



I en digital hverdag genererer mennesker, smarte ting og systemer enorme mengder data som kan samles inn, analyseres og gi ny innsikt. Hvordan påvirkes hverdagen vår av innsikten stordata kan gi? Illustrasjon: Astibuag/Shutterstock.

Når alt vi gjør registreres og systematiseres:

Stordata i hverdagen

Hedvig Tønnesen, Lisa Reutter og Melanie Magin

Stadig mer av hverdagen vår og omgivelsene rundt oss er koblet på internett, og vi legger fra oss mengder med digitale spor. Til sammen blir disse digitale sporene til stordata som kan brukes til å skape nye verktøy, tjenester og offentlige goder. Stordata er et komplekst fenomen som bringer med seg både muligheter og utfordringer for samfunnet vårt, og som spår å endre måten vi lever, jobber og tenker på.



Hedvig Tønnesen er PhD-stipendat i samfunnsvitenskap ved Norges teknisk-naturvitenskapelige universitet (NTNU) i Trondheim. Hun forsker på politisk påvirkning i sosiale medier og er tilknyttet det internasjonale forskningssamarbeidet DigiWorld (Digital Election Campaigning Worldwide). Hedvigs forskningsinteresser inkluderer retorikk, teknologi (herunder digitale medier, stordata og algoritmer), medier og demokrati. (Foto: Geir Mogen)



Lisa Reutter er PhD-stipendiat i samfunnsvitenskap ved Norges teknisk-naturvitenskapelige universitet (NTNU) i Trondheim. Hun er en del av NTNUs tverrfaglige satsning Digital Transformation og tilknyttet Digital Infrastructures and Citizen Empowerment (DICE). Lisa forsker på datadrevet offentlig forvaltning og er opptatt av samspillet mellom teknologi, samfunn og demokrati.



Melanie Magin er førsteamanuensis i mediesosiologi ved Norges teknisk-naturvitenskapelige universitet (NTNU) i Trondheim og har doktorgrad i kommunikasjonsvitenskap fra JGU Mainz (Tyskland). Hennes forskning fokuserer på samfunnsrolle og innflytelse av tradisjonelle og nye medier, samt mulighetene og risikoen forbundet med dem. Hun jobber i krysset mellom politisk kommunikasjon, online kommunikasjon (spesielt sosiale medier og søkemotorer) og sammenlignende forskning, hovedsakelig basert på kvantitative metoder.

Hver av oss legger igjen enorme mengder med digitale spor i hverdagen. Det skjer hver gang vi sender en e-post eller skriver en gratulasjonsmelding på Facebook, nettshopper eller trekker kortet i kassa på Rema, eller bare går rundt med mobilen i lomma og deler posisjonen vår. Siden stadig mer av hverdagen er koblet på internett, legger vi fra oss stadig mer informasjon om hvor vi befinner oss, hva vi gjør, hva vi skriver og hva vi søker etter. Det samme gjelder tingene rundt oss: TV-en, termostaten og den app-styrte lyspæren. Den bevegelsesstyrte gatelykten utenfor er sannsynligvis også på nett. Sammen blir disse digitale sporene til gigantiske mengder data som kan samles inn, analyseres og brukes på forskjellige måter. Stordata spås å endre måten vi lever, jobber og tenker på.¹ Dette kapitlet vil ta for seg hva stordata er og hvordan det påvirker hverdagen vår på både positive og negative måter.

Hva er stordata?

Data er digitale representasjoner av informasjon. Dette kan være alt fra personlige data som alder, kjønn og personnummer til en Twitter-melding, et bilde på Facebook, GPS-koordinater og hvor mange kWh strøm et borettslag forbruker. Så når blir egentlig «data» store?

