, i hvert fall hva angår mulighetene for offentlig sektor selv. I en rapport fra et svensk utredningsinstitutt i 2013[228] påpekes det at det kan synes som om Sverige –relativt til andre land,skyver en større del av forventningene og ansvaret på området over på privat sektor. Av debatten knyttet til EU-kommisjonens strategi for stordata fra 2014 fremgår det at man først og fremst har vært opptatt av de forskningspolitiske aspektene.

3.8 USA

USA er ledende også på stordata-området. Myndighetene har i mange år samarbeidet tett med teknologileverandører (IBM, EMC, SAS, Amazon m.fl.) for å utvikle teknologier og infrastruktur for analyse av store, sammensatte datamengder. Viktige områder for amerikanske myndigheters innsats på området har vært å understøtte større grad av transparens, og å imøtekomme behovet for prediktiv analyse og forebyggende innsats på viktige samfunnsområder som kriminalitet, skatt, sikkerhet og helse. Etableringen av datahotellet Data.gov i 2009, med hundretusener av åpne datasett og som omhandler mange samfunnssektorer og temaer, har slik sett vært et godt utgangspunkt for mange applikasjoner i både privat og offentlig sektor –også for fremveksten av stordata-anvendelser. Det amerikanske statistiske sentralbyrået (US Census Bureau[229]) har gjort en stor innsats på å lenke opp stedfestede data om bedrifter og individer, og utviklet APIer for andres utnyttelse av disse datasettene.

Anvendelser

Obama-administrasjonen lanserte i 2012 en FoU-strategi for stordata som utgjorde startskuddet for en samlet tilnærming til dette nye teknologiområdet –som en felles satsing fra det hvite hus, det nasjonale forskningsrådet (NSF), det nasjonale helseforskningsinstituttet (NIH), Forsvarsdepartementet, militære forskningsprogrammer, Energidepartementet, Helsedepartementet (HHS) og de amerikanske Geologiske undersøkelser. Fokus for satsingen som fikk en bevilgning på $200 millioner var først og fremst å videreutvikle state-of-the-art teknologi for stordata, styrke forskningen innen naturvitenskap og teknologi, og å styrke den nasjonale sikkerheten. Videre var det et selvstendig mål å stimulere til utdannelse av flere med kompetanse innen utvikling og bruk av stordata-teknologier. I årene som har fulgt etter dette har bl.a. NIH samlet enorme datamengder for forskning og analyse –uten selv å måtte

[228] (Tilväxtanalys, 2013)
[229] (United States Census Bureau, 2015)

bygge opp noe særlig datakapasitet. Et eksempel på dette er hvordan de har benyttet seg av Amazon Web Services (AWS)-prosjektet «1000 Genom» som etter hvert er blitt et allment kjent stordata-prosjekt[230] i USA.

USA kan gjennom sin tidlige start på området fremvise ganske mange piloter på flere områder. Og det er ikke bare føderale myndigheter som har tatt i bruk stordata, USA kan også vise til en rekke lokale case. I byen Syracuse i staten New York har byens myndigheter – igjen sammen med IBM, tatt initiativ til et «Smarter City» -prosjekt med fokus på effektiv forvaltning av boligeiendommer, og hvor det å predikere og forhindre fraflytting og opphopning av ledige boliger er et viktig formål. IBM har de siste 4-5 årene gått tungt inn i flere tilsvarende prosjekter, med en sterkt sosial profil –og med bruk av sine teknologier, gjennom sin satsing «Smarter Cities® Challenge»[231].

Lokale frivillighetsarbeid nyter også godt av de nye teknologiene. Byen New York tok i 2011 initiativ til et «Center for Innovation through Data Intelligence» med sikte på å bruke stordata i arbeidet med å støtte ungdomsarbeidet i byen, i samarbeid med frivilligsektoren. Initiativet innebærer i dag at tilbudene og ordningene for utsatt ungdom er mer målrettede. De «siloene» som tradisjonelt vil prege oppgaveløsningen til ulike myndigheter, helsetjenester og organisasjoner, overvinnes ved hjelp av teknologi. Data fra et stort antall administrative databaser settes sammen med kartdata og analysene gis en lettfattelig visualisering for å gi råd om husvære, utdanningsmuligheter, helsetjenestetilbud og jobbmuligheter. Verktøyene gir også mulighet til å følge opp ungdommenes videre utvikling basert på åpne datakilder, og kan dermed gi kunnskap som igjen kan benyttes til å videreutvikle aktuelle sosiale støtteprogrammer og ordninger.

Også andre føderale myndigheter har i årene etter dette lansert sine egne stordata-programmer. Et eksempel som ofte trekkes fram er de amerikanske føderale skattemyndighetene (IRS) som har utviklet avanserte analyser av store datamengder for å avdekke skatteunndragelser og bedrageri[232]. I det såkalte Return Review Program (RPR) har IRS i samarbeid med teknologi fra IBM[233], også på delstatsnivå benyttet stordata til å analysere skattedata og mulig skattesvindel[234].

På helse området benyttes stordata i utstrakt grad både til forskning på sykdommer, men også til for eksempel dynamisk flåtestyring av ambulanser. Homeland security har også finansiert utvikling av stordataverktøy for tidlig varsling av bioterror/epidemier med grunnlag i alarmsentral (AMK) data.[235]

Andre områder for stordata i USA er bl.a. miljøområdet. Det nasjonale «Oceanic and Atmospheric Administration» kjørte i 2014 en omfattende RFI i markedet, hvor de stilte

[230] (1000 Genomes Project and AWS, 2015)
[231] (IBM's Smarter Cities Challenge, 2011)
[232] (Kim & Trimi, Big-Data Applications in the Government Sector. Communications of the ACM. Vol. 57, no. 3, 2014)
[233] (CPA Practice Advisor, 2015)
[234] (Governing.com, 2015)
[235] (Homeland Security News Wire, 2013)

leverandørmarkedet en lang rekke spørsmål om hvordan de burde realisere sine ambisiøse planer om å ta i bruk stordata.

En samlet oversikt over pågående stordata-prosjekter i regi av føderale myndigheter per 2012 finnes på nettsidene til det hvite hus[236].

En offentlig utredning fra mai 2014 – «Big data: Seizing opportunities, preserving values» anbefaler en fortsatt offensiv satsing på stordata, og at en kan imøtekomme de utfordringene samfunnet står ovenfor. Utgangspunktet er at det er riktig og viktig å satse offensivt, samtidig som tradisjonelle verdier som personvern skal ivaretas. Utredningen peker bl.a. på følgende områder som særlig interessante for stordata-anvendelser[237]:

- Helse og omsorgssektoren (prediktiv medisinering, diagnostisering, personlig helse, etc.)
- Skole og utdanning (såkalt learning analytics og optimalisering av utdanningstilbudet)
- Samfunnssikkerhet (bekjempelse av terrorisme, analyser av sammensatte trusselbilde basert på mange informasjonskilder)
- Kriminalitetsforebygging (organisert kriminalitet, predikering basert på analyser av omgivelsesdata, analyser av profiler til mistenkte etc.)

Rammevilkår

Personvern er på mange måter en relativt mindre utfordring i USA enn i mange andre land. Med en svakere personvernlovgivning enn i de fleste europeiske land er terskelen for dataanalyse på alle typer data lavere i USA. Det er ikke slik at personvernet er fraværende i amerikansk politikkutvikling; ofte er et sterkt personvern bygget spesifikt inn i de ulike sektorlovgivningene. Men det generelle fokuset er for USA sitt vedkommende i større grad på *forbrukervern*.

Det hvite hus har engasjert seg sterkt på området de siste 5-6 årene, og det legges fortløpende fram policydokumenter rettet mot stordata på områder som prising av offentlig informasjon og personvern. En av de ferskeste sakene er en redegjørelse for muligheter, utfordringer og dilemmaer knyttet til det å benytte stordata innen høyere utdanning, til å avdekke svak framdrift blant studentene (og derved kunne sette inn de riktige støttetiltak i tide).

På policyutviklingssiden er det som i andre land stor oppmerksomhet rettet mot tilgang til og viderebruk av offentlige data, som kilde for stordata-anvendelser. Åpne data er en tydelig politikk også i USA, men kanskje med noe mer av sitt rasjonale knyttet til transparens og kontroll med myndighetene enn det som er vanlig i Europa -hvor verdiskapingspotensialet har vært det helt sentrale. Det opphavsrettslige rammeverket i USA utfordres i mindre grad av nye teknologier som stordata, bl.a. fordi det har unntak for «fair use» som for eksempel dekker «data mining». USA har også et svakere vern av databaser enn det som er vanlig i EU-området.