Det er ikke enighet om en definisjon på «stordata» i dag. Men ofte snakker vi om stordata hvis et datasett (1) omfatter **store mengder** data (tera- eller zettabytes), (2) produseres i **stor fart**, (3) blir satt sammen fra mange ulike kilder (**stor variasjon**), (4) omfatter data om alle som bruker en tjeneste (**fullstendig utvalg**), (5) har et **høyt detaljnivå** og **standardiserte**, lett identifiserbare måter å sette opp data på, og (6) er **fleksibelt**, noe som betyr at en enkelt kan kombinere data fra flere kilder i datasettet.²

Når vi kaller data for «stordata», er det altså ofte fordi vi har å gjøre med så *mye* data, så mange *ulike typer data* eller så *nye* datakilder at det er utfordrende å ta dem i bruk, og at det kreves kraftige datamaskiner og intelligente analysemetoder (og gjerne en god dose nytenkning!) for å kunne utnytte dem til å finne nye sammenhenger og innsikter.

Stordata er på mange måter bindeleddet mellom flere av teknologiene som beskrives i denne boken, for eksempel sensorteknologi og kunstig intelligens. Sensorer er små dingser som kan festes på ting, og så registrerer ulike forhold, for eksempel bevegelse, fuktighet eller lyd. For å analysere de enorme mengdene data fra sensorer trenger vi avanserte analysemetoder, for eksempel metoder innen kunstig intelligens. Stordata har hatt et oppsving de siste årene, først og fremst fordi stadig mer data produseres, men også på grunn av teknologiske fremskritt innen prosesseringskraft, lagringskapasitet og overføringshastighet.

Stordata i samfunnet: Status, muligheter og utfordringer

Stordata kan bidra til ny innsikt om og samtidig påvirke hvordan folk lever, jobber, kommuniserer og bruker ulike tjenester. Denne innsikten kan hjelpe både myndigheter og private bedrifter med å skape

¹ Mayer-Schönberger, V. and K. Cukier (2017). *Big data : The essential guide to work life and learning in the age of insight*. London, John Murray.

² Laney, D. (2011). *3D Data Management: Controlling Data Volume, Velocity and Variety*. Publisert på analysebyrået Gartners blogg. Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. SAGE Publications Ltd. <https://doi.org/10.4135/9781473909472>

Se gjerne Gulla, J.A. (2017). Stordata. I A. Rolstadås, A. Krokan & L.T. Dyrhaug (red.). *Teknologien endrer samfunnet*. Fagbokforlaget. <https://www.gammel.ntva.no/Teknologibok2017/kap6/>

nye og bedre verktøy, tjenester og offentlige goder – kort sagt til å forbedre hverdagen vår. Men som med enhver ny teknologi er det ikke bare muligheter, men også utfordringer ved bruk av stordata.

Informasjon, nyheter, underholdning og reklame

Stordataenes inntog i hverdagen blir ofte assosiert med internettplattformer. Mange av oss er innom plattformer som Facebook, Instagram og Google flere ganger om dagen. Når vi klikker på en sak på Facebook, følger en bruker på Instagram eller googler etter noe, legger vi fra oss mange digitale spor. På sikt vokser de digitale sporene våre til rikholdige registre om hver enkelt av oss. Internettplattformene bruker de store mengdene med data for å sortere og anbefale innhold som er relevant for hver enkelt bruker, og dette kan uten tvil være nyttig. Å bruke Facebook og Google er gratis: Grunnen til at de likevel er blant verdens rikeste selskaper, er at de tjener penger på å selge innsikten de store mengdene med data gir om oss mennesker.



Figur 7.1 Dataene vi legger fra oss når vi bruker sosiale medier, apper og søkemotorer kan analyseres og brukes til å skreddersy innhold, reklamer og tjenester. Foto: GaudiLab/Shutterstock.

At internettplattformene kjenner oss godt, er gode nyheter for de som lager underholdning og reklame, som kan nå frem til de som er mest interessert i produktene og tjenestene de tilbyr, i stedet for å kaste bort tid og penger på folk som egentlig ikke er interessert. Det kan også være veldig praktisk for oss som forbrukere å få opp reklamer for produkter vi faktisk er interessert i, eller anbefalinger om å se en serie som viser seg å være vår nye favorittserie. Dataene vi legger fra oss når vi er på internett, kan analyseres og brukes til å skreddersy innhold og tjenester som holder på oppmerksomheten vår, og får oss til å bruke mer tid på plattformene – noe som gjør at selskapene kan samle inn enda mer data. Mange digitale plattformer bruker disse dataene til å finne ut hvordan vi kan påvirkes til å skrolle mer på Facebook, dele en *story* på Snapchat, *binge* en serie på Netflix eller trykke på en annonse for joggesko.

Dette er ikke nødvendigvis negativt i seg selv, men påvirkningen kan bli usynlig: Vi er ikke bevisste på hvor mye data som samles inn om oss og hvordan dataene benyttes. For det første gir mange av oss

fra oss data uten engang å ha tenkt på det eller lagt merke til det. Vet du for eksempel hvor mange av appene på mobilen din som sporer lokasjonsdata om hvor du oppholder deg når du bruker dem? For det andre kan det være vanskelig å legge merke til og tenke over konsekvensene av at stordata-drevne algoritmer påvirker hvilke nyheter vi leser, hva slags musikk vi hører på og hvilke klesbutikker eller universiteter vi vurderer.³

Værvarsling

Når du sjekker værvarelet på Yr.no, ser du resultatene av at enorme mengder data fra mange ulike datakilder kvernes gjennom avanserte værvarslingsmodeller ved hjelp av kraftige datamaskiner, slik at værvarelet for én million steder automatisk kan oppdateres minst én gang i timen. Bruken av stordata har revolusjonert norsk værmelding. Blant annet på 1980-tallet med bilder fra satellitter, og i 2018 da data fra private værstasjoner ble tatt med som en ny datastrøm. I dag suppleres data fra 1 000 tradisjonelle værstasjoner rundt omkring i Norge med data fra om lag 60 000 private værstasjoner i Skandinavia, noe som gjør det lettere å lage lokale værvarsler. De private værstasjonene er ment å gi eieren bedre informasjon om inne- og uteklimaet hjemme, og er et godt eksempel på hvordan data fra den enkeltes smart hjem-enheter kan utnyttes i stor skala for å lage bedre tjenester for samfunnet som helhet. Nye datakilder kan også være aktuelle å ta i bruk for å forbedre værmeldingene: Hva om busser kan utstyres med sensorer som avgir værdata? Eller hva med å hente ut data fra sosiale medier for å finne ut om folk opplevde at værvarelet traff?

Mye faglig kompetanse er bygget inn i algoritmene som jobber under panseret på Yr. Dette er spesielt viktig for utfordringen med å melde været i et stadig mer uforutsigbart klima. Når et datasett ikke viser hvordan været påvirkes av klimaendringer, kan ikke datasettet i seg selv benyttes for å spå dette fremover heller. Da må klimamodeller inn i analysene.

Når satellitt-bilder og data fra private værstasjoner benyttes for å lage værmeldingen, er dette data som ikke opprinnelig ble samlet for dette formålet. Stordata handler gjerne om slik *gjenbruk* av data. Det er mange problemstillinger rundt gjenbruk av data. På et praktisk nivå kan nye datastrømmer fra kilder man ikke selv har kontroll over være ustabile, noe som fordrer at man tenker utenfor boksen. For eksempel kan en værstasjon være plassert slik at den blir altfor varm når sola står rett inn, og varsler hetebølge på en vanlig norsk sommerdag. Når sola *ikke* står rett inn, kan værstasjonen likevel være en god datakilde. I stedet for å fjerne hele værstasjonen som en kilde, fjernes heller de feilaktige observasjonene fra værstasjonen.

Datasettene og kildekoden bak Yr er åpne og fritt tilgjengelige for alle. Det betyr at det er fritt frem for andre aktører å lage tjenester ved hjelp av disse dataene, mens det i andre land er kostbart å kjøpe værdata. Datasettene ligger ikke «bare» åpent tilgjengelig, mye arbeid har blitt lagt ned i å gjøre dataene forståelige og brukervennlige. Det gjør at f.eks. små oppstartsbedrifter kan leke med datasettene og prøve å lage nye løsninger uten å måtte investere svimlende summer i kraftige datamaskiner.

Helse

Helse er et av mange områder hvor potensialet for å utnytte stordata er stort. I Norge har vi lenge samlet enorme mengder med helsedata i register. I tillegg har vi en mengde smarte dingser, som

³ Av og til refereres dette til som «filter-bobler» og «ekko-kammere», som du kan lese mer om i kapittel 18 i denne boken.

kontinuerlig registrerer og samler inn data om kroppen vår. Stordataanalyser kan hjelpe hver enkelt av oss å ta gode valg i hverdagen, og når vi blir syke, kan stordata hjelpe leger med å stille mer presise diagnoser og tilpasse behandlingen til hver enkelt.