Personvernet i USA er forankret i sektorlovgivning – ikke i et generelt juridisk rammeverk som her i Norge. Det konstitusjonelle personvernet omfatter kun myndighetenes eventuelle

[236] (Whitehouse.gov, 2012)
[237] (Whitehouse.gov, 2014)

inngripen i privatsfæren, ikke andre private aktørers inngripen. Sammenlignet med EU innebærer dette at gjen- og viderebruk av data i mye mindre grad begrenses av personvernlovgivningen. Amerikanske myndigheter publiserer også mye mer personrelaterte data enn det europeiske myndigheter gjør, noe som ytterligere underbygger amerikanske verdiøkere og analysemiljøers muligheter for å samle, sammenslå og utnytte data fra et bredt spekter av datakilder.

3.9 Sør-Korea

Sør-Koreas arbeid med stordata skjøt fart i 2012 da myndigheter, industriaktører, akademia og forskningsmiljøer kom sammen for å utvikle en strategi. Et «Big Data Forum» ble etablert, med deltakere fra både offentlig og privat sektor. Senere samme år vedtok man en «Big Data Master Plan», hvor kjernen er en trinnvis innføring av stordata på områder hvor fordeler og gevinster er mest åpenbare. Fokus er ellers støtte til politikkutvikling (ved å lenke og analysere åpne data fra ulike myndighetsområder, og koble dem med data fra private kilder). Planen innebærer også at det etableres en nasjonal infrastruktur for stordata, deling av erfaringer, teknisk support og utdanning av eksperter.

Anvendelser

Det har videre vært en del av planen å igangsette piloter fra 2013-2014. Disse tidlige pilotene inkluderer predikering og varsling av skogbranner, smart katastrofe- og beredskapskartlegging og analyser av medisinsk informasjon, og vil bli videreutviklet fram mot 2017 –sammen med ovennevnte felles infrastruktur. Koreas statistikkbyrå vil ha en viktig oppgave i denne pilotfasen. Fra 2015 utvides satsingen til helsetjenesteområdet og ytterligere 10 nye tjenesteområder. I 2016-2017 vil fiskeri og maritim virksomhet og ytterligere nye samfunnssektorer inkluderes i satsingen.

De case som omtales fra offentlig sektor omfatter for øvrig et bredt spekter av anvendelsesområder. Korrupsjon er en stor utfordring i mange asiatiske land. Et av de anvendelsesområder som vises til i flere sammenhenger er nettopp det å avdekke korrupsjon. Det nasjonale anti-korrupsjonsbyrået (ACRC) benytter stordata-teknikker for å analysere de ca. 4 millionene klager og tips de årlig får inn fra publikum[238]. Andre anvendelsesområder er værmeldingstjenester (med teknologi fra IBM)[239], analyser av klager på offentlige transporttilbud, boligmarkedstiltak, arbeidsløshetstiltak, vurderinger av industriell kapasitet og råvareforsyningen. Et område som har visse likhetstrekk med tidligere omtalte case i New York, er lokale myndigheters bruk av stordata for å analysere oppvekstsvilkår (basert på mange parametere / datakilder) for å forutsi mulige sosiale problemer i et gitt byområde.

Rammevilkår

Som en oppfølging av Government 3.0-initiativet i 2013 har koreanske myndigheter strevd for å bli en transparent, kompetent og serviceorientert statsforvaltning. Som en del av arbeidet med å bli mer transparent har myndighetene åpnet og delt en stor del av de offentlige dataene, og i det hele tatt åpnet seg mot publikum.

[238] (ACRC Korea annual Report 2014, 2014)
[239] (IBM, 2013)

Som et ledd i å bli mer kompetent er også de tidligere silogrensene mellom ulike deler av statsforvaltningen fjernet, for å muliggjøre interdepartementalt samarbeid. Stordata er introdusert som «vitenskapelig» verktøy og metode gjennomgående i administrasjonen.

3.10 Australia

Australske myndigheter har vært offensive på området åpne data, og gjennom lanseringen av Data.gov.au og andre satsinger framstått som et av de mer avanserte landene i undersøkelsen. De lanserte i 2013 også en egen Big data-strategi[240]. Sentralt i politikken står dataanalyse som verktøy i samfunnsplanleggingen.

Anvendelser

Australias økonomi er bl.a. avhengig av gruvedrift og landbruk, som begge er sektorer som i prinsippet kan ha nytte av stordata-analyser –i tillegg til helsetjenesten og øvrige deler av offentlig sektor[241]. Det er imidlertid vanskelig å finne mange konkrete caser som beskriver implementeringer av stordata i Australia. Mye av det som er tilgjengelig av informasjon om konkrete områder handler om framtidige muligheter. Et av unntakene i privat sektor er det globale gruvedriftselskapet Rio Tintos vellykkede bruk av sanntids dataanalyse fra driften av deres anlegg til fortløpende å identifisere driftsmessige avvik og foreslå prosessforbedringstiltak (ref. Rio Tintos Processing Excellence Center)[242].

Mye tyder på at utbredelsen av stordata har kommet kort i Australia. Myndighetene har i tråd med dette fokus på utprøving, kunnskapsutvikling og erfaringsutveksling.

Rammevilkår

Det er etablert lagringsfasiliteter i offentlig regi for å imøtekomme behovene for lagring av store datamengder.

Myndighetene har også lansert en beste praksis-veileder[243] for stordata-satsinger, som også vektlegger forutsetningene for å håndtere stordata (kapasitet, kompetanse etc.), informasjonssikkerhet og personvern -samt viktigheten av ansvarlighet i analysene. Et av de sentrale tiltakene for å adressere kompetanse- og innovasjonsutfordringene på området har vært å etablere et «Data Analytics Centre of Excellence» som et tverrsektorielt, fellesoffentlig og eksplorativt analysemiljø. Senteret som fysisk er organisert under australske skattemyndigheter, ble etablert i 2013 og har til oppgave å tilby kompetanse, rammeverk, verktøy for analyser og erfaringsutveksling på stordataområdet på tvers av departementsområdene.

På lovgivningsområdet har Australia gjort lite for å legge bedre til rette for stordata-anvendelser. Lite er så langt gjort med IPR-lovgivningen for å tilpasse denne til en digital økonomi. Det er heller ikke gjort unntak i opphavsrettslovgivningen for tekst- og «data mining», noe som naturligvis er et potensielt hinder.

[240] (Australian Government, 2013)
[241] (PWC Australia, 2014)
[242] (Rio Tinto, 2014)
[243] (Australian Government, 2015)

Personvernlovgivningen omfatter prinsipper som ligner EUs, med formålsbegrensninger, samtykke som basis for lovlig prosessering, transparens, og adgang til å se og eventuelt korrigere opplysninger om seg selv. I Australias Big Data Strategy er det definert 6 prinsipper. Et av prinsippene er «privacy by design». Et annet er dataintegritet og prosesstransparens. Dette innebærer et ansvar på alle som er involvert med hensyn til å ivareta personvernhensyn, sikre kvalitet i kildedata, opprettholde kontroll med personlige og sensitive data langs hele verdikjeden, og plikten til å gjennomføre risiko- og sårbarhetsvurderinger, kvalitetssikre («peer reviews») og gjennomføre høringer før man implementerer en ny stordata-anvendelse.

3.11 Utvalgte EU-initiativer

Under møte i Europarådet den 24. – 25. oktober 2013 ble det konkludert med at «det er et behov for å øke potensialet for stordata og datadreven innovasjon: Teknologier relatert til stordata er viktige for å øke produktiviteten og bedre tjenesteytingen»[244]. I juli 2014 lanserte Kommisjonen en ny strategi for «veien mot en blomstrende, datadrevet økonomi»[245].

De strategiske målene for initiativene i regi av EU er:

- Gripe mulighetene stordata tilbyr, og skape en «datarevolusjon» til nytte for innbyggere, økt økonomisk vekst, flere og bedre jobber, skreddersydde personlige produkter og tjenester av høy kvalitet.
- Ta igjen etterslepet i Europa når det gjelder evnen til å dra nytte av stordata
- Ivareta europeiske verdier i forhold til personvern

De prioriterte utfordringene er:

- Mangel på skalering i EU på området forskning og innovasjon.
- Fragmentert lovverk
- Mangel på kunnskap
- Forsinkelser i tilnærming til datateknologier
- Mangelen på et dataøkosystem med et samspill mellom flere økonomiske/samfunnsaktører.

Og de anbefalte fremgangsmåter er:

- Vektlegge datatilgjengelighet – tekniske standarder og lovverk
- Utvikle kompetanse – kompetansesentre og nettverk
- Samspill – fasilitere partnerskap og samarbeid på tvers av EU
- Teknologi – besvare utfordringer som måtte oppstå på hvert nivå i verdikjeden
- Prosessering og datalagring – støtte distribuerte e-infrastrukturer
- Personvern og beskyttelse av data – et balansert juridisk rammeverk for å øke tilliten til den digitale økonomien.

Med dette utgangspunktet har EU-kommisjonen finansiert en rekke prosjekter og initiativ hvor det jobbes for å oppnå de strategiske målene for stordata. Mot slutten av 2014 var det i

[244] (European Commission, 2014)
[245] (European Commission, 2014)

underkant av 150 aktive EU-prosjekter relatert til stordata. Det ble bevilget 50 millioner euro i 2014 og 89 millioner euro til dette formålet i 2015.