Figur 7.2 Helse i hver byte? Aktivitetsklokker og treningsapper kan motivere til en sunnere livsstil, samtidig som bruken av dem genererer en mengde helsedata som tidligere ikke har vært tilgjengelig. Foto: KANUT PHOTO/Shutterstock.

I dag baseres mye av behandlingen av sykdom på hva som fungerer for en gjennomsnittspasient. Ved hjelp av stordata kan forskere og leger få mer innsikt i hvordan sykdommer rammer den enkelte, og i større grad skreddersy behandlingen. Ved hjelp av stordata fra pasientjournaler fant svenske forskere for eksempel ut at det ikke finnes *to*, men *fem* ulike typer diabetes. Kunnskapen fra analysene gjør at pasienter kan få mer tilpasset medisin.⁴

Apper som tilpasses enkeltpersoner kan hjelpe pasienter med å håndtere sin egen sykdom. Et eksempel på dette er Propeller Health, som bruker stordata for å gjøre det lettere for mennesker med astma å håndtere sykdommen. Ved hjelp av data fra sensorer som festes til astmainhalatorer, samt data om symptomer og medisinbruk, vær og luftkvalitet, lærer Propellers plattform mer om den enkeltes symptomer, og gir astmapasientene persontilpassede råd og påminnelser om å ta medisinen. Dette har bedret hverdagen til mange astmatikere, samtidig som alle dataene som samles inn kan hjelpe forskere med å forstå sykdommen bedre.

Helsedata blir sett på som svært verdifulle, og norske myndigheter ser mange muligheter i utnyttelse av dataene. Samtidig er stordata om helse svært sensitive, og bruken er derfor ikke uproblematisk. Et spennende eksempel ser vi i problematikken rundt appen Smittestopp.

⁴ Teknologirådets notat «Kunstig intelligens og norske helsedata» fra 2019 gir en fin oversikt over hvilke muligheter som ligger i bruk av stordata og kunstig intelligens for den norske helsetjenesten: https://teknologiradet.no/wp-content/uploads/sites/105/2020/01/KI-og-helsedata_m-lenker2.pdf

Smittevern

Stordata har allerede blitt et viktig verktøy i kampen mot trusler som naturkatastrofer og sykdomsutbrudd. I april 2020, like etter at koronapandemien hadde kommet til Norge, lanserte Folkehelseinstituttet appen Smittestopp. Den første versjonen ble trukket tilbake i juli samme år og kom i en ny versjon påfølgende desember. Forskjellen mellom de to appene viser at det er ulike avveininger vi kan ta rundt bruken av stordata.

Begge versjonene av Smittestopp-appene brukte sporing av mobiltelefoner for å varsle brukerne om at de hadde vært i nærheten av andre app-brukere som senere hadde fått påvist koronaviruset. Men den første appen brukte i tillegg mobiltelefonenes innebygde GPS-sporing. Dermed sporet appen brukerne hvorhen de gikk. Det var det mange som reagerte på; de mente appen kom for tett på folks privatliv og ville ha den stoppet fordi den ikke tok vare på den enkeltes rett til personvern. Dataene fra appen skulle anonymiseres slik at det ikke var mulig å koble bevegelser til en person, men når man samler inn lokasjonsdata, er anonymisering lettere sagt enn gjort. Hvis du har data om hvilken adresse en person befinner seg på hver natt og hvilken adresse personen befinner seg på i ukedagene, da er det lett å skjønne hvem det er snakk om.

En annen vesentlig forskjell mellom de to versjonene av appen var hva dataene kunne brukes til. I den første appen kunne Folkehelseinstituttet analysere data fra appene for å bli bedre i stand til å forutse utviklingen i smitteutbrudd, og se hvordan koronatiltakene fungerte. Her ble Folkehelseinstituttet kritisert for at de brukte dataene til mer enn det som var nødvendig for smittesporing.

I den andre appen fjernet Folkehelseinstituttet GPS-sporing og gikk vekk fra å bruke innsamlede data til analyse og forskning. Dermed begrenset de muligheten til å ta i bruk stordataanalysene for å hjelpe Norge og andre land i kampen mot en global pandemi. Smittestopp-appene reiser viktige spørsmål rundt stordata. Hvordan skal vi balansere samfunnsnytte mot personvern? Bør det være en samfunnsplikt å dele data om oss selv, eller er dette et overtramp mot privatlivet som vi ikke burde godta?

Smarte byer, kommuner og transport

Med mer tilgjengelige data om hvordan vi lever, interagerer og flytter på oss, kan vi skape byer som er bedre å bo i, og som styres mer effektivt og bærekraftig. I dag jobber mange norske byer og kommuner nettopp med prosjekter som har med «smarte byer» eller «smarte kommuner» å gjøre.⁵ Nøkkelen er å lage løsninger som baserer seg på hvordan byer og kommuner faktisk brukes, og å kunne bruke denne innsikten i sanntid. Her er stordata sentralt, gjerne i form av data fra sensorer og lokasjonsdata fra innbyggernes mobiler.

Ved å utstyre alt fra veier, bygninger og parker til søppeldunker og gatelys med sensorer, kan vi generere enorme mengder data som gir et øyeblikksbilde av tilstanden i en by eller kommune. Denne innsikten kan bidra til at kommunale tjenester kan styres på andre måter. For eksempel: I stedet for at renovasjonsbilene kjører faste ruter hvor de tømmer halvfulle søppeldunker, kan de kjøre dit det er behov for tømming av søppeldunker. Eller innbyggerne kan få informasjon om hvor trafikken flyter best og hvor det er ledig parkering. Og ved å blande data fra sensorer, aktivitetsklokker og

5

https://www.regjeringen.no/contentassets/d6fa05005d5d4ea3a45f62286c2ba2fe/kartlegging_av_smarte_byer.pdf

mobiltelefoner, kan både byplanleggere og private selskaper finne ut hva som motiverer oss nordmenn til komme oss opp fra sofaen og ut i parken eller marka – og dermed hvordan byer og kommuner kan utformes slik at vi beveger oss mer.



Figur 7.3 Oslo er blant de mange norske kommunene som jobber med å bli «smartere» ved hjelp av blant annet bedre utnyttelse av stordata til planleggings- og styringsformål. Foto: photovideoworld/Shutterstock.

Norges største byer blir stadig større, og det gjør at det potensielt blir kaos i trafikken – og mer utslipp. Hvordan sikrer vi at folk kan komme seg raskt og smertefritt fra a til b, enten det er med bil, buss, tog eller el-sparkesykkel? Her kan stordata hjelpe oss med å finne nye løsninger og å sørge for at eksisterende kollektivtilbud fungerer bedre. Kanskje blir det slutt på at toget er forsinket på grunn av sporfeil. Ved hjelp av massevis av sensorer ute i togsporene og rask analyse av disse dataene kan man rette opp feilene *før* de oppstår.⁶

Hvis vi vil at flere skal ta bussen i stedet for bilen for å komme seg på jobb, kan vi bruke informasjon om hvordan folk beveger seg ved å spore mobiler for å vurdere om vi trenger flere bussavganger eller nye bussruter. Selskaper som Ruter har store forventninger til bruk av stordata.

Løsninger bygget på stordata kan også hjelpe bilister med å unngå kjedelige og forsinkende køer. For eksempel samler den Google-eide appen Waze inn data fra millioner av mobiler om hvor det er kø, hvor det har vært ulykker på veien og så videre, og kan gi brukerne tips om de raskeste kjørerutene.

Offentlige tjenester

Statistiske analyser har alltid vært en viktig del av den moderne staten. Den norske velferdsstaten har samlet enorme mengder med data om sine innbyggere i lang tid. Også offentlig sektor har vist økt interesse for stordata de siste årene. Bruk av stordata i offentlig sektor kan føre til en generell

⁶ Dette jobber Bane NOR med i dag. Ved hjelp av sensorovervåking, ny programvare og stordata ønsker de å drive smartere vedlikehold. Tiltakene har allerede gitt god effekt: de reduserte forsinkelser på grunn av sporfeil fra 730 minutter i november 2019 til 38 minutter i januar 2020. Meld. St. 20 (2020–2021), s. 61.

effektivisering av offentlig forvaltning og bedre innbyggertjenester, ifølge flere rapporter og utredninger.