De fleste prosjektene er igangsatt innen sektorene helse, transport, miljø, jordbruk, tjenesteyting og salg[246].

Helse

Et eksempel på et prosjekt innen helse er DAPHNE[247]. Prosjektet baseres i hovedsak på sensorteknologi og hvor målet for initiativet er å skape en stordatastøttet plattform for å redusere inaktivitet og usunne vaner i befolkningen og konkrete pasientgrupper.

Transport og smarte byer

Transport-aspektet er nært knyttet til konseptet «smart cities» som bl.a. innebærer bruk av sensordata og GPS-data i samspill med for eksempel sosiale media for å optimalisere trafikkflyt og offentlig transport. VIAJEO[248] er et prosjekt med hensikt å bedre trafikkflyten i byer. Problemstillingen i prosjektet er at utfordringene i forbindelse med trafikkøkningen i mange europeiske byer ikke kan tilfredsstilles med utbygging av infrastruktur alene. Prosjektet ser for seg at i framtiden vil man i økende grad vil være avhengig av å kunne bruke stordata til å styre trafikkflyten via for eksempel alternative ruter eller at det kan settes inn mer kollektiv transport hvor og når det trengs.

Miljø

Stordata gir ny innsikt i miljømessige utfordringer. Med bredere tilgang på globale, nasjonale og lokale datasett kan forskere og beslutningstakere gjøre informerte og kvalitativt bedre beslutninger knyttet til for eksempel klimaendringer og lokale luftforurensninger. Miljøperspektivet har også en viktig plass i satsingene på «smart cities» - bl.a. med tanke på energisparing og mer effektiv bruk av fornybar energi. DC4Cities[249] er et initiativ for mer miljøvennlige datasentre, samt å gjøre de eksisterende sentrene mer miljøvennlige.

Landbruk

Med sensor- og stordatateknologi kan bønder ha tilgang til sanntidsdata på tilstanden på egen maskinpark, historiske værmønstre, topografi på eiendommen og status for avlingene. Dette vil kunne øke produktiviteten og inntektene i landbruket. E-agri[250] er et eksempel på EU-prosjekt hvor man benytter seg av stordata for å overvåke avlinger og predikere avkastningen. Prosjektet prøves ut i Marokko, Kina og Kenya.

Produksjon av varer, tjenester og salg

Stordata gir muligheter til å optimalisere prosesser relatert til produksjon og servicetilbud, til dels også i sanntid der dette er relevant. Stordata gjør det også mulig å svare mer direkte og presist på forbrukernes behov og ønsker, hvilket vil kunne øke lønnsomheten for en rekke

[246] (European Commission, 2015)
[247] (DAPHNE, 2015)
[248] (European Commission, CORDIS, 2012)
[249] (DC4Cities, 2015)
[250] (E-AGRI, 2015)

type virksomheter. Det er også et stort potensiale for høyere kvalitet og økt kvalitet under produktutviklingsprosessene dersom man kan benytte seg av stordata.

LinkedDesign[251] er et eksempel på et initiativ (med deltakelse fra norske Sintef, NTNU og Aker Solutions) som skal se kundebehov, design og produksjon av produkt i sammenheng.

Horizon 2020

Horizon 2020[252] er EUs største forsknings- og innovasjonsrammeprogram noensinne. Initiativet har et budsjett på oppimot 80 milliarder euro og skal forme og definere satsningen på forskning frem mot og etter 2020. Hovedpilarene i initiativet er «Excellent Science», «Industrial Leadership» og «Societal Challenges», herunder store satsninger innen stordata. I utkastet til arbeidsprogrammet for 2016-2017[253] legges det særlig vekt på satsninger mot informasjons- og kommunikasjonsteknologier. Det indikeres at det vil bli lagt til rette for større pilotprosjekter samtidig som at det vil bli lagt ned et vesentlig arbeid i å utrede rammevilkårene (for nye tjenester og produkter), slike som for eksempel personvern og teknologiske utfordringer.

Det kommer tydelig frem at Europarådet anser satsninger innen stordata som helt nødvendige. Dette for at man skal kunne konkurrere og oppnå økt verdiskapning i et verdenssamfunn hvor verdiskapning i økende grad vil avhenge av et svært raskt voksende informasjonsgrunnlag[254].

[251] (LinkedDesign, 2015)
[252] (Horizon 2020 Projects, 2015)
[253] (Horizon 2020 Work Programme 2016-2017, 2015)
[254] (European Council, 2013)

Vedlegg 4: Definisjoner

Anonyme opplysninger	Anonyme opplysninger kan defineres som opplysninger som det ikke er mulig å knytte til et identifiserbart individ, når man tar i betraktning alle de hjelpemidlene som med rimelighet kan tenkes brukt for å identifisere vedkommende. (Datatilsynet 2015).
Anonymisering	Å gjøre personopplysninger anonyme (Datatilsynet 2015).
Avidentifisering	At alle personentydige kjennetegn er fjernet fra opplysningene, slik at de ikke lenger kan knyttes til en enkeltperson. (Datatilsynet 2015).
Business Intelligence	Business Intelligence (BI) er en samlebetegnelse som omfatter applikasjoner, infrastruktur, verktøy og beste praksis som gjør det mulig å få tilgang til (via analyse) informasjon for å forbedre og optimalisere beslutninger og ytelse. (Gartner, u.d.) Business Intelligence er en samlebetegnelse som refererer til en rekke programmer som brukes til å analysere en organisasjons rådata. BI som fag består av flere beslektede aktiviteter, inkludert «data mining», online analytisk prosessering (OLAP), spørring og rapportering. Selskaper bruker BI til å forbedre beslutningsprosesser, kutte kostnader og identifisere nye forretningsmuligheter. (Mulchay, 2007)
Cloud computing	«Cloud computing» er en modell for praktisk, «on-demand» nettverkstilgang til en felles pool av konfigurerbare dataressurser (f.eks. nettverk, servere, lagring, applikasjoner og tjenester) som raskt kan klargjøres med minimal interaksjon med tjenesteleverandøren. Denne skymodellen er sammensatt av fem grunnleggende egenskaper, tre tjenestemodeller, og fire distribusjonsmodeller. Grunnleggende egenskaper: • On-demand selvbetjening (Tilgang ved behov, uten menneskelig interaksjon med tjenestetilbyder) • Bred nettverkstilgang (Funksjoner er tilgjengelige over nettverket) • Ressursgruppering (Leverandørers dataressurser slås sammen for å tjene flere brukere) • Rask elastisitet (Funksjoner kan bli elastisk klargjort, i noen tilfeller automatisk) • Målt tjeneste (Automatisk kontroll og optimalisering av ressursbruken) Tjenestemodeller: • SaaS (Software as a Sercice): Programvare som en tjeneste (Applikasjoner og tilhørende data er tilgjengelig som tjeneste via Internett) • PaaS (Platform as a Service): Plattform som en tjeneste (Skybasert dataplattform for utvikling og utrulling av programvare i nettskyen)

	• IaaS (Infrastructure as a Service): Infrastruktur som en tjeneste (Virtuelle eller fysiske hardwareressurser tilbys som en standardtjeneste via nettskyen) Distribusjonsmodeller: • Privat nettsky (Leveransemiljø bygget på nettskyprinsipper dedikert til en spesifikk kunde) • Nettskyfellesskap (Flere virksomheter går sammen om en felles nettskyløsning) • Offentlig nettsky (Tjenester tilgjengelig for alle virksomheter og levert fra en felles plattform) • Hybrid nettsky (En kombinasjon av modellene over) (NIST - National Institute of Standards and technology, 2011)
Data mining	«Data mining» er prosessen med å finne avvik, mønstre og sammenhenger i store datasett for å forutsi ulike utfall. Ved hjelp av et bredt spekter av teknikker, kan man bruke denne informasjonen til å øke inntekter, redusere kostnader, forbedre kunderelasjoner, redusere risiko osv. «Data mining» gir oss mulighet til å sile gjennom all kaotisk og repetitiv støy, forstå hva som er relevant og deretter bruke denne informasjonen til å vurdere sannsynlige utfall. (SAS Institute, u.d.)
Datadrevet forvaltning	Innenfor datadrevet forvaltning brukes analyse til å planlegge og effektivisere måten innbyggere betjentes på, blant annet ved å individualisere tjenester, målrette innsats og optimalisere håndtering av henvendelser og saksbehandling. Næringslivet vil også dra nytte av nye teknologier og muligheter som datadrevet forvaltning bringer. Blant annet fordi flere offentlige data gjøres tilgjengelig for private selskaper - med borgere og selskapers rett til privatliv og sikker lagring av data. Men også fordi det skaper mulighet for å redusere produktivitetsgapet mellom offentlig og privat sektor, for eksempel i form av nye tjenesteleveransemodeller. (Deloitte, u.d.)
Datadrevet innovasjon	Datadrevet innovasjon dreier seg om å utnytte interne og eksterne data bedre. Dette kan for eksempel være å bruke salgs- og kundedata til å segmentere kunder, få oversikt over deres preferanser og målrette markedsføringen. Stordata kan i så måte være en sentral faktor i utviklingen av nye produkter, tjenester og forretningsmodeller. I takt med at store mengder data tilgjengeliggjøres øker mulighetene for å samle mer nøyaktig og detaljert informasjon, -gjøre smalere segmenteringer og forbedre beslutninger ved mer avanserte analyser. Når selskaper systematiserer og strukturerer egne data kan det også oppstå forretningsmuligheter knyttet til salg (eller frigivelse av data fra et åpne data-perspektiv). (Viden Metro, 2014) (Manyika, et al., 2011)
Datavarehus	Et datavarehus er en temaorientert, integrert, tidsvarierende og ikke-flyktig samling av data til støtte for ledelsens beslutningsprosess. I et datavarehus samler man inn, kvalitetssikrer og konsoliderer data fra de ulike kildesystemene i en bedrift. Dette gir et helhetlig