Stordata har potensial til å oppnå bedre kvalitet og aktualitet i beslutninger, hjelpe til å forutse uventede hendelser og bidra i planleggingen av samfunnet vårt. Mange trekker frem at bruken av stordataanalyser kan redusere responstiden og skreddersy tjenester til innbyggerne. Bruk av stordata i offentlig sektor handler om å bruke nye datakilder, sammenstille data fra ulike offentlige aktører og sette søkelys på gjenbruk av data. I dag blir mye av dataene som samles inn både i register og i generell saksbehandling, arkivert og lagret i siloer. Flere offentlige virksomheter jobber derfor med å tilgjengeliggjøre dataene sine på analyseplattformer, slik at disse kan sammenstilles og brukes til å forbedre organisasjonen og tjenestene.

Stordata kan for eksempel brukes av Skatteetaten til å gjøre sanntidsanalyser av ulike risikofaktorer og dermed avdekke risiko for skatteunndragelser og svart økonomi. Eller vi kan bruke stordataanalyser for å predikere sannsynligheten for at en sykmeldt arbeidstaker kommer til å bli langtidssykemeldt og dermed trenger tettere oppfølging av NAV. Eller hva om vi setter ut tusenvis av målepunkter langs norske veier, slik at Vegvesenet kan bruke stordata for å forutse vedlikehold og gi bedre informasjon til både syklistene og andre trafikanter? Norsk regnesentral har gått sammen med flere norske offentlige aktører til å etablere Big Insight i 2017 for å videre utforske mulighetene som ligger i stordataanalyser for offentlig sektor. Datamengdene er store og anvendelsesmulighetene mange, likevel møter vi også på en rekke utfordringer når vi skal bruke stordata i offentlig forvaltning.

Forvaltningsloven legger strenge føringer for hvordan norsk offentlig forvaltning organiseres. Loven stiller for eksempel krav om gjennomsiktighet og forklarbarhet av beslutningene. I tillegg skal alle behandles likt, og offentlig forvaltning har strenge krav om sikkerhet og personvern. Bruk av stordata kan innebære at vi drar med oss fordommer. Ofte finnes det skjevheter i datagrunnlaget som så kan påvirke beslutningene som tas basert på stordataanalyser. Dataene vi har samlet inn i offentlig sektor, er tross alt ikke perfekte.

Vi kan for eksempel bruke store mengder med data for å finne mønster i data om sykefravær og arbeidsledighet og forsøke å lage mer persontilpasset oppfølging av sykemeldte og arbeidsledige. NAV har gjort flere forsøk på dette, men har fort funnet ut at det finnes mange skjevheter i dataene, og at data ofte hadde lav kvalitet. Det er viktig at analysene er like treffsikre for forskjellige grupper i samfunnet. I arbeidet med arbeidsoppfølgingen oppdaget NAV for eksempel at analysene ofte var dårligere på å klassifisere arbeidssøkere over 50 år.

Akkurat hva som er en rettferdig beslutning, er ikke alltid like lett å forstå. Tidligere forsøk på bruk av slike prediktive analyser basert på stordata i offentlig sektor har ofte blitt kritisert for å være urettferdige. Analysene ble her ansett som diskriminerende. Data er aldri helt perfekte, og vi må være klar over at det ofte ligger skjevheter i stordata som så reproduseres når dataene brukes i beslutningsprosesser. I tillegg må beslutningsgrunnlaget være forståelig både for innbyggerne og saksbehandleren. EU har stilt høye krav om rettferdighet, gjennomsiktighet, sikkerhet og godt personvern i bruk av stordata i offentlig sektor, og det er fortsatt en god del ubesvarte spørsmål og utfordringer når det kommer til gjenbruk av store mengder med data i offentlig sektor.

Et datadrevet samfunn?

Stordata og stordataanalyser har allerede blitt en viktig del av hverdagen. Vi har nok bare sett begynnelsen på hva stordata kan brukes til, både når det gjelder de store og små spørsmålene. Som vi har beskrevet, kan stordata hjelpe oss med noen av de største utfordringene vi står overfor i verden i dag, og også hjelpe oss med å få bedre liv og lokalsamfunn. Digital teknologi og stordata kan gjøre selskaper og offentlige aktører – og oss selv – bedre i stand til å møte våre individuelle behov og interesser. Men samtidig medfører stordata noen utfordringer samfunnet må være bevisst på og prøve å løse.⁷



Figur 7.4 Bruk av stordata har sine potensielle skyggesider, blant annet når det gjelder personvern og det asymmetriske maktforholdet som følger av at store mengder data er samlet på få hender. Foto: Andrew V Marcus/Shutterstock.

Vi ser allerede en mengde politiske diskusjoner i Norge og EU om hvordan innsamling, eierskap og benyttelse av stordata kan reguleres for å balansere samfunnsnytte med hensynet til enkeltpersoners personvern og autonomi. Disse diskusjonene vil også være sentrale i fremtiden. Her er noen av problemstillingene vi må ta stilling til:

Hvordan balansere hensyn til samfunnsnytte og personvern?

Ofte er det personvern og anonymisering som står sentralt når vi diskuterer etiske hensyn ved stordata. Personvern handler om retten vi har til privatliv, og retten til å bestemme over bruk og spredning av personopplysninger.⁸ For at data skal kunne kalles personopplysninger, må dataene direkte eller indirekte kunne knyttes til en enkeltperson. Derfor anses anonymisering av data gjerne som en riktig og viktig løsning på personvernproblematikken stordata bringer med seg.

⁷ Teknologirådet, som skal gi råd til regjeringen om digitale løsninger, utvikler ulike fremtidsscenarier for ulike sektorer (f.eks. helse og omsorg, transport og smarte byer) som tar opp i seg slike problemstillinger.

⁸ <https://www.datatilsynet.no/rettigheter-og-plikter/hva-er-personvern/>

Det er ikke alltid like lett å sikre anonymisering av data, slik at ikke digitale spor kan spores tilbake til oss som enkeltmennesker. Noen ganger er det til og med umulig, f.eks. med historiske data om geografisk plassering. Men bør vi alltid etterstrebe å anonymisere mest mulig? Dette har mange perspektiver.

Norge har en lang tradisjon for personvern og høy tillit, og vi er i dag omfattet av EUs personvernregelverk, som setter strenge krav til anonymisering av data og hva data kan brukes til. Men det er delte meninger om hvor strenge vi egentlig burde være med å anonymisere.

Noen mener at de strenge kravene til anonymisering kan frarøve oss (og andre land) verdifull innsikt. Norge har i utgangspunktet gode forutsetninger for å være med på utviklingen innen stordata. Vi har allerede unike og oppdaterte data i ulike offentlige registre. I tillegg har Norge alltid vært tidlig ute med digitalisering, og svært mange av oss bruker internett og digitale verktøy hyppig. Og siden analyser av stordata dreier seg om å bruke data om hvordan samfunnet er i dag for å lære noe om hva som vil skje i fremtiden, kan data fra den norske konteksten utgjøre et godt analysegrunnlag for å utvikle innsikt, verktøy og tjenester.

Som vi var inne på i diskusjonen om Smittestopp-appen: Bør det rett og slett være en samfunnsplikt å dele data om oss selv som kan benyttes i kampen mot klimaendringer, sykdommer og kriser? Dette er i seg selv et problematisk spørsmål. Enkeltpersoner bør ikke presses til å gi fra seg persondata. For selv om deling av data i prinsippet skulle være frivillig, er det ikke mye fri vilje igjen når man nærmest gir enkeltpersoner som ikke ønsker å gi fra seg persondata ansvaret for kriser.

Andre mener tvert imot at Norges strenge krav til personvern, åpenhet og forsvarlighet godt lar seg kombinere med stordata, og at dette vil være et konkurransefortrinn: Vi kan bli best på å balansere personvern, tillit og datainnsamling, og tjene på å eksportere løsninger som befolkningen i andre land vil ha fordi vi tar personvern og forsvarlighet på alvor.⁹

Det er viktig å huske at ikke-personidentifiserende data kan misbrukes av andre. Det vil si at anonymiserte data kan gjenbrukes på måter som har uheldige utfall for samfunnet. Ofte er gode stordataanalyser ikke avhengig av personidentifiserende data for å skape meningsfulle resultater for interesserte parter.