	og pålitelig grunnlag for videre analyse og rapportering, som igjen sikrer god og konsistent informasjon ut til beslutningstakere. Datavarehus er fundamentet for konsistent og kvalitetssikret Business Intelligence. (BedreInnsikt.no, u.d.)
DevOps	DevOps er ofte et forum (kan være basert på forskjellige medier) bestående av software utviklere og andre IT-ressurser. Formålet er å skape en plattform for å øke samarbeid og kommunikasjon hvor man automatiserer prosessene rundt programvare-releaser og infrastrukturendringer. Målet er å skape en kultur og miljø hvor utvikling, test og utgivelse av software kan skje raskere, oftere og sikrere. (Wikipedia)
Hadoop	Hadoop er et åpen-kildekode-rammeverk for lagring av data og kjøring av applikasjoner på klynger av maskinvare. Det gir massiv lagringsplass for alle typer data, enorm prosessorkraft og evne til å håndtere nesten ubegrensede oppgaver samtidig. (SAS Institute, u.d.)
Interoperabilitet	Virksomheten og dens IT-løsninger må ved behov kunne samhandle med andre relevante virksomheter og deres IT-løsninger på et hensiktsmessig nivå. Det deles inn i tre ulike typer interoperabilitet: • Organisatorisk interoperabilitet: Innebærer samordning av arbeidsprosesser, avtaleverk og endringer av organisatoriske forhold nødvendig for samhandling • Semantisk interoperabilitet: Innebærer å avklare meningsinnholdet for informasjonselementene som utveksles. • Teknisk interoperabilitet: Innebærer å bruke tekniske standarder som legger til rette for veldefinerte grensesnitt, overføringsprotokoller og formater. (Difi, 2012)
Metadata	Metadata er data som definerer og beskriver andre data og prosesser, og dataene blir metadata når de brukes på denne måten. Det sett av omstendigheter og formål for hvilket enkelte data er brukt som metadata kalles kontekst. (OECD, 2002)
NoSQL	NoSQL omfatter et bredt spekter av forskjellige databaseteknologier som ble utviklet som svar på en økning i volumet av data lagret om brukere, gjenstander og produkter, frekvensen der denne informasjonen er tilgjengelig, og ytelses- og behandlingsbehov. NoSQL benyttes ofte til lagring og gjenfinning av data i XML-baserte dokumenter. Relasjonsdatabaser, derimot, ble ikke laget for å takle de skala- og smidighetsutfordringene som møter moderne applikasjoner, og de var heller ikke bygget for å dra nytte av billig lagringsplass og prosessorkraft tilgjengelig i dag. (MongoDB, u.d.)
OLAP (Online Analytical Processing)	OLAP (Online Analytical Processing) er teknologien bak mange Business Intelligence-applikasjoner som datavarehus. OLAP er en kraftig teknologi for dataoppdagelse, inkludert muligheter for grenseløs rapportvisning, komplekse analytiske beregninger, og logisk «hva hvis»-scenarioplanlegging av budsjett og prognose. (OLAP, u.d.)
Ontologi	Ontologi er i datateknologien og informasjonsvitenskap en formell representasjon av et sett av begreper innenfor et domene,

	samt hvordan disse begrepene er relatert til hverandre. En ontologi brukes til å markere dataene sin betydning og lette smart dataprosessering av dataprogrammer i domenet (Wikipedia).
Pseudonymisering	At enkelte direkte identifiserende parametere erstattes med pseudonymer, som fremdeles vil være unike indikatorer (Datatilsynet 2015).
Stordata (Big Data)	*Analyse av massive samlinger av data (volume), med stor variasjon i datakilder og formater (variety), hvor datasettet oppdateres med høy frekvens (velocity) og hvor grunnlagsdataenes opprinnelse og kvalitet er avklart (veracety) og hvor analysen av data samlingen gir økt verdi i forhold til rådataene (value).*
Structured Query Language (SQL)	«Structured Query Language (SQL) er et standardisert databasespråk for tilgang til relasjonsdatabaser, og i senere tid også objekt-relasjonsdatabaser. Standarden er definert både av ISO og ANSI, men omtales av sistnevnte som Standard Query Language.» (itpro.no)
Strukturerte og ustrukturerte data	Strukturerte data kan beskrives som data som følger en formell struktur av datamodeller, slik som rader og kolonner i en relasjonsdatabase, for eksempel et kunderegister. Ustrukturerte data refererer til data som mangler en gjenkjennelig struktur. For eksempel regnes bilder, video, e-post, Word-dokumenter og ren tekst for å være ustrukturerte data innen et datasett eller fil. Semistrukturerte data er en form for strukturerte data som ikke følger den formelle strukturen av datamodeller. Deler har fast format og deler består av fritekst. Det er ofte kjent som skjemaløs eller selvdokumenterende struktur. Semistrukturerte data er gjerne XML-filer, bøker, websider, e-post og EDI – ofte data innsamlet fra sensorer og maskiner. (Datatilsynet, 2013)
Text mining	Text mining er tekstanalyse og kjennetegnes ved at man henter ut informasjon av høy kvalitet fra ulike tekstbaserte datakilder slik at man kan påvise mønstre og sammenhenger (fungerer i stor grad på samme måte som «data mining»).
Tingenes internett	Tingenes internett (The Internet of Things (IOT)) er nettverket av fysiske objekter eller "ting" innebygd med elektronikk, programvare, sensorer og nettverkstilkobling, som gjør at disse objektene kan samle inn og utveksle data. Tingenes internett gjør at objektene kan identifiseres og fjernstyres over eksisterende nettverksinfrastruktur og skape muligheter for mer direkte integrasjon mellom den fysiske verden og datamaskinbaserte systemer, og som resulterer i økt effektivitet, nøyaktighet og økonomiske gevinster. (Wikipedia (engelsk) 2015).
Web Crawling	En «web crawler» eller web krabber er en web eller internett robot som systematisk søker gjennom internett med hensikt å indeksere innholdet. Benyttes av søkemotorer el.l. for å bedre søkehastighet m.m. (Wikipedia) .
Åpne data	Åpne data er gjort tilgjengelig slik at de kan leses og tolkes av både maskin og menneske, slik at alle som vil kan viderebruke dataene direkte i sine egne datasystem. Dataeiere kan ikke kreve

	mer enn at man navngir og/eller deler videre på samme vilkår. De viktigste kravene til åpne data er: • Tilgjengelighet og tilgang: Dataene må være tilgjengelige i sin helhet, helst som gratis nedlasting via Internett. Dataene må også være tilgjengelige i en enkel form og mulig å redigere. • Deling og viderebruk: Dataene må gjøres tilgjengelige under vilkår som tillater deling og viderebruk, også når det kombineres med andre datasett. • Universell deltakelse: Alle skal ha muligheten til å bruke og dele – det skal ikke være noen diskriminering av enkelte arbeidsområder, enkeltindivid eller grupper. Det er ikke tillatt med restriksjoner som hindrer kommersiell bruk, eller restriksjoner for enkelte bruksområder. • Med data mener vi enkle biter av informasjon. Når mange slike biter med data struktureres og settes sammen, får vi et datasett. • Offentlige data er all data som er organisert og arkivert i offentlig sektor. Offentlige data er enten samlet inn utenfra eller blir utarbeidet av de ansatte i en statlig virksomhet. Slike data kan for eksempel være næringslivsregistre, kartdata, organisasjonsmodeller, budsjett, årsregnskap og lignende. • Maskinlesbare data er data i et format som kan leses og bearbeides av en datamaskin og dermed lett kan deles på tvers av IT-systemer. Maskinlesbare dataformater, som XML, JSON og CSV, egner seg derfor godt til viderebruk. Eksempler på dataformater som ikke er maskinlesbare, er bildefiler og skannede dokumenter. (Difi, 2012)

Litteratur

1000 Genomes Project and AWS. (2015). Hentet fra https://aws.amazon.com/1000genomes/

abc. (2015, 08 26). Hentet fra http://www.abc.net.au/radionational/programs/futuretense/can-software-help-police-predict-and-prevent-crime/6723222

ACRC Korea annual Report 2014. (2014). *ACRC Korea annual Report 2014.* Hentet fra http://www.acrc.go.kr/eng/file/file.do?command=downFile&encodedKey=MjkwOTVfMg%3D%3D

Algemene Rekenkamer. (2014). Hentet fra http://www.courtofaudit.nl/english/Publications/Audits/Introductions/2014/03/Open_Data_Trend_Report

Algemene Rekenkammer. (2014). *Open Data Trend Report.* Hentet fra http://www.courtofaudit.nl/english/Publications/Audits/Introductions/2014/03/Open_Data_Trend_Report

Altinn, A-ordningen. (2015). Hentet fra https://www.altinn.no/en/a-ordningen/

Amazon. (2015). *Amazon.com*. Hentet fra aws.Amazon.com: http://aws.amazon.com/elasticmapreduce/

Amazon Simpledb. (2015). Hentet fra https://aws.amazon.com/simpledb/

Amsmarterdam City. (2015). Hentet fra http://amsterdamsmartcity.com/projects/theme/label/big-open-data

Amsterdam Smart City. (2015). Hentet fra http://amsterdamsmartcity.com/projects/theme/label/big-open-data

Apache. (2015). *Hadoop*. Hentet fra Hadoop Apache: http://hadoop.apache.org/

Apache. (2015). *Mahout*. Hentet fra http://mahout.apache.org//

Apache Cassandra. (2015). Hentet fra http://cassandra.apache.org/

Apache Hbase. (2015). Hentet fra https://hbase.apache.org/

Australian Government. (2013, 08). *The Australian Public Service Big Data Strategy.* Hentet fra http://www.finance.gov.au/sites/default/files/Big-Data-Strategy_0.pdf

Australian Government. (2015). *Australian Public Service Better Practice Guide for Big Data.* Hentet fra http://www.finance.gov.au/sites/default/files/APS-Better-Practice-Guide-for-Big-Data.pdf?v=1

Barlow, M. (2014). *Data Vizualisation - A New Language For Storytelling.* Bejing, Cambridge, Farnham, Køln, Sebastopor, Tokyo : O'Railly Media.

BedreInnsikt.no. (u.d.). *www.BedreInnsikt.no*. Hentet fra Datavarehus: http://www.bedreinnsikt.no/tags/datavarehus

Big Data Alliance. (2015). Hentet fra http://www.bigdata-alliance.org/aboutbda/

Big Data Alliance bda. (2015). Hentet fra http://www.bigdata-alliance.org/aboutbda/

Big Data for Europe, EU. (2013, 11 07). Hentet fra http://europa.eu/rapid/press-release_SPEECH-13-893_en.htm

Bull-Berg, H., & Olsson, N. (2014). *Nye datakilder i evaluering av store statlige investeringer – et potensial for Big Data?* Trondheim: SINTEF, Cocept programmet. Hentet fra http://www.ntnu.no/documents/1261860271/1262021752/054_Rapportutkast_Endelig%20versjon.pdf

Byte. (2014). *Big Data Policies.* Hentet fra http://byte-project.eu/wp-content/uploads/2014/09/D1.2-Big-Data-Policies-FINAL.pdf

BYTE, Universitetet i Oslo. (2015). Hentet fra www.byte-project.eu

Capgemini Consulting. (u.d.). *The Open Data Economy; Unlocking economic Value by Opening Government and Public Data.* Hentet fra https://www.capgemini-consulting.com/resource-file-access/resource/pdf/opendata_pov_6feb.pdf

Chris Yiu. (2012). *The Big data Opportunity.* Hentet fra http://www.policyexchange.org.uk/images/publications/the%20big%20data%20opportunity.pdf

CISCO. (2015). *Visual Networking Index (VNI) Global Forecast, 2014 - 2019*. Hentet fra Visual Networking Index (VNI): http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html

Clay Labs. (2015). Hentet fra http://claylabs.com/portfolio

Cloudera. (2015). Hentet fra http://www.cloudera.com/content/www/en-us.html

Computer Weekly. (2014). *Swedens' Klarna and Spotify design big data architectures to aid business growth*. Hentet fra http://www.computerweekly.com/feature/Swedens-Klarna-and-Spotify-design-big-data-architectures-to-aid-business-growth

Computer, how stuff works. (2015). Hentet fra http://computer.howstuffworks.com/internet/basics/search-engine1.htm

Copernicus. (2015). Hentet fra http://www.copernicus.eu/

CPA Practice Advisor. (2015). Hentet fra http://www.cpapracticeadvisor.com/news/12117971/irs-return-review-program-improves-detection-of-fraud

DAPHNE. (2015). Hentet fra http://www.daphne-fp7.eu/

Data Norge. (2015). *Veileder i tilgjengeliggjøring av offentlige data*. Hentet fra http://data.norge.no/document/del-og-skap-verdier-veileder-i-tilgjengeliggj%C3%B8ring-av-offentlige-data/hva-er-%C3%A5pne-data

DATA.GOV.UK. (2015). Hentet fra www.data.gov.uk

Data.Norge.No. (2015). *http://data.norge.no/data/statens-vegvesen/nasjonal-vegdatabank-api*. Hentet fra http://data.norge.no/data/statens-vegvesen/nasjonal-vegdatabank-api: http://data.norge.no/data/statens-vegvesen/nasjonal-vegdatabank-api

Datatilsynet. (2013). *Big Data - personvernprinsipper under press.* Oslo: Datatilsynet. Hentet fra https://www.datatilsynet.no/Global/04_planer_rapporter/Big%20Data_web.pdf

Datatilsynet. (2013, 09 26). *Big Data - Personvernprinsipper under press.* Hentet fra https://www.datatilsynet.no/Teknologi/Big-Data/

Datatilsynet. (2015). *ANONYMISERING AV PERSONOPPLYSNINGER, VEILEDER.* OSLO.

Datavisualization.ch. (2015). *https://datavisualization.ch/about/*. Hentet fra https://datavisualization.ch/about/: https://datavisualization.ch/about/

DC4Cities. (2015). Hentet fra http://www.dc4cities.eu/en/

Dean, J., & Ghemawat, S. (2004). *MapReduce: Simplified Data Processing on Large Clusters.* Hentet fra Googleusercontent: http://static.googleusercontent.com/media/research.google.com/no//archive/mapreduce-osdi04.pdf

Deloitte. (u.d.). *www.deloitte.com*. Hentet fra Indsigt - Datadrevet offentlig forvaltning: http://www2.deloitte.com/dk/da/pages/deloitte-analytics/articles/Datadrevet-offentlig-forvaltning.html

Difi. (2012). *Overordnede IT-arkitekturprinsipper for offentlig sektor.* Oslo, Leikanger: Direktoratet for forvaltning og IKT.

Difi. (2012, Oktober 18). *www.data.norge.no*. Hentet fra Open Knowledge Foundation, oversatt av Difi: http://data.norge.no/aktuelt/2012/10/hva-menes-med-åpne-data

Difi, Informasjonsforvaltning i offentlig sektor. (2013). Hentet fra https://www.difi.no/sites/difino/files/rapport-informasjonsforvaltning-i-offentleg-sektor-2013-10-10.pdf

Difi, LivsIT. (2007). Hentet fra http://los.difi.no/sites/los/files/H%C3%A5ndbok%20for%20innf%C3%B8ring%20av%20Los.pdf

Difi, Åpne data. (2015). Hentet fra https://www.difi.no/veiledning/ikt-og-digitalisering/apne-data

DIGILE. (2015). Hentet fra http://digile.fi/en/services/

Direktoratet for forvaltning og IKT. (2015, oktober). *www.datanorge.no*. Hentet fra http://data.norge.no/kategori/transport-og-kommunikasjon?type=3

Dygraphs. (2015). *http://dygraphs.com/index.html*. Hentet fra http://dygraphs.com/index.html: http://dygraphs.com/index.html

E.F.Codd. (1970). *A Relational Model of Data for Large Shared Data Banks.* New York: Communication with ACM. Hentet fra http://www.seas.upenn.edu/~zives/03f/cis550/codd.pdf

E-AGRI. (2015). Hentet fra http://www.e-agri.info/

Elastic, Elasticsearch. (2015). Hentet fra https://www.elastic.co/products/elasticsearch

Elasticsearch. (2015). *elastic.co*. Hentet fra elastic.co/guide: https://www.elastic.co/guide/index.html

Elsevier. (2010). *Strategic foresight in public policy: Reviewing the esperiences of the UK, Singapore, and the Netherlands.* Hentet fra https://saudeamanha.fiocruz.br/sites/saudeamanha.fiocruz.br/files/pdf%20-%20propspec%C3%A7ao.pdf

ePOLICE. (2013). Hentet fra https://www.epoolice.eu/

EU. (2014). *Towards a data-driven economy.* Hentet fra http://ec.europa.eu/digital-agenda/en/towards-thriving-data-driven-economy

EU. (2015). *What can big data do for you?* Hentet fra http://ec.europa.eu/digital-agenda/en/what-big-data-can-do-you

EU ITS Action Plan. (2015). Hentet fra http://ec.europa.eu/transport/themes/its/road/action_plan/

EU, ITS Directive. (2010, 07 07). *DIRECTIVE 2010/40/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL.* Hentet fra http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:207:0001:0013:EN:PDF

European Commission. (2014). Hentet fra http://ec.europa.eu/digital-agenda/en/towards-thriving-data-driven-economy

European Commission. (2014, 09 10). *Recent EU initiatives in the field of (big) data: Communication on data-driven economy.* Hentet fra http://webcache.googleusercontent.com/search?q=cache:P3Mwfi33zQcJ:ec.europa.eu/newsroom/dae/document.cfm%3Fdoc_id%3D7275+&cd=2&hl=en&ct=clnk&gl=no

European Commission. (2015). Hentet fra http://ec.europa.eu/digital-agenda/en/what-big-data-can-do-you

European Commission. (2015). Hentet fra https://ec.europa.eu/digital-agenda/en/connecting-europe-facility

European Commission. (2015). *Connecting Europe - Supporting document for the Information session on CEF building block - Digital Service Infrastructures (DSI).* Oslo, Februar 2015: European Commission.

European Commission, CORDIS. (2012). Hentet fra http://cordis.europa.eu/result/rcn/89972_en.html

European Council. (2013, 10 25). Hentet fra http://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ec/139197.pdf

Finnish Ministry of Finance. (2015). *Open Data Policy in the Finnish Government.* Hentet fra http://vm.fi/documents/10623/360816/Avoimen+tiedon+tavoitteet+2015-2020/c7e9c09c-c492-4f04-ac52-2449b0c4973b

Forbes. (2014, 06 26). Hentet fra http://www.forbes.com/sites/emc/2014/06/26/the-hottest-jobs-in-it-training-tomorrows-data-scientists/

Forskningsrådet. (2014, 10 20). Hentet fra http://www.forskningsradet.no/prognett-nevronor/Nyheter/_Ingenting_er_mer_interessant_enn_hjernen/1254001149420/p1224698072633

Forskningsrådet. (2014, 10 20). Hentet fra http://www.forskningsradet.no/prognett-nevronor/Nyheter/_Ingenting_er_mer_interessant_enn_hjernen/1254001149420/p1224698072633

FusionCharts. (2015). *http://www.fusioncharts.com/*. Hentet fra http://www.fusioncharts.com/: http://www.fusioncharts.com/

Gartner. (u.d.). Hentet fra http://www.gartner.com/technology/home.jsp

Gartner. (2013, 06 12). Hentet fra http://www.gartner.com/newsroom/id/2513615

Gartner. (2013, 06 12). *Gartner says by 2019, 90 percent of organizations will have personal data on IT systems they don't own or control.* Hentet fra http://www.gartner.com/newsroom/id/2513615

Gartner. (2015, 06). *Gartner.com*. Hentet fra Big Data Survey.

Gartner. (u.d.). *IT Glossary*. Hentet fra Business Intelligence: http://www.gartner.com/it-glossary/business-intelligence-bi/

Getoor, L., & Taskar, B. (2009). *Introduction to Statistical Relational Learning.* Boston: MIT Press. Hentet fra www.cs.umd.edu/srl-book/.

GitHub. (2015). *https://github.com/tornadoweb/tornado*. Hentet fra https://github.com/tornadoweb/tornado: https://github.com/tornadoweb/tornado

Google. (2015). *Google Cloud Platform*. Hentet fra https://cloud.google.com/prediction/docs?csw=1

Governing.com. (2015). Hentet fra http://www.governing.com/columns/tech-talk/gov-states-big-data-tax-fraud.html

Helse og omsorgs departementet. (2011). *NOU: Innovasjon i omsorg.* Oslo: Helse og omsorgs departementet.

Helse SørØst. (2014). *Nasjonal utredning av persontilpasset medisin i helsetjenesten.* Hentet fra http://www.helse-sorost.no/fagfolk_/forskning_/persontilpasset-medisin_/Documents/juridiske%20og%20etiske%20vurderinger.pdf

Helsedirektoratet. (2012). *Velferdsteknologi. Fagrapport om implementering av velferdsteknologi i de kommunale helse- og omsorgstjenestene 2013-2030.* . Oslo: Helsedirektoratet.

Helsedirektoratet. (2015, Mars 20). *Forskrift om krav til og organisering av kommunal legevaktordning, ambulansetjeneste, medisinsk nødmeldetjeneste m.v. (akuttmedisinforskriften) Fastsatt ved kgl.res. 20. mars 2015.* Hentet fra Regjeringen.no:

https://www.regjeringen.no/contentassets/359efd7f3b6441488ba8d260e47a7ae0/forskrift_akuttmedisinforskriften_200315.pdf

HM Government. (2013). Hentet fra Seizing the data opportunity: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/254136/bis-13-1250-strategy-for-uk-data-capability-v4.pdf

HM Government. (2013). *Seizing the data opportunity.* Hentet fra https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/254136/bis-13-1250-strategy-for-uk-data-capability-v4.pdf

HM Government. (2015). Hentet fra https://www.gov.uk/government/collections/foresight-projects#horizon-scanning-reports

HM Government. (u.d.). *UK vision for the EU's digital economy.* Hentet fra https://engage.number10.gov.uk/digital-single-market/uk-vision-for-eu-digital-economy-single-market.pdf

Homeland Security News Wire. (2013, 07 29). Hentet fra http://www.homelandsecuritynewswire.com/dr20130728-north-carolina-s-biothreat-warning-system-receives-funding

Horizon 2020 Projects. (2015). Hentet fra http://horizon2020projects.com/

Horizon 2020 Work Programme 2016-2017. (2015). Hentet fra https://ec.europa.eu/programmes/horizon2020/sites/horizon2020/files/05i.%20LEIT-ICT_2016-2017_pre-publication.pdf

Houses of Parliament. (2014). *Big Data, Crime and Security.* Hentet fra http://researchbriefings.files.parliament.uk/documents/POST-PN-470/POST-PN-470.pdf

IBM. (2010). Hentet fra https://www-03.ibm.com/press/us/en/pressrelease/29903.wss

IBM. (2012). *Big Data & Analytics Hub.* Hentet fra Travel and Transportation in the Age of Big data: http://www.ibmbigdatahub.com/blog/travel-and-transportation-age-big-data

IBM. (2013). *IBM Big Data Technology Helps South Korea's Meteorological Administration Increase Accuracy of weather Forecasting.* Hentet fra https://www-03.ibm.com/press/us/en/pressrelease/40411.wss

IBM. (2013). *IBM Big Data Technology Helps South Korea's Meteorological Administration Increase Accuracy of Weather Forecasting.* Hentet fra https://www-03.ibm.com/press/us/en/pressrelease/40411.wss

IBM. (2014). *Big Data Beyond the Hype: A Guide to Conversations for Today's Data Center.* San Francisco: IBM. Hentet fra http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

IBM. (2015). *Bringing big data to the enterprise*. Hentet fra http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

IBM. (2015, 05 01). *Harnessing Big Data, The Cloud And Watson Build The Car Of The Future*. Hentet fra http://www.forbes.com/sites/ibm/2015/05/01/harnessing-big-data-the-cloud-and-watson-build-the-car-of-the-future/

IBM Informix. (2015). Hentet fra https://www-01.ibm.com/software/data/informix/

IBM's Smarter Cities Challenge. (2011). *Syracuse Summary Report.* Hentet fra http://smartercitieschallenge.org/wps/wcm/connect/61d30e804127b68480bde55e2785d2db/SmarterCities-Syracuse.pdf?MOD=AJPERES&CACHEID=61d30e804127b68480bde55e2785d2db

Innlandsutvalget. (2015). Hentet fra http://innlandsutvalget.no/

INSPIRE, European Commision. (2015). Hentet fra http://inspire.ec.europa.eu/

InstantAtlas. (2015). *http://www.instantatlas.com/*. Hentet fra http://www.instantatlas.com/: http://www.instantatlas.com/

Invest In Finland. (u.d.). Hentet fra http://www.investinfinland.fi/industries/rd-and-innovation/big-data-in-finland/123

Invest in Finland. (2015). Hentet fra http://www.investinfinland.fi/industries/rd-and-innovation/big-data-in-finland/123

IQmulus. (2015). Hentet fra www.iqmulus.eu

issuu; Predictive Policing. (2015). Hentet fra http://issuu.com/rutgerrienks/docs/predictive_policing_rienks_uk

ITpro. (2004). Hentet fra http://itpro.no/artikkel/7162/definisjon-hva-er-sql/

James Manyika, M. C. (2013). *Open data: Unlocking innovation and performance with liquid information.* London, San Francisco, Shanghai, Washington DC, New Jersey : The McKinsey Global Institute. Hentet fra http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information

Kim, G.-H., & Trimi, S. &.-H. (2014). *Big-Data Applications in the Government Sector. Communications of the ACM. Vol. 57, no. 3.*

Kim, G.-H., & Trimi, S. &.-H. (2014). *Big-Data Applications in the Government Sector. Communications of the ACM. Vol. 57, no. 3.*

Kystverket "Havbase". (2011). *"Havbase"*. Hentet fra http://havbase.kystverket.no/havbase_report/doc/Havbase.pdf

Kystverket. (2015, mai 27). *www.kystverket.no*. Hentet fra http://www.kystverket.no/Maritime-tjenester/Meldings--og-informasjonstjenester/Meldingstjenesten-SafeSeaNet-Norway1/Et-nasjonalt-meldingssystem/

Kystverket AIS. (2012). *AIS.* Hentet fra http://havbase.kystverket.no/havbase_report/doc/AIS.pdf

LinkedDesign. (2015). Hentet fra http://www.linkeddesign.eu/

Loukides, M. (2011). Data Science and Data Tools. I O. R. Team, *Big Data Now.* O'Railly Press.

Lovdata. (2002, 01 01). Hentet fra https://lovdata.no/dokument/LTI/lov/2001-05-18-24

Lovdata. (2003, 06 24). *Forskrift om innsamling og behandling av helseopplysninger i Nasjonalt vaksinasjonsregister*. Hentet fra https://lovdata.no/dokument/SF/forskrift/2003-06-20-739

Luftfartstilsynet. (2015). *Luftfartstilsynets flysikkerhetsstatistikk*. Hentet fra http://www.luftfartstilsynet.no/flysikkerhet/

Løvoll, G., Vega-Gorgojo, G., Akerkar, R., Grumbach, S., Faravelon, A., Finn, R., . . . Bigagli, L. (2015). *Understanding and mapping big data.* BYTE: Big data roadmap and cross-disciplibary community for addressing societal externaliities. Hentet fra www.byte-project.eu

Management Events. (2015). Hentet fra http://managementevents.com/news/the-use-of-big-data-proceeds-in-the-finnish-government-organizations/

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011, Mai). *mckinsey.com.* Hentet fra Big data: The next frontier for innovation, competition, and productivity: http://www.mckinsey.com/insights/business_technology/Big_data_the_next_frontier_for_innovation

Marr, B. (2015, September 20). Big Data: 20 Mind-Boggling Facts Everyone Must Read. *Forbes magazine*, s. 2.

Marr, B. (2015, April 27). Big Data-As-A-Service Is Next Big Thing. *FORBES Magazine*, s. 2. Hentet fra http://www.forbes.com/sites/bernardmarr/2015/04/27/big-data-as-a-service-is-next-big-thing/

Marta Nagy-Rothengass, EU. (2014, 10 17). Hentet fra https://ec.europa.eu/digital-agenda/en/news/recent-eu-initiatives-field-big-data-communication-data-driven-economy-presentation-marta-nagy

MathWorks. (2015). *se.mathworks.com*. Hentet fra http://se.mathworks.com/products/connections/product_detail/product_35366.html

McKinsey. (2011, 05). *Big data: The next frontier for innovation, competition, and productivity.* Hentet fra http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Metsäteho. (2015). *Potential Business Models for Forest Big Data.* Hentet fra http://metsate1.asiakkaat.sigmatic.fi/wp-content/uploads/2015/03/Raportti_235_Potential_Business_Models_for_Forest_Big_Data_pv_ym.pdf

Ministry of Transport and Communications. (2014). *Big data applications have export potential.* Hentet fra http://www.lvm.fi/pressreleases/4418044/big-data-applications-have-export-potential

Ministry of Transport and Communications, Finland. (2014). Hentet fra http://www.lvm.fi/pressreleases/4418044/big-data-applications-have-export-potential

MongoDB. (u.d.). *www.mongodb.com*. Hentet fra NOSQL Database Explained: https://www.mongodb.com/nosql-explained

Mulchay, R. (2007, Mars 6). *www.CIO.com*. Hentet fra Business Intelligence Defintions and Solutions: http://www.cio.com/article/2439504/business-intelligence/business-intelligence-definition-and-solutions.html

Nasjonal Kommunikasjonsmyndighet. (2015, 09 24). *Bredbåndsdekning 2015: Stadig bedre tilbud av bredbånd med høye hastigheter*. Hentet fra http://www.nkom.no/aktuelt/nyheter/bredbandsdekning-2015-stadig-bedre-tilbud-av-bredband-med-hoye-hastigheter

Nasjonal Reiseplanlegger. (2015). Hentet fra http://www.nasjonalreiseplanlegger.com/rutebanken-as.html

Network Computing. (2015). Hentet fra http://www.networkcomputing.com/applications/big-data-deployment-models-open-vs-proprietary/a/d-id/1319627

NIST - National Institute of Standards and technology. (2011). *The NIST Definition of Cloud Computing.* Washington: US Department of Commerce. Hentet fra http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

ODI Open Data Institute. (2015). Hentet fra http://theodi.org/

OECD. (2002, November 7). *www.stats.oecd.org*. Hentet fra Glossary of Statistical Terms: https://stats.oecd.org/glossary/detail.asp?ID=5136

OECD. (2013). *Supporting Investment in Knowledge Capital, Growth and Innovation.* Paris: OECD Publishing. Hentet fra http://dx.doi.org/10.1787/9789264193307-en

OECD. (2015). *Data-Driven Innovation.* Hentet fra http://www.keepeek.com/Digital-Asset-Management/oecd/science-and-technology/data-driven-innovation_9789264229358-en#page1

OLAP. (u.d.). *www.olap.com*. Hentet fra What is the definition of OLAP?: http://olap.com/olap-definition/

Open Data Barometer. (2013, 10 31). *Open Data barometer, 2013 Global Report.* Hentet fra http://www.opendataresearch.org/dl/odb2013/Open-Data-Barometer-2013-Global-Report.pdf

Optique. (2015). Hentet fra www.optique-project.eu

O'Railly Radar Team. (2011). *Big Data now.* Bejing, Cambridge, Farnham, Köln, Sebastopol: O'Railly Strata. Hentet fra orailly.com

O'Railly Radar Team. (2015). *Big Data Now 2014 edition.* Bejing, Cambridge, Farnham, Köln, Sebastopol, Tokyo: O'Railly Media. Hentet fra http://www.oreilly.com/data/free/files/big-data-now-2014-edition.pdf

People 4 Smarter Cities. (2015). Hentet fra http://people4smartercities.com/series/how-dutch-captured-flow-data-manage-water-1

Politidirektoratet. (2014). *Datakrimstrategien.* Hentet fra https://www.regjeringen.no/contentassets/4d2ba37bf2ae4cc9ac39afdf20a2f41b/datakrimstrategi_2015.pdf

PROBA samfunnsanalyse. (2013). *Trygdesvindel i Norge - en kartlegging av fem stønadsordninger.* Oslo: PROBA samfunnsanalyse. Hentet fra http://www.proba.no/wp-content/uploads/sites/32/2013/04/Rapport-2013-05-Trygdesvindel-i-Norge.pdf

Prodatamarket. (2015). Hentet fra http://blog.prodatamarket.eu/about/

proDataMarket. (2015). *http://blog.prodatamarket.eu/about/*. Hentet fra http://blog.prodatamarket.eu/about/: http://blog.prodatamarket.eu/about/

Productive and inventive Finland. (2011). *Digital agenda for 2011-2020.* Hentet fra http://www.oph.fi/download/135323_productive_and_inventive_finland.pdf

PWC Australia. (2014). Hentet fra www.pwc.com.au/consulting/assets/publications/Data-drive-innovation-Sep14.pdf

PyBrain - The Pyton Machin Learning Library. (2015). Hentet fra http://pybrain.org/

QPR International. (2015). *Hva er ITIL?* Hentet fra QPR International: http://www.qrpmmi.no/index/itil/what-is-itil

Regjeringen. (2005). *Kultur for deling - strategi for utdanningsportaler*. Hentet fra https://www.regjeringen.no/no/dokumenter/kultur-for-deling-strategi-for-utdanning/id106469/

Regjeringen. (2015, 11 20). *Digitaliseringsrundskrivet*. Hentet fra https://www.regjeringen.no/no/dokumenter/digitaliseringsrundskrivet/id2462793/

Regjeringen.no. (u.d.). *Utvalg som skal revidere forvaltningsloven.* Hentet fra https://www.regjeringen.no/no/dep/jd/org/styre-rad-og-utval/tidsbegrensede-styrer-rad-og-utvalg/forvaltningsloven-skal-bli-mer-brukervennlig/id2458957/

Regjeringskansliet. (2015, 04 02). Hentet fra http://www.regeringen.se/rapporter/2011/10/n2011.12/

Reimsbach-Kounatze, C. (2015). *The Proliferation of Big Data and Implications for Official Statistics and Statistical Agencies.* Paris: OECD. Hentet fra http://dx.doi.org/10.1787/5js7t9wqzvg8-en

Rijksoverheid. (2014). Hentet fra https://www.rijksoverheid.nl/onderwerpen/doorbraakprojecten-met-

ict/documenten/publicaties/2014/01/13/routekaart-doorbraakproject-open-geodata-als-grondstof-voor-groei-en-innovatie

Rio Tinto. (2014). *Delivering greater value for shareholders.* Hentet fra http://www.riotinto.com/documents/RT_BoAML_2014_slides.pdf

Sage Journals. (2015). Hentet fra http://bds.sagepub.com/content/1/2/2053951714564227

SAS Institute. (u.d.). *www.sas.com*. Hentet fra "Data Mining": http://www.sas.com/en_us/insights/analytics/data-mining.html

SAS Institute. (u.d.). *www.sas.com*. Hentet fra Hadoop - What is it and why does it matter?: http://www.sas.com/en_us/insights/big-data/hadoop.html

SAS Institute, Tech Partnership. (2014). *Big Data Analytics, Assessment of Demand for Labour and Skills 2013-2020.*

ScienceDirect. (2009). *Strategic foresight in public policy: Reviewing the experiences of the UK, Singapore, and the Netherlands.* Hentet fra http://www.sciencedirect.com/science/article/pii/S0016328709001402

SINTEF. (2013, 05 22). *Big Data - for better or worse*. Hentet fra http://www.sintef.no/en/corporate-news/big-data--for-better-or-worse/

Skatteetaten, A-Ordningen. (2015). Hentet fra http://www.skatteetaten.no/se/Om-skatteetaten/Presse/Nyhetsrommet/Pressemeldinger/Pressemeldinger-2015/A--ordningen-pa-lufta--Betydelig-forenkling-for-220-000-arbeidsgivere/

Skatteetaten, MAG. (2013). Hentet fra https://www.skatteetaten.no/upload/PDFer/Skatteinfo-3-2013.pdf

Solr. (2015). *http://lucene.apache.org/solr/*. Hentet fra http://lucene.apache.org/solr/: http://lucene.apache.org/solr/

SSB. (u.d.). Hentet fra http://www.ssb.no/omssb/styringsdokumenter/lover-og-prinsipper/fns-prinsipper-for-offisiell-statistikk

Statens Vegvesen. (u.d.). Hentet fra http://www.vegvesen.no/fag/Trafikk/ITS

Statens vegvesen. (2014). *Handlingsplan for ITS 2014 - 2017.* Oslo: Statens vegvesen.

Statens vegvesen. (2015). *Nasjonalreiseplanlegger.com*. Hentet fra Rutebanken AS: http://www.nasjonalreiseplanlegger.com/rutebanken-as.html

Statens vegvesen. (2015, Oktober 10). *NorTraf*. Hentet fra NorTrafKom: https://nortrafkom.triona.no/faces/front.jsp;jsessionid=2DAAC2AA4E5BCF45C8D1CC5BDAB2DB2A

Statens vegvesen. (2015, Oktober 12). *Vegkart*. Hentet fra vegvesen.no/vegkart: https://www.vegvesen.no/vegkart/vegkart/#kartlag:geodata/vegreferanse:224099.52189696:6831323.3889489/hva:(~)/hvor:()/@388909,6807715,4

Statens Vegvesen Vegkart. (2015). *https://www.vegvesen.no/vegkart/*. Hentet fra https://www.vegvesen.no/vegkart/: https://www.vegvesen.no/vegkart/

Stavangerregionen. (2014, 10 28). Hentet fra http://stavangerregion.no/2014/10/28/sammen-bygger-vi-en-smart-stavangerregion/

Swedish ICT SICS. (2015). Hentet fra http://www.sics.se

Tableau. (2015). Hentet fra http://www.tableau.com/learn/whitepapers/top-7-trends-big-data-2015

Tekes. (2015). Hentet fra http://www.tekes.fi/en/programmes-and-services/tekes-programmes/

TEKES. (2015). Hentet fra http://www.tekes.fi/en/programmes-and-services/tekes-programmes/

TEKES, Industrial Internet. (2015). Hentet fra https://www.dreambroker.com/channel/wn1d8vnu/cjhiwjf6

Teknologirådet. (2015). *Forutseende Politi.* Hentet fra http://teknologiradet.no/wp-content/uploads/sites/19/2013/08/ForebyggendeAnalyse_endelig_WEB.pdf

Teknologirådet. (2015). *FORUTSEENDE POLITI - Kan dataanalyser hjelpe politiet til å være på rett sted til rett tid?* Oslo: Teknologirådet.

The Praliament Magazine. (2015, 11 06). Hentet fra https://www.theparliamentmagazine.eu/articles/news/eu-commission-sets-out-guidelines-interim-safe-harbour

Tie Kinetix. (2015). Hentet fra http://tiekinetix.com/en-us/company

Tilväxtanalys. (2013). *Stora data & öppna data.* Hentet fra http://www.tillvaxtanalys.se/download/18.1694230714d5b805006b0453/1432287031372/wppm_2013_16.pdf

Ulrike Deetjen, E. T. (2015). *Big Data for Advancing Dementia Research.* Oxford: OECD. Hentet fra http://dx.doi.org/10.1787/5js4sbddf7jk-en

UN Economic and Social Council. (2014). *Big data - an opportunity or a threat to official statistics.* Paris: ECE/CES.

UNECE Statistics. (2015). Hentet fra http://www1.unece.org/stat/platform/display/bigdata/Big+Data+in+Official+Statistics.

United States Census Bureau. (2015). Hentet fra http://www.census.gov/data.html

Universitetet i Stavanger. (2013). Hentet fra https://www.uis.no/forskning/vi-forsker-paa/informasjonsteknologi/biomedisinsk-dataanalyse/prosjekter/?s=10491

Universitetet i Stavanger. (2014). Hentet fra https://www.uis.no/forskning-og-ph-d-studier/forskningssentre/cipsi-senter-for-ip-basert-serviceinnovasjon/

Universitetet i Tromsø. (2014). Hentet fra https://uit.no/om/enhet/aktuelt/nyhet?p_document_id=399245

Universitetet i Tromsø. (2014, 12 22). Hentet fra https://uit.no/om/enhet/aktuelt/nyhet?p_document_id=399245&p_dimension_id=88136

Vegdirektoratet. (2012). *Handlingsplan Trafikkdata 2013 - 2018.* Oslo: Statens vegvesen.

Viden Metro. (2014, Mars 24). *www.evv.dk.* Hentet fra Big Data Om værdien af datadreven innovation: http://evv.dk/wp-content/uploads/VidenMetro_BigDataPlatformen_.pdf

Vinnova. (2015). Hentet fra www.vinnnova.se

Vinnova. (2015). *Big Data Analytics*. Hentet fra http://www.vinnova.se/sv/Var-verksamhet/Gransoverskridande-samverkan/Samverkansprogram/Strategiska-innovationsomraden/Strategiska-innovationsagendor/Forteckning-agendor-2013/Big-Data-Analytics/

Vinnova. (2015). *Big Data Analytics.* Hentet fra http://www.vinnova.se/sv/Var-verksamhet/Gransoverskridande-samverkan/Samverkansprogram/Strategiska-innovationsomraden/Strategiska-innovationsagendor/Forteckning-agendor-2013/Big-Data-Analytics/

Vinnova. (2015, 07 04). *FFI - Fordonsstrategisk Forskning och Innovation.* Hentet fra http://www.vinnova.se/PageFiles/751328373/Projektkatalog%20Bada%202015-04-07.pdf

Whitehouse.gov. (2012, 03 29). *Big Data Across the federal Government.* Hentet fra https://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_fact_sheet.pdf

Whitehouse.gov. (2014, 05). *Big Data: Seizing Opportunities, Preserving Values.* Hentet fra https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf

Wikipedia. (2015). Hentet fra Alan Turing Institute: https://en.wikipedia.org/wiki/Alan_Turing_Institute

Wikipedia. (2015). Hentet fra www.wikipedia.com

Wikipedia, Exascale Computing. (2015). Hentet fra https://en.wikipedia.org/wiki/Exascale_computing

Wikipedia, Internet of Things. (2015). Hentet fra https://en.wikipedia.org/wiki/Internet_of_Things

Wikipedia, Nettskyen. (2015). Hentet fra https://no.wikipedia.org/wiki/Nettskyen

Wikipedia, Privacy by design. (2015). Hentet fra https://en.wikipedia.org/wiki/Privacy_by_design

Wikipedia, Text Mining. (2015). Hentet fra https://en.wikipedia.org/wiki/Text_mining