Hvem bør eie dataene om oss?

Stordata og stordataenes påvirkningskraft er ofte usynlige for oss i hverdagen. Det ligger mye makt hos de som samler inn og eier dataene om oss. De som eier dataene kontrollerer også tilgang til dataene, og kan legge premisser for offentlige og private aktører som ønsker å benytte seg av dem. I dag forvalter private selskaper som Facebook og Google mye av denne makten. De fleste av oss aksepterer (mer eller mindre bevisst) at slike selskaper samler inn og selger data om livene våre i bytte mot at vi får benytte tjenestene deres gratis.

At private selskaper og offentlige aktører prøver å påvirke oss er ikke noe nytt, ei heller at statistikk benyttes for å finne ut hvordan påvirkningen kan bli mest mulig treffsikker. Men aldri før har så

⁹ Hør eksempelvis på podcast-episoden *Ansvarlig bruk av stordata* med assisterende forskningssjef ved Norsk Regnesentral, Anders Løland: <https://www.lorn.tech/podder/ansvarlig-bruk-av-stordata>

omfattende og inngripende informasjon om så mange mennesker vært samlet på så få hender.¹⁰ Dette asymmetriske maktforholdet i samfunnet handler om makt til både å samle inn, eie, dele, manipulere, selge og analysere store mengder med data. Det gir en stor forskjell i tilgang til data, og i makten som ligger i tilgangen til kunnskapen vi kan få ut av dem.

Hvem får tilgang til den innsikten, kunnskapen og makten adgang til stordata kan gi? Kan vi stole på at selskaper som Facebook og Google er ansvarlige nok til å sitte på så stor makt? Én ting er at makten kan benyttes til å selge oss varer og tjenester og tjene kommersielle interesser vi egentlig ikke er enige i. Men hva om denne påvirkningsmakten brukes politisk?

Avslutning

Stordataanalyser sorterer hva vi ser på nett, foreslår hvem vi skal være venn med på Facebook, gir oss informasjon om været og bidrar i navigering gjennom GoogleMaps. Stordata har blitt en viktig ressurs i samfunnet og en sentral del av nye forretningsmodeller. Vi har nok bare sett begynnelsen på hvordan stordata kan komme til å prege samfunnet vårt, fra hverdagslivet til hvilke rettigheter og valgmuligheter vi har som innbyggere.

Stordataanalyser gir håp om bedre oversikt og nye tjenester i et komplekst samfunn preget av informasjonsoverflod. Samtidig bringer dette fenomenet med seg en rekke utfordringer for oss som enkeltmennesker og samfunnet i sin helhet. Siden stordata gjerne dreier seg om store datamengder som analyseres av kompliserte statistiske metoder, kan det fremstå som at stordata gir en slags nøytral, sannferdig innsikt. Mange forskere har imidlertid vist at heller ikke stordata kan gi oss en objektiv representasjon av samfunnet. Dette er spesielt viktig å ta hensyn til når stordata skal benyttes inn i automatiske systemer, siden slike systemer kan være vanskelige å endre når de først er satt ut i livet.

Akkurat hvordan stordata vil prege hverdagen vår i årene som kommer, avhenger i stor grad av hvordan vi velger å bruke og utvikle denne teknologien.

¹⁰ Flere har ytret stor bekymring om hvordan data, kunnskap og makt samles hos et knippe teknologiselskaper, og mener dette truer friheten vår og demokratiet. En av de fremste er Harvard-professor Shoshana Zuboff med den monumentale *Overvåkingskapitalismens tidsalder: kampen for en menneskelig framtid ved maktens nye frontlinje*, utgitt på norsk hos Spartacus i 2020. For en noe kortere, men likefullt viktig og engasjerende diskusjon, se Hilde Nagells *Digital revolusjon – hvordan ta makten og friheten tilbake*, utgitt i 2020 hos Res Publica